

Lõpparuanne

Uuring „Lingitud Eesti“

Tallinn 2014



Uuringu autorid

Peep Küngas
Hele-Mai Haav
Jaak Tepandi
Lauri Tepandi
Mihkel Lauk
Triin Tars

Uuringu juhtrühm

Juhtrühma liige	Organisatsioon/asutus/ettevõte
Uuno Vallner	Majandus- ja Kommunikatsiooniministeerium
Kaisa-Maarja Pärtel	Riigikantselei
Siim Sikkut	Riigikantselei
Hannes Kiivet	Riigi Infosüsteemi Amet
Priit Parmakson	Riigi Infosüsteemi Amet
Margus Püüa	Siseministeeriumi infotehnoloogia- ja arenduskeskus
Peep Küngas	SOA Trader OÜ
Hele-Mai Haav	SOA Trader OÜ
Jaak Tepandi	Tepinfo OÜ
Lauri Tepandi	Tepinfo OÜ
Mihkel Lauk	PricewaterhouseCoopers Advisors AS
Triin Tars	PricewaterhouseCoopers Advisors AS

Täname uuringu autoreid, juhtrühma ja uuringus osalejaid.

ISBN 978-9949-33-722-4

PricewaterhouseCoopers Advisors AS

Tallinn 2014

Uuringu algatasid Majandus- ja Kommunikatsiooniministeerium ning Eesti Infotehnoloogia ja Telekommunikatsiooni Liit. Uuring viidi läbi Riigikantselei tarkade otsuste fondi ja Euroopa Sotsiaalfondi toel.

Sisukord

Mõisted ja lühendid	8
Kokkuvõte	10
Valdkonna globaalne areng	10
Linkandmete kasutuselevõtmine	12
Linkandmete kasutamise eesmärgid ja tulemused	12
Olulisemad trendid, Eesti võimekused ja võimalused	13
Aktiivsus rahvusvahelistes initsiatiivides	13
Pilootprojektide käivitamine	14
Linkandmete alase võimekuse tõstmine	14
Linkandmete alase teadlikkuse tõstmine	15
Linkandmetega kaasnevad võimalikud ohud ja riskid	15
Vajalikud muudatused õigusruumis	16
Ettepanekute elluviimise korraldamine	16
Koosvõime arhitektuuri arendamine	17
Linkandmetega seotud avatud standardid	17
Tehnilise infrastruktuuri lahendused	18
Koosloomel põhinevate lahenduste loomine	18
Vajalikud semantikaressursid	18
Linkandmed avaandmete publitseerimisel	19
Sisutöötlus veebisaitides ja -liidestest	19
Organisatsioonilise koosvõime arendamine	20
Linkandmete tehnoloogiate monitoorimine, õpetamine ja propageerimine	20
Linkandmed Eesti IKT sektori lahendustes	20
Avaliku sektori ülesanded infosüsteemide tellijana	21
Soovitusi erasektorile infosüsteemide arendamisel	21
Vabasektor ja kodanikkonna kaasamine	22
Tööjaotus eri huvipoolte vahel	23
Andmetöötlusprotsessid ja andmete täieliku elutsükli toetamine	23
Sissejuhatus	25
1 Uuringu metoodika	26
1.1 I etapp – Ettevalmistus	26
1.2 II etapp – Lähteandmete kogumine ja esialgne analüüs	26
1.3 III etapp – Intervjuud, küsitlused ja andmete analüüs	27
1.4 IV etapp – Uuringutulemuste valideerimine, esialgsete soovituste koostamine	28
1.5 V etapp – Uuringutulemuste kokkuvõtte koostamine ja Tellijale esitamine	28

2	Linkandmete ja seonduvate tehnoloogiate ning praktikate analüüs	29
2.1	Linkandmete ja nendega seotud tehnoloogiarendide analüüs	30
2.1.1	Linkandmete tehnoloogia	30
2.1.2	Semantilised tehnoloogiad.....	36
2.1.3	Avaandmed	39
2.1.4	Suurandmete töötlemise ja hõive tehnoloogiad	44
2.1.5	Koosloome tehnoloogiad.....	48
2.1.6	Tulevikuinternet	50
2.1.7	Asjade internet.....	52
2.1.8	Linkandmete seos teiste tehnoloogiatega	55
2.2	Rahvusvaheliste initsiatiivide, tegevuste ja standardite analüüs	57
2.2.1	Majanduskoostöö ja arengu organisatsiooni tegevused (OECD)	58
2.2.2	Euroopa Komisjoni ja tema peadirektoraatide algatused	59
2.2.3	Euroopa Komisjoni raamprogrammid	62
2.2.4	Standardid	65
2.2.5	Soovitused Eestile.....	67
2.3	Andme- ja teabevärvade analüüs.....	68
2.3.1	Ava- ja linkandmete portaalid	68
2.3.2	Portaalide ja kataloogide hetkeseis	68
2.3.3	Soovitused Eestile.....	70
2.4	Valdkonna juhtivate riikide parimate linkandmete alaste praktikate analüüs	71
2.4.1	Suurbritannia (UK)	71
2.4.2	Ameerika Ühendriigid (USA).....	76
2.4.3	Juhtivate riikide edutegurid linkandmete kasutuselevõtul	81
2.4.4	Kasulikke õppetunde välisriikide parimatest praktikatest Eestile	82
3	Arhitektuurianalüüs	83
3.1	Infosüsteemide valim	83
3.2	Referentssüsteem juhtumiuuringuks	84
3.3	Juhtumiuuringud	88
3.3.1	DBpedia Live.....	88
3.3.2	Riiklik ehisregister	89
3.3.3	TWC LOGD	90
3.4	Vastavus Linked Data Platform 1.0 nõuetele	92
3.5	Ettepanekud	92
4	Õigusruumi analüüs.....	94
4.1	Analüüsi eesmärgid, ulatus ja piirangud	94
4.1.1	Õigusruumi analüüsi üldised eesmärgid.....	94
4.1.2	Ava- ja linkandmete kasutuselevõtu võimalike tugevuste, nõrkuste, võimaluste ja ohtude analüüs.....	95

4.1.3	Analüüsi piirangud: hõlmataavad tehnoloogiad ja õigusruum	96
4.2	Avaandmete/linkandmete publitseerimine	98
4.2.1	Avalikustatavad andmed ja avalikustamise tähtsajad	98
4.2.2	Avalikustatavate andmete masinloetavuse tase	99
4.2.3	Andmete avalikustamise viis ja nõuded	99
4.2.4	Ettepanekud	100
4.3	Avaliku teabe taaskasutamise tasustamine	101
4.3.1	Ettepanekud	102
4.4	Isikuandmete kaitse ja privaatsuse tagamine linkandmete kasutuselevõtu raames	102
4.4.1	Isikuandmete esitamine avaandmetena	102
4.4.2	Isikuandmete töötlemine linkandmetena	104
4.4.3	Isikuandmete kaitse ja privaatsuse tagamine seoses linkandmetega	105
4.4.4	Ettepanekud	106
4.5	Õigused, vastutused ja kohustused	106
4.5.1	Intellektuaalomandi õiguste kaitse tagamine	107
4.5.2	Vastutus linkandmete ja nende kaudu esitatavate teenuste kvaliteedi eest	107
4.5.3	Ettepanekud	108
4.6	Omandi, kriitilise taristu ja riigi kaitse	108
4.6.1	Omandi andmete kaitse	108
4.6.2	Kriitiline taristu ja riigi kaitse	109
4.6.3	Ettepanekud	110
5	Kvalitatiivuuringu tulemused	111
5.1	Linkandmete ja seonduvate tehnoloogiate kasutuselevõtu võimalused ja takistajad	111
5.2	Lingitud Eesti koosvõime arhitektuur	116
5.3	Organisatsiooniline koosvõime	118
6	Kvantitatiivuuringu tulemused	119
6.1	Vastanute profiil	119
6.2	Vastused küsimustele	121
7	Vastused uurimisküsimustele, ettepanekud ja poliitikasoovitused	127
7.1	Linkandmete kasutuselevõtu üldised ettepanekud	127
7.1.1	Linkandmete tehnoloogiate kasutamise eesmärgid ja tulemused	127
7.1.2	Olulisemad trendid, Eesti võimekused ja võimalused	129
7.1.3	Aktiivsus rahvusvahelistes initsiatiivides	130
7.1.4	Pilootprojektide käivitamine	132
7.1.5	Linkandmete alase võimekuse tõstmine	135
7.1.6	Linkandmete alase teadlikkuse tõstmine	136
7.1.7	Linkandmete kasutuselevõtuga kaasnevad võimalikud ohud ja riskid	138
7.1.8	Vajalikud muudatused õigusruumis	139

7.1.9	Ettepanekute elluviimise korraldamine	140
7.2	Lingitud Eesti koosvõime arhitektuuri arendamise ettepanekud	141
7.2.1	Linkandmetega seotud avatud standardid.....	141
7.2.2	Tehnilise infrastruktuuri lahendused.....	142
7.2.3	Koosloomel põhinevate lahenduste loomine	144
7.2.4	Vajalikud semantikaressursid.....	145
7.2.5	Linkandmete tehnoloogiad andmete avaandmetena publitseerimisel	146
7.2.6	Sisutöötlus veebisaitides ja -liidestest	148
7.3	Lingitud Eestit toetava organisatsioonilise koosvõime arendamise ettepanekud	149
7.3.1	Linkandmete tehnoloogiate monitoorimine, õpetamine ja propageerimine	149
7.3.2	Linkandmete tehnoloogia Eesti IKT sektori väljatöötatavates lahendustes	151
7.3.3	Eesti avaliku sektori täiendavad ülesanded infosüsteemide tellijana	152
7.3.4	Soovitusi erasektorile oma infosüsteemide tellimisel ja arendamisel.....	153
7.3.5	Vabasektor ja kodanikkonna kaasamine.....	155
7.3.6	Tööjaotus eri huvipoolte vahel.....	156
7.3.7	Andmetöötlusprotsessid ja andmete täieliku elutsükli toetamine	158
8	Kasutatud kirjandus.....	162
9	Lisad.....	173
9.1	Kvantitatiivküsitluse ankeet.....	173

Joonised

Joonis 1. Avalike linkandmete elutsükli mudel	35
Joonis 2. Semantilised tehnoloogiad	36
Joonis 3. Tõusvad tehnoloogiad ja nende seosed	56
Joonis 4. Linkandmete tehnoloogia kui erinevate andmete ja tehnoloogiate linkija	57
Joonis 5. OECD riikide tsentraalsetes portaalides publitseeritud avaandmehulkade arvud (2013)	59
Joonis 6. Avalike andmete tehnoloogiline teekaart	63
Joonis 7. UK avalike teenuste kasutuselevõtu kontseptuaalne mudel poliitilise konteksti jaoks	73
Joonis 8. UK avaandmete strateegia.....	74
Joonis 9. Referentsarhitektuur valimi lahenduste arhitektuuridest.....	87
Joonis 10. DBpedia Live System Architecture.....	88
Joonis 11. TWC LOGD portaali dünaamiliselt genereeritud esileht	91
Joonis 12. Vastanute profiil: sihtrühmade jaotus	119
Joonis 13. Tehnoloogiate olulisus Eestile linkandmete kontekstis	122
Joonis 14. Sihtrühmade teadlikkus valitsuse plaanidest linkandmete vallas.	122
Joonis 15. Eesti võimekus tehnoloogiarendide raames	122
Joonis 16. Eesti arenguvajadus tehnoloogiarendide raames	123
Joonis 17. Potentsiaalsed kasud linkandmete kasutamisest	123
Joonis 18. Linkandmete loomist ja kasutamist pärssivad tegurid	124
Joonis 19. Ohtude suurus seoses linkandmete laialdase kasutuselevõutuga	125
Joonis 20. Linkandmete avamise potentsiaalne kasu valdkonniti	125
Joonis 21. Projektide ja tegevuste mõju Eestile	126
Joonis 22. Lingitud avaandmete kasutusvõimaluste atraktiivsus	126

Tabelid

Tabel 1. Tähtsamad rahvusvaheliste organisatsioonide avaandmete portaaliid	69
Tabel 2. Linkandmete kasutuselevõtu edutegurid	81
Tabel 3. Soovitused Eestile teiste riikide edutegurite alusel (tähtsuse järjekorras)	83
Tabel 4. Arhitektuurianalüüsi valim	83
Tabel 5. Linked Data Platform 1.0 nõuete kaetus valimi lahenduste poolt	92
Tabel 6. Linkandmete kasutuselevõtu SWOT-analüüs.....	96
Tabel 7. Ankeetküsitluse sihtrühmade koosseis.....	119
Tabel 8. Vastanute profiil: teadmised tehnoloogiatest	120
Tabel 9. Vastanute profiil: praktilised kogemused	120
Tabel 10. Avalike linkandmete elutsükli tegevused ja nende toetamine	159

Mõisted ja lühendid

Mõiste	Mõiste inglise keeles (vajadusel)	Selgitus
ADS		Aadressiandmete süsteem
Avaandmed	<i>open data</i>	<p>Andmed, mis on kõigile antud vabalt kasutada, taaskasutada ja jagada, kitsendatult maksimaalselt viitamise või samadel tingimustel taasjagamise nõudega (<i>A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike</i>)¹.</p> <p><u>Märkus 1.</u> Mõiste „avaandmed“ räägib õiguslikust koosvõimest, mõiste „linkandmed“ - semantilisest ja tehnilisest koosvõimest².</p> <p><u>Märkus 2.</u> Avaandmed võivad olla linkandmed ja võivad ka mitte olla.</p> <p><u>Märkus 3.</u> Vt ka „Avaandmete viie täрни süsteem“.</p>
Avaandmete viie täрни süsteem	<i>5 * deployment scheme for open data</i>	Tim Berners Lee poolt pakutud vormingute ja kodeeringute süsteem avaandmete „kasutajasõbralikkuse“ hindamiseks ³
Avalik teave	<i>public sector information, government data</i>	<p>Mis tahes viisil ja mis tahes teabekandjale jäädvustatud ja dokumenteeritud teave, mis on saadud või loodud avaliku sektori poolt seaduses või selle alusel antud õigusaktides sätestatud avalikke ülesandeid täites.</p> <p><u>Märkus:</u> juurdepääsu avalikule teabele võib seaduses sätestatud korras piirata (sõnal „avalik“ on eesti keeles mitu tähendust, siinkohal kasutatakse neist ühte).</p>
Avaliku sektori avaandmed	<i>open government data</i>	Avaliku sektori poolt avaandmetena publitseeritud teave
Avaliku sektori linkandmed	<i>linked government data</i>	Avaliku sektori poolt linkandmetena publitseeritud teave
Avalikustatud teave	<i>published data</i>	Avalikkusele kättesaadavaks tehtud teave
EL	<i>EU</i>	Euroopa Liit

¹ <http://opendefinition.org/>, <http://www.w3.org/2005/Incubator/lld/wiki/Scope>

² <http://www.w3.org/2005/Incubator/lld/wiki/Scope>

³ <http://5stardata.info/>

Mõiste	Mõiste inglise keeles (vajadusel)	Selgitus
Linkandmed	<i>linked data</i>	<p>Struktureeritud andmed, mis on kirjeldatud ja avaldatud viisil, mis lubab neid automaatselt seostada (vt ka mõiste „Avaandmed“ ja selle märkused).</p> <p><u>Märkus 1.</u> Linkandmed võivad olla avaandmed ja võivad ka mitte olla.</p> <p><u>Märkus 2.</u> Avaandmete viie tärni süsteemis on linkandmed viiendal tasemel.</p> <p><u>Märkus 3.</u> W3C avaliku sektori linkandmete töögrupp (<i>W3C Government Linked Data Working Group</i>) toetub linkandmete määratlemisel spetsiifilistele tehnoloogiatele ja standarditele (URI, HTTP, RDF, SPARQL jt)⁴. Käesolevas dokumendis on kasutatud laiemat mõistet, et mitte kitsendada käsitlust vaid neile tehnoloogiatele.</p>
Masinloetav vorming	<i>machine-readable format</i>	<p>Failivorming, mis on struktureeritud selliselt, et tarkvararakendused suudavad spetsiifilisi andmeid, sealhulgas üksikuid faktiväiteid, ja nende sisemist struktuuri kergelt tuvastada, ära tunda ja välja lugeda⁵.</p> <p><u>Märkus.</u> Vrd ka USA valitsusasutuste jaoks kasutatav mõiste⁶.</p>
Mosaiikandmed		Andmed, millelt on eemaldatud isikustamist võimaldavad tunnused, ent mitmest andmekogust pärinevate andmete linkimisel ja koosesitamisel tekib võimalus andmeid taasisikustada.
RIA		Riigi Infosüsteemi Amet
RISO		Majandus- ja Kommunikatsiooniministeeriumi riigi infosüsteemide osakond

⁴ Linked data, <https://dvc.w3.org/hg/gld/raw-file/default/glossary/index.html#linked-data>

⁵ EUROOPA PARLAMENDI JA NÕUKOGU DIREKTIIV 2013/37/EL, 26. juuni 2013, millega muudetakse direktiivi 2003/98/EÜ avaliku sektori valduses oleva teabe taaskasutamise kohta

⁶ http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/s200.pdf

Kokkuvõte

„Lingitud Eesti“ all mõistetakse IKT-lahenduste toel loodud kogu Eesti inforuumi hõlmavat hajutatud isikute, organisatsioonide, tarkade seadmete, teadmiste, infosüsteemide ning andmete koosvõimelist võrgustikku.

Linkandmed on struktureeritud andmed, mis on kirjeldatud ja avaldatud viisil, mis lubab neid automaatselt seostada. Linkandmete tehnoloogia võimaldab seni hajutatult eksisteerivad andmebaasid, sotsiaalmeedia, dokumendid, faktid, seadmed, teadmised jms integreerida ühtseks tervikuks ning avab seeläbi riikidele, era- ja vabasektorile seni võimatud olnud ehk põhimõtteliselt uued võimalused uut tüüpi teenuste loomiseks. Kuid selle võrgustiku loomine eeldab läbimurret mitmetes suundades nagu ülikiire andmesidevõrk ja targad seadmed, koosvõimelised infosüsteemid ja teadmusvõrgustikud, semantilised võrgustikud – kõige olulisemana aga linkandmete kasutuselevõtt.

Linkandmete kontseptsiooni järgi saavad lisaks dokumentidele otselingitavaks ka andmed ise ja kõik füüsilised kui ka abstraktsed „asjad“. Kui praegu saavad brauserite abil teavet lugeda eelkõige inimesed (andmed on inimloetavad), siis kasutades linkandmeid toetavaid tehnoloogiaid, saavad arvutid andmeid lugeda – ning mis olulisemgi, eri kohtades paiknevaid andmeid siduda, nendes otsida ja neid töödelda. Linkandmete tehnoloogia võimaldab loobuda traditsiooniliste andmekogude loomisest ja nende seostamisest: kasutades linkandmete tehnoloogiaid on see funktsionaalsus andmete avaldamisega vaikinisi saavutatav.

Linkandmete kasutuselevõtu initsiatiiv edendab läbipaistvust, koostööd riikide tasandil ja laiemalt, võimaldab luua uusi, uuenduslikke teenuseid ning tõsta otsuste tegemise kvaliteeti.

Linkandmed kui tehnoloogiline suundumus on tihedalt seotud selliste ülemaailmsete trendidega nagu tulevikuinternet (*Future Internet, FI*) ja semantiline veeb (*Semantic Web*), asjade Internet (*Internet of Things*) ja suurandmed (*Big Data*), avaandmed (*Open Data*) ja koosloome (*Crowdsourcing*).

Eestis puudub terviklik, Eesti konteksti ja vajadusi arvestav uuring linkandmete kasutusvõimaluste osas, millel võiks põhineda linkandmete kasutuselevõtu poliitika. Käesoleva uuringu tulemused on otseseks sisendiks riigi IKT-poliitika kujundamisele: järgmiste aastate infoühiskonna arengukava rakendusplaanide koostamisele, struktuurifondide rahastamisvoorude ettevalmistamisele ja koosvõime raamistiku dokumentide täiendamisele.

Uuringu eesmärgiks on loodava „Infoühiskonna arengukava 2020“ eesmärkide saavutamiseks vajalike rakendusplaanide jaoks sisendi andmine lingitud andmete loomise ja kasutamise osas ning vastavate tegevuste efektiivseks elluviimiseks vajalike muudatuste tegemiseks Eesti õigusruumis ja poliitilistes meetmetes.

Uuringus pakume välja lähenemise linkandmete tehnoloogiate laiemaks kasutuselevõtuks Eestis, tuues välja puudused, mis pärsivad lingitud andmete efektiivset loomist ja kasutamist. See hõlmab soovitusi avaliku sektori, erasektori ja vabasektori andmekogude, veebisaitide ja muude inforessursside ümberkorraldamiseks ning neid toetavateks poliitikameetmeteks riigi poolt. Järgnevas toome välja ettepanekud ja poliitikasoovitused. Nende põhjendused ja täpsem kontekst on esitatud uuringu täispikas aruandes.

Valdkonna globaalne areng

Kokkuvõtvalt võib öelda, et linkandmetega seotud tehnoloogiatest on avaliku sektori avaandmed ja NoSQL andmebaasisüsteemid (siia alla kuuluvad ka RDF andmebaasid) ajaliselt kõige lähemal produktiivsuse platoole (2-5 a). Suurandmete hõive ja analüütika tehnoloogiad on kasutuselevõtu faasi alguses. Gartneri

2012. a suurandmetega seotud tehnoloogiate kasutuselevõtu graafikus oli MapReduce jt suurandmete tehnoloogiad kasutuselevõttust 2-5 aasta kaugusel. Teised seonduvad tehnoloogiad nagu semantiktehnoloogiad ja asjade internet on produktiivsuse platoost rohkem kui 10 aasta kaugusel.

Avaandmete arengul on otsene mõju avalike linkandmete kasutuselevõtule ja linkandmete kasutuselevõtule üldse, sest ka erasektor saaks oma andmetega linkida avalikke linkandmeid ja seega linkandmete mõju veelgi suurendada. Tuleb tähtsustada rahvusvaheliste ja rahvuslike initsiatiivide ja tegevuste mõju avaliku sektori andmete avalikustamisel ja avalike linkandmete edendamisel.

2013. a oli avaandmete seisukohast väga tähtis aasta: avaldati G8 riikide avaandmete harta ja USA Presidendi otsekorralduse riigivalitsemise uue digitaalstrateegia kohta, W3C avaldas linkandmete platvormi parimad praktikad ja juhised, ODI laienemine (Prantsusmaale, Dubaisse, Kanadasse, USA-sse, Argentiinasse ja Venemaale) ja avaldati Eesti avaandmete rohelise raamatu mustand.

Nii OECD kui EK oma peadirektoraatide kaudu on pühendanud palju initsiatiive avalikele avaandmetele ja nende linkimisele ning seotud tehnoloogiate toetamisele. Eraldi võib märkida OECD avaandmete projekti ja avaandmete indikaatorite liigitamist OECD 2013. a ülevaatesse „*Government at a Glance*“. Euroopa Komisjoni tegemistest võib esile tõsta DG CONNECT EL digitaalarengu tegevuskava 2020, DIGIT töötas välja vabavara strateegia perioodiks 2011-2013 ja haldab ISA programmi.

Euroopa Komisjon on toetanud tervet rida teadus- ja arendusprojekte oma FP7 IKT tööprogrammi raames selleks, et vähendada nii linkandmete loojate kui kasutajate sisenemistõkkeid valdkonda. Nende projektide hulgas on suurema mõjuga LOD2, LATC, PlanetData, TELEIOS, BIG ja IKS projektid. EL uus teaduse ja innovatsiooni raamprogramm Horizon 2020 perioodiks 2014-2020 rahastab aastatel 2014-2015 palju ava- ja linkandmetega seonduvaid teadus- ja arendusprojekte. Põhiline fookus on selles suurandmete teema sidumisel avalike avaandmetega. Eriti palju tähelepanu saavad 2014-2015 tegevusprogrammis tulevikuinterneti ja asjade interneti teemad.

Ava- ja linkandmete rakendamisel on tähtis roll rahvusvahelistel standarditel ja soovitel, mida annab välja rahvusvaheline veebikonsortsium W3C. Linkandmete rakendamisega seoses on tähtsad rahvusvahelised sõnastikud ja ontoloogiad. Praegu on W3C poolt soovitatud ontoloogiate hulgas juba inimeste ja organisatsioonide standardontoloogiad.

PwC EU Services avaliku sektori lingitud avaandmete ärimudelite alase uuringu järeldus oli, et avalikud linkandmed liiguvad edasi tehnoloogia kasutuselevõtu elutsükli olles näiteks raamatukogunduse/arhiivinduse alal jõudnud juba varajase kasutuselevõtu faasi. Lõppjäreldusena toodi uuringus välja, et kuna põhilised andmete pakkujad on tootmas suures mahus avalikke linkandmeid, siis on ette näha, et ka nende andmete taaskasutajad leiavad lähitulevikus oma võimaluse innovatsiooniks.

Kui poliitiline motiiv avaandmete osas oli elavdada soiku jäänud majandust, siis suurandmete osas on poliitika stimuleerida andmeanalüüsil põhinevat otsustusprotsessi ning kasutada mittestruktureeritud veebiandmeid (põhiliselt sotsiaalmeediast) riiklike missioonikriitiliste rakenduste loomiseks.

Kasulikkuse aspektist on suurandmete tehnoloogia oma kasulikkust teatud määral juba tõestanud. Suurandmete tehnoloogiate kasutuselevõtt ja selle majanduslik kasu ettevõtluses ja avalikus sektoris kasvab vastavalt sellele, kuidas ollakse võimelised neid tehnoloogiaid kasutama keeruliste ülesannete lahendamiseks. Siin ei ole kitsaskohaks niivõrd tehnoloogia kui andmetöötajate analüütiline kompetentsus. Just personali suurandmete analüütika alast ebakompetentsust peetakse põhiliseks suurandmete tehnoloogiate kasutuselevõtu piirajaks.

SAS-i uuringu „*Big-Data-in-Big-Companies*“ järgi on selge, et suured tööstusettevõtted lähevad üle suurandmete töötlemise ja analüütika tehnoloogiatele kuigi see võtab aega ja nõuab täiesti uut tehnoloogilist lähenemist, muutuvat organisatsiooni ja juhtimisstruktuuri ning uusi oskusi. Suurandmete peamine väärtus tuleb just nende töötlemisest ja analüüsist ja toodetest ning teenustest, mis lähtuvad selle analüüsi tulemustest.

Kuigi on üsna selge, et suurandmete tehnoloogiate kasutuselevõtt võimaldab suurendada majanduslikku kasvu, peab siiski pöörama tähelepanu ka selle sotsiaalsetele aspektidele. Nendeks on autoriõiguste tagamine, turvalisus ja privaatsus. Seega saab suurandmete tehnoloogiate kasutuselevõtu kriitiliseks eduteguriks balansseerimine kättesaadavate andmete ja nende üha suureneva linkimise ning analüüsimise soovi ning võimaliku sotsiaalse häirumuse vahel.

Kaubamärkide omajad lasevad praegu iga nädal välja tooteid, teenuseid ja sisu, mis on inspireeritud ja loodud internetikogukondade poolt. Aastal 2014 loodetakse näha koosloome efektiivsuse tõestust turunduse ja innovatsiooni valdkondades. Gartner ennustab, et koosloomet kasutavate firmade kasum tõuseb 1% võrreldes firmadega, kes koosloomet ei kasuta ja seda juba aastaks 2015.

Praegu on murranguline periood interneti arengus, sest kommunikatsioonile orienteeritud internet peaks vastavalt muutunud nõuetele ja vajadustele muutuma sisule orienteeritud internetiks. Seoses sellega muutub aktuaalseks uute internetiarhitektuuride arendus. Pragmatilisest vaatepunktist lähtuvalt on praegu maailmas tähtis üleminek uuele internetiprotokollile IPv6, sest IPv4 aadresside hulk on ammendumas. IPv6 võimaldab kasutusele võtta uue aadressruumi ja lülitada internetti üha rohkem asju s.o sensoreid jms.

Praeguseks olemegi juba jõudnud olukorran, kus rohkem seadmeid (asju) kui inimesi on liitunud internetiga ja ennustatakse, et internetti ühendatud asjade hulk ületab peagi inimpopulatsiooni arvu, moodustades 20-50 biljonit ühendatud tarka seadet. See tähendab ka ühtlasi seda, et interneti protokoll IPv4, mis on mõeldud 4 biljoni aadressi jaoks, ammendub.

Aastaks 2020 ennustatakse Internetti lülitatud seadmete arvuks 50 biljonit kuni 1 triljon. Sensor- ja võrguvõimekuse lisamine asjadele on kaotanud oma tehnoloogilised ja hinnabarjäärid. Praegu toimub asjade interneti tehnoloogiate läbimurre tarbijaturule ja toodetakse üha rohkem tarku laiatarbetooteid. Personaalsete jälgimisvahendite ja kantava tehnoloogia turg peaks kasvama ülikiiressti just sel aastal. Personaalsete jälgimisvahendite hinnad langevad ja erinevad targad kellad, tervise ja inimkeha seisukorra näitajate jälgimise vahendid jms tooted tulevad turule. Asjade interneti laialdasel kasutuselevõtul on ka varjuküljed. Võrgustatud asjad võivad saada spioneerimise objektideks. Samas saab seda tehnoloogiat kasutada ka vastuluureks ja militaarsetel eesmärkidel.

Maailma juhtivad riigid nagu UK ja USA on oma riiklike informatsiooniliste infrastruktuurilahenduste alal üle minemas pilvearvutuse infrastruktuurile ja linkandmete kasutamisele info integreerimiseks. Vastavad strateegilised dokumendid anti välja põhiliselt 2012. aastal. Tulemused on juba praegu näha.

Uusimad trendid USA-s ja EL-is onaju-uuringutel põhinevate biokiipide loomine ja nende ühendamine võrkudeks, mille produktiivset kasutuselevõttu ennustab Gartner 10 aasta pärast. Järgnevad aastad on ilmselt murrangulised vaadeldud tõusvate tehnoloogiate osas, muutes oluliselt maailma meie ümber ja sees.

Linkandmete kasutuselevõtmine

Linkandmete kasutamise eesmärgid ja tulemused

Linkandmete tehnoloogiad võimaldavad seostada omavahel isoleeritud andmeid ja seeläbi luua uusi teenuseid ja teadmisi ning edendada innovatsiooni. Näiteks on uuringutes toodud positiivsete võimalustena välja valitsemise läbipaistvus ja efektiivsus, innovatsioon, 21. sajandi infrastruktuur. Valitsustele annavad linkandmete tehnoloogiad võimaluse nutikamate ja efektiivsemate avaliku sektori teenuste ja rakenduste väljatöötamiseks. Seega **linkandmete tehnoloogiate kasutuselevõttu tuleb soodustada**.

Linkandmete kasutamise eesmärkide ja tulemuste hulka kuuluvad riigi ja kodaniku vahelise suhtluse paranemine (nt LEO, Läbipaistev Eesti Omavalitsus); „nähtamatud“ teenused, millele inimesele on seadusandlik õigus ning mida ta peab saama sõltumatult sellest, kas ta tellib teenust või mitte (nt lapse riikliku sünnitoetuse teenus või riiklik vanaduspension); andmete kiirem analüüs (valedete otsustele on võimalik kiiremini reageerida ning uusi otsuseid langetada); monitooringu võimaldamine elutähtsate

teenuste osutajate jaoks (lubab saada oluliselt kiiremini infot, kui elutähtis teenus on häiritud); avaliku sektori kulude vähenemine, sh seoses teabenõuete vastamise vajaduse vähenemisega ja kodanike kaasamisega (nt tänavaaukude kaardistamine mobiilirakendusega); elanikkonna elukvaliteedi tõstmine ja tervishoiu valdkonna kulude vähendamine (nt geenivaramu andmete sidumisel e-Tervise andmetega).

Linkandmete ning nendega seotud tehnoloogiate temaatika tuleks ajakohastatult ja täiendatult lülitada riigi infosüsteemi koosvõime raamistikku ja sellega seotud dokumentidesse. Arvestada tuleb, et linkandmete puhul on tegemist on tüüpilise innovatsiooniprotsessiga, mille puhul uue tehnoloogia kuluefektiivne rakendamine eeldab eelnevat mahukat investeerimist toimiva taristu loomisesse, linkandmete eelnevasse publitseerimisse, kvalifitseeritud arendajate koolitamisest jne.

Olulisemad trendid, Eesti võimekused ja võimalused

Käesolevas uuringus läbi viidud intervjuudes peetakse riigi infosüsteemide seisukohalt oluliseks peamiselt linkandmete ja semantilise veebi tehnoloogiaid. Neid kahte tehnoloogiat nimetatakse sageli paaris, kuna semantika loob linkandmete kasutuselevõtuks vajalikud eeldused. Samuti on nendes tehnoloogiates Eestil täna esialgne võimekus olemas ning arenguvõimalused väga head, kui arvestada olemasolevat riigi infosüsteemi arhitektuuri, läbiviidud avaandmete tootmise pilootprojekte ning andmete ja teenuste semantilise metainfo esitamise juba olemasolevaid kogemusi. Linkandmete kasutamise eelisena tuuakse välja riigi paindlikumat reageerimist muutustele ning kvaliteetsemat poliitikakujundamist. Informatsiooni kättesaadavuses ja kasutamises ei nähta konkurentsieelist mitte ainult erasektorile, vaid ka riigile. Asjade interneti, suurandmete ja koosloome osas arvamused hajuvad, kuna kogemus ja teadmised nende tehnoloogiate kasutamisel ei ole ekspertidel sarnased. Kokkuvõttes, kõige tähtsamaks linkandmete kasutuselevõttu peetakse andmete avamist ja semantiliste tehnoloogiatele kasutuselevõttu, milleta ei ole linkandmete kasutuselevõtt võimalik. Samuti tundub avaandmete osas olevat mahajäämus suurim võrreldes teiste linkandmetega seotud tehnoloogiatega. Ülejäänud tehnoloogiate järjestuses tekivad eriarvamused ning ei joonistu välja selgeid prioriteete.

Aktiivsus rahvusvahelistes initsiatiivides

Aktiivsus rahvusvahelistes initsiatiivides võib anda uut kogemust, aidata levitada oma tulemusi, parandada riigi mainet ja tuua kaasa lisaressursse. Samas nõuab see omapoolset raha ja kõrge kvalifikatsiooniga inimeste ajaressursside panustamist. Eesti riigi väiksuse ja ressursside piiratuse tõtu tuleks initsiatiive hoolega valida. Ankeetides hinnati Eesti osalemise mõju rahvusvahelistes linkandmete ja avaandmete projektides/algatustes keskmiseks, võrreldes teiste uuritud valdkonna alaste tegevustega.

Eesti peaks olema aktiivne järgmistes linkandmetega seotud initsiatiivides (prioriteetsuse järjekorras): Euroopa Komisjoni initsiatiivid ja programmid (sh strateegiliste dokumentide loomine, ISA programm, Joinup foorum, Horizon raamprogramm, eSENS projekt jt), standardimistegevused, OECD avaliku sektori avaandmete (*Open Government Data*) projekt, Euroopa andmete foorumi konverentside seeria, Euroopa avaliku sektori infotribüün ePSIplatform, avatud valitsemise partnerlusprogramm jne.

Osavõtt Euroopa Komisjoni tegevusest vajab üldist link- ja avaandmete teemalist kompetentsi; osavõtt EL raamprogrammist Horizon 2020 eeldab põhjalikku tehnilist koosvõime, linkandmete ja teadustöö alast kompetentsi; osavõtt standardimisest eeldab väga head kompetentsi standarditavas valdkonnas, teadmisi standardimise protsessist ja üsna palju tööaega. Arvestades Eesti väiksust ning ressursside nappust tuleks linkandmetega seotud tegevus võimalusel ühendada muude tegevustega vastavate initsiatiivide raames.

Eesti erasektori esindajad (näiteks ITL aga ka erafirmad, eriti idufirmad) võiksid astuda ODI toetajaliikmeks (liikmemaks 720 Inglise naela aastas). ODI partnerliikmeks saamine ja olemine on suhteliselt kallis (70000 Inglise naela aastas), mis pole ilmselt jõukohane ja ei vasta ka ootustele.

Pilootprojektide käivitamine

Linkandmete kasutuselevõtt eeldab järgmisi etappe: andmete avamist (kättesaadavaks tegemist), nende puhastamist ja korrastamist (vajadusel), viimist vähemalt neljanda või viienda järni tasemele, linkimist, taaskasutamist (nt päringute võimaldamist) ja haldamist⁷. Kõigil neil etappidel on vaja lahendada suur hulk uurimisküsimusi ning näiteks EL uus teaduse ja innovatsiooni raamprogramm Horizon 2020 perioodiks 2014–2020. a rahastab aastatel 2014–2015 kokku enam kui poolteise miljardi euro maksumuses ITK alaseid initsiatiive, sealhulgas palju ava- ja linkandmetega seonduvaid teadus- ja arendusprojekte.

Eesti siseriiklikke linkandmete kasutuselevõtu pilootprojekte on soovitatav kavandada nii, et nad arvestaksid Eesti eripära (nt X-tee ja RIHA olemasolu), ei dubleeriks mõjuva vajaduseta rahvusvahelisi suuremahulisi uurimisprojekte, looksid võimaluse kolmandatele osapooltele linkandmete rakenduste loomiseks, tekitaksid lõppkasutajale või üldsusele demonstreeritava kasuliku näitetulemuse („*proof of concept*“), annaksid asjaosalistele uusi kogemusi ja oleksid jätkusuutlikud. Näiteks võiks suhtuda ettevaatlikult linkandmete publitseerimise pilootprojektide algatamisse, mis ei demonstreeri selgelt saadavat kasu ning mis ei kindlusta selle kasu jätkusuutlikkust, sealhulgas loodud linkandmete uuendamist ja haldust.

Üheks pilootprojektide käivitamise eesmärgiks on kogemuste andmine uute tehnoloogiatega töötamisel. Anketeerimise tulemustest selgus, et kogemused avaandmetega puudusid 41% vastajatest, kogemused ülejäänud vaadeldavate tehnoloogiatega – 57–68% vastajatest.

Otstarbekas on käivitada orienteeruvalt kuni kuus tehnilist pilootprojekti alternatiivsete linkandmete tehnoloogiate järeleproovimiseks praktikas erinevate seotud tehnoloogiate võtmes valdkonnaüleste infoühiskonna arengukavas väljatoodud oluliste mõõdikute parendamise võtmes. Ühe pilootprojekti maksumuse suurusjärguks võib pakkuda kuni sada tuhat eurot. Pilootprojektid võiksid olla soovitatavalt tervishoiu, georuumilises, hariduse või transpordi, e-äri ja infrastruktuuri valdkondades, kusjuures väga oluline on vastavate organisatsioonide valmisolek projektide algatamiseks. Pilootprojekte on soovitatav kavandada nii, et nad arvestaksid Eesti eripära, ei dubleeriks mõjuva vajaduseta rahvusvahelisi suuremahulisi uurimisprojekte, looksid võimaluse kolmandatele osapooltele linkandmete rakenduste loomiseks, tekitaksid lõppkasutajale või üldsusele demonstreeritava kasuliku näitetulemuse, annaksid asjaosalistele uusi kogemusi ja oleksid jätkusuutlikud.

Järgmine suurem programm oleks enamkasutatavate andmete põhjal linkandmete tuumiku loomine, kuhu ülejäänud andmeid saab vastavalt vajadusele juurde linkida. See programm peaks algama õiguslike aspektide analüüsi sisaldava pilootprojektiga ning sisaldama mitmeid järjekordseid projekte oluliste andmekogude linkimiseks. Programmi mahtu saab täpsemalt hinnata peale tehniliste pilootprojektide läbiviimist.

Lisaks ülalmainitud tehnilistele pilootprojektidele tuleks kasutada linkandmete platvorme samaliigiliste andmete erinevatest allikatest kokkukogumiseks, näiteks kohalike omavalitsuste andmete ühendamiseks, üldistades kohalike omavalitsuste raamatupidamise avaandmete rakendust. Selliste projektide maksumus jääb tehniliste pilootprojektide maksumuse ja linkandmete tuumiku programmi maksumuse vahepeale.

Asjade Interneti ja tulevikuinterneti arendamiseks oleks esmajoonel vaja pöörata tähelepanu internetiprotokolli 6. versiooni kasutusele võtmisele Eestis. Riigi poolne initsiatiiv võiks seisneda ülemineku käivitamises IPv6-le avalikus sektoris. Tuleb arvestada, et rahalised ja tööjõu kulutused IPv6-le üleminekuks üle tuhande töötajaga ministereeriumi valitsemisalas võivad olla märkimisväärsed ning need tuleks eelarvesse sisse planeerida või rahastada eraldi projektidena.

Linkandmete alase võimekuse tõstmine

Võimekuse tõstmine on kompleksne teema. Nii näitab UK kogemus, et linkandmete RDF-kujul kättesaadavaks tegemisest ei piisa nende laialdaseks tarbimiseks rakenduste kaudu. Seetõttu loodi Ühendkuningriikides vahevara linkandmete kättesaadavaks tegemiseks ka JSON ja XML andmekirjelduskeeltes. Lisaks loodi valdkonnaspetsiifilised API-d vastavate andmete kasutamiseks. Samuti

⁷ <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>

toob aruanne välja linkandmete tootjate vajaduse konkreetsete eeskirjade ja muustrite järele linkandmete tootmiseks ning vajaduse juhendamise järele linkandmete tootmisel ja haldamisel. Eestis näitavad esialgsed Ehtisregistri linkandmete kogemused RDF-i kasutamise osas nii õppetöös kui erasektori tarbijate poolt, et piisava huvi olemasolu korral ei saa linkandmete tehnilised detailid tarkvaraarendajatele rakenduste loomisel takistuseks. Pigem on küsimus teadlikkuse tõstmises ja ajendi tekitamises linkandmete loomise ja kasutamise motiveerimiseks vahendite piiratust arvestades.

Linkandmete alase võimekuse tõstmiseks tuleks eelkõige paralleelselt käivitada linkandmete teemalised kursused ja koolitused, algatada pilootprojektid ning luua tehnilise infrastruktuuri lahendused. Lisaks tuleks osaleda rahvusvahelistes initsiatiivides, maandada riske ning korrastada õigusruumi.

Linkandmete alase teadlikkuse tõstmine

Tavaelanikud võivad tänu linkandmete kasutamisele saada paremaid teenuseid, kuid nad pole eeldatavalt linkandmete nn toorel kujul kasutajate peamine sihtgrupp. Seda arvestades peaks peamine teavitustöö linkandmete olemuse osas toimuma sihtgruppides, kes loovad, haldavad või kasutavad otseselt linkandmeid (sh andmete omanikud ja rakenduste loojad).

Anketeerimine näitas, et teadlikkuse tõstmine on Eestis oluline – 77-st laekunud täidetud ankeedist vastati kahekümne ühes (27%), et vastajal ei olnud teadmisi ühestki küsitletud tehnoloogiast. Nende seas, kes olid mõnest tehnoloogiast teadlikud, oli jaotus tehnoloogiate lõikes küllaltki ühtlane – 9-25% vastajatest ei teadnud vastavast tehnoloogiast midagi, 38-53% vastajatest olid üldisel tasemel teadmised, 20-32% vastajaid hindas oma teadmisi keskmiseks.

Linkandmete alase teadlikkuse tõstmiseks tuleks alustada infopoliitika ja selle rakendusplaanide uuendamisest, koosvõime raamistiku täiendamisest ja juhendite koostamisest. Tuleks käivitada esmased pilootprojektid, mis demonstreerivad linkandmete kasulikkust ja mida saab koolitustel näidetena kasutada. Koolituste esimene etapp peaks keskenduma üldistele (orienteeruva mahuga üks päev) ja süvendatud (tehniline lisakoolitus orienteeruva mahuga kaks päeva) koolitustele. Teavitustöö linkandmetega seotud võimalustest ja ohtudest võiks jälgida häid praktikaid, mis on tekkinud ISKE turvameetmete süsteemi või riigi IT koosvõime raamistiku alase teadlikkuse tõstmise kampaaniate käigus.

Lisaks tuleks välja valida andmete linkimiseks suurema potentsiaaliga andmekogud ning aidata ja motiveerida nende omanikke andmete avamisel ja linkandmete ettevalmistamisel. Samuti on suurema mõjuga selliste andmete avamine või linkimine, mis on ka teistes riikides avatud või lingitud.

Linkandmetega kaasnevad võimalikud ohud ja riskid

Ankeetides hinnati kõige suuremateks andmete vale interpreteerimise, privaatsuse vähenemise ja keerukate sõltuvuste ning andmete muutumise ohtusid. Muuhulgas, linkandmete puhul võib piisata isikuandmetest ühe lingi küljes selleks, et tuvastada ülejäänud andmete seotus vastava isikuga; näiliselt ebaolulised isikulised detailid võivad kombineeritult anda isikust põhjaliku pildi; andmete linkimine võib laiendada võimalusi statistilise üksuse kaudseks tuvastamiseks; isikustatud andmed kombineeritult isikustamata andmetega võivad kaasa tuua isikuandmete kaitse nõuete riive; linkandmed võimaldavad seostada isikuandmeid avaliku sektori andmete ja muude allikate (meedia, sotsiaalmeedia jm) kombineerimisel; oht väärarusaamade, põhjendamatute hinnangute, inimesele suunatud vaenulike kampaaniate jms tekkeks on suurem. Seetõttu tuleb isikuandmete kaitseks tagada, et ükski avalik lingitud andmete komplekt ei sisaldaks isikuandmeid.

Täpsustada sellised isikuandmete kaitse aspektist küsitavad olukorrad, kus ühe õigusakti alusel saab isikukode või muid isikuandmeid avalikustada, AvTS alusel saab seda teavet taaskasutada ning isikuandmete kaitse seaduse alusel ei kohaldata tulemusele isikuandmete seadusega võimaldatavat kaitset. Realiseerida andmekogusid sätestatavates õigusaktides täielikumalt IKS § 11 ja 12 pakutavad andmesubjektid

poolse isikuandmete avalikustamise võimalused. Täpsustada, millal rakendub isikuandmete kaitse seaduse § 17 (automaatsed otsused).

Lisaks õigusruumi täiendamisele nähtub eelnevalt linkandmeid rakendanud riikide kogemusest, et linkandmete loojad vajavad kvaliteetsete andmete loomiseks ja rämpsandmete osakaalu vähendamiseks konkreetseid eeskirjasid ja juhendamist⁸. Nt Ühendkuningriikide ja Belgia kogemus (üleriigilise ja regionaalsete aadressiregistrite andmete ühendamine), FP7 (LOD) ja teised rahvusvahelised linkandmete projektid annavad hea sisendi parimate praktikate näol kvaliteetsete linkandmete loomise, haldamise (sh versioneerimise) jm eeskirjade jaoks.

Rämpsandmete vähendamiseks on vaja kehtestada andmekogude põhimäärustes ja juhendmaterjalides avalikustatavate andmete kvaliteedi nõuded ja määratleda vastutus andmete uuendamise eest. Publitseerida nende nõuete ja nende vastavuse taseme hinnangud vastavas andmekogus või teha nad kättesaadavaks RIHA kaudu. Lisada avaliku teabe seadusesse säte, mis selgelt määratleb vastutuse väliste osapoolte töödeldud andmete ja nende osutatud teenuste kvaliteedi eest.

Vajalikud muudatused õigusruumis

Linkandmete kasutuselevõtt hõlmab endas kokkuleppeid koosvõime neljal tasemel - tehnilisel, semantilisel, organisatoorsel ja õiguslikul tasemel. Muudatused õiguslikul tasemel peavad toetama ka ülejäänud tasemeid ja vastupidi. Näiteks uuring „*Open data: an international comparison of strategies*“⁹ toob välja nn suletud valitsemiskultuuri kui ühe barjääridest andmete avalikuks tegemisel – nimelt ametnikul organisatsioonis on lihtsam andmeid mitte avada kui riskida sellest tulenevate sanktsioonidega. Taoliste barjääride mahavõtmine eeldab muutusi nii organisatsioonides kui õigusruumis. Vajalikud muudatused õigusruumis, sealhulgas piirangud ja reguleerimist vajavad teemad, on kokku võetud allpool.

- Viia AvTS ning andmekogude aluseks olevad vastavusse taaskasutuse printsiipidega ja direktiividega 2003/98/EL ning 2013/37/EL. Kehtestada direktiivi 2013/37/EL järgimiseks vajalikud õigus- ja haldusnormid, muuhulgas täiendades RIHA määrust avaandmete korrastamisega seotud sätetega.
- Vaadata läbi ja vajadusel uuendada riigi andmekogude kohta käivad õigusaktid, kus sätestatakse andmete väljastamise tasumäärad, eriti piirkulusid ületava tasu võtmise osas.
- Sätestada avalikustatava teabe litsentseerimise kord ja litsentsitingimused või litsentsitüübid avaliku teabe seaduses, võimalusel soovitades seejuures kasutada sobivat litsentsi *creative commons* litsentsiperest. Kehtestada andmekogude põhimäärustes ja juhendmaterjalides avalikustatavate andmete kvaliteedi nõuded ning publitseerida nende nõuete ja nende vastavuse taseme hinnangud vastavas andmekogus või teha nad kättesaadavaks RIHA kaudu.
- Sätestada konkreetsete andmekogude kohta käivates õigusaktides, võimalusel ka avaliku teabe seaduses (nt AvTS § 31, § 4 või § 8), millised on õigused päringute tulemusena saadavate andmete publitseerimiseks. Andmete (linkandmetena) publitseerimisel tuleks arvesse võtta riske (sh isikutele, omandile, seadmetele, kriitilisele taristule ja riigi julgeolekule), mis tekivad nende andmete kombineerimisel juba olemasolevate publitseeritud andmetega. Võimalik andmete publitseerimise kitsendamine selliste riskide tõttu peaks siiski olema pigem osaline (enamikes andmekogudes on andmeid, mida saab suurema riskita avalikustada), põhjalikult läbi analüüsitud ning detailselt põhjendatud.

Ettepanekute elluviimise korraldamine

Kuna linkandmete puhul on tegemist on tüüpilise innovatsiooniprotsessiga, mille puhul uue tehnoloogia kuluefektiivne rakendamine eeldab eelnevat mahukat investeerimist, siis tuleks leida poliitiline tugi

⁸ http://events.linkedata.org/ldow2010/papers/ldow2010_paper14.pdf

⁹ http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

linkandmete ja avaandmete tehnoloogiate soodustamiseks. Sellise toe motiveerimiseks võib tugineda rahvusvahelisele praktikale (nt USA-s president Obama poliitiline tugi avaandmetele ja eelmise Ühendkuningriigi peaministri Gordon Browne osalemine andmete avamise poliitikate loomisel) ning avaliku sektori andmete avalikustamisest saadava võimaliku kasu suuruse kaudsetele hinnangutele.

Eraldi struktuuri loomist Eesti valitsusasutuste juurde vaid selleks, et ellu viia Lingitud Eesti suunas viivaid ettepanekuid, ei pea käesolev uuring otstarbekaks, välistamata sellise keskuse võimalikkust laiemate ülesannete täitmiseks. Samas tuleks tähtsustada suunatud strateegiat ja konsensuslikku tegevuskava, viia tegevuskava ametnikeni ning kontrollida selle täitmist, luua valitsuse juurde ava- ja linkandmete töögrupp ning teha rahvusvahelist koostööd ODI-ga.

Andmete avamisel ja lingitaval kujul esitamisel tuleks eristada kolme erinevat tüüpi tegevusi. Spetsiifilised tegevused ava- ja linkandmete loomiseks tuleks läbi viia nende asutuste organisatsiooni ja vahenditega, kes andmeid toodavad ja haldavad, vastavalt õigusaktidele. Eelnevaid tegevusi toetavad ettevõtmised, näiteks koolitused, tuleks finantseerida ja tellida keskselt, nt RIA, RISO või Riigikantselei poolt. Konkreetseid arendusprojekte võib läbi viia nii asutuste vahenditega kui ka keskselt; neid tuleks võimalusel teha laiemate projektide kontekstis, mis viivad asutuste või kodanike jaoks nähtavate kasulike tulemusteni.

Era- ja vabasektori kaasamiseks tuleks kehtestada kõikidele tellitavatele infosüsteemidele koosvõimelisuse, avaandmete ja linkandmete osa kohustuslikkuse ning semantiliste klassifikaatorite ja URI-de olemasolu nõuded, propageerida edukaid linkandmete projekte nii era- kui ka avalikus sektoris ning arendada tarkvara linkandmete esitamiseks ja päringute tegemiseks ning pakkuda Eesti vabasektori ja kodanikkonna kaasamiseks linkandmete alast teavitamist ja koolitamist, huvipakkuvaid rakendusi, suhteliselt lihtsaid tööriistu ja kättesaadavaid linkandmeid. Nende jaotiste, samuti intervjuude ja ankeetide põhjal ei saa prognoosida olulist era- ja vabasektori rahastuse kaasamist Lingitud Eesti suunas viivate ettepanekute elluviimisesse.

Koosvõime arhitektuuri arendamine

Linkandmetega seotud avatud standardid

Uuring „Standardiseerimise majanduslik mõju“¹⁰ toob välja, et ligi 25% iga-aastasest Prantsusmaa SKP kasvust tuleneb standardimisest. Uuringus viidatakse ka analoogilistele tulemustele teiste Euroopa tööstusriikide Saksamaa ning Ühendkuningriigi kohta. Lisaks otsestele tuludele suurendab standardimine ettevõtte väärtust, võimaldab kiiremat innovatsiooni, parandab vastavust regulatsioonidega, muudab tulemused rahvusvaheliselt kasutatavateks ning soodustab toodete ja teenuste kvaliteeti. (Rahvusvahelise) koostöö võimaldamine on standardimise üks olulisemaid tulemusi. Näiteks on Eestis kokku 24003 kehtivat standardit, neist vaid 275 ehk 1,1 % on algupäraseid Eesti standardid (8.01.2014 seisuga). Ülaltoodud põhjustel on investeeritud standardite kasutuselevõttu suhteliselt suure tulususega.

Linkandmete ja semantilise veebi standardite kasutuselevõtmine ei tähendaks olulisi muutusi riigi infosüsteemi arhitektuuris – standardid RDF, RDFa, microdata, OWL, SA-WSDL, URI, URL, HTTP jt on riigi infosüsteemi koosvõime raamistikus¹¹ juba olemas ning piisavalt küpsed kasutuselevõtuks. Pigem on neist osade puhul probleemiks nende rakendamine praktikas, nagu ka linkandmete puhul üldiselt. Kuna linkandmete standardite rakendamine ei ole eesmärk omaette, tuleks seda korraldada seoses linkandmetega seotud projektide, koolituste ja muude algatustega. Ankeetide vastustes märgitakse, et avatud ja lingitud andmed ei ole väärtuslikud „asjana iseeneses“ ning mingi standardi eelistamine ei tähenda ilmtingimata rahalist säästu.

¹⁰ <http://www.iso.org/iso/fr/home/standards/benefitsofstandards/benefits-detail.htm?emid=7>

¹¹ <http://www.riso.ee/et/koosvoime/raamistik>

Tehnilise infrastruktuuri lahendused

Linkandmete tootmiseks on mõistlik võimalusel ära kasutada olemasoleva riigi infosüsteemi arhitektuuri poolt pakutavaid võimalusi. Riikliku Ehitisregistri linkandmete pilootprojektis kasutati näiteks X-tee andmeteenuseid linkandmete komplekti ehitamiseks. Kuna X-tee kaudu on riigi infosüsteemide andmed kättesaadavad, on X-tee kasutamine linkandmete tootmiseks igati otstarbekas. Selline kasutamine võimaldab paremini korraldada ka linkandmete uuendamist. Samas on X-tee andmeteenustekasutamiseks linkandmete loomisel vaja koostada andmeteenuste disainimise metoodika ja teha selle jälgimine kohustuslikuks, läbi Riigi infosüsteemi haldussüsteemi määruse, et uute andmeteenuste loomisel tekiks linkandmete loomiseks vajalikud andmeteenused ning samal ajal pareneks andmeteenuste taaskasutus. Lisaks on vaja luua avatud linkandmete tootmiseks infrastruktuur, kus saaks avaandmeid pakkuvaid X-tee jaoks loodud andmeteenuseid rakendada, sh linkandmete tootmiseks, ilma turvaservereid kasutamata.

Tuleks edasi arendada Eesti avaandmete portaali opendata.riik.ee ning avatud linkandmete portaali, seostades neid teiste andmeallikatega, nagu Eesti geoportaal, Maa-ameti geoportaal, Keskkonnaagentuuri portaal ning Statistikaameti andmebaas. Kuna paljud avaliku sektori asutuste andmed on portaalis registreerimata, tuleks käivitada projekt Eesti avaliku sektori avaandmete loetelu loomiseks ning publitseerimiseks.

Suurema sünergia saavutamiseks nii linkandmete vallas kui seotud tehnoloogiates on mõistlik tellida ideekonkursi kaudu avatud tehnilise pilveplatvormi (PaaS) loomine ja haldamine linkandmete *kasutamise lihtsustamiseks* tehnoloogilistes pilootides (nt linkandmete loomine ja haldamine koosloome kaudu, linkandmete reaalaajaanalüütika infoühiskonna arengukava meetrikate pidevaks mõõtmiseks, valdkonnaüleline linkandmete kasutamine e-äri lahendustes, jm) ja mis uuendaks linkandmeid linkandmete portaalist ning oleks avatud linkandmetele, funktsioonidele ja linkandmete rakendustele. Platvorm peab toetama valdkondlike REST teenuste loomist andmete publitseerijate, firmade või kogukonna poolt nii kommertseesmärkidel kui tasuta kasutamiseks. Avatud tehnilise platvormi jätkusuutlikkuse tagamiseks on vaja eralda eelarvelised vahendid selle organisatsiooni ülalpidamiseks.

Koosloomel põhinevate lahenduste loomine

Riigi seisukohast on koosloome oluline instrument kodanikuühiskonna kujundamisel, kuna see lihtsustab indiviidide panustamist suurte probleemide lahendamisse. Koosloome elluviimiseks on kriitilise tähtsusega mitmeotstarbelise koosloome platvormi (sh liitunud inividid) loomine, mis lihtsustab isikute kaasamist erinevatesse initsiatiividesse. Sellised platvormid eeldavad ühiste mõistete ja probleemihalduse kasutamist, mida saab korraldada nii linkandmete kasutamisega kui ka ilma selleta.

Riik saab läbi viia pilootprojektid koosloome platvormi ja selle näidisrakenduste loomiseks kas siis üldiste konkurssidena või konkreetsete teenuste moderniseerimise eesmärgil (nt seotult julgeoleku või keskkonna seirega). Selliste pilootprojektide raames tuleks üheks eesmärgiks seada sellise platvormi loomine, mida saaks kasutada ka väljaspool avaliku sektori teenuste pakkumist. See lihtsustaks tunduvalt uute koosloome lahenduste loomist.

Selliste konkursside algatamisel tuleks tugineda vähemalt ühele sisukale rakendusele, kus linkandmete kasutamine annaks koosloome protsessile lisaväärtust. Kuna koosloome eelduseks olevate ühiste mõistete ja probleemihalduse kasutamist saab paljudel juhtudel realiseerida ka ilma linkandmeteta, pole sellise rakenduse leidmine triviaalne. Eesti valitsus võiks koondada ava- ja linkandmete kasutajate grupid, kes kasutades koosloome platvormi annaksid riigile nõu, milliseid andmeid avada ja linkida.

Vajalikud semantikaressursid

Linkandmete kasutamiseks tuleks olemasolevaid semantilisi varasid täiendada, ühtlustada ja/või kooskõlastada, eriti valdkonnaüleste rakenduste puhul. Lisaks olemasolevate semantiliste varade

täiendamisele vastavalt kehtivale semantilise koosvõime raamistikule tuleb kaasajastada semantilise koosvõime raamistiku järgmist versiooni linkandmete ja seotud tehnoloogiate vaatevinklist.

Esmajoones tuleks keskenduda juriidilisi ja füüsilisi isikuid, aadresse, kinnistuid ning ruumiandmeid kirjeldavate ontoloogiate loomisele, sest need andmed on esindatud enim riigi infosüsteemides, neil on suurim ühisosa riigi infosüsteemide lõikes ja nad on eelduseks andmete linkimisele. Seoses linkandmete kasutamisega peaks nende üleilmse lingitavuse parandamiseks võtma kasutusele nii palju standardontoloogiaid kui võimalik. Ka Eesti riigi infosüsteemi ontoloogiaid tuleks linkandmete loomise aspektist üle vaadata, eriti just avalike linkandmete loomise nõuetest lähtuvalt. Lisaks tuleks luua rohkem valdkonnaüleseid ontoloogiaid.

Riigi infosüsteemide semantilise koosvõime raamistikus on viimaste esitamiseks ette nähtud valdkonnasõnastikud, klassifikaatorid, ontoloogiaid, nendega seotud veebiteenuste semantilised kirjeldused ja muud semantilised varad, mis tuleb publitseerida RIHA-s. Kuigi RIHA määruses on öeldud, et kõik uued infosüsteemid tuleb varustada semantiliste kirjeldustega, pole selge, kas see on nii läinud ja kas toimub vastav järelevalve. Seda küsimust tuleks eraldi uurida näiteks auditi raames.

Linkandmed avaandmete publitseerimisel

Kuna linkandmete üks põhilisi ideid on võimaluse andmine kolmandatele osapooltele uute teenuste ja rakenduste loomiseks, siis suur osa linkandmete tehnoloogiate kasutuselevõtu raskusest langeb kolmandatele osapooltele. Ankeetides hinnati, et **avaandmete rakenduste loomine kuulub projektide rühma, millel on Eestile kõige suurem mõju.**

Registrites, infosüsteemides ja andmekogudes (sh andmete avaandmetena publitseerimisel) tuleks linkandmete tehnoloogiate kasutusele võtmiseks jälgida W3C linkandmete publitseerimise parimaid praktikaid: modelleerida andmed, nimetada asju püsivate URI-dega, taaskasutada sõnastikke niipalju kui võimalik, avaldada avaandmete kohta nii inim- kui masinloetavad kirjeldused (metaandmed), teisendada andmed RDF-i, määrata sobiv litsents, avalikustada avaandmed linkandmetena ja kuulutada nad välja. Mitteavalike andmete puhul on mõistlik eelkõige lingitavana hoida see osa andmetest, mille juures on näha vajadust kombineerida neid andmetega välistest allikatest.

Andmed tuleks publitseerida vastavalt kehtivatele õigusaktidele, masinloetaval kujul ja avatud litsentsiga, sätestades vastutused andmete kvaliteedi ning nende haldamise osas, järgides riigi infosüsteemi koosvõime raamistikus toodud standardeid andmete ja metaandmete publitseerimiseks, kasutades valitud linkandmete kasutuselevõttu toetavaid tehnilise infrastruktuuri lahendusi ning soovituslikult pakkudes andmete kasutajale andmeid harjumuslikus andmekirjelduskeeles üle kasutaja pöördumiste jaoks suhteliselt lihtsa vahekihi, kasutades sobivaid semantikaressursse andmete metamudelite kirjeldamiseks vastavalt semantilise koosvõime raamistikule, vajadusel olemasolevaid kirjeldusi täiendades ja korrastades.

Sisutöötlus veebisaitides ja -liidestest

Linkandmete kasutuselevõtt veebisaitides ja -liidestest peaks algama uute vajaduste selgitamisest, mida linkandmete kasutuselevõtt rahuldab, eelistest, mida linkandmed seejuures pakuvad ning linkandmete kasutuselevõtu tulu-kulu analüüsist. Eriti tuleks olemasolevate veebisaitide puhul hinnata, kas nende ümberkorraldamise ja linkandmete kasutuselevõtu positiivsed efektid ületavad tehtavaid kulutusi.

Linkandmete kasutuselevõtt veebisaitides ja -liidestest peaks algama RDFa, mikroandmete või *microformats* märgendite lisamisest HTML dokumentidele. Konkreetne märgendite keel ja sõnavara valitakse pragmaatilistest kaalutlustest lähtuvalt. Avaliku sektori veebides on esmajärjekorras oluline märgendada organisatsiooni töötajate kontaktid ja sündmused. Suuremate organisatsioonide puhul on mõistlik selleks laiendada olemasoleva sisuhaldussüsteemi mooduleid, mis töötajate kontakte ja uudiseid veebi väljastavad, selliselt, et väljastatud andmetele lisatakse märgendid automaatselt. KOVide puhul annaks suurima mõju KOVTP lahendusele analoogselt toe lisamine. Otstarbekas on kasutada esitust, mille puhul mitmete

valdkondade jaoks on loodud vastavad metamärgendite komplektid, nt mikroandmete jaoks loodud <http://schema.org>, <http://purl.org/goodrelations> jt. Linkandmete kasutuselevõtul tuleks joondada sisutöötlus LOD2 Stack poolt pakutud linkandmete elutsükliga ning järgida veebide koosvõime raamistiku soovitusi ja W3C linkandmete publitseerimise parimaid praktikaid. Üheks linkandmete kasutamise rakenduseks sisutöötluse vaatevinklist oleks KOV-ide ja asutuste veebidest sisu automaatne kokkukogumine ja esitamine suuremates portaalides, näiteks eesti.ee.

Kirjeldatud nõuete ja tehnoloogiate rakendamiseks tuleks olemasolevate veebisaitide puhul käivitada linkandmete kasutuselevõttu toetavad projektid, mis võivad olla iseseisvad või seotud linkandmete kasutuselevõttu pilootprojektidega. Uute veebisaitide projektide hangetesse tuleks lülitada linkandmete kasutuselevõttu toetavad sisutöötluse nõuded, muuhulgas seoses metaandmete lisamise, semantilise kirjeldamise ja andmete agregeerimisega.

Organisatsioonilise koosvõime arendamine

Linkandmete tehnoloogiate monitoorimine, õpetamine ja propageerimine

Linkandmete, semantilise veebi jt käesolevas uuringus käsitletud tehnoloogiate temaatikat on seni Eesti ülikoolides käsitletud väheste kursuste ja tudengiprojektide raames õppejõudude initsiatiivil. Ülikoolid peaksid seda initsiatiivi võimendama läbi temaatiliste õppe-, koolitus- ja teadusprojektide, mille raames kaasatakse nii eksperte välismaalt kui kohalikke tudengeid. Ka on vaja ülikoolide õppekavasid uute trendide valguses regulaarselt kaasajastada.

Koolituste esimene etapp peaks keskenduma üldistele (orienteeruva mahuga üks päev, koolitatavate suurusjärg viissada inimest ehk orienteeruvalt 20 koolituspäeva) ja süvendatud (tehniline konkreetse probleemi linkandmetega lahendamise lisakoolitus orienteeruva mahuga kaks päeva - suuremate andmekogude haldajatele, kes avaldavad oma andmeid ava- ja linkandmetena, orienteeruvalt 200 inimest ehk kokku 20 süvendatud koolituse päeva) koolitustele.

Erinevad uuringud on üksmeelel selle kohta, et ava- ja linkandmetel on positiivsed mõjud – näiteks, võimaldades pakkuda paremaid teenuseid, luues paremat investeerimiskliimat ja vähendades korruptsiooni. Riik peaks selliseid mõjusid toetama, isegi kui täpsed kulude/tulude uuringud ei ole kättesaadavad. Kindlasti õigustavad pakutud hinnangud ülaltoodud koolituste läbiviimist – nende orienteeruv esimese etapi kogumaht jääb suurusjärku 50 000 eurot aastas.

Linkandmed Eesti IKT sektori lahendustes

Intervjuudest nähtub, et Eesti IKT sektori võimalusena nähakse eelkõige teenuste loomist. Innovatsiooni väga ei usaldada ning kõige olulisemate tegevustena tuuakse välja riigi poolset eestvedamist ning vastutuse võtmist. Erasektori kaasamiseks tuleks riigihangetele lisada ava- ja linkandmeid käsitlevad reeglid, mis on kohustuslikud iga infosüsteemi tellimisel. Ka ankeedivastustes hinnati avaandmete rakenduste mõju (näiteks teenustes) kõrgelt.

Samal ajal näitavad globaalsed trendid kasvavat turgu (suur)andmete analüüsimise lahendustele. Globaalsed trendid viitavad märkimisväärsele ja kasvavale puudusele andmete töötlemise ja analüüsimise oskustega tööjõu osas lähima 5 aasta perspektiivis. Sellest tulenevalt on suurenenud (suur)andmeid kasutavate analüütiliste tööriistade ja SaaS lahenduste tootmine nii globaalses mastaabis. Lingitud ja puhastatud andmete kasutamine lihtsustab taoliste lahenduste loomist. Veelgi enam, kasutades globaalselt aktsepteeritud metamudeleid linkandmetes muutub lihtsamaks nende lahenduste eksportimine teistele turgudele. Eesti IKT sektor saab tõsta oma ekspordivõimekust linkandmetel põhinevat tarkvara tootes ning

avalik sektor saab aidata sellele kaasa nõudes enda poolt tellitud või toetatud toodetes järjepidevalt linkandmete kasutamist.

Linkandmete kasutuselevõtu suurendamisele aitavad kaasa temaatilised teadusarendus-, tootearendus- ja innovatsioonitoetused ja Garage48-tüüpi ettevõtmised. Need võivad aidata kaasa olukorra muutumisele Eesti IKT sektoris, kus on suhteliselt vähe sektorisisest tootearendust ning riskide maandamiseks tehakse üldiselt tööd klientide tellimisel. Temaatilised rakendusuuringud ja tootearendusprogrammid võiksid tagada tootearendusele suunatud projektide jätkusuutlikkuse. Samal ajal saab avalik sektor tellijana suunata IKT sektorit linkandmetesse rohkem panustama, nõudes uutes või jätkuprojektides mõistlikkuse piires linkandmete kasutamist. Erasektor saab tellijana suunata IKT sektorit linkandmeid kasutama, kui avalikult sektorilt vajaminevad andmed on linkandmetena kättesaadavad.

Lisaks tuleks võtta kasutusele meetmed avatud, valdkondlike linkandmete PaaS platvormide loomise ja tööshoidmise või valdkondlike/valdkonnaüleste avatud teenuste olemasolevale avatud linkandmete pilveplatvormile loomise motiveerimiseks (sealhulgas e-äri jaoks nt CRMid, finantsanalüüsi, turundusanalüütika ja toote planeerimise lahendused, mis kasutavad linkandmeid).

Vastavalt avaliku teabe seadusele tuleb avalikustamisele kuuluv teave avalikustada ajakohasena ning viisil ja vormis, mis võimaldab selle allalaadimist masinloetaval kujul. Seepärast ei too täiendav linkandmetena publitseerimise nõue, kui see lisatakse kohe projekti algatamise faasis, kaasa väga suuri lisakulutusi. Innovatsioonitoetuste ja Garage48-tüüpi meetmete maht sõltub pakutavatest toetust vajavatest projektidest ning seda saab hoida valitud piires, näiteks samas suurusjärgus koolituste ja teavitamise mahuga. Arvestades eelmises jaotises toodud linkandmete kasude hinnanguid võib ülalpakutud meetmeid seega hinnata kuluefektiivseiks.

Avaliku sektori ülesanded infosüsteemide tellijana

Intervjuudes rõhutati, et avalik sektor peaks kehtestama kõikidele tellitavatele infosüsteemidele koosvõimelisuse, avaandmete ja linkandmete osa kohustuslikkuse ning semantiliste klassifikaatorite ja URI-de olemasolu nõuded. Nõuete järelevalve peaks struktuurifondidest rahastatavate projektide puhul olema seotud rahade väljamaksmisega. Muude projektide üle võiks järelevalvet korraldada RISO, kes kooskõlastaks projekti rahade väljamaksmist sõltuvalt nõuete täidetusest.

Eriti oluline on ontoloogiade loomine ja uute linkandmete tekitamine, sest linkandmete väärtus kasvab hüppeliselt võrreldes lingitud andmete mahu kasvuga. Seetõttu võiks iga uus projekt, kus andmeid kasutatakse ja/või uusi andmeid luuakse, lisada uusi linkandmeid läbi tekkinud andmete linkimise. Ontoloogiade loomisele on kaasa aidanud tellijapoolsed nõuded täitjatele luua uute infosüsteemide jaoks ka sobivad ontoloogiad. Lisaks on veel teisigi näiteid, nt ehitussektoris, kus tellija nõuete kaudu saab suunata valdkonna arengut.

Avalik sektor peaks suurendama analüütiliste rakenduste osakaalu, kus kasutatakse andmeid ka väljastpoolt nende endi haldusala ja kasu linkandmetest on otseselt näha. Taoliste lahenduste tekkimise soodustamiseks on mõistlik infosüsteemide lähteülesannete püstitustes eraldi välja tuua väljaspool haldusala paiknevate andmete võimalik mõju rakendusele valdkondlike andmekomplektide ja avaandmete kasutamise lõikes. Sel juhul saab infosüsteemide loomise finantseerimise otsuseid tehes arvestada nende valdkondadele mõju.

Soovitusi erasektorile infosüsteemide arendamisel

Intervjuude ja ankeetide põhjal ei ole erasektori esindajad oma rolli hindamisel eriti optimistlikud. Ei usuta, et uued tehnoloogiad toovad suurt edu – innovatsiooni on tehtud, kuid see on enne pooleli jäetud, kui tulemus tekib. Kõige olulisemate tegevustena tuuakse välja riigi poolset eestvedamist ning vastutuse võtmist. Samuti nimetatakse edulugude vajadust, mis võimaldaks näidata innovatiivset probleemilahendust ning mida Euroopasse skaleerida. Erasektori rolli nähakse eelkõige teenuste loomises, taandades avalik sektor

teenuseosutaja rollist. Ka ankeetides on turunduse ja müügi, erasektorile tüüpiliste avaandmete kasutusvõimaluste mõju hinnatud keskmisest madalamana.

Samas on avatud standardite kasutamine ning standardiseerimises osalemine arenenud tööstusriikides andnud olulise osa SKP kasvust. Erasektorile võiks soovitada linkandmete käsitlemist avatud standardina ning nende positiivse mõju arvestamist sarnaselt valdkondlikele metamudelitele, nagu nt SID, HR-XML jne. Tuleks olla kursis linkandmete tehnoloogia arenguga, osaleda koolitustel, jälgida linkandmete edulugusid ja kasutada avaliku sektori poolt publitseeritavaid linkandmeid. Eestis on näiteid linkandmete kasutuselevõttust, näiteks on Riikliku Ehitisregistri lingitud kujul andmed leidnud kasutust kinnisvaraga seotud portaalides.

Kuna erasektori esindajad ei ole oma rolli hindamisel linkandmete kasutuselevõttus eriti optimistlikud, tuleks soovitada riigi poolset eestvedamist ning vastutuse võtmist ning edulugude esiletoomist. Erasektorile võiks soovitada linkandmete käsitlemist avatud standardina ning nende positiivse mõju arvestamist. Tuleks olla kursis linkandmete tehnoloogia arenguga, osaleda koolitustel, jälgida linkandmete edulugusid ja kasutada avaliku sektori poolt publitseeritavaid linkandmeid. Erasektor saab linkandmete tehnoloogiaid oma infosüsteemide tellimisel ja arendamisel kasutada, propageerides edukaid linkandmete projekte nii era- kui ka avalikus sektoris, arendades tarkvara linkandmete esitamiseks ja päringute tegemiseks, pakkudes avaliku sektori andmete publitseerimise ja linkimise teenust ning näidislahendusi, pakkudes ekspordipotentsiaaliga linkandmetel põhinevaid andmeanalüüsi ja pilveteenuseid, kasutades ava- ja linkandmeid teenuste pakkumiseks, rakendades linkandmete tehnoloogiat oma ettevõtte andmete haldamisel (linkandmed ei pruugi olla avaandmed), rakendades linkandmete tehnoloogiat firmade vaheliseks andmete vahetamiseks ning kasutades võimalusi, mida pakuvad innovatsioonitoetused ja Garage48-tüüpi ettevõtmised.

Vabasektor ja kodanikkonna kaasamine

Vabasektor ja kodanikkond on üle maailma olnud linkandmete alal üsna aktiivne. Näiteks on DBpedia Live koosloome initsiatiiv, mille eesmärk on teha Wikipedia andmed linkandmetena kättesaadavateks ja kasutatavateks. Eestis on koosloome läbi vastavate rakenduste leidnud kasutuse isikute poolt linkandmete loomisesse. Näiteks lisatakse veebilehtedele juba praegu märgendeid (nt *microdata* kujul), mida otsimootorid kasutavad linkimiseks, tehes samas saadud andmed otsitulemustes kättesaadavaks.

Üldiselt saab vabasektori ja eraisikute huvi (link)andmete vastu tulla nende seesmistest ajenditest, näiteks soovist teha midagi paremaks, isiklikust huvist, soovist saada tuntuks vms. Linkandmete tehnoloogia on suhteliselt keeruline ja kompetentsi puudumine avaldab siinkohal negatiivset mõju. Inimesed, kes seda tehnoloogiat hästi valdavad, on suure tõenäosusega juba mingis organisatsioonis hõivatud ja tegelevad sellega oma tööülesannete raames. Lisaks on vabasektori ja kodanikkonna jaoks olemas ka palju muid huvitavaid teemasid.

Eelnevast ning intervjuudest nähtub, et Eesti vabasektori ja kodanikkonna kaasamiseks oleks vaja linkandmete alast teavitamist ja koolitamist, huvipakkuvaid rakendusi, suhteliselt lihtsaid tööriistu ja kättesaadavaid linkandmeid. Teavitamist, koolitust ja linkandmete loomist tuleks teha avaliku sektori tööülesannete täitmise ning eelmistes jaotistes toodud algatuste raames, näiteks eraldades koolitustele teatud protsendi kohti vabasektori ja kodanikkonna jaoks. Sellise eraldamise mõistlik maht selgub koolituste käigus ning sellest lähtudes võib arvestada vajalikke lisaressursse.

Lisaks on vaja disainida meetmed selliste valdkondlike ja valdkonnaüleste avatud linkandmete ökosüsteemide loomise või kohandamise võimaldamiseks, kus on toetatud nii avatud kui privaatsete/konfidentsiaalsete linkandmete kooskasutamine ilma andmete privaatsust/konfidentsiaalsust ohustamata. Vastavad ökosüsteemid koondaks erinevaid linkandmete komplekte valdkondlike või valdkonnaüleste probleemide lahendamiseks.

Tööjaotus eri huvipoolte vahel

Kirjanduse analüüs, nt TNO avaandmete strateegiate uuring¹² ning käesolevas uuringus läbi viidud intervjuud toovad välja avaliku sektori juhtiva rolli ava- ja linkandmete kasutuselevõtus. See on ka arusaadav, sest erasektor näeb andmeid kui oma strateegilisi varasid ning on väga ettevaatlik nende avalikustamisega, vabasektor ja kodanikkond omavad aga suhteliselt vähe andmeid. Samuti on näha tugeva poliitilise juhtfiguuri positiivne mõju. Ka see on mõistetav, sest avaliku sektori töötajad ei pruugi iseenesest olla motiveeritud andmete avalikustamiseks.

Toodud üldine rollijaotus kehtib ka Eestis. Avalik sektor on enamasti initsiaator, sätestab õigusakte, soovib või kehtestab standardeid, publitseerib andmeid, arendab esmaseid teenuseid, tellib pilootprojekte, koolitab ja teavitab teisi osapooli. Erasektor, vabasektor ning kodanikkond on pigem järgijate, teenusepakujate ja andmete kasutajate rollis. Toodud üldisest rollijaotusest saab tuletada ka järgneva detailsema tööjaotuse eri huvipoolte vahel Eestis.

Täpsemalt, RISO haldab ja arendab riigi infosüsteemi semantilise koosvõimega seotud dokumentatsiooni, semantika töörühma materjale, semantilise koosvõime raamistikku ning muid riigi infosüsteemi strateegilisi dokumente¹³. Riigi Infosüsteemi Amet koordineerib riigi infosüsteemi arendamist ja haldamist, teostab järelevalvet riigi infosüsteemi haldamist reguleerivatest õigusaktidest tulenevate nõuete täitmise üle ja peab riigi infosüsteemi haldussüsteemi. RISO ja RIA koostöös koordineeritakse metaandmete, sealhulgas valdkondade vaheliste metaandmete loomist, kasutades selleks valdkondade vahelisi töörühmi ning vajadusel hankides uuringuid või pilootprojekte. MKM infoühiskonna teenuste arendamise osakond (ITAO) koordineerib linkandmete kasutamise aspekte, mis on seotud avalike teenuste ning avaliku sektori asjaajamise ja elektroonilise dokumendihalduse arendamisega.

Avaliku sektori (täpsemalt, riigi, kohaliku omavalitsuse või muu avalik-õigusliku juriidilise isiku või avalikke ülesandeid täitva eraõigusliku isiku) andmekogude vastutavad või volitatud töötajad registreerivad andmekogud ning loovad ja registreerivad vastavad semantilised varad riigi infosüsteemi haldussüsteemis. Samuti vastutavad avaliku sektori asutused oma infosüsteemide jaoks vajalike semantikavarade väljatöötamise ning andmete publitseerimise eest vastavalt avaliku teabe seadusele, Vabariigi Valitsuse määrusele „Riigi infosüsteemi haldussüsteem“ ja muudele õigusaktidele.

Erasektor, vabasektor ning kodanikkond hoiab ennast ava- ja linkandmete valdkonnaga kursis, leiab kasulikke, tulusaid ja/või huvitavaid rakendusi, pakub teenuseid ja kasutab andmeid. Avaliku ja erasektori koostööd tuleb igati soodustada, kuid riigihangete seadusele vastavaid PPP projekte linkandmete valdkonnas võib lugeda Eestis vähetöenäolisteks ning suure mahu ja kestvuse tõttu peaksid nendele eelnema koolitused, pilootprojektid ja muud eelmistes jaotistes pakutud tegevused.

Toodud loomulik tööjaotus taaskasutab Eesti juba olemasolevaid organisatsiooni ja protsesse ning ei too kaasa täiendavaid kulutusi lisaks juba eelpool mainitutele.

Andmetöötlusprotsessid ja andmete täieliku elutsükli toetamine

Üldjuhul on linkandmete loomine mitte põhieesmärk, vaid vahend põhieesmärkide saavutamiseks. Andmeid kogutakse ja hallatakse rahuldamiseks mingit vajadust ning saamaks mingit sisulist tulemust. Vajadusel ja võimalusel publitseeritakse need andmed ava- või linkandmetena. Seega peaks kasutatavate andmetöötlusprotsesside valik eelkõige lähtuma andmekogu loomise eesmärkidest. Valitud protsesse täiendatakse linkandmete loomiseks ja kasutamiseks vajalike tegevustega.

Linkandmete andmetöötlusprotsesside eripäraks andmete elutsükli mõttes on andmete semantiline hajusus. Seetõttu on avaliku sektori linkandmete jaoks pakutud erinevaid elutsükli mudeleid. Need pakuvad erinevat elutsükli ulatust (nt kas mudel käsitleb vaid andmeid või ka äriprotsesse), etappide valikut ning etappide organisatsiooni (nt järjestikune või iteratiivne). Ühte kõikide poolt tunnustatud linkandmete elutsükli

¹² http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

¹³ <http://www.riso.ee/et/koosvoime/semantika>

modelit ei ole. Olukord on analoogiline tarkvara arendusega, kus kasutatakse sõltuvalt vajadusest mitmesuguseid elutsükli mudeleid.

Käesoleva töö autorid soovivad Eesti riigi infosüsteemis kasutada linkandmete tootmiseks ja publitseerimiseks iteratiivset elutsükli modelit, mis hõlmab järgmisi etappe: spetsifitseerimine, modelleerimine, genereerimine, publitseerimine ja kasutamine. See on heas kooskõlas nii TNO avaandmete elutsükli mudeliga kui ka W3C soovitustega.

Sissejuhatus

„Lingitud Eesti“ all mõistetakse IKT-lahenduste toel loodud kogu Eesti inforuumi hõlmavat hajutatud isikute, organisatsioonide, tarkade seadmete, teadmiste, infosüsteemide ning andmete koosvõimelist võrgustikku.

Linkandmed on struktureeritud andmed, mis on kirjeldatud ja avaldatud viisil, mis lubab neid automaatselt seostada. Linkandmete tehnoloogia võimaldab seni hajutatult eksisteerivad andmebaasid, sotsiaalmeedia, dokumendid, faktid, seadmed, teadmised jms integreerida ühtseks tervikuks ning avab seeläbi riikidele, era- ja vabasektorile seni võimatud olnud ehk põhimõtteliselt uued võimalused uut tüüpi teenuste loomiseks. Kuid selle võrgustiku loomine eeldab läbimurret mitmetes suundades nagu ülikiire andmesidevõrk ja targad seadmed, koosvõimelised infosüsteemid ja teadmusvõrgustikud, semantilised võrgustikud – kõige olulisemana aga linkandmete kasutuselevõtt.

Linkandmete kontseptsiooni järgi saavad lisaks dokumentidele otselingitavaks ka andmed ise ja kõik füüsilised kui ka abstraktsed „asjad“. Kui praegu saavad brauserite abil teavet lugeda eelkõige inimesed (andmed on inimloetavad), siis kasutades linkandmeid toetavaid tehnoloogiaid, saavad arvutid andmeid lugeda – ning mis olulisemgi, eri kohtades paiknevaid andmeid siduda, nendes otsida ja neid töödelda. Linkandmete tehnoloogia võimaldab loobuda traditsiooniliste andmekogude loomisest ja nende seostamisest: kasutades linkandmete tehnoloogiaid on see funktsionaalsus andmete avaldamisega vaikinisi saavutatav.

Linkandmete kasutuselevõtu initsiatiiv edendab läbipaistvust, koostööd riikide tasandil ja laiemalt, võimaldab luua uusi, uuenduslikke teenuseid ning tõsta otsuste tegemise kvaliteeti.¹⁴

Linkandmed kui tehnoloogiline suundumus on tihedalt seotud selliste ülemaailmsete trendidega nagu tulevikuinternet (*Future Internet, FI*) ja semantiline veeb (*Semantic Web*), asjade Internet (*Internet of Things*) ja suurandmed (*Big Data*), avaandmed (*Open Data*) ja koosloome (*Crowdsourcing*).

Eestis puudub terviklik, Eesti konteksti ja vajadusi arvestav uuring linkandmete kasutusvõimaluste osas, millel võiks põhineda linkandmete kasutuselevõtu poliitika. Käesoleva uuringu tulemused on otseseks sisendiks riigi IKT-poliitika kujundamisele: järgmiste aastate infoühiskonna arengukava rakendusplaanide koostamisele, struktuurifondide rahastamisvoorude ettevalmistamisele ja koosvõime raamistiku dokumentide täiendamisele.

Uuringu eesmärgiks on loodava „Infoühiskonna arengukava 2020“ eesmärkide saavutamiseks vajalike rakendusplaanide jaoks sisendi andmine lingitud andmete loomise ja kasutamise osas ning vastavate tegevuste efektiivseks elluviimiseks vajalike muudatuste tegemiseks Eesti õigusruumis ja poliitilistes meetmetes.

Uuringus pakume välja lähenemise linkandmete tehnoloogiate laiemaks kasutuselevõtuks Eestis, tuues välja puudused, mis pärsivad lingitud andmete efektiivset loomist ja kasutamist. See hõlmab soovitusi avaliku sektori, erasektori ja vabasektori andmekogude, veebisaitide ja muude inforessursside ümberkorraldamiseks ning neid toetavateks poliitikameetmeteks riigi poolt.

¹⁴ Ding, L., Peristeras V., Hausenblas M., "Linked Open Government Data," IEEE Intelligent Systems, vol. 27, no. 3, pp. 11-15, May-June 2012

1 Uuringu metoodika

Käesoleva uuringu metoodika lähtub järgmistest ettenähtud uurimisülesannetest:

- Lingitud Eesti esmase arengustrateegia väljatöötamine.
- Lingitud Eesti koosvõime arhitektuuri arenduse ettepanekute väljatöötamine.
- Lingitud Eestit toetava organisatsioonilise koosvõime arendamise ettepanekute väljatöötamine.

Nimetatud uurimisülesannete läbiviimiseks viisime projekti ellu viies etapis:

- | | |
|-------------|--|
| ▪ I etapp | Ettevalmistused |
| ▪ II etapp | Lähteandmete kogumine ja esialgne analüüs |
| ▪ III etapp | Intervjuud, küsitlused ja andmete analüüs |
| ▪ IV etapp | Esialgsete soovitude koostamine ja aruteluseminar |
| ▪ V etapp | Uuringutulemuse kokkuvõtte koostamine ja Tellijale esitamine |

Uurimisülesandeid käsitlesime paralleelselt läbi kõigi etappide ning iga ülaltoodud etapi kohta esitame kasutatud metoodika.

1.1 I etapp – Ettevalmistus

Eesmärk. Esimene etapp käsitles uuringu eesmärkide ja ulatuse ning uurimismetoodika täpsustamist, andmekogumistegevuste ettevalmistamist ja planeerimist ning intervjuueeritavate ja anketeeritavate valimite moodustamist.

Metoodika. Relevantse dokumentatsiooni kaardistamise käigus kogusime projektimeeskonda kaasatud ekspertide ning tellija abiga kokku projekti tausta puudutava dokumentatsiooni (sh linkandmete ning sellega seonduvate mõistete nt suurandmed, avaandmed, tulevikuinternet jmt) tähenduse, rakendamise hetkeseisu ja tulevikutrende käsitlevate uuringute, statistika, analüüside jmt leidmine) fookusega Eestil ja muul maailmal parimate praktikate väljatoomiseks.

1.2 II etapp – Lähteandmete kogumine ja esialgne analüüs

Eesmärk. Teine etapp oli pühendatud uuringu teostamiseks vajalike alusmaterjalide kogumisele ja nendega tutvumisele ning sobiliku dokumentatsiooni sõelumisele. Esialgse analüüsi eesmärk oli peamiste trendide ja probleemistiku tuvastamine linkandmete ja nendega seonduvate tehnoloogiate valdkonnas.

Metoodika. Uuringu teises etapis analüüsisime temaatilist kirjandust, sh Tellija soovitatud täiendavad materjalid, et tuvastada peamisi trende ja probleemistikku linkandmete ja seonduvate tehnoloogiate valdkonnas. Dokumentatsioonianalüüsil ja praktilistel lahendustel baseerub arhitektuurianalüüs, mille läbiviimiseks kooskõlastasime Tellijaga valimi, mis koosnes kolmest juhtumiuuringust. Dokumentatsioonianalüüsi täiendati kogu projekti kestel.

Dokumentatsioonianalüüsi teostasime kolme fookusega:

- Eesmärgiga mõista valdkonda, kujundada taust – varasema kirjanduse ja statistika analüüsis põhinesime uuringutel ja analüüsidel, mis sisaldavad ülevaateid ja järeldusi linkandmete probleemistiku ja kasutamise ning poliitikate kohta nii Euroopa riikide kui ka teiste riikide tasemel.

- Eesmärgiga anda esialgne hinnang linkandmete rakendamise võimalustele, soodustavatele teguritele, takistavatele asjaoludele ja kaasnevatele riskidele.
- Eesmärgiga kirjeldada parimaid praktikaid.

Arhitektuurianalüüs koosnes järgnevatest etappidest:

- *Analüüsi kaasatavate infosüsteemide valimi koostamine*, mille eesmärk oli tagada, et uuringusse valitakse arhitektuursed lahendused, mis annavad vastused linkandmete arhitektuuriliste küsimuste osas. Uuringus analüüsisime kolme laiemat kasutust leidva lahenduse arhitektuuri. Analüüsitud lahendused on valitud suuremahulisi linkandmete komplekte kasutavate rakenduste seast, mille kohta on piisavalt dokumentatsiooni.
- *Referentssüsteemi valik/loomine infosüsteemide arhitektuuriliste lahenduste juhtumiuuringuks*, mille eesmärk oli analüüsi läbiviimise pidepunkti valik, et analüüs annaks vastused arhitektuurilistele uuringuküsimustele. W3C mustandis *Linked Data Platform 1.0*¹⁵ on toodud välja nõuded linkandmete platvormidele. Linkandmete platvormide osas analüüsisime nende vastavust väljatoodud nõuetele. Lisaks platvormi arhitektuurile hindasime platvormide elujõulisust arvestades nende kasutajate ja arendajate kvantitatiivseid näitajaid, täienduste tegemise intervallide ning kolmandate osapoolte komponente, mida need platvormid toetavad.
- *Juhtumiuuring*, mille eesmärk oli arhitektuuriliste lahenduste hindamine. Juhtumiuuringu (*case study*) puhul on peamiseks analüüsiühikuks üksikjuhtum. Sellisel puhul kogutakse võimalikult palju relevantseid fakte analüüsitava juhtumi kohta ning analüüsitakse neid süvitsi. Juhtumiuuringu tulemused ei ole üldistatavad, kuid võimaldavad mõista fenomeni põhjuseid. Juhtumiuuringu meetodikat kasutasime nii Eesti kui ka rahvusvaheliste linkandmete valdkonna lahenduste edulugude ja ebaõnnestumiste analüüsil konkreetse lahenduse tasandil. Analüüsi läbiviimisel uurisime valimisse kuuluvate lahenduste dokumentatsiooni vastavalt referentssüsteemile.

1.3 III etapp – Intervjuud, küsitlused ja andmete analüüs

Eesmärk. Kolmanda etapi eesmärgiks oli tagada uuringu lõppjäreldest erapooletus, tegelikkusele vastavus ja terviklikkus.

Metoodika. Kolmandas uuringuetapis viisime läbi 8 intervjuud ja 2 fookusgruupiintervjuud valdkondlike ekspertidega (siltrühmade esindajatega, kes on loonud või kavatsevad lähiajal luua linkandmete tehnoloogiatel põhinevaid lahendusi, ning IKT-poliitikakujundajatega). Intervjuudest saadud tulemit kasutasime sisendina fookusgruupiintervjuude läbiviimiseks. Intervjuude ja fookusgruppide tulemusi kasutasime ankeetküsitluse väljatöötamisel.

Individuaalintervjuud. Intervjuu formaadi valikul eelistasime personaalset intervjuud, kuna see formaat võimaldab intervjuueeritaval anda lisainfot, mis ankeetküsitlusele vastates muidu edasi andmata jääks. Intervjuud viisime läbi pool-standardiseeritult, mis tähendab, et intervjuueerijal on ees nimekiri võtmeküsimustega, mida ta saab küsida avatud küsimustena. See annab võimaluse saada vastused põhiküsimustele, kuid võimaldab koguda ka lisainfot, mida ei oleks võimalik koguda struktureeritult. Kokku viisime läbi 9 intervjuud. Intervjuude järgselt koondasime kokku ka intervjuueeritavate poolt viidatud relevantseid lisamaterjalid ning analüüsisime neid.

Fookusgruupiintervjuud. Fookusgruupi eesmärk oli koguda infot linkandmete kasutajatelt ja pakkujalt, kes on loonud või kavatsevad lähiajal luua linkandmete tehnoloogiatel põhinevaid lahendusi, sh uurida võimalusi linkandmete ulatuslikumaks rakendamiseks Eestis (sh põhimõtteline, tehnoloogiline, organisatoorne, juriidiline ja finantsiline valmidus) ning anda ulatuslikum sisend ankeetküsitluse läbiviimiseks. Samuti soovisime teada saada, millised on selles osas võimalikud vajadused, barjäärid ja soodustavad tegurid. Viisime läbi 2 fookusgruupi intervjuud.

¹⁵ <http://www.w3.org/TR/ldp/>

Ankeetküsitlus. Ankeetide sihtrühmad olid organisatsioonide esindajad, kes on kaalunud linkandmete tehnoloogiate kasutuselevõttu või kelle infosüsteemides on kasutuselevõtt otstarbekas. Samuti kaasasime muude organisatsioonide esindajaid, kes kuuluvad ülalnimetatud sihtrühmadesse (muuhulgas organisatsioonide esindajaid, kes on loonud või kavatsevad lähiajal luua linkandmete tehnoloogiatel põhinevaid lahendusi). Ankeetide küsimustiku koostamise üheks aluseks võtsime intervjuude läbiviimisel saadud tagasiside. Samuti võtsime arvesse juhtgrupi töökoosolekul ja vaheseminaril pakutavaid arvamusi ja ettepanekuid ning EL projekti LOD2¹⁶ poolt läbi viidud linkandmete alase uuringu küsimusi. Ankeetküsitluse läbiviimise kohta leiab põhjalikuma informatsiooni jaotises 6.

Esialgsete tulemuste analüüs. Kogutud andmete põhjal viisime läbi esialgse andmete analüüsi, mille käigus tuvastasime peamised linkandmete valdkonna arenguvisionid ja trendid maailmas (sh Eesti kontekstis), esimesed vastused uurimisküsimuste lõikes.

Andmete kvaliteedikontroll. Intervjuude puhul vaatasime läbi intervjuude tulemused ning hindasime nende järeluste erapooletust, kontrollisime pisteliselt kogutud andmestikku, et leida üles vigu andmekogumises või esitamises. Kahtluse korral kontrollisime andmed üle algallikate abiga.

1.4 IV etapp – Uuringutulemuste valideerimine, esialgsete soovitude koostamine

Eesmärk. Neljanda etapi eesmärgiks oli välja töötada järelused ja poliitikasoovitused, mida oleks võimalik kasutada uute suundade seadmiseks linkandmete valdkonnas.

Metoodika. Uuringutulemuste valideerimiseks kasutasime aruteluseminare. Viisime läbi 2 aruteluseminari, millest võtsid osa lisaks sihtrühmade esindajatele ka uuringu juhtrühma liikmed.

Aruteluseminar. Aruteluseminar sarnaneb oma olemuselt fookusgruupiintervjuuga, kus igal osalisel on võimalik avaldada käsitletaval teemal oma arvamust ning esitada täiendavaid küsimusi probleemistiku paremaks avamiseks. Aruteluseminari põhiülesanne käesoleva uuringu raames oli avada osalejatele uuringu tulemusi ning valideerida tulemused arutelu käigus. Aruteluseminari puhul ei eeldata, et iga osaleja peab tingimata omapoolse sisendi andma küsimusi esitades või täiendavaid ettepanekuid tehes, vaid lähtutakse osalejate isiklikust aktiivsusest ja tahtest. Seetõttu ei seata aruteluseminarile osalejate arvu piiranguid nii rangelt kui fookusgruppidele. Arendusdiskussiooni käigus kogusime tagasisidet uuringu peamistele järeldestele ja tulemitale ning testisime järeldusi toetavate argumentide tugevust. Saadud tagasiside sünteesiti kõigi eelnevalt kasutatud metoodikate kasutamisel saadud tulemustega lõpparuandesse.

1.5 V etapp – Uuringutulemuste kokkuvõtte koostamine ja Tellijale esitamine

Eesmärk. Neljanda etapi eesmärgiks on vastata kõikidele uurimisküsimustele ja prioritseerida soovitusel.

Metoodika. Täiendasime uuringu lõpparuannet vastavalt aruteluseminaride käigus saadud ning juhtrühmalt saadud tagasisidele.

¹⁶ <http://lod2.eu>

2 Linkandmete ja seonduvate tehnoloogiate ning praktikate analüüs

Linkandmete ja seonduvate tehnoloogiate ning rahvusvahelise praktika analüüsi eesmärgiks on leida vastuseid käesoleva projekti uurimisküsimustele, mis on seotud kas linkandmete või nendega seonduvate tehnoloogiatrendide kasutuselevõtuga Eestis ning vastavate rahvusvaheliste poliitikate, initsiatiivide ja praktikate kasutamisega Eesti linkandmete strateegia ja poliitika välja töötamisel.

Käesolev tehnoloogiatrendide ja nendega seonduvate rahvusvaheliste poliitikate ning praktikate analüüs on seotud põhiliselt järgmiste uurimisküsimustega:

- *Millised järgnevatest trendidest on Eestile esmaselt olulised linkandmete kasutuselevõtu plaanis: tulevikuinternet, semantiline veeb, asjade Internet, suurandmed, avaandmed, koosloome?*
- *Millistes rahvusvahelistes initsiatiivides peaks Eesti olema aktiivne, sh riiklikul tasemel (prioriteetsuse järjekorras)? Millised kompetentsid ja ressursid peaksid iga sellise initsiatiivi lõikes Eestil osalemiseks olemas olema?*
- *Millised linkandmetega seotud avatud standardid on Eesti jaoks olulised (sh piisavalt küpsed kasutuselevõtuks)?*
- *Kas Eesti peaks looma juurde või täiendama mingeid tehnilise infrastruktuuri lahendusi (nt avaandmete X-tee või avaandmete portaali arendus, andmeanalüüsi töövahendid vm), et toetada linkandmete kasutuselevõttu ja kogu elutsükli protsessi?*
- *Kas ja kuidas korraldada Eestis linkandmete kasutuselevõttuga seotud koosloomel põhinevate lahenduste loomist?*
- *Kuidas korraldada Lingitud Eesti suunas viivate ettepanekute elluviimine? Kas ja milliseid organisatsioonilisi (ümber)korraldusi selle realiseerimiseks on vaja teha, sh riiklikul tasandil?*
- *Millised semantikaressursse Eesti vajab linkandmete tehnoloogia rakendamiseks?*
- *Missugune peaks olema linkandmete kasutuselevõtu puhul andmetöötlusprotsessid, sh andmete täieliku elutsükli toetamine?*

Lisaks sellele on käesoleva analüüsi tulemused abiks ka teistele uurimisküsimustele vastamisel pakkudes rahvusvahelist praktikat ja mõõdet.

Käesolev jaotis annab ülevaate erinevate linkandmetega seotud valdkondade arengust ja hetkeseisust maailmas, analüüsib vastavaid rahvusvahelisi initsiatiive, parimaid praktikaid, initsiatiivide käigus avatud andmeportaale ja standardeid Euroopas ning mujal. Lisaks käsitletakse linkandmete alaseid uuringuid nii strateegiate, elutsükli mudelite, kasutuselevõtu kui mõju osas. Eraldi analüüsitakse UK ja USA poliitikaid ja parimaid praktikaid linkandmete kasutuselevõtul. Iga alamjaotise lõpus tuuakse välja järeldused ja õppetunnid Eesti jaoks ning esitatakse üldised rahvusvahelisest praktikast lähtuvad hüpoteetilised vastused eespool toodud uurimisküsimustele. Nimetatud uurimisküsimuste täielikud vastused, mis võtavad arvesse ka uurimuse jooksul Eestis läbi viidud intervjuude ning kvantitatiivuuringu tulemusi esitatakse peatükis 7 *Vastused uurimisküsimustele*.

2.1 Linkandmete ja nendega seotud tehnoloogiarendide analüüs

Linkandmed kui tehnoloogiline suundumus on tihedalt seotud selliste ülemaailmsete tehnoloogiarendidega nagu tulevikuinternet (*Future Internet*), semantiline veeb (*Semantic Web*), asjade Internet (*Internet of Things*), suurandmed (*Big Data*), avaandmed (*Open Data*) ja koosloome (*Crowdsourcing*).

Paljud teaduse ja tehnoloogiarendide uuringud näitavad, et ülikiire interneti toel on tekkimas tuleviku veeb, mis on lingitud andmete ja (nende kaudu) lingitud seadmete (asjade) veeb, kus on võimalik reaajas agregeerida ja analüüsida ülisuuri andmehulki.^{17,18,19,20,21} Viimased on tekitanud nii inimeste kui seadmete (nt sensorite) poolt. Taolise veebi toimimiseks on tähtis kõigi ülalnimetatud tehnoloogiate areng ja kasutuselevõtt.

Kuigi tehnoloogiarendide analüüsid on toodud käesoleva dokumendi eraldi jaotistes on vaadeldavad tehnoloogiad läbipõimunud. Seetõttu käsitletakse linkandmete tehnoloogia seoseid teiste tehnoloogiatega sellele spetsiaalselt pühendatud jaotises (vt 2.1.8 *Linkandmete seos teiste tehnoloogiatega*).

Iga tehnoloogiarendi analüüsi lõpus esitatakse üldised järeldused projekti „Lingitud Eesti“ jaoks näidates vajadusel ära ka vastava uurimisküsimuse. Neid järeldusi on täpsustatud ja kasutatud peatükis 7 *Vastused uurimisküsimustele* esitatud uurimisküsimustele vastamisel.

2.1.1 Linkandmete tehnoloogia

Definitsioon. Linkandmete tehnoloogia on üks semantilisi tehnoloogiaid. Linkandmed kujutavad endast hajusaid, seotud ja koosvõimelisi andmehulki, mis on esitatud ühtses formaadis (nt RDF), on publitseeritud veebis ja omavad juurdepääsu läbi päringupunktide ehk lüüside (nt SPARQL). Üks kuulsamatest linkandmete printsiipi kasutatavatest projektidest on DBpedia projekt²², mis teeb Wikipediast²³ välja eraldatud info RDF-andmetena veebis kättesaadavaks. Linkandmete tehnoloogia on üles ehitatud olemasolevatele standardsetele veebitehnoloogiatele. Ta pakub minimaalset vajalikku konsensust andmete esitusel kasutades selleks URI-sid²⁴ ja RDF-i²⁵ ning tagades andmetele juurdepääsu läbi HTTP. URI-sid ei kasutata niivõrd inimestele veebilehtede serverimiseks kui pigem informatsiooni jagamiseks sellisel kujul, mis võimaldab arvutitel informatsiooni lugeda. See omakorda võimaldab ühendada erinevatest allikatest tulevaid andmeid ning esitada neile päringuid²⁶.

Tim Berners-Lee esitas linkandmete 4 printsiipi 2006. aastal²⁷. Need printsiibid ei seosta linkandmeid avaandmetega. Taoline seostamine (vt ka 2.1.3 *Avaandmed*) algas koos W3C lingitud avaandmete projektiga ja hiljem võimendus EL LOD2 projektiga²⁸, olles praeguseks ajaks saanud väga levinuks, eriti avalikus sektoris. Seega võivad linkandmed olla nii kinnised (näiteks firmasisesed intranetis olevad andmed jt) kui avatud. Viimaseid käsitleme käesoleva töö avaandmeid käsitlevas jaotises (vt 2.1.3 *Avaandmed*), sest linkandmed on avaandmete kontekstis lihtsalt üks avaandmete esitusviise.

¹⁷ Gartner, Hype Cycle for Big Data, 2012, Published: 31 July 2012

¹⁸ PwC, Spinning a data Web. Making Semantic Web connections, PwC Technologyforecast, Spring 2009

¹⁹ Gartner, Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data. <http://www.gartner.com/it/page.jsp?id=1731916>, Gartner 2012

²⁰ PwC, Internet of Things: Evolving transactions into relationships, PwC Technologyforecast.Issue 1, 2013

²¹ PwC, Harnessing the Power of Crowdsourcing, PwC Technologyforecast, 2011

²² <http://dbpedia.org/>

²³ <http://www.wikipedia.org/>

²⁴ <http://www.w3.org/TR/uri-clarification>

²⁵ <http://www.w3.org/RDF>

²⁶ Bizer, Heath & Berners-Lee, Linked Data—The Story So Far". International Journal on Semantic Web and Information Systems 5 (3): 1–22. doi:10.4018/jswis.2009081901.

²⁷ <http://www.w3.org/DesignIssues/LinkedData.html>

²⁸ <http://lod2.eu/>

Motivatsioon: Linkandmete põhiliseks eeliseks on erinevates kohtades paiknevate andmehulkade agregeerimise võimalus ilma andmeladustamiseta ja nende andmete märksa otsesem ja võimekam analüüs, kui seda lubavad traditsioonilised meetodid. Linkandmete ja eriti avalike linkandmete (LOD)²⁹ paradigma kasutamine on paljutõotav, kui tegemist on erilaadilistest allikatest pärinevate andmete integreerimise vajadusega. Näiteks on võimalik linkida väga erinevates formaatides andmeid nii andmebaasidest, andmeaitadest kui mittestruktureeritud allikatest.

Hetkeseis ja tulevikutrend. Tehnoloogiliselt ja seoses kasutuselevõtuga avalikus sektoris on linkandmete tehnoloogia läbinud mitmeid arengupunkte, muutudes alates 2009. aastast lihtsaks võimaluseks parendada süsteemide vahelist andmevahetust.

Järgnevalt on esitatud linkandmete tehnoloogia tähtsamad arenguetaapid:³⁰

- 2004 – W3C publitseerib RDF soovituset;³¹
- 2006 – Tim Berners-Lee publitseerib nn „5 Stars of Linked Open Data“³² printsiibi;
- 2008 – SPARQL päringukeel RDF publitseeritakse kui W3C standard;
- 2009 – USA ja UK avavad oma avaandmete portaali ja UK esitab esimesed avalike linkandmete näited;
- 2010 – Euroopa 7nda raamprogrammi projektid LOD2 ja LATC, mis on pühendatud avalike linkandmete uurimisele, saavad rahastuse;
- 2011 – W3C loob linkandmete töögrupi;³³
- 2012 – Euroopa Komisjoni avaliku halduse koostöö programmi (ISA) liikmesriikide esindus kiidab heaks avalike linkandmete spetsifikatsiooni;³⁴
- 2013 – W3C avaldas linkandmete platvormi parimad praktikad ja juhised;³⁵
- 2013 – W3C avaldas SPARQL 1.1 soovituset.³⁶

Linkandmete tehnoloogia hetkeseisu iseloomustab W3C aktiivsus nii päringukeele SPARQL uue versiooni kui ka parimate praktikate ja juhiste väljaandmisel 2013. aastal. Lisaks on õnnestunud standardiseerida mõned vajalikud sõnastikud (ontoloogiad) nagu näiteks isiku ja organisatsiooni ontoloogiad (vt 2.2.4 Standardid). Seega võib öelda, et linkandmete tehnoloogilised alused on praeguseks olemas, kuid kasutuselevõttu takistab linkandmete elutsüklit toetavate töövahendite nappus. Linkandmete kasutuselevõtu probleeme analüüsime allpool.

- Linkandmete tulevikutrend on seotud eelkõige semantiliste tehnoloogiate tulevikuga, kuid ka suurandmete ja avaandmete arenguga. Praeguseks momendiks on nimetatud tehnoloogiad ja nende kasutuselevõtt erinevas arengujärgus ja erinevas küpsuse astmes. Gartneri uuringu „Hype Cycle for Big Data“³⁷ järgi on linkandmete tehnoloogiatega seotud tehnoloogiates avaliku sektori avaandmed ja NoSQL andmebaasisüsteemid (siia alla kuuluvad ka RDF andmebaasid) ajaliselt kõige lähemal produktiivsuse platoole (2-5 a). Sellest võib teha järelduse, et avaandmete linkandmetena publitseerimine muutub lähiaastatel aktuaalseks ja kasulikuks. Semantikatehnoloogiad ja asjade

²⁹ <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

³⁰ <https://joinup.ec.europa.eu/community/semic/document/case-study-how-linked-data-transforming-egovernment>

³¹ <http://www.w3.org/RDF>

³² <http://5stardata.info>

³³ <http://www.w3.org/2011/gld>

³⁴ <https://joinup.ec.europa.eu/news/isa-member-state-representatives-endorse-key-specifications-e-government-interoperability>

³⁵ <https://dvc.w3.org/hg/ldpwg/raw-file/default/ldp-bp/ldp-bp.html#bib-LDP-PRIMER>

³⁶ <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>

³⁷ Gartner, Hype Cycle for Big Data, 2012, Published: 31 July 2012

internet on sellest rohkem kui 10 aasta kaugusel. Semantikatehnoloogiatest, mis toetavad lingitud avaandmete kasutuselevõttu, tuleb eraldi esile tõsta ontoloogiate haldust ja tehnikaid³⁸, mis mängivad olulist rolli linkandmete arenduses ja kasutamises.

Gartneri tippennustused aastaks 2014³⁹ toovad linkandmete tehnoloogiaga seoses esile asjade interneti kui inimeste ja arvutite vahelise silla loomise vahendi ning neuroarvutuse (*deep neural network technologies*). Neurovõrkudel baseeruv arvutus võib saada üheks linkandmete tehnoloogia trendi jätkutehnoloogiaks, vastavaid ajuuuringute projekte käsitleme jaotises 2.1.7 *Asjade internet*.

Gartner ennustab, et neuroarvutuste tehnoloogia kasutuselevõtu tulemusena juba 2017. aastal 10% arvutitest pigem õpivad kui töötlevad informatsiooni. See saab olema revolutsiooniline, sest arvutite programmeerimise asemel tuleb neid hakata õpetama.

Kasutuselevõtt. Linkandmete tehnoloogiat kasutatakse tavaliselt kui informatsioonilist infrastruktuuri andmete integreerimiseks erinevatest allikatest. Võrreldes peavoolu moodustavate relatsiooniliste andmebaasisüsteemide tehnoloogiaga on linkandmete tehnoloogial eelised just seal, kus relatsioonilised andmebaasid on aeglasel ja ebaefektiivsel, s.o mingi objekti kohta käivate andmete ühendamisel (ehk ühendi operatsioon). Peale selle võimaldab linkandmete tehnoloogia edukalt automaatselt ühendada RDF andmebaasi andmeid andmebaasi väliste andmetega (näiteks veebiandmetega, sensorandmetega jms). Samas on RDF andmebaasid praegu üldjuhtudel (va ühendi operatsioon) aeglasemad kui kommertsiaalsed relatsioonilised andmebaasid kuigi leidub häid kommertsiaalseid RDF andmebaase (AllegroGraph, Big OWLIM, Marklogic jt) ja ka avatud koodiga RDF andmebaasisüsteeme nagu Virtuoso⁴⁰, 4store, Bigdata jt.⁴¹

Võrreldes veebiteenuste tehnoloogiaga, mis ka võimaldab andmeintegratsiooni teenuseid luua, on linkandmete tehnoloogia kergem, paindlikum, kiirem ja seega odavam. Kuna linkandmete tehnoloogia kasutab juba olemasolevaid veebistandardeid (ja põhiliselt vabavaralisi RDF andmebaase), siis on selle kasutuselevõtu võimalus ka paljudel väikestel firmadel ja üksiküritajatel, kel ei pruugi olla näiteks vahendeid veebiteenuste-põhise andmeintegratsiooni lahenduse ehitamiseks.

LOD2 projekti materjalide järgi on ainult 5% andmete veebist lingitud, mis tähendab, et loodud linkandmete kogumid pole omavahel ühendatud linkide kaudu. Selle põhjuseks näeme üldkasutatavate ja avalike taksonoomiate (ontoloogiate) puudumist. W3C on soovitanud mõned üldised vabad standardiseeritud ontoloogiad nagu isiku ja organisatsiooni ontoloogiad (vt 2.2.4 *Standardid*), mida saab kasutada teatud linkandmete kogumite linkimiseks aga seda on väga vähe ja lingid on seega väga piiratud. Parem seis on linkandmeid kasutavate infosüsteemide sees, sest seal luuakse andmete integreerimiseks ja linkimiseks vajalik ontoloogia süsteemi ülesehitamisel. Teisalt on raske luua standardiseeritud üldisi ontoloogiaid, isegi valdkondlike ontoloogiate loomine on osutunud väga vaevaliseks. Põhjuseks on töö keerukus, teadmus- ja ajamahukus. Meie arvates pole ilmselt reaalne varustada nn andmete veeb vajalike ontoloogiatega selleks, et linkida enamus andmete veebi andmetest. Seda probleemi püütakse lahendada ontoloogiate masinõppe meetodite kasutamisega. Paraku on ka selles valdkonnas veel pikk tee usaldusväärsete masinate poolt loodud ontoloogiate loomiseni. Tulevikus võibki ehk näha Gartneri poolt ennustatud neuroarvutuse võidukäiku, mis asendab või täiendab praegust linkandmete tehnoloogiat.

Linkandmed on oma olemuselt avatud andmed ja seega on avaliku sektori avaandmed ideaalne testikeskkond linkandmete tehnoloogiale. Käesoleva töö autorite arvates on see ka üheks avaandmete linkandmetena esitamise ja linkimise initsiatiivide motivatsiooniks. Kuna linkandmete laiem kasutuselevõtt on seotud nn käivitusprobleemidega (st on vaja kriitilist massi linkandmeid, et saaks tekkida nende eksponentsiaalne kasv), siis on selline linkandmete loomise poliitika arusaadav.

Linkandmete tehnoloogia võimaldab integreerida ettevõtte andmeid vastavalt selle väärtusahelale. Sõnastikel (ontoloogiatel) põhinev järjepidev andmete linkimine vähendab oluliselt andmete integreerimise kulusid

³⁸ Gómez-Pérez, Asunción, Mariano Fernández-López and Oscar Corcho. 2004. Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer Heidelberg.

³⁹ Gartner Top Predictions 2014: Plan for a Disruptive, but Constructive Future, Gartner 2013

⁴⁰ <http://virtuoso.openlinksw.com/rdf-quad-store/>

⁴¹ <http://www.slideshare.net/lod2project/the-semantic-data-web-sren-auer-university-of-leipzig>

ettevõtetes ja loob eeldused rakenduste tegemiseks. Linkandmete tehnoloogia sobib hästi meediarakenduste loomiseks. Positiivne näide on selles vallas BBC.

Linkandmete kasutuselevõtuks on tähtis toetada linkandmete elutsükli mudelit. Erinevates projektides ja erinevate autorite poolt on välja pakutud erinevaid linkandmete elutsükli mudeleid. Näiteks on EL LOD2 projektis loodud suhteliselt keeruline elutsükli mudel, mille iga etappi projekt proovib automatiseerida niipalju kui võimalik ja varustada vastava tarkvara toega. Lihtsam ja mitte niivõrd automatiseerimisele rõhuv linkandmete elutsükli mudel on pakutud välja nii W3C⁴² kui ka B. Villazon-Terrazas jt poolt⁴³.

Kaasajal kasutatakse linkandmete tehnoloogiat nii era- kui riigisektoris.

Erasektoris kasutavad firmad nii kinniseid linkandmeid kui ka võimalust täiendada oma kinniseid linkandmeid avaandmetega või avalike linkandmetega või neist andmekaeve abil saadud sobilike andmetega. Selliste nõuete täitmiseks on linkandmete tehnoloogia sobiv. Üks esimesi linkandmete kasutajaid oli BBC, kes kasutas linkandmeid olümpiaprogrammi koostamisel. Finantssektorist võib nimetada firmat Garlik⁴⁴, kes kasutab ekstensiivselt kinniseid eriliselt turvatud linkandmeid. IBM ja Pearson kasutavad kinniseid linkandmeid. Sindicetech⁴⁵ aitab firmadel luua oma linkandmete pilvi, mis on kinnised. Näiteks on nende klientideks kirjastus Elsevier, ravimifirma AstraZeneca jt.⁴⁶ USA firma MarkLogic⁴⁷ pakub oma klientidele, kelle hulgas on palju erafirmasid, linkandmete tehnoloogial põhinevaid lahendusi.

Paljud firmad kombineerivad oma kinniseid (link)andmeid avaandmetega. Näiteks Fujitsu Euroopa kombineerib avatud ja kinniseid andmeid oma sensoritel põhinevate tervishoiu alaste rakenduste loomisel.⁴⁸

Kuigi paljud erasektori ettevõtted maailmas on juba võtnud linkandmete tehnoloogia kasutusele kui efektiivse andmete integratsiooni infrastruktuuri, leidub ka pessimistlikumaid arvamusi, mis tunnistavad, et IT firmad pole linkandmete tehnoloogiat nii hästi vastu võtnud kui suurandmete tehnoloogiat⁴⁹. IBM, Microsoft jt ei ole linkandmete tehnoloogiast seni eriti huvitatud olnud, sest see on omaette tehnoloogiline platvorm ja IT firmad ei näe seetõttu võimalusi tulusaks koostööks oma toodetega. Siiski Oracle pakub RDF andmebaasi lahendust⁵⁰. Samas on IBM jt integreerinud oma tehnoloogiatega sellised suurandmete töötlemise vabavaravahendid nagu Hadoop ja MapReduce. Teisalt on tekkinud palju spetsiaalseid linkandmete tehnoloogiat pakkuvaid firmasid nagu näiteks MarkLogic, OpenLinks Software⁵¹, Franz⁵² jt.

Peale selle publitseerivad paljud firmad oma RDF andmeid veebis. Näiteks Facebook, The New York Times, Bestbuy.com, Tesco, Volkswagen, BBC, jt⁵³. Seda võib ka käsitleda kui linkandmete ärist kasutust andes võimaluse näiteks idufirmadele, kes soovivad töödelda (lingitud) veebiandmeid ja alustada oma äri. Samas kasutavad veebis saadaval olevaid linkandmeid oma äriks ka veebiotsingut pakuvad firmad nagu Google.

Andmebaasi tehnoloogia ja linkandmete tehnoloogia ei võistle omavahel, sest nad on mõeldud lahendama erinevaid ülesandeid. Andmebaasitehnoloogial on endiselt ettevõtete infosüsteemides oluline osa. Linkandmete tehnoloogia on mõeldud integreerima andmeid erinevatest (ka sõltumatutest) allikatest ja esitama neile päringuid. Oma linkandmeid veebis publitseerides antakse võimalus nende taasintegreerimiseks. RDF oli algselt mõeldud kui metaandmete formaat ja selliselt esitatud metaandmeid pidi kasutatama dokumentide jm anoteerimiseks. Kaasajal on RDF muutunud universaalseks andmeformaadiks, kuid leidub küllalt palju rakendusi, kus teda kasutatakse ikkagi tema algselt planeeritud eesmärgil s.o metaandmete esitamiseks ja linkimiseks.

⁴² <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

⁴³ Villazón-Terrazas, B. et al., Methodological Guidelines for Publishing Government Linked Data. In D. Wood, ed. Linking Government Data. Springer

⁴⁴ <http://www.garlik.com/>

⁴⁵ <http://www.sindicetech.com/>

⁴⁶ http://semanticweb.com/sindicetech-helps-enterprises-build-private-linked-data-clouds_b30454

⁴⁷ <http://www.marklogic.com>

⁴⁸ Open Data on The Web, 2013, Report, <http://www.w3.org/2013/04/odw/report#closed>

⁴⁹ <http://blog.semantic-web.at/2013/06/04/theres-money-in-linked-data/>

⁵⁰ <http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/rdfsemantic-graph-1902016.html>

⁵¹ <http://www.openlinksw.com/>

⁵² <http://www.franz.com/>

⁵³ <http://answers.semanticweb.com/questions/14029/are-semantic-technologies-ready-for-commercial-use>

Riigisektoris kasutatakse linkandmete tehnoloogiat nii kinniste kui avaandmete või nende kombinatsiooni baasil töötavate infosüsteemide loomiseks. Näiteks, kinniseid linkandmeid kasutatakse julgeoleku ja riigikaitse valdkondades. Ka USA tervishoiu osaliselt kinnise portaali www.healthcare.gov hiljutisel uuendamisel kasutati linkandmete tehnoloogiat⁵⁴. Seal kasutati firma MarkLogic lahendust. MarkLogic pakub linkandmete, semantiliste ning suurandmete tehnoloogiate lahendusi ka valitsuste missioonikriitiliste süsteemide loomiseks.⁵⁵

Ettevõtetele kasulike avalike linkandmete näiteks võib tuua OpenCorporates⁵⁶ linkandmete pilve, mis sisaldab andmeid 50000 ettevõtte kohta üle maailma. Teine näide on Product Ontology ⁵⁷, mis klassifitseerib ja sisaldab infot üle 1 miljoni toote kohta.

USA statistikaamet näiteks konverteeris osa 2000. aasta andmetest RDF-i saades kokku üle miljoni RDF kolmiku⁵⁸. Nende andmete hulgas oli rahvastiku statistika andmed erinevate geograafiliste tasandite lõikes (osariigid, linnad jne). Nad avasid ka vastava SPARQL päringupunkti (löösi) ⁵⁹. Kaasajal on loodud portaali nimega TheDataWeb⁶⁰, mis võimaldab kergesti leida statistilisi andmeid ja mis kasutab andmete publitseerimiseks ja integreerimiseks muu hulgas ka linkandmete tehnoloogiat.

Seoses avaandmetega on viimastel aastatel avalike avaandmete linkandmetena esitamine ja linkimine teiste andmetega muutunud aktuaalseks ja trendikaks. Selles seoses viis *PwC EU Services ISA* programmi raames läbi uuringu avaliku sektori lingitud avaandmete ärimudelite alal.⁶¹

Uuringu eesmärgiks oli uurida avaliku sektori lingitud avaandmete väärtust äri- ja kodanikele ja avaliku sektori administratsioonile; nende esitamise kulude struktuure ja tarbimise kasumivooge; võimalusi ja takistusi seoses nende andmete väärtuste loomisega. Uuring esitab avaliku sektori lingitud avaandmete ökosüsteemi analüüsimise teoreetilise raamistiku. Töö toob esile 37 kasutusjuhtumit, kus avaliku sektori administratsioonid on kasutanud avaliku sektori lingitud avaandmeid selleks, et pakkuda oma andmeid kui veebiteenust. Kasutusjuhte analüüsitakse uuringus esitatud raamistiku aspektide valguses. 14 juhtumit 37-st valiti välja põhjalikumaks analüüsiks.

Uuring jõudis järgmistele järeldustele: Avalikud linkandmed liiguvad edasi tehnoloogia kasutuselevõtu elutsüklis olles näiteks raamatukogunduse/arhiivinduse alal jõudnud juba varajase kasutuselevõtu faasi. Avalike linkandmete pakkumine välistele taaskasutajatele ei ole vaadeldud linkandmete loojate kasutusjuhtumites esmatähtis eesmärk, pigem kasutatakse linkandmeid sisemise andmeintegratsiooni efektiivsuse tõstmisel või andmete integreerimiseks oma partnerorganisatsioonidega.

Avalike linkandmete pakkumist nähakse kui avalikku ülesannet, mida finantseeritakse avaliku sektori rahadega ja mõnede grantidega väljastpoolt ja mille tulemusena loodud andmed on tasuta vabalt saadaval. Uuringus vaadeldud juhtumites ei esinenud linkandmete kasutamist kolmandate osapoolte poolt uute teenuste loomiseks. Sellest teevad nimetatud uuringu autorid järelduse, et linkandmete taaskasutus kolmandate osapoolte poolt on alles innovaatorite faasis.

Lõppjäreldusena tuuakse uuringus välja, et kuna põhilised andmete pakkujad on tootmas suures mahus avalikke linkandmeid, siis on ette näha, et ka nende andmete taaskasutajad leiavad lähitulevikus oma võimaluse innovatsiooniks. See järeldus on korrelatsioonis eelpool toodud Gartneri uuringu tulemustega avaandmete ja semantiliste tehnoloogiate trendide osas ja meie järeldusega nende trendide koosmõjust.

Käesolev uuring käsitleb avalikke linkandmeid põhjalikumalt jaotistes *2.1.3 Avaandmed* ja *2.1.8 Linkandmete seos teiste tehnoloogiatega*.

⁵⁴ http://www.nytimes.com/2013/11/23/us/politics/tension-and-woes-before-health-website-crash.html?_r=0

⁵⁵ <http://www.marklogic.com/solutions/government/>

⁵⁶ <http://opencorporates.com/>

⁵⁷ <http://www.productontology.org/>

⁵⁸ <http://datahub.io/dataset/2000-us-census-rdf>

⁵⁹ <http://www.rdfabout.com/demo/census/sparql.xpd>

⁶⁰ <http://thedataweb.rm.census.gov/TDW.html>

⁶¹ Phil Archer, Makx Dekkers, Stijn Goedertier, Nikolaos Loutas, Study on Business Models for Linked Open Government Data - BM4LOGD, PwC EU Services, <https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Soovitused ja järeldused Eestile. Arvestades lühiajalist (2-5 a) linkandmete arengu trendi koos avaliku sektori avaandmete trendiga, võib teha järelduse, et avalikule sektorile muutub just avaandmete linkandmetena publitseerimine aktuaalseks ja kasulikuks. Seega peaks Eesti avalik sektor nimetatud trendiga kaasa minema.

Rahvusvaheliste linkandmete trendide põhjal võib Eestile pakkuda järgmisi soovitusi. Näiteks võiks REGREL projektis kasutada erinevate andmete integreerimisel linkandmete tehnoloogiat (nagu näiteks tegi USA Statistikaamet). Eeskuju võiks võtta ka DRI⁶² Iiri kohanimede linkandmetena esitamise projektist⁶³ ja ka Eesti kohta midagi sellist teha. Selline kohanimede esitamine masinloetaval kujul võimaldab kõigil huvigruppidel seda vabalt kasutada oma rakenduste või teadustööde jms tegemisel.

Eesti firmades kasutatakse praegu andmete integratsiooniks ja analüüsiks põhiliselt andmeaitade tehnoloogiat. Teatud juhtudel on selle asemel linkandmete tehnoloogia rakendamine ilmselt efektiivsem, paindlikum ja odavam.

Võrreldes veebiteenuste tehnoloogiaga on linkandmete tehnoloogia paindlikum, lihtsam standardite poolest (st kasutab ainult tavapärasest HTTP-d ja lisaks RDF-i ning SPARQL-i) ja odavam. Probleemiks on aga linkimist (koosvõimet) toetavate sõnastike (ontoloogiate) olemasolu. See on üleilmne probleem nagu eespool juba mainitud. Eesti peaks linkandmete loomisel kasutama niipalju kui võimalik üleilmseid standardontoloogiaid (nii valdkondlikke kui W3C standardeid). See tagaks meie linkandmete parema lingitavuse teiste poolt loodud linkandmete kogumitega.

B. Villazon-Terrazas jt^{64, 65} pakuvad oma kogemuste põhjal avalike linkandmete elutsükliks välja järgmisel joonisel (vt *Joonis 1*) toodud iteratiivse elutsükli mudeli, mida käesoleva töö autorid soovivad ka Eesti riigi IS-s kasutada linkandmete tootmiseks ja publitseerimiseks, sest see mudel on heas kooskõlas nii TNO avaandmete elutsükli mudeliga kui ka W3C soovitustega.



Joonis 1. Avalike linkandmete elutsükli mudel

Peatükis 7.3.7 vastatakse linkandmete elutsükliga seotud uurimisküsimustele selle soovitatud mudeli alusel.

Linkandmetel baseeruvate rakenduste loomise ärimudelina võiks Eestis kasutada taimekaitsevahendite pilootprojekti⁶⁶ ärimudeli kavandit taimekaitsevahendite linkandmete infrastruktuuri haldamiseks. Selles

⁶² <http://www.dri.ie/projects>

⁶³ <http://apps.dri.ie/locationLODer/>

⁶⁴ http://delicias.dia.fi.upm.es/wiki/images/7/7a/07_MGLD.pdf

⁶⁵ Villazón-Terrazas, B. et al., Methodological Guidelines for Publishing Government Linked Data. In D. Wood, ed. Linking Government Data. Springer

⁶⁶ Nikolaos Loutas, Christophe Colas, Stijn Goedertier, Linking data about applications and decisions for authorisation of plant protection products, PwC EU Services, 2013, https://joinup.ec.europa.eu/asset/core_business/document/linking-data-about-applications-and-decisions-authorisation-ppp

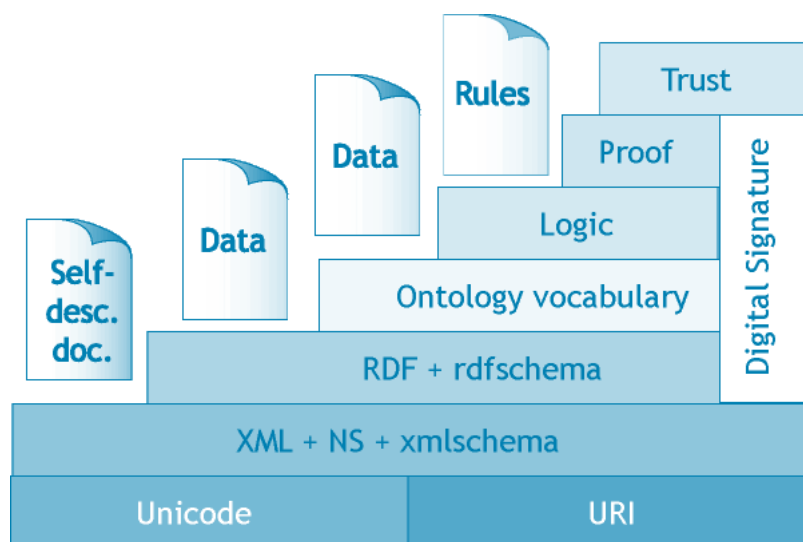
toodi välja tulevase ärimudeli järgmised komponendid, mis on piisavalt üldised taaskasutamiseks ka meil Eestis:

- Andmete pakkumuse ja nõudluse identifitseerimine (Andmete pakkumuse osas on riigid üle minemas vaikimisi avaandmete pakkumusele. Siiski mingi konkreetse rakenduse ärimudeli loomisel on oluline andmete nõudluse identifitseerimine);
- Kulumudel;
- Hinnamudel ja litsentsi raamistik;
- Rahastamise mudel;
- Kasualanaliis;
- Riskialaliis.

2.1.2 Semantilised tehnoloogiad

Definitsioon. Semantilised tehnoloogiad võimaldavad lisada andmetele tähendust olemasoleva sõnastiku (või ontoloogia) baasil või andmete tähendust andmetest automaatselt välja eraldada ning esitada andmeid koos nende tähendustega masinloetaval kujul. Põhilised kasutusküpsed tehnoloogiad on järgmised: ressursside identifikaatorid (*Uniform Resource Identifiers*-URI), ressursside kirjeldamiseks kasutatavad keeled (*The Resource Description Framework*-RDF⁶⁷, RDF/XML ja RDF Schema-RDFS⁶⁸) ning ontoloogiatega (mõistete süsteemide või metaandmete) semantilise kirjeldamise keel *Web Ontology Language*-OWL⁶⁹. RDF on graafipõhine andmete esitus, mis on defineeritud W3C poolt koos SPARQL⁷⁰ päringukeele ja ontoloogiatega kirjelduskeele OWL-iga, moodustamaks aluse semantilisele andmete veebile. Viimasega seondub linkandmete mõiste.

Semantilisi tehnoloogiaid vaadeldakse tihti koos ja kihtidena nagu näha järgneval joonisel (vt *Joonis 2*).



Joonis 2. Semantilised tehnoloogiad⁷¹

⁶⁷ <http://www.w3.org/RDF/>

⁶⁸ <http://www.w3.org/TR/rdf-schema/>

⁶⁹ <http://www.w3.org/TR/owl-features/>

⁷⁰ <http://www.w3.org/TR/rdf-sparql-query/>

⁷¹ Allikas: T. Berners Lee <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-o.html>

Jooniselt on näha, et eespool vaadeldud linkandmete tehnoloogia aluseks olev RDF keel on mõeldud eelkõige andmevahetuseks. RDF on paindlik ja efektiivne, kuid nõuab ontoloogiate olemasolu, et defineerida ja klassifitseerida olemeid ja nendevahelisi seoseid. Ontoloogiate kirjeldamise keel *The Web Ontology Language* (OWL) baseerub RDF-il, andes ette sõnastiku kirjeldamiseks objekte (olemeid), nende klasse ja seoseid. See on vajalik tuletamiseks uusi fakte olemasolevate faktide baasil. Nii nagu inimsuhtluse korralgi, võib ka semantilise veebi korral sõnastik tekkida vastavalt kasutusvajadusele. See võib kirjeldada mingit kasutusvaldkonda või mingeid üldisi seoseid andmete või teabe vahel. Näiteks on loodud *The Simple Knowledge Organisation System* (SKOS), mis baseerub ka RDF-il, kuid on suunatud just hierarhilise teabe esitamiseks. Lingitud andmetele (RDF andmetele) päringute esitamiseks kasutatakse päringukeelt SPARQL, mis on praeguseks saanud *de-facto* standardiks (vt ka 2.2.4 Standardid).

Motivatsioon. Masinate võimetus aru saada veebilehtede sisust viis mitmete seda kitsendust ületada proovivate initsiatiivideni. Algselt oli semantilise veebi eesmärk veebilehtede annoteerimine metaatribuutidega selleks, et võimaldada masinatel osaliselt interpreteerida tekstis sisalduvat teadmust ja paigutada see teatud konteksti. Mikroformaadid ja RDFa (*Resource Description Framework in attributes*) kannavad edasi veebilehtede annoteerimise ideed. Märghendamist on õnnestunud ainult osaliselt kasutada – näiteks schema.org standardsõnastikku kasutavad Google ja teised otsimootorid. Ka ontoloogiate kui veel keerukamate metaandmete kasutamine veebilehtede annoteerimiseks pole üldiselt vilja kandnud. Põhjus on töö keerukus ja ajamahukus.

Annoteerimist kasutavad lähenemisviisid proovivad parendada inimeste jaoks loodud veebilehtede masinloetavust. Selline lähenemine on piiratud, sest annoteerida saab ainult veebilehte ennast ja tema teatud elemente. Peale veebilehtede on kaasajal olemas suur hulk struktureeritud andmehulki, mis sisaldavad väga palju infot. Selleks, et hõivata veebi ka need andmehulgad tekkis idee nn andmete veebist⁷². Kui need andmehulgad oleksid semantiliselt kirjeldatud ja lingitud, siis saaksid masinad sellise veebi andmehulgad läbi käia ja nendest infot koguda sõltumata veebilehtedest.

Hetkeseis ja tulevikutrend. Semantiliste tehnoloogiate hetkeseis on sarnane linkandmete tehnoloogiate seisuga, sest viimased moodustavad suure osa semantilistest tehnoloogiatest (vt 2.1.1 *Linkandmete tehnoloogia*). Gartneri eelpool mainitud ennustuste järgi on semantiliste tehnoloogiate produktiivse kasutuselevõtu viimased 10 aastat.

Gartner tõi 2013. aasta ja järgnevatel aastatel trendide ennustuses välja 10 trendi, milles semantilised tehnoloogiad olid kolmandal kohal ja NoSQL andmebaasid (sh RDF ja graafandmebaasid) 5ndal kohal⁷³.

Gartner selgitab semantiliste tehnoloogiate tähtsust järgmiselt: semantiliste tehnoloogiate tähtsus tõuseb seoses uute ärinõuetega informatsiooni kui strateegilise vara käibele laskmiseks. Suurandmete töötlus ja analüüs nõuavad semantilisi tehnoloogiasid selleks, et inimesed saaksid aru andmetest, või selleks, et automatiseerida otsustusprotsessi.

Ontotext⁷⁴ on üks tuntud semantilisi tehnoloogiasid pakkuv firma, kelle poolt pakutud lahendused said 2012. aastal kiire kasutuselevõtu osalisteks sellistes sektorites nagu publitseerimine, farmaatsia, e-riik, töökohtade vahendamine ja kultuuripärand.

Esile võiks tõsta metaandmete üha suurenevat tähtsust, seda just koosvõimeliste süsteemide ehitamisel. Metaandmeid esitatakse masinloetavate ontoloogiatena. Ontoloogiate alal võib nimetada kolme trendi:

- Üleilmselt aktsepteeritud standardontoloogiate loomine. See suund on seotud linkandmete kasutuselevõutuga.
- Ontoloogiakeelte ja standardite lihtsustumine (näiteks SKOS, palju kasutatakse linkandmete kontekstis ka lihtsalt RDFSi).

⁷² Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American, 284(5), 28-37 (2001)

⁷³ <http://www.ontotext.com/news/gartner-predicts-exciting-2013-for-semantic-tech>

⁷⁴ www.ontotext.com

- Ontoloogiade loomine masinõppe meetodeid kasutades. See suund püüab lahendada ontoloogiade loomise teadmuse- ja töömahukuse probleemi asendades inimeksperdid masinõppega.

Kasutuselevõtt. Semantiliste tehnoloogiade kasutuselevõtt on juba vaikselt alanud, sest kasutusele on võetud linkandmed ja nende linkimiseks vajalikud ontoloogiad (vt 2.1.1 *Linkandmete tehnoloogia*). Metaandmetest (eriti ontoloogia kujul) on saanud omaette intellektuaalne vara ja kaitse objekt.

Samas leidub juba ka positiivseid näiteid, kus semantikatehnoloogiade kasutamine süsteemide koosvõime tagamisel on andnud majanduslikku kasu. Näiteks, Wal-Mart teatas, et nende 2012. aastal kasutusele võetud e-ärisüsteemi semantilise koosvõime lahendus (semantiline toodete ostimootor ja soovitaja) andis juba paari kuuga 15% müügis kasvu⁷⁵.

Semantilisi tehnoloogiaid kasutavad edukalt ka sellised firmad nagu Google, IBM ja Facebook⁷⁶. Facebook-i avatud graaf on hea näide semantilistest metaandmetest. Google'i teadmiste graaf (Knowledge Graph) loodi 2012. aastal ja semantilisi tehnoloogiaid kasutatakse seal objektide vaheliste seoste leidmisel, toodete ja teenuste vaheliste linkide loomisel jms. Google kasutab graafi ka otsingutulemuste personaliseerimiseks, kasutades semantilist märgendamist standardse schema.org sõnastiku alusel. Ka Bing, Yahoo ja Yandex kasutavad schema.org standardit. IBM kasutab semantilisi tehnoloogiaid arvamuste analüüsiks ja ennustamiseks.

Eesti hetkeseis ja järelused. Eestis on teatud tasemel eeltöö tehtud selleks, et kasutusele võtta eespool mainitud semantikatehnoloogiad, eriti linkandmed. Näiteks on olemas riigi infosüsteemi ontoloogiade loomise⁷⁷ ja semantilise kirjeldamise⁷⁸ meetodid. On loodud nõuded riigi infosüsteemi ontoloogiatele⁷⁹. Ontoloogiade loomise meetodikat⁸⁰ on levitatud ka EL vastavates initsiatiivides. RIHA repositooriumis on juba mitmete riigi IS-ide ontoloogiad (näiteks Äriregistri, ADS ontoloogiad jt). Eesti osaleb EL eSENS projektis⁸¹. Üheks teemaks selles projektis on semantikatehnoloogiade kasutamine selleks, et anda ühine tähendus avalikus sektoris kasutatavatele definitsioonidele. Konkreetsemalt on projekti eesmärgiks töötada välja ja piloteerida EL piiriüleste e-teenuste taristu – nii taristu arhitektuur kontseptuaalsel tasandil kui ka lahendused, mille abil saaks piiriüleseid teenuseid realselt kasutada.

Lisaks võib nimetada veel riigi infosüsteemide koosvõime raamistikke, s.o veebide⁸², infoturbe, tarkvara ja Eesti IT koosvõime raamistikke⁸³, milledest viimane näeb ette semantikatehnoloogiade arendamise ja kasutuselevõtu lähimate aastate jooksul. Veebide koosvõime raamistikus on käsitletud linkandmete teemat veebiportaali aspektist.

Lähtuvalt ülalpool toodud Eesti praegusest tasemest, tuleks semantikaressursside loomisel keskenduda vajalike ontoloogiade loomisele, sest need on eelduseks andmete linkimisele. Seoses linkandmete kasutamisega peaks nende üleilmse lingitavuse parandamiseks võtma kasutusele niipalju standardontoloogiaid kui võimalik. Ka Eesti riigi infosüsteemi ontoloogiaid tuleks linkandmete loomise aspektist üle vaadata, eriti just avalike linkandmete loomise nõuetest lähtuvalt. Lisaks tuleks luua rohkem valdkonnaüleseid ontoloogiaid.

⁷⁵ [http://www.computerworld.com/s/article/9230801/Walmart rolls out semantic search engine sees business boost](http://www.computerworld.com/s/article/9230801/Walmart_rolls_out_semantic_search_engine_sees_business_boost)

⁷⁶ <http://www.cmswire.com/cms/customer-experience/semantic-technology-the-future-of-search-for-digital-marketers-022209.php>

⁷⁷ Haav, H.-M. 2010. Ontoloogiade loomise meetodika. (A methodology for ontology development) http://www.ria.ee/public/RIHA/ontoloogiade_loomise_metoodika_v4.PDF (accessed February 03, 2012). (in Estonian).

⁷⁸ Küngas, P. 2010. Semantilise kirjeldamise juhised vo.4. (Guidelines for semantic enrichment of data and services) http://ftp.ria.ee/pub/riha/Semantilise_kirjeldamise_juhis_vo4.pdf (accessed February 01, 2013). (in Estonian).

⁷⁹ Haav, H.-M. 2010. Nõuded RIHA ontoloogiatele. (Requirements for ontologies in RIHA) http://www.ria.ee/public/RIHA/nouded_riha_ontoloogiatele_r26.PDF (accessed February 03, 2013). (in Estonian).

⁸⁰ Haav, H.-M. 2011. A practical methodology for development of a network of e-government domain ontologies. In Building the e-World Ecosystem: 11th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2011, Revised Selected Papers, ed. T. Skersys, R. Butleris, L. Nemuraite, and R. Suomi, 113, Springer Heidelberg.

⁸¹ Electronic Simple European Networked Services, <http://www.esens.eu/home.html>

⁸² <http://www.riso.ee/et/koosvoime/veebide-raamistik.odt>

⁸³ <http://www.riso.ee/et/koosvoime/raamistik>

Kuigi RIHA määruses on öeldud, et kõik uued infosüsteemid tuleb varustada vastavate ontoloogiatega, siis pole selge, kas see on nii läinud ja kas toimub vastav järelevalve. Seda küsimust tuleks eraldi uurida näiteks auditi raames.

2.1.3 Avaandmed

Definitsioon. Avaandmete (*Open Data*) all mõistetakse kõigile vabalt ja avalikult kasutamiseks antud masinloetavas formaadis andmeid, millel puuduvad kasutamist ning levitamist takistavad piirangud. Avaandmed on tasuta kättesaadavad (lubatud on võtta tasu kopeerimiskulude katmiseks) kõigile isikutele mistahes kasutuseesmärgil, digitaalsed, masinloetavad ja riskasutatavad teiste andmetega, litsentseeritud kitsendusteta kasutamiseks ja edasilevitamiseks. Seega pole avaandmed otseselt tõusev tehnoloogiarend, vaid lihtsalt andmed, mis on masinloetaval kujul avalikuks tehtud.

Avaandmed võivad olla esitatud linkandmetena (*Linked Open Data*). Avalikud linkandmed on avalikud avaandmed, mis on publitseeritud vastavalt linkandmete printsiipidele.

Vastavalt UK valitsuse dokumendile *Open Data whitepaper – Unleashing the Potential*⁸⁴, peavad avalikud avaandmed vastama järgmistele tingimustele:

- Andmed peavad olema veebis kättesaadavad ilma identiteedile ja kavatsustele seatavate piiranguteta ning maksimaalselt nende paljundamise hinna eest.
- Andmed peavad olema digitaalses masinloetavas vormingus, võimaldades koosvõimet teiste andmetega.
- Andmete kasutuslitsents ei tohi piirata andmete kasutamist ja taaslevitamist.

Motivatsioon. Riigiasutused üle maailma koguvad väärtuslikku andmestikku, mida oleks võimalik kasutada nii äriettevõtete kui kodanike poolt. Seoses nende andmetike ökosüsteemide mitmekesisuse ja keerukusega ei ole kodanikel ega äridel kerge ja odav leida, linkida ja taaskasutada riigi poolt kogutud andmeid (isegi nende enda kohta kogutud andmeid).

Riigisektori andmete avalikustamine oli poliitiline otsus (USA, seejärel EL) selleks, et suurendada riigivalitsemise läbipaistvust (toetada nn avatud valitsemise mudelit) ja aktiveerida majandust andes äri sektorile võimaluse pakkuda avalike andmete baasil loodud teenuseid kodanikkonnale.

2009. aastat võib pidada avaliku sektori andmete avamise aastaks, sest USA ja seejärel UK avasid oma riigisektori andmed vabalt taaskasutamiseks spetsiaalses avaandmete portaalis (data.gov ja data.gov.uk). Pärast seda on paljud riigid, eriti EL-i riigid, avanud riigisektori andmed (näiteks transpordi, statistika, geo- ja keskkonnaandmed jm) vabaks kasutamiseks (vt näiteks <http://open-data.europa.eu/>, <http://publicdata.eu>, CensusHub⁸⁵). Näiteks, USA-s on loodud lisaks avaandmete veebivärvale (data.gov) veel palju valdkondlikke avaandmete veebivärvaid energietika, hariduse, avaliku julgeoleku, tervishoiu jms alal.⁸⁶ Paljud EL riigid on järginud EL-i initsiatiivi ja Euroopa Liidu direktiivi avaliku sektori informatsiooni taaskasutuse alal (nn PSI-direktiivi⁸⁷) ning teinud oma andmed avalikuks.⁸⁸ Suurbritannia rakendas seda direktiivi juba 2005. aastal ja seal on avaandmete ökosüsteem hästi arenenud (vt näiteks <http://data.gov.uk>).

⁸⁴ <http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential>

⁸⁵ CensusHub 2013. https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Census_Hub

⁸⁶ B. Obama, Technology, <http://www.whitehouse.gov/issues/technology>, 2011

⁸⁷ PSI-directive. 2003. DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information, http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf

⁸⁸ <http://open-data.europa.eu/>

TNO poolt 2011. aastal läbi viidud uuring USA, Inglismaa, Taani, Hispaania ja Austraalia avaliku sektori strateegiate osas⁸⁹ näitas, et tähtsaimad motiveerivad jõud on teiste eesrindlike maade strateegia ja kogemused, poliitiline juhtimine, regionaalsed, kodanike ja äri initsiatiivid. Alles seejärel omab tähtsust tehnoloogiline areng ja õiguslikud regulatsioonid. Põhilisteks tõketeks on suletud valitsemise struktuur, privaatsusega seotud õigusaktid, andmete kvaliteedi probleemid, mittepiisav kasutajamugavus ja info üleküllus ning avaliku sektori andmete avamisel.

Hetkeseis ja tulevikutrend. Euroopa avaliku sektori infotribüün ePSIplatform⁹⁰ publitseerib igakuiseid raporteid avatud valitsuse ja avaliku sektori erinevate teemade kohta⁹¹. Platform publitseerib ka avaliku sektori edinemise tablood⁹², millelt saab võrrelda avaliku sektori edenemist erinevates riikides. Üldise tulemuse järgi on hetkel 5 esimest: Inglismaa, Hispaania, Holland, Austria ja Iirimaa. Viis viimast on Horvaatia, Eesti, Bulgaaria, Malta, Slovakkia.

*Open Data Barometer*⁹³ on avaliku sektori uurimisvõrgustiku⁹⁴ projekt, mis uurib avaliku sektori globaalseid trende, initsiatiive, rakendusi ja mõju. Avaliku sektori baromeetri uuring⁹⁵ viidi läbi avaliku sektori avaliku sektori osas 77 riigis üle maailma ja selle tulemused esitati 31. oktoobril 2013. a.

Uuring keskendus avaliku sektori avaliku sektori valmisolekule, kasutuselevõtule ja ilmnevatele mõjudele kasutades ekspertküsitlusi, avatud andmehulki ja teiseid indikaatoreid vaadeldavates maades. Allpool toome lühikokkuvõtte selle uuringu tulemustest.

Avaliku sektori poliitika on viimastel aastatel kiirelt levinud ja jõudnud 55% vaadeldud 77-st riigist. Samas on avaliku sektori initsiatiividel väga erinevad vormid alates üksikutest avaliku sektori portaalist kuni üldriiklike avaliku sektori rakendusteni. Seega on avaliku sektori valdkonna arengul veel palju potentsiaali, sest vaatamata poliitika levikule on avaliku sektori kättesaadavus väga madalal tasemel. Vähem kui 7% vaadeldud andmehulkadest olid kättesaadavad masinloetavas vormingus ja avatud litsentsiga. Seetõttu on kasutajatel väga raske töödelda ja kasutada avaliku sektori andmeid ning äri sektoril puudub kindlus, et avaliku sektori kasutamine ärilisel eesmärgil on legaalne.

Uuringu tulemuste järgi on 6 esimest eesrindlikku riiki UK, USA, Rootsi, Uus-Meremaa ning Taani ja Norra (mõlemad 5ndal kohal).

Uuringu järgi on vaadeldud maade hulgas kõige vähem avatud andmehulgad maaregister ja äriregister, mis näitab, et avaliku sektori avaliku sektori initsiatiivid ei taga poliitiliselt tähtsate andmehulkade avamist. Enamuses riikides ei ole need andmehulgad veel avatud või kui on, siis pole nad standardises vormingus. Isegi sellises standardiseeritud valdkonnas nagu transport on uuringu järgi ainult 25% riikidest need andmed avatud masinloetaval kujul. Isegi kaardiandmed on tihti saadaval mittedigitaalses vormingus või digitaalne vorming on saadaval eraldi tasu eest. Kuigi statistilised andmed peaksid tõenäoliselt olema kättesaadavad veebis, on nad tihti avatud liiga agregeeritud vormis või on nende litsents piirav.

Uuring ei tuvastanud kindlaid tõendeid avaliku sektori avaliku sektori mõju osas. Uuring käsitles 6 tüüpi mõju, kuid keskmine mõju indeks 10-st oli 1,7. Uuringus tõdeti, et kuigi avaliku sektori visioon on tekkinud, nõuab tema reaalsusse viimine palju pingutusi nii õiguslikul, sotsiaalsel, majanduslikul, tehnilisel, organisatsioonilisel kui poliitilisel tasandil.

Samas pole avaliku sektori reeglina lingitud, vaid kasutatakse traditsioonilisi lahendusi ja andmeformaate (andmebaasid, pdf, xls, csv, xml jms), mis ei võimalda andmeid automaatselt linkida.

⁸⁹ Noor Huijboom, Tijs Van den Broek, Open data: an international comparison of strategies, European Journal of ePractice, www.epracticejournal.eu N° 12 March/April 2011

⁹⁰ <http://epsiplatform.eu/>

⁹¹ <http://epsiplatform.eu/topicreports>

⁹² <http://epsiplatform.eu/content/european-psi-scoreboard>

⁹³ <http://www.opendataresearch.org/barometer>

⁹⁴ Open Data Research Network, <http://www.opendataresearch.org/>

⁹⁵ Tim Davies, Open data barometer, 2013 Global Report, 31. oktoober, 2013, ODI and Web Index, <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>

Seoses semantiliste tehnoloogiate arenguga ja eriti lingitud andmete tehnoloogia kasutatavuse tõusuga (vastavad standardid, arendusvahendid jms on juba olemas) on riigid asunud viimastel aastatel oma avaandmeid esitama lingitud andmete kujul. Eesrindlikud on selles osas UK (data.gov.uk/linked-data) ja USA (data.gov).

Kaasajal on avalike linkandmete paradigma laialdaselt aktsepteeritud paljude riikide valitsuste poolt. Avaliku sektori andmete publitseerimine lingitud avaandmetena võimaldab neid andmeid (taaskasutada nii kodanikeühenduste, avaliku kui erasektori organisatsioonide poolt ja seega vähendada andmete integratsiooni kulusid.

Avaandmete linkandmeteks üleviimisel on võimalik järgida sammhaaval ülemineku strateegiat vastavalt Tim Berners-Lee nn 5 tärni printsiibile.⁹⁶ Avalike linkandmete ökosüsteem on avatud veebipõhine süsteem, mis ühendab nii andmete tootjad (näiteks avalik sektor), andmete töötlemise teenuse pakkujad (IT äri) kui ka andmete kasutajad (äri ja kodanikud). Näiteks L. Ding eristab oma artiklis⁹⁷ 3 järgmist avalike linkandmete andmetöötluste taset (märgitud rohelise alaga):

- **Avamise tase**, kus avalik sektor publitseerib veebis oma andmed taaskasutatavas vormingus ja haldab tsentraalset andmekataloogi aitamaks kodanikel vajalikke andmeid leida.
- **Linkimise tase**, kus huvigrupid (näiteks tööstus ja akadeemia) aitavad parandada avatud andmete kvaliteeti luues ontoloogiaid, sõnastikke, linke jm kasutades selleks nii inimtööjõudu kui automaatseid süsteeme.
- **Taaskasutuse tase**, kus arendajad koondavad publitseeritud andmestikud kokku selleks, et luua kõrge väärtusega rakendusi (nt *mashups*). Tulevikus, andmete turu arenedes peaks selline vabatahtlik andmetele lisaväärtuse lisamine saama kasumlikuks ärisektoriks.

Vaatleme järgnevalt 3 avaandmete tulevikku mõjutavat tegurit: riikide ja riikide ühenduste poliitilised otsused, digitaalse äri arengutendentsid ja sotsiaalsed trendid.

Esimeseks avaandmete tulevikku mõjutavaks teguriks on riikide ja riikide ühenduste poliitilised otsused. Näiteks, avaandmete mahtu ja andmete kasutuselevõttu hakkab tulevikus tugevasti mõjutama G8 avaandmete harta (G8 Open Data Charter), mis publitseeriti koos tehniliste lisadega 12. juunil 2013. aastal.⁹⁸

Harta esitab 5 printsiipi, mille järgimises G8 riigid on kokku leppinud ja mis on järgmised:

- vaikimisi avaandmed;
- andmete kvaliteet ja kvantiteet;
- andmed peaksid olema kasutatavad kõigile;
- andmete vabastamine paremaks valitsemiseks;
- andmete vabastamine uuenduslikele tehnoloogiatele.

G8 riigid leppisid kokku, et nende eesmärkide saavutamiseks koostavad G8 riigid siseriiklikud tegevuskavad, kontsentreerivad oma tähelepanu kõrge kasutusväärtusega valdkondadele ja lepivad kokku metaandmete vastavustabelites. G8 riigid toovad oma dokumendis välja 14 prioriteetset valdkonda, millest esimesed on statistikaandmed, ruumiandmed, valimisandmed ja eelarveteave.

⁹⁶ 5 Stars of Linked Open Data, <http://5stardata.info>

⁹⁷ Li Ding Qualcomm, Vassilios Peristeras and Michael Hausenblas. Linked Open Government Data, IEEE Intelligent Systems, May/June 2012 (Vol. 27, No. 3) pp. 11-15 1541-1672/12/© 2012 IEEE
<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>

⁹⁸ <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

Teiseks mõjuteguriks on digitaalse äri arengutendentsid. Digitaalse äri valdkonnas ennustab Gartner⁹⁹, et nii firmad kui valitsused loobuvad 2020. aastaks 75% oma praegu sensitiivsete andmete kaitsest ja annavad need andmed avalikuks kasutamiseks. Põhjuseks on see, et kõigi andmete kaitse on mitte ainult ebareaalne, vaid ka mittevajalik, sest muutub arusaam, mis on sensitiivne andmestik ja mis seda pole.

Kolmandaks mõjuteguriks on üha kasvav sotsiaalne trend avatuse suunas, mis hõlmab nii kodanikke, kes ise jagavad oma isiklikke andmeid meelsasti teistega, kui ka ettevõtteid ja valitsusi, kes jagavad oma andmeid kodanikega. Gartneri uurimuse järgi pole paljud andmed, mida praegu peetakse tundlikeks, seda enam aastal 2020. Vabanevad ressursid suunatakse väikese hulga väga sensitiivsete andmete väga heatasemeliseks kaitseks.

Kasutuselevõtt. Eespool toodud analüüsist ja uuringutest järeldub, et paljud riigid on aktiivsed avaandmete loomises, eriti Euroopas ja Ameerikas. Praeguseni on põhirõhk olnud avaliku sektori andmete avamises ja nende masinloetavas vormingus esitamisel. Vähem on tegeletud avaandmete linkimisega. Alles aastatel 2012 ja 2013 on selles osas toimunud murrang ning paljudes riikides on hakatud looma linkandmetel põhinevaid rakendusi ja pakkuma linkandmehulki avatud litsentsi alusel.

Lisaks on mitmed riigid läbi viinud võistlusi saamaks avaandmete taaskasutuse baasil tehtud rakendusi. Näiteks, USAs oli võistlus „Apps for Democracy“, mis kestis 30 päeva ja mille ROI-d hinnatakse 4 000%. Norras oli initsiatiiv *Nettskap 2.0*, mille tulemusel loodi 135 avaandmete rakendust.¹⁰⁰

Osa valitsusi panustavad kodanike ja riigiametnike ning avaliku sektori asutuste vahelisele koostööle, kasutades koosloome ja koostöö vahendeid. Näiteks Hollandis loodi vastav riigiametnike veebivõrgustik *Ambetnaar 2.0*.¹⁰¹ EL-i *Data Challenge* võistlus 2011. a päädis 430 eri tüüpi tulemusega alates ideedest, rakendustest ja lõpetades andmehulkade ja visualiseerimisega.¹⁰²

Capgemini Grupi¹⁰³ avaandmete ja selle majandusliku mõju analüüs pakub välja, et peamised avaandmete initsiatiivide trendide loojad on UK, USA, Prantsusmaa, Kanada ja Austraalia.

EPSIPlatform viis 2013. a läbi uuringu avaliku sektori avaandmete võimalikust mõjust ja selle aspektidest.¹⁰⁴

Uuringu tulemusena sai selgeks, et avaandmete mõjust arusaamine, selle struktureerimine ja mõõtmine on olnud väljakutse paljudele valitsustele ja organisatsioonidele. Mõju-uuring on keerukas ja praeguseks pole loodud häid võrdlemist võimaldavaid meetodikaid. Seetõttu on vajalik viia läbi uurimistöid avaandmete potentsiaalse mõju struktureerimiseks ja mõõtmiseks.

Nimetatud aruanne vaatleb 3 mõju: majanduslikku, poliitilist ja sotsiaalset. Majanduslike mõjudena vaadeldakse uute töökohtade rajamise potentsiaali, uute teenuste ja toodete loomist, teadusmajanduse kasvu, avalike teenuste efektiivsuse kasvu ja seotud turgude kasvu. Poliitiliste mõjudena käsitletakse läbipaistvust ja vastutavust, kodanikuosalust, poliitilist teadlikkust ja informatsioonile juurdepääsu. Sotsiaalse mõju faktoritena nähakse suuremat kaasatust ja aktiivset osalemist, juurdepääsu infole ja kodanikuosalust.

Töö toob välja ka võtmeindikaatorid mõõtmaks avaandmete taaskasutuse kõiki kolme liiki mõjusid ja pakub välja, kuidas neid indikaatoreid üldiselt mõõta.

⁹⁹ Gartner Top Predictions 2014: Plan for a Disruptive, but Constructive Future, Gartner 2013

¹⁰⁰ Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing. <http://dx.doi.org/10.1787/5k46bj4f03s7-en>, <http://www.oecd-ilibrary.org/docserver/download/5k46bj4f03s7.pdf?expires=1385822932&id=id&accname=guest&checksum=62792E49ED3863A2CF9B8E9046597CB7>

¹⁰¹ <http://ambtenaar20.ning.com/page/welkom-op-ambtenaar-20>

¹⁰² Open Data Challenge Website <http://opendatachallenge.org/>

¹⁰³ Capgemini Group, www.campegmini.com

¹⁰⁴ Karolis Granickas, Understanding the Impact of Releasing and Re-using Open Government Data, European Public Sector Information Platform, Topic Report No. 2013 / 08, http://www.epsiplatform.eu/sites/default/files/2013-08-Open_Data_Impact.pdf

Siiski tehakse uuringus järeldus, et avaandmete taaskasutuse mõjude hindamine ja sellest arusaamine jäävad rasketeks ülesanneteks. Vastuste leidmine nendele rasketele küsimustele nõuab paljude siht- ja huvigruppide koostööd ning pingutusi. Seega soovitatakse edendada avaandmete mõju-uuringuid ja mõju haldamist selleks, et luua efektiivsemaid avaandmete poliitikaid ja jätkusuutlikumaid avaandmete taaskasutamise püüdlusi.

OECD uuring „*Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*”, mis avaldati 2013. a mais¹⁰⁵, analüüsib avaliku sektori avaandmete mõisteid, initsiatiive, nende initsiatiivide rakendamise printsiipe ja õppetunde.

Allpool on esitatud selles uuringus välja toodud mõned avaandmete kasutuselevõtu õppetunnid ja soovitusel.

Strateegia. Paljudel riikidel puudub avaandmete strateegia, kuigi see võimaldaks kindlaks määrata, milliseid andmehulki avada ja mis eesmärgil. Samuti näitaks strateegia viisid, kuidas valitsus planeerib mõjutada turgu avaandmeid kasutama.

Avaandmete portaaliid ja andmetele juurdepääs. Paljud valitsused annavad praegu kõrge prioriteedi avaandmete portaali arendusele, jättes tahaplaanile avalike andmete avamise tehnilise infrastruktuuri arendamise. Seetõttu võib avaandmete tõeline väärtus jääda varju ja kättesaamatuks. Rahvuslike avaandmete portaaliid aadressil tehtud kriitika on seotud valitsuste huviga esitada avaandmeid portaalis teatud kindlal viisil, mis piirab andmete potentsiaalsetel kasutajatel avaandmete kasutamist oma eesmärgil.

Topeltnaksumustamine. Üks tõsisemaid probleeme on asjaolu, et avaliku sektori avaandmete taaskasutamine võib kehtestada topeltnaksumustamise. Üheltpoolt kehtib printsiip, et maksumaksja raha eest loodud avaliku sektori informatsioon on alati tasuta kättesaadav (ainus erand on selle info kopeerimise ja edastamise tasu). Teisalt aga litsentside tasud PSI Direktiivi all välja lastud andmete eest võivad ulatuda miljonitesse eurodesse. Aastal 2009 tehtud EL uuring¹⁰⁶ leidis, et näiteks 10 ruutkilomeetrise maatüki aerofoto maksumus võib keskmiselt olla 0-292 euro piires. Tõsine rakendus aga nõuab palju selliseid fotosid. Sama uuring näitas, et 27 geoinfo valdajat 24-st EL riigist said taoliste toodete müügi eest 2007. a tulu 356 miljonit eurot. Seega tundub, et osa avaliku sektori andmete kättesaamine on kallis.

Eesti hetkeseis ja järeldused. Eestis rakendus PSI-direktiiv 1. aprillil 2013. Juba varem oli loodud Eesti avaandmete portaal, opendata.riik.ee, ja avatud linkandmete portaal opendata.riik.ee/en/linked-data. Positiivsetest näidetest väärib märkimist Eesti geoportaal¹⁰⁷ ja Maa-ameti geoportaal¹⁰⁸, mille kaudu avalikustatakse ja tehakse kättesaadavaks Eesti riigi ja kohalike omavalitsuste ning teiste avalik-õiguslike juriidiliste isikute haldusalas olevad ruumiandmed. Keskkonnaagentuuri portaalis¹⁰⁹ on avalikult kättesaadavad 11 registri andmed. Statistikaameti andmebaas¹¹⁰ esitab riiklikku statistikat. Registreeritud kasutajal on võimalik päringu tulemusena saada andmeid alla laadida Exceli, CSV-, XML- jm vormingutes. Eestis on valminud ka lingitud avaandmete pilootprojekt (Ehitisregistri avaandmed). Eelkõige on tegemist ühe konkreetse tehnilise fookusega projektiga X-tee teenuste ja RDF andmebaasi kasutamise uurimise osas. Peale selle on T. Tammeti meeskonna poolt loodud põhimõttel opereriv NoSQL andmebaasi C teek nimetusega WhiteDB¹¹¹, millel on RDF tugi.

¹⁰⁵ Ubaldi, B. (2013), “Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives”, OECD Working Papers on Public Governance, No. 22, OECD Publishing. <http://dx.doi.org/10.1787/5k46bj4f03s7-en>

¹⁰⁶ MICUS Management Consulting report on – Assessment of the Re-use of Public Sector Information (PSI) in the Geographical Information, Meteorological Information and Legal Information Sectors, published March 2009, www.micus.de/pdf/MICUS-Studie_PSI_EU_March_2009.pdf

¹⁰⁷ <http://inspire.maaamet.ee/>

¹⁰⁸ <http://geoportaal.maaamet.ee>

¹⁰⁹ <http://www.keskkonnainfo.ee/>

¹¹⁰ <http://pub.stat.ee/px-web.2001/dialog/statfile2.asp>

¹¹¹ <http://whitedb.org>

Eesti on avaandmete portaalide osas alustajate hulgas. TNO avaandmete strateegia uuringu järgi puudus Eestil 2011. a arvestatav avaandmete strateegia. 2014ndal aastal on Eestis kavas vastu võtta Avaandmete Roheline Raamat¹¹², mis saab olema esimene avaandmete strateegia dokument Eestis.

Avaandmete temaatika on seotud seadusandlusega ja regulatsioonidega. Eriti tähtis on tõik, et Eesti avaliku teabe seadus (AvTS) nõuab, et avalik teave, sealhulgas registritesse kogutud andmed, oleks „vaikimisi avalikustatud“.¹¹³ Seega on Eesti valinud vaikimisi avatud avalike andmete tee nagu näiteks USA (vt täpsemalt jaotis 2.4), mis tähedab, et Eesti riik on selle seadusega (AvTS) võtnud vastava kohustuse.

Vastavalt ülaltoodud Eesti hetkeseisule ja maailma trendidele avaandmete valdkonnas pakume allpool järgmisi soovitusi andmete avalikustamise strateegiate, tootmise ja linkimise osas:

Avaandmete tootmine ja kasutuselevõtt maailmas liigub praegu avaandmete mahu suurenemise ja kvaliteedi olulise paranemise suunas (vt näiteks G8 harta). Alates 2012. aastast on uueks trendiks avaandmete linkimine või avalike linkandmete tootmise ja kasutuselevõtu kasv. Et Eestis on rakendunud PSI Direktiiv, siis on Eestil kohustus selle sätteid järgida. Arvestades trende, võiks Eesti keskenduda avalike linkandmete publitseerimisele valdkondades, kus see on võimalik, kasulik ja turvaline. Selle tegevuse mõju linkandmete kasutuselevõtule Eestis oleks väga suur. Avalikud linkandmed võimaldaksid kodanikel ja ettevõtetel luua endale sobivaid rakendusi.

Arvestades avaandmete kasutuselevõtu edutegureid teistes riikides, peaks Eestil eelkõige tekkima hea riiklik avaandmete kasutuselevõtu strateegia ja tegevusplaan. Luua tuleks toimiv avaandmete portaali ja kindlaks määrata andmete maksustamispoliitika. Tuleks soodustada avaandmete linkimist ja taaskasutust läbi vastavate rakenduste.

Gartneri ennustust arvestades peaksid ka ettevõtted valmistuma osa oma andmete avalikustamiseks. Eesti kodanikud teevad seda juba praegu, kasutades Facebook-i ja muid sotsiaalse võrkustluse vahendeid.

PWC uuring „E-äri ja e-kaubanduse kasutamine Eestis ja kasutamise laiendamise võimalused 2013“¹¹⁴ tõi välja, et avaliku sektori andmeid kasutab äritegevuses 52% Eesti ettevõtetest. Avalikud andmed leiavad kasutust eelkõige raamatupidamises, müügitegevuses, teenuste osutamisel ja kliendiandmete haldamisel, väiksemal määral ka turunduses. Kahjuks piirab andmete kättesaadavust ja kasutamist nende hajus paiknemine erinevates andmekogudes (nt kohalike omavalitsuste andmekogud), mille puhul andmed võivad küll kättesaadavad olla tasuta, kuid nende komplekteerimine ja täiendamine nõuab ettevõtjalt täiendavat ressursi. Siit tuleneb omakorda nõudlus ühtse avaandmete portaali loomise järele.

2.1.4 Suurandmete töötlemise ja hõive tehnoloogiad

Definitsioon. Suurandmeid iseloomustavad järgmised omadused: mahukus, muutuvus, mitmekesisus ja lisaks ka keerukus. Viimasel ajal on lisandunud suurandmetele ka tõepärasuse omadus. Andmemahutuste suurus mõõdetakse suurandmete korral tavaliselt peta-baitides. Üldiselt räägitakse suurandmetest sel juhul kui nende maht on nii suur, et traditsiooniliste andmebaasisüsteemide võimekus pole piisav nende haldamiseks. Muutuvus iseloomustab andmete kiiret ajas genereerimist, käibele laskmist ja analüüsimiseks kohale toimetamist tavaliselt reaajas või selle lähedaselt. Mitmekesisus näitab, et andmed, mida on vaja integreerida ja analüüsida, on väga erinevat tüüpi. Mõningatel juhtudel vaadeldakse lisaks ka andmete keerukust. Näiteks andmehulka, mis sisaldab 30 biljonit RDF kolmikut, võib vaadelda kui suurandmeid. Andmete tõepärasus viitab sellistele andmete omadustele nagu usaldusväärsus, vigade puudumine, kooskõllalisus jms.

Suurandmetega seotud tehnoloogiad võib jagada kaheks:

- Suurandmete hõive ja säilitamise tehnoloogiad.

¹¹² Avaandmete roheline raamat, AARR, Avaliku teabe avamine taaskasutamiseks, versioon 0.4 2013-11-21, Eesti Vabariigi valitsus, <http://www.riso.ee/et/avaandmete-roheline-raamat>

¹¹³ Avaliku teabe seadus: <https://www.riigiteataja.ee/akt/119122012005?leiaKehtiv>

¹¹⁴ http://www.mkm.ee/public/Lopparuanne_-_E-ari_ja_e-kaubandus_1_6_avalik_2013.pdf

- Suurandmete analüütika, mille käigus uuritakse kogutud andmeid, leidmaks nendest peidetud mustreid, seniteadmata korrelatsioone ja muud huvitavat teavet, mida saab kasutada paremate otsuste langetamiseks. Need tehnoloogiad on kõrge jõudlusega andmekaevandus, ennustav analüütika, tekstikaevandus, prognoosimine ja optimeerimine.

Motivatsioon. Gartner¹¹⁵ ennustas 2010. aastal, et ettevõtete igasuguses formaadis andmete maht kasvab kuni 650% kümne järgmise aasta jooksul. IDC¹¹⁶ andmetel kahekordistub maailmas andmete maht iga 18 kuu järel. Sellise andmete mahu kasvuga kaasaskäimine on väljakutse andmetööstustehnoloogiatele.

Võrreldes relatsiooniliste andmebaasidega pole suurandmed ühtse ja hästi konstrueeritud struktuuriga. Suurandmeid tekitatakse nii masinate poolt (näiteks sensorite poolt) kui ka väga erineva kogemusega inimeste poolt (näiteks sotsiaalmeedias). See seab uued nõuded suurandmete töötlemise tehnoloogiatele, analüüsi vahenditele, andmekaevandamise algoritmidele ja andmete visualiseerimise tehnikatele. Et suurandmed on kaasajal olemas praktiliselt kõigis majandussektorites (näiteks pangandus, tervishoid, kindlustus, tootmine, turundus, transport jne), siis on tekkinud turunõudlus suurandmete töötlemise ja analüüsi tehnoloogiate järele, mis on positiivselt mõjunud vastava infotehnoloogiavaldkonna arengule. Üks põhilisi infotehnoloogilisi väljakutseid selles vallas on andmete mastaapsuse haldamine ja vastavate skaleeruvate arhitektuuride loomine, sest traditsioonilised andmebaaside tehnoloogiad selleks ei sobi. Üks suurandmete lahendamata probleeme on semantiline koosvõime, sest andmete mitmekesisus on väga suur ja andmed pärinevad väga erinevatest allikatest. Suurandmete koosvõime on saanud üha tähtsamaks probleemiks suurandmete efektiivsel haldamisel. Et suurandmed on põhiliselt struktureerimata, siis muutub ülitähtsaks metaandmete (andmed andmete kohta) ja koosvõime seisukohast ka semantiliste metaandmete olemasolu. Linkandmete tehnoloogia võimaldab siduda erinevaid suurandmeid andes võimaluse tekitada laiem vaade andmetele.¹¹⁷

Teisalt on tähtis suurandmete analüütiline võimekus, sest just see annab suurandmetele ärilise väärtuse võimaldades ammutada olemasolevatest andmetest sügavamad teadmused.

Tassilo Pellegrin, vaatleb oma artiklis „*Economics of Big Data: A Value Perspective on State of the Art and Future Trends*“¹¹⁸ suurandmete makro- ja mikromajanduslikku väärtust. Ta leiab, et makrotasemel on suurandmed pigem poliitika kui tehnoloogia teema sarnaselt avaandmetega, mis said alguse USA poliitikast aastal 2009. Avaandmete poliitika kandus seejärel EL-i, kus see sai Euroopa digitaalarengukava aluseks. Poliitiline motiiv avaandmete osas oli elavdada soiku jäänud majandust, suurandmete osas aga stimuleerida andmeanalüüsil põhinevat otsustusprotsessi ning kasutada mittestruktureeritud veebiandmeid (põhiliselt sotsiaalmeediast) riiklike missioonikriitiliste rakenduste loomiseks.

Seetõttu esitas USA 2012. aastal väga jõuliselt oma suundumuse ülisuurte andmemahude töötlemise tehnoloogiate uurimisele, arendamisele ja kasutamisele nn *Big Data Initiative* raames¹¹⁹. Vastav EL-i riikide initsiatiiv hõlmab nii EL 7 ja Horizon2020 raamprogrammide vastavaid kutseid ja projekte kui ka laiemaid foorumeid nagu BIG Project.¹²⁰

Aastal 2014 aga võime juba rääkida uuest suurandmete tehnoloogiate kasutamise motiivist, mis on seotud asjade interneti üha kiirema kasutuselevõtuga ja sellega seotud sensorandmete plahvatusliku kasvuga (vt 2.1.7 *Asjade internet*).

Hetkeseis ja tulevikutrend. Käesoleval ajal võib öelda, et suurandmete hõive ja analüütika tehnoloogiad on kasutuselevõtu faasi alguses. Gartneri 2012. a suurandmetega seotud tehnoloogiate kasutuselevõtu analüüsis oli MapReduce jt tehnoloogiad kasutuselevõttust 2-5 aasta kaugusel. Aastal 2014 võimegi öelda, et Gartneri ennustus on täide läinud. See muidugi ei tähenda, et suurandmete tehnoloogiatel poleks

¹¹⁵ http://www.gartner.com/it/content/1258400/1258425/january_6_techtrends_rpaquet.pdf

¹¹⁶ www.idc.com

¹¹⁷ Mitchell, I., Wilson, M., 2012. Linked Data. Connecting and exploiting big data. Fujitsu White Paper, March 2012. <http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-%28v1.0%29.pdf>

¹¹⁸ Tassilo Pellegrin, Economics of Big Data: A Value Perspective on State of the Art and Future Trends, In: Big Data Computing, R. Akerkar (Ed), pp 344-367, 2014 by Taylor & Francis Group, LLC

¹¹⁹ Whitehouse 2012, Big Data Press Release,

http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf

¹²⁰ <http://www.big-project.eu>

arenguvajadust. Suurandmete töötlemise *vabavaralised* baastehnoloogiad nagu Hadoop¹²¹, MapReduce¹²² ja nendega seotud rakendused nagu Cloudera¹²³ ja Hive¹²⁴ ning nende adopteerimine suurte tarkvarafirmade nagu IBM, Oracle ja IBM jt poolt on andnud oma panuse suurandmete tehnoloogiate kasutuselevõtule ja ka usalduse tekkimisele nende suhtes.

Kaasajal on suurandmete analüüsiks kasutusel põhiliselt tehnoloogia, kus paralleelsele või NoSQL andmebaasile lisatakse Hadoop ühendus. Hadoop-i kasutatakse mittestruktureeritud suurandmete töötlemiseks. Näiteks kasutab seda Amazon. Põhjuseks on Hadoop süsteemi skaleeruvus nii andmete mahu kui arvutite võrku lisandumisel ja asjaolu, et tegemist on vabavaraga. Viimane omadus loob head võimalused firmade vajadustele vastavate rakenduste loomiseks.

Gartneri hiljutise 720 firmat hõlmanud uuringu tulemuste¹²⁵ järgi plaanisid 64% organisatsioonidest 2013. aastal suurandmete projekti, eesotsas meedia ja kommunikatsioonifirmade ning pankadega. Gartner kutsus seetõttu 2013. aastat isegi *suurandmete suureks eksperimenteerimise ja varase kasutuselevõtu aastaks*. Üle kolmandiku meedia- ja kommunikatsioonifirmadest ütles, et nad on juba investeerinud suurandmete analüütika projekti.

Samas nendib Gartner, et kasutuselevõtt on siiski varases staadiumis, sest vähem kui 8% kõigist vastanutest ütlesid, et nende organisatsioon juba kasutab suurandmete analüütika lahendusi. Samas 20% vastanutest eksperimenteerivad, 18% arendavad strateegiat, 19% koguvad teadmisi ja ülejäänud ei oma plaane või ei tea.

Tekib küsimus, et mis siiski hoiab firmasid tagasi suurandmete tehnoloogiate kasutamisel? Gartneri uurimuse järgi on probleemiks see, et paljud firmad ei näe suurandmete tehnoloogiate ärilist väärtust. Lisaks saabuvad suurandmete tehnoloogiate kasutamise viljad firmadeni mitte kohe pärast projekti investeerimist, vaid pika aja järel. Samas pole investeringud suurandmete projekti väikesed. Lisaks nõuab suurandmete analüütika kõrgepalgaliste analüütikute palkamist situatsioonis, kus tööjõuturg on neist tühi. See aga on hea äri konsultatsioonifirmadele. Teisalt teeb konsultatsioonifirmade teenuste kasutamine suurandmete projekti firmadele veelgi kallimaks, blokeerides ilmselt paljude projektide käivitamise.

Kasutuselevõtt. Iga tehnoloogia kasutuselevõtu edukust mõjutavad nii selle tunnetatud kasulikkus kui ka kasutamise lihtsus. Kasulikkuse aspektist on suurandmete tehnoloogia ilmselt oma kasulikkust teatud määral juba tõestanud. Suurandmete tehnoloogiate kasutuselevõtt ja selle majanduslik kasu ettevõtluses ja avalikus sektoris kasvab vastavalt sellele kuidas ollakse võimelised neid tehnoloogiaid kasutama keeruliste ülesannete lahendamiseks. Siin ei ole kitsaskohaks niivõrd tehnoloogia kui andmetöötajate analüütiline kompetentsus. Just personali suurandmete analüütika alast ebakompetentsust peetakse põhiliseks suurandmete tehnoloogiate kasutuselevõtu piirajaks. Selline situatsioon tekitab tööjõuturul nõudluse analüütikute järele. UK Majanduse ja Ettevõtlusuuringute Keskus ennustab, et aastaks 2017 luuakse suurandmete analüütika alal UKs 36000 uut idufirmat ja 60000 uut töökohta. See tõstab suurandmete analüütika kasutuselevõtu määra 2011. aasta keskmiselt 34%-lt hinnanguliselt 54%-ni aastaks 2017.¹²⁶

Suurandmete analüüsi kasutusvaldkondadest ettevõtluses võiks nimetada ettevõtete ressursside planeerimist, teadus- ja arendustegevusi, turundust ja riskimaandamise valdkondi. Eriti lai kasutusala on suurandmete analüütikal jaemüügi valdkonnas, kus analüüsi kasutatakse ostjate segmenteerimisel jm. Ka paljud veebisuhtlust pakkuvad ettevõtted nagu Google, Twitter, Facebook, LinkedIn jt rakendavad suurandmete töötlust ja analüütikat oma kasutajabaasi laiendamiseks.

Paljud andmebaaside ja andmeaitade halduse tooteid müüvad tarkvarafirmad (näiteks IBM, Oracle, Teradata, SAP EMC. HP. Amazon, MS, Google jpt¹²⁷) on integreerinud on toodetega suurandmete töötlemise võimalused, põhiliselt Hadoop ja MapReduce vabavara.

¹²¹ <http://hadoop.apache.org>

¹²² <http://research.google.com/archive/mapreduce.html>

¹²³ <http://www.cloudera.com>

¹²⁴ <http://hive.apache.org>

¹²⁵ <http://www.gartner.com/newsroom/id/2593815>

¹²⁶ Tassilo Pellegrin, Economics of Big Data: A Value Perspective on State of the Art and Future Trends, In: Big Data Computing, R. Akerkar (Ed), pp 344-367, 2014 by Taylor & Francis Group, LLC

¹²⁷ <http://www.networkworld.com/slideshow/114134/15-most-powerful-big-data-companies.html#slide1>

Teine leer on spetsiaalselt suurandmete töötamise ja analüüsi tehnoloogiatele orienteerunud firmad nagu näiteks HaDapt¹²⁸, Platfora¹²⁹, YarcData¹³⁰, SiSense¹³¹, Space-Time Insights¹³², Zettaset¹³³ jt. Suured firmad nagu eBay, Booking.com, Samsung, Intuit, Carlsberg jpt (ka NASA) kasutavad SiSense suurandmete tehnoloogiat.

IBM Research lõi hiljuti San Josese suurandmete uurimise laboratooriumi Accelerated Discovery Lab¹³⁴. Laboratooriumil on hallata juba 15 suurandmete projekti. IBM suurandmete laboratooriumi eesmärk on leida seoseid suurtes andmehulkades ja kasutada neid analüüsiks ja andmetest aru saamiseks¹³⁵. Laboratooriumil on projekte ja kliente nii tervishoiu kui farmaatsia alal, geoinfosüsteemide, veeressursside juhtimise, maavarade kaevandamise ja sotsiaalmeedia analüüsi alal.

SAS uuringu „Big-Data-in-Big-Companies“¹³⁶ järgi on selge, et suured tööstusettevõtted lähevad üle suurandmete töötlemise ja analüütika tehnoloogiatele kuigi see võtab aega ja nõuab täiesti uut tehnoloogilist lähenemist, muutuvat organisatsiooni ja juhtimisstruktuuri ning uusi oskusi. Suurandmete peamine väärtus tuleb just nende töötlemisest ja analüüsist ja toodetest ning teenustest, mis lähtuvad selle analüüsi tulemustest. Suurettevõtted nagu UPS, General Electric jt kasutavad suurandmete tehnoloogiaid oma tootmisprotsessi optimeerimiseks kogudes andmeid sensorite abil.

IBM Research arendab muuhulgas ka kognitiivset arvutust kasutavaid vahendeid, mis kombineerivad masinõpet ja tehisintellekti (eriti loomuliku keele töötlust)¹³⁷. Need on süsteemid, mis õpivad ja suhtlevad inimestega loomulikult moel selleks, et laiendada kas inimese või masina võimekust. Nad aitavad inimekspertidel võtta vastu paremaid otsuseid hõlvates suurandmete keerukust. Üks kuulsamaid kognitiivseid arvuteid on IBM Watson¹³⁸. Kognitiivsest arvutusest räägime täpsemalt jaotises 2.1.7 *Asjade internet*. Selles kontekstis on üks tähtsaid mõisteid kollektiivne intelligents, mis tähendab inimeste ja arvutite kombinatsiooni, mis on võimeline mõtlema viisil, mis pole omane ei inimestele ega arvutitele eraldiseisvalt.¹³⁹

USA luure keskagentyur (CIA) on huvitatud suurandmete tehnoloogiast¹⁴⁰. Luureorganisatsiooni huviorbiidil on suurandmete tehnoloogiade rakendamine nii luurandmete kogumisel kui nende analüüsil. Avaandmed (eriti sotsiaalmeedia jm) võimaldavad vaba juurdepääsu ka luureorganisatsioonidele, mis annab neile võimaluse automaatselt andmete korjeks ja inimefaktoriga seotud riskide maandamiseks oma organisatsioonis. Automaatne andmekaeandus ja andmeanalüüs võimaldavad leida peidetud seoseid eritüübiliste andmete ja sündmuste vahel. Loomulikult on luureorganisatsioonide huvides nende tehnoloogiate kiire kasutuselevõtt ja nende baasil tehtud rakenduste töökiirus ja turvalisus.

IT-teenuste ettevõtte EMC avaldas hiljuti uuringutulemused¹⁴¹, milles küsitleti 106 IT-juhti Eestis, Lätis ja Leedus seoses väljakutsete ning võimalustega, mida suurandmed ja vastav andmetöötlus nende ettevõtetele ja organisatsioonidele võiks pakkuda. Uuringu tulemused näitasid, et kõigest 22% vastanutest on saavutanud praeguseks konkurentsieelise tänu suurandmete analüütilistele tehnoloogiatele, kuid 44% neist uskus, et edu saavutavad need majandusharud, mis kasutavad just selliseid vahendeid. 44% vastanutest oli nõus, et suurandmete tehnoloogia osutub hädavajalikuks küberrünnakute identifitseerimiseks ja nende ära hoidmiseks.

¹²⁸ <https://hadapt.com/>

¹²⁹ <http://www.platfora.com/>

¹³⁰ <http://www.yarcdata.com/>

¹³¹ <http://www.sisense.com/>

¹³² <http://www.spacetimeinsight.com/>

¹³³ <http://www.zettaset.com/>

¹³⁴ <http://venturebeat.com/2013/10/09/ibm-launches-research-lab-for-big-data-applications/>

¹³⁵ <http://www.research.ibm.com/client-programs/accelerated-discovery-lab/index.shtml>

¹³⁶ <http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf>

¹³⁷ <http://www.research.ibm.com/cognitive-computing/index.shtml#fbid=-pHjUq5w8zr>

¹³⁸ <http://www-03.ibm.com/innovation/us/watson/>

¹³⁹ <http://www.research.ibm.com/cognitive-computing/why-cognitive-systems.shtml#fbid=-pHjUq5w8zr>

¹⁴⁰ <https://www.cia.gov/news-information/speeches-testimony/2012-speeches-testimony/in-q-tel-summit-remarks.html>

¹⁴¹ <http://e24.postimees.ee/2054986/uuring-kolmveerand-balti-ettevotetest-leiab-et-suurandmed-aitavad-paremaid-otsuseid-teha>

Uuringu tulemustel on Balti riikides suurandmed märgatavalt parandanud otsuste tegemist ja hakanud mõjutama ettevõtete eristumist konkurentidest, samuti nende võimet riske vältida. 76% vastanutest ütles, et nende organisatsioonis vastuvõetavaid otsuseid võiks parandada andmete parem kasutamine, ning 40% nõustus, et nende juhtkonnad usaldavad suurandmete ülevaateid, tegemaks muutustele orienteeritud äriotsuseid.

Siiski pole antud uuringu kontekstis päris selge, kuidas ülalpool nimetatud uuring suurandmete tehnoloogiaid defineeris.

Big Data Scoring OY, mis kuulub Eestis registreeritud Social Media Holding OÜ-le, tõi hiljuti esimesena turule vaid sotsiaalmeedia andmetele tugineva laenuvõime hindamise (*credit scoring*) mudeli Euroopa turgude jaoks¹⁴². „Pangad üle maailma otsivad üha enam viise, kuidas andmete hiidkogumeid oma töös paremini rakendada. Esimestena maailmas on meil neile pakkuda läbiproovitud ja töötav lahendus Euroopa turgude jaoks, mis kasutab klientide laenuvõime hindamiseks suurt hulka internetis ja sotsiaalmeedias leiduvat infot,“ ütles Postimehe teatel ettevõtte juht ja kaasasutaja Erki Kert.

Kuigi on üsna selge, et suurandmete tehnoloogiate kasutuselevõtt võimaldab suurendada majanduslikku kasvu, peab siiski pöörama tähelepanu ka selle sotsiaalsetele aspektidele. Nendeks on autoriõiguste tagamine, turvalisus ja privaatsus. Seega saab suurandmete tehnoloogiate kasutuselevõtu kriitiliseks eduteguriks balansseerimine kättesaadavate andmete ja nende üha suureneva linkimise ning analüüsimise soovi ning võimaliku sotsiaalse häirumuse vahel.

Järeldused Eestile. Nagu eespool toodud andmetest ja analüüsist järeldub, on ka Eestis vaikselt algamas suurandmete analüütika ajajärk. Teatud nõudlus on tekkinud ja nii mõnedki idufirmad pakuvad juba seda teenust. Näiteks DeciderLab¹⁴³, Iteraction¹⁴⁴ jt.

Siiski on teadlikkus nendest tehnoloogiatest Eestis veel madalal tasemel ja seda peaks kindlasti aitama riiklike vahenditega tõsta. Teadlikkus on vajalik nii erasektorile kui avalikule sektorile paremaks andmeanalüüsi kasutamiseks otsustusprotsessi parendamisel. Riigisektoris on potentsiaalsed suurandmete analüütika vahendite kasutajad siseministeeriumi haldusalas. Näiteks, Los Angeles-is kasutatakse seda tüüpi analüütikat kuritegevuse ennetamiseks kasutades vastavaid arvuti poolt tehtud prognoose võimalike kuritegevuse toimepaneku piirkondade kohta.

Ka Eesti teaduse- (nt keskkonnauuringud, kosmoseuuringut jm) ja ärivaldkondades (nt telekommunikatsioon, meedia, pangad, kindlustusfirmad jm) on kogutud suurandmeid, mille töötluks ja analüüsiks saaks kasutada suurandmete tehnoloogiaid.

Eesti riigikaitse jaoks on muidugi oluline omada teadlikkust suurandmete korje ja analüütika turvalisuse aspektidest, seda eriti seoses avalike andmete publitseerimisega.

2.1.5 Koosloome tehnoloogiad

Definitsioon. Koosloome (*Crowdsourcing*) on veebipõhine hajutatud ülesannete lahendamise ja toodete ning teenuste loomise mudel.¹⁴⁵ See on protsess, kus töö tellitakse või rahastus saadakse internetis tegutsevalt rahvahulgalt.

Koosloome läbiviimise viisid võib jagada koosrahastuseks, mikroülesannete lahendamiseks ja avatud innovatsiooniks¹⁴⁶:

- Koosrahastus tähendab rahvahulkade kasutamist projekti rahastamiseks annetuste teel.

¹⁴² <http://e24.postimees.ee/1225740/erki-kerdi-firma-hakkab-sotsiaalmeedia-pohjal-laenuvoimet-hindama>

¹⁴³ <http://deciderlab.com/>

¹⁴⁴ <http://iteration.ee/>

¹⁴⁵ Brabham Daren, "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", *Convergence: The International Journal of Research into New Media Technologies* 14 (1): 75–90, 2008

¹⁴⁶ <http://dailycrowdsource.com/training/crowdsourcing/what-is-crowdsourcing>

- Mikroülesannete lahendamine tähendab, et suur ülesanne jagatakse väikesteks ülesanneteks ja siis antakse need rahvahulgale lahendamiseks. Iga väikse ülesande lahendamise eest on ette nähtud pisike rahaline hüvitis.
- Avatud innovatsiooni mudel on selleks, et kasutada rahvahulgast tulnud ideesid mingi uue äri, tootedisaini või idee ellukutsumiseks.

Motivatsioon. Tarbekaupade tootjad on kasutanud koosloome tehnika elemente juba enne 2006. aastat, mil termin „koosloome“ kasutusele võeti. Põhjuseks on see, et neil on suurem vajadus kui teistel tootmisharudel olla kursis tarbijate üha muutuvate vajaduste ja harjumustega.

Koosloome kasutamisega võib saada väga hea kvaliteediga tulemuse, sest saab kaasata palju rohkem andekaid inimesi, kui see muidu oleks võimalik. Näiteks võib tuua Wikipedia, kuhu paljud inimesed üle maailma teevad kaastööd vabatahtlikult. Kaasajal on koosloomel väga suur tähtsus ja seda kasutatakse nii informatsiooni kui loominguliste ideede kogumiseks, ülesannete lahendamiseks ja infoanalüüsiks. Koosloome tagamiseks kasutatakse vastavaid veebipõhiseid koosloome/koostöö vahendeid ja tehnoloogiaid. Põhilisteks trendideks on koosloome kasutamine ärilistel eesmärkidel ja koosloome vahendite loomine mobiili/nutitelefonide platvormidele.

Hetkeseis ja tulevikutrendid. Gartneri hiljutiste ennustuste kohaselt peaks aastaks 2017 rohkem kui 50% tarbekaupade tootjaid saama 75% innovatsioonist ning teadus- ja arendustegevusest koosloome lahenduste kaudu¹⁴⁷. Kaubamärkide omajad lasevad praegu iga nädal välja tooteid, teenuseid ja sisu, mis on inspireeritud ja loodud internetikogukondade poolt. Aastal 2014 loodetakse näha koosloome efektiivsuse tõestust turunduse ja innovatsiooni valdkondades. Gartner ennustab, et koosloomet kasutavate firmade kasum tõuseb 1% võrreldes firmadega, kes koosloomet ei kasuta ja seda juba 2015. aastaks.

Kuna firmad pöörduvad koosloome kasutamise poole, siis hakkab seda tegema ka kogu seotud ökosüsteem. On tekkinud palju koosloome firmasid – näiteks TopCoder, Appirio, DesignCrowd, Freelancer.com, 99designs jt.

Koosloome tagavad kogukonnad, mis on kasvanud kiiresti viimase 5 aasta jooksul. Eyekas Francois Petavy hindab oma uuringus „Five predictions for crowdsourcing in 2014“, et praegu panustavad koosloomesse umbes 10 miljonit inimest üle maailma.¹⁴⁸ Ta toob välja, et maailm muutub multi-modaalseks, kus inimesel ei pea olema ainult üks töö, vaid ta võib vabalt pärast tavatööd panustada mõnesse loovasse koosloomeprojekti. On tekkinud juba mõiste pilvetööjõud (*cloud labor*). Need on virtuaalselt hajutatud töötajad, kes on saadaval vastavalt vajadusele nii väikeste kui ka keerukamate ülesannete täitmiseks. Samas tekivad selle tööjõu kasutamisel ka autoriõiguste, sotsiaalsete tagatiste jm probleemid.

Deloitte toob oma industriaalse koosloome uuringus välja 6 koosloome platvormi (Gigwalk, oDesk, Kaggle, Tongal, Quirky ja Kickstarter) ja nendega seotud kogukondade suurused ning läbiviidud projektide arvu.¹⁴⁹

Näiteks olid kõige suurema kogukonna arvuga Kickstarter (5,4 miljonit) ja oDesk (4,5 miljonit) samas kui täidetud ülesannete arv oli suurim Gigwalk-il (4 miljonit). Gigwalk loodi 2011 aastal mobiilseteks välitöödeks. Deloitte'i uuringus tehakse järeldus, et koosloome on praegu varases arengustaadiumis, kuid kaasaegsed koosloome platvormid on piisavalt keerukad, võimaldamaks lahendada palju erinevat tüüpi ülesandeid. Siiski on esmatähtis, et organisatsioonid ise võtaksid omaks uusi ideid, mida võib genereerida koosloome initsiatiivide abil.

Kasutuselevõtt. Kaasajal kasutavad paljud firmad uute innovatiivsete toodete loomisel rahvahulki esitades küsimusi kasutajate foorumite, väärtusahela koostöö saitide ja virtuaalsete kasutajate fookusgruppide kaudu. Massilist koosvõime kasutamist võimaldavad reklaamindus, internetikogukonnad, teaduslike probleemide lahendamine, uute toodete ideede arendamine, loov probleemide lahendamine, tarbijate poolt loodud tooted jm. Tarbijad ise kasutavad ka oma elu korraldamiseks koosloomet, eriti nutitelefonide abil.

¹⁴⁷ <http://www.gartner.com/newsroom/id/2603215>

¹⁴⁸ <http://www.crowdsourcing.org/editorial/eyekas-francois-petavy-five-predictions-for-crowdsourcing-in-2014/30116>

¹⁴⁹ Tech Trends 2014, Deloitte, 2014, <http://dupress.com/articles/2014-tech-trends-crowdsourcing/?coll=6210>

Koosloome ärimudeleid võib jagada järgmistesse klassidesse¹⁵⁰: meedia ja andmed (nt <http://data.com>), turuplatsid (nt <http://freelancer.com>), platvormid (nt <http://kluster.com>), rahvahulkade ettevõtmised (nt <http://my3p.com>), rahvahulkade teenused (nt <http://thinkspeed.com>), mitteäriine kasutusmudel (nt <http://kiva.com>), rahvahulkade protsessid (nt liveops), sisu ja toodete turud (nt <http://made.com>).

Koosloomet on kasutatud selleks, et saada ideid avaandmete rakenduste loomiseks ja linkandmete halduseks. Näiteks on kasutatud mikroülesandeid andmete linkimiseks ja ontoloogiate joondamiseks. Ka on koosloomet kasutatud linkandmete kvaliteedi tagamiseks ja vigaste RDF kolmikute leidmiseks¹⁵¹. Koosvõimetööriist *TripleCheckMate*¹⁵² on mõeldud avalike linkandmete kvaliteedi hindamiseks.

Järeldused Eestile. Koosloomel on suur potentsiaal nii Eesti avaliku kui erasektori ning vabakonna jaoks. Eestis saaks rakendada erinevaid koosloome ärimudeleid ja koosloome tööriistu nii ava- kui linkandmete kvaliteedi hindamiseks. Näiteks võiks läbi viia koosloome võistlusi avaandmete ja linkandmete rakenduste ideede saamiseks aga ka rakenduste loomiseks. Koosloome kasutamine võimaldab kaasata probleemide lahendamisse huvilisi kogu maailmast, see aga oleks Eesti väiksust arvestades oluline.

Eesti valitsus võiks koondada ava- ja linkandmete kasutajate grupid, kes kasutades koosloome platvormi annaksid riigile nõu, milliseid andmeid avada ja linkida.

Ka riikliku linkandmete poliitika kujundamiseks saab kasutada koosloome tehnoloogiaid.

2.1.6 Tulevikuinternet

Definitsioon. Tulevikuinternet (*Future Internet*) on üldine termin tähistamaks ülemaailmseid teadusuuringuid ja arendustegevusi, mis on pühendatud Interneti edasiarendamiseks. Tähtsamad Interneti tulevikuperspektiivid on seotud uute Interneti infrastruktuuridega, Internetis lingitud andmete tohutu kasvuga, Interneti turvalisusega, Internetipõhise koostoimega (sh sotsiaalne suhtlus aga ka koostoime seadmetega jms), Internetipõhiste seadmete, süsteemide ja teenustega, Interneti vahendatud innovatsiooniga.

Motivatsioon. Interneti edasiarendamine on väga tähtis ja sel on suur mõju indiviididele, ärile ja ühiskonnale tervikuna. Tulevikuinterneti teemadega tegeletakse nii USA-s kui EL-i riikides. Selle teema uurimisele ja arendustegevusele on pühendatud palju EL teadus- ja arendusprojekte¹⁵³.

Tulevikuinterneti teemade ring on väga lai ja osaliselt on nendest teemadest juttu kõigis käesoleva aruande tehnoloogiatrende käsitlevates osades. Selles osas keskendume kahele väga olulisele teemale: uued interneti arhitektuurid ja uus internetiprotokoll IPv6.

Praegu on murranguline periood Interneti arengus, sest kommunikatsioonile orienteeritud internet peaks vastavalt muutunud nõuetele ja vajadustele muutuma sisule orienteerituks Internetiks. See toetaks dokumendiveebi muutumist andmete veebiks, millest oli juttu jaotises 2.1.1 *Linkandmete tehnoloogia*. Seoses sellega vaatlemegi käesolevas aruandes tulevikuinterneti uute arhitektuuride trende.

Pragmatilisest vaatepunktis lähtuvalt on praegu maailmas tähtis üleminek uuele internetiprotokollile IPv6, sest IPv4 aadresside hulk on ammendumas. IPv6 võimaldab kasutusele võtta uue aadressruumi ja lülitada internetti üha rohkem asju s.o sensoreid jms.

Hetkeseis ja tulevikutrend. Teadlased tegelevad uute Interneti arhitektuuridega, mis peaksid muutma viise, kuidas teavet Internetis jagatakse tehes Interneti kiiremaks ja turvalisemaks. Näiteks loodi EL projekti

¹⁵⁰ <http://www.resultsfromcrowds.com/features/crowd-business-models/>

¹⁵¹ Dimitris Kontokostas, Sören Auer and Jens Lehmann, Crowdsourcing Linked Data quality assessment, http://jens-lehmann.org/files/2013/iswc_crowd_qa.pdf

¹⁵² <http://aksw.org/Projects/TripleCheckMate.html>

¹⁵³ Euroopa tulevikuinterneti portal <http://www.future-internet.eu/>

Pursuit¹⁵⁴ käigus pilootlahedus Interneti protokollid IP kihi muutmiseks nii, et see oleks intelligentsem ja kus kasutajad ei pea vajalikule teabele juurdepääsemiseks omama otsest ühendust serveriga, milles see teave oli algselt salvestatud. Selle asemel on Interneti ühendatud arvutid ise võimelised kopeerima ja taaspublitseerima nõutavat sisu või selle fragmente väga erinevatest kohtadest. See on sarnane nn „peer-to-peer“ lähenemisviisiga, mida kasutatakse failide jagamise jm süsteemides aga sel juhul toimiks ta üle kogu Interneti¹⁵⁵.

Uue Interneti arhitektuuri projekt Nebula¹⁵⁶ aga lähtub pilvearvutusest ja loodab pakkuda välja turvalise ja elastse Interneti arhitektuuri selleks, et toetada praegusi ja tulevase pilvearvutuse rakendusi. Turvalisus, konfidentsiaalsus ja terviklikkus on selle arhitektuuri põhiomadused¹⁵⁷.

Mõned teadlased pakuvad välja nn linkandmete Interneti arhitektuuri, kus Interneti rakenduste kiht asendatakse kahe kihiga: linkandmete kiht ja rakenduste kiht.^{158,159}

Internetiprotokollid IPv6 kasutuselevõtt on tähtis seoses asjade Interneti levikuga, mida käsitleme järgmises jaotises. EL projekti IoT6¹⁶⁰ andmetel, võib aastat 2012 vaadelda kui globaalset IPv6-le ülemineku algust. Google'i kasutajatest 1% kasutab IPv6. Euroopas on samuti IPv6-le üleminek tähtis eesmärk ja poliitika. Seda toetab ka EK tegevuskava. Rumeenia on liider IPv6 kasutuselevõtus 8,43% kasutusmääraga, järgnevad Prantsusmaa 4,69 %-ga. Põhja Ameerikas on see määr 1,97%. See teeb kokku 3,5 miljonit kasutajat, mis on suurim kasutajate arv maailmas.

Aasia maad (Hiina, Jaapan, Lõuna Korea) stimuleerivad IPv6 kasutuselevõttu asjade Interneti rakendustega. Hiina suurim IPv6 võrk loodi Pekingi olümpiamängudeks aastal 2008. See ühendab miljoneid seadmeid, kasutajaid, turva- ja transpordivõrke üle kogu Hiina.

Eestis on IPv6 edendamine Infoühiskonna 2020 aasta arengukavas¹⁶¹ meetme 1 all: Interneti kättesaadavuse suurendamine. Seal on ette nähtud järgmine tegevus: „Muu hulgas käivitatakse üleminek IPv6-le avalikus sektoris ning ärgitatakse seda tegema ka erasektori teenusepakkujaid.“

Tarmo Hanga RIA-st esines 05.06.2012 IPv6 Forum-il ettekandega „(Kas) riik on IPv6 kasutuselevõtuks valmis?“¹⁶². Ta ütles oma ettekandes, et riigiasutuste vahelisel magistraalvõrgul ASOnet on IPv6 võimekus olemas.

Iga klient saaks soovi korral IPv6 vahemiku, kuid erilist aktiivsust klientide poolt IPv6 aadressiruumi aktiivsema kasutuselevõtu osas polevat näha. Suuremat IPv6 kasutuselevõttu prognoositakse alles siis, kui tekib massiline vajadus uute aadresside järele – näiteks targa maja lahendustega või uute nutikate teenustega.

Tamo Hanga väitis, et niisugust olukorda, kus Eesti kontekstis IPv4 privaatne aadressiruum ammenduks ja see nõuaks IPv6 kasutuselevõttu, ei oska momendil riigisektori poolel ette näha.

Eesti hetkeseis ja järelused. Käesolev jaotis keskendus uutele Internetiarhitektuuridele ja IPv6 kasutuselevõtule. Meile teadaolevalt ei tegeleta Eestis uute Internetiarhitektuuride väljatöötamisega ja seoses vastava kompetentsi puudumisega ei saa Eesti ka selles suunas tegutseda. Linkandmete kasutuselevõtu aspektist oleks hea kui uus Interneti arhitektuur toetaks andmete veebi. Seega, Eesti tegutsemine linkandmete valdkonnas võib avaldada toetust ka uute sobivate Internetiarhitektuuride loomisele.

¹⁵⁴ <http://www.fp7-pursuit.eu/>

¹⁵⁵ <http://www.cam.ac.uk/research/news/future-internet-aims-to-sever-links-with-servers>

¹⁵⁶ <http://nebula-fia.org/>

¹⁵⁷ T. Anderson et al, The NEBULA Future Internet Architecture, Galis and A. Gavras (Eds.): FIA 2013, LNCS 7858, pp. 16–26, 2013

¹⁵⁸ <http://linkeddata.future-internet.eu/images/3/3f/DeckerIntro.pdf>

¹⁵⁹ http://linkeddata.future-internet.eu/index.php/Main_Page

¹⁶⁰ IoT6 European research project, Deliverable D1.5: “IoT6 Reference Model”, White Paper (2012), <http://www.iot6.eu>

¹⁶¹ http://www.riso.ee/sites/default/files/elfinder/article_files/infoyhiskonna_arengukava_2020_f.pdf

¹⁶² (Kas) riik on IPv6 kasutuselevõtuks valmis?, Tarmo Hanga, 05.06.2012 – IPv6 Forum, IEEE Eesti sektsioon

Internetiprotokoll IPv6 kasutuselevõtt on eelkõige seotud asjade internetiga (vt järgmine jaotis 1.1.7). Eestis tuleks IPv6 vajadusel kasutusele võtta ja mobiilioperaatorid peaksid hakkama ka seda pakkuma selleks, et toetada asjade interneti kasutuselevõttu Eestis. Internetiprotokolli IPv6 koolituspraktikum toimus just hiljuti.¹⁶³ Ilmselt on IPv6 kasutusala laiendes vaja ka jätkata koolitustega.

2.1.7 Asjade internet

Definitsioon. Asjade internet (*Internet of Things*) tähendab füüsiliste objektide (näiteks esemete, seadmete) varustamist sensoritega, mis on lingitud traadiga või traadita võrgu abil, kasutades selleks tihti Interneti protokoll (IP). Sellised nn sensorvõrgud või füüsilised informatsioonisüsteemid tekitavad suure hulga andmeid, mis saadetakse arvutitesse analüüsimiseks. Asjade internet on üks osa tulevikuinterneti tehnoloogiatest.

Motivatsioon. Kevin Ashton oma kuulsas 2009. a artiklis, „*That 'Internet of Things' Thing*”,¹⁶⁴ milles ta võttis kasutusele termini „*asjade internet*“, on väga hästi toonud välja asjade interneti motivatsiooni, mille siinkohal teatud lähenduses taasesitame. Internet ja veeb on peaaegu täiesti sõltuvad inimestest, kes loovad veebis kättesaadavaid andmeid. Inimestel on piiratud ajaressurss, tähelepanu ja täpsus – seega pole nad väga head reaalse/füüsiliste asjade kohta käivate andmete kogujad. See on probleem, sest nii inimesed kui nende keskkond on füüsilised. Selge on see, et ideed ja info on olulised, kuid kaasaegne infotehnoloogia on liiga sõltuv andmetest, mis pärinevad inimestelt ja seega teavad meie arvutid rohkem ideedest kui asjadest. Kui meil oleksid arvutid, mis teaksid kõike, mis on vajalik asjade kohta ning kui neid andmeid kogutakse ilma inimeste abita, siis meil oleks võimalik jälgida asju ja nende asjade näitajate muutumist. Me saaksime näiteks teada, millal asju on vaja asendada, parandada, kas nad on värsked või riknenud jms. Asjade internetil on potentsiaali muuta maailma nii nagu tegi seda Internet. Ehk isegi rohkem.

Hetkeseis ja tulevikutrend. Internetti oli aastal 2013 hinnanguliselt lülitunud 39% maa 7,1 biljonist elanikust¹⁶⁵. Arvatakse, et aastaks 2015 on 3 biljonit inimest ühendunud Internetti. Praeguseks aga oleme jõudnud olukorrani, kus rohkem seadmeid (asju) kui inimesi on liitunud Internetiga ning ennustatakse, et Internetti ühendatud asjade hulk ületab peagi inimpopulatsiooni arvu moodustades 20-50 biljonit ühendatud tarka seadet. See tähendab ühtlasi ka, et Interneti protokoll IPv4, mis on mõeldud 4 biljoni aadressi jaoks, ammendub.¹⁶⁶

Aastaks 2020 ennustatakse Internetti lülitatud seadmete arvuks 50 biljonit kuni 1 triljon. Sensor- ja võrguvõimekuse lisamine asjadele on kaotanud oma tehnoloogilised ja hinnabarjäärid. Näiteks lisatakse selline võimekus igapäevaselt kasutatavatele asjadele nagu hambahari selleks, et jälgida suuhügieeni harjumusi (vt Beam Brush¹⁶⁷) jms.

PwC nimetab oma hiljutises asjade interneti tehnoloogia uuringus¹⁶⁸ asjade interneti arengut väga kiireks ja vaatleb seda koosnevat kolmest eri kihist:

- Sensorid on asetatud seadmetesse või füüsilisse keskkonda ja koguvad vajalikke andmeid või informatsiooni sündmuste kohta.
- Võrgu- ja arvutusplatvorm jagab sensoritega informatsiooni ja toimib vastavalt sellele informatsioonile mõjutades keskkonda.

¹⁶³ https://www.ria.ee/ipv6-koolituspraktikum/?op=training_detailview

¹⁶⁴ Kevin Ashton: *That 'Internet of Things' Thing*. In: *RFID Journal*, 22 July 2009. <http://www.rfidjournal.com/articles/view?4986>

¹⁶⁵ http://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users

¹⁶⁶ Sébastien Ziegler1., et al, *IoT6 – Moving to an IPv6-Based Future IoT*. Galis and A. Gavras (Eds.): FIA 2013, LNCS 7858, pp. 161–172, 2013.

¹⁶⁷ <http://www.beamtoothbrush.com/toothbrush/>

¹⁶⁸ Vinod Baya and Bo Parker, *The Thing Stack: Technologies that guide customers to their goals*, in Fred Cripe, *Internet of Things: Evolving transactions into relationships*, PwC Technology Forecast, 2013, No 1

- Teenusplatvorm teenindab lõppkasutajaid agregeerides ja analüüsides andmeid ning määratledes ja orkestreerides üldist kasutaja kogemust.

Iga ülaltoodud kiht areneb kiiresti nii uute tehnoloogiate kui lahenduste suunas.

Sensoreid on aastaid kasutatud autotööstuses, tervishoius, tootmises ja mujal. Praeguseks aga on sensorid muutunud väga väikesteks ja odavateks, mis teeb võimalikuks nende asetamise nii kõigisse seadmetesse kui mistahes keskkonda. Sensorite müük kasvab väga kiiresti, umbes 18% aastast aastatel 2011-2016. Sensorite turgu veab nende paigutamine nutitelefonidesse, kus nad koguvad kontekstuaalset infot nagu näiteks asukoht, liikumine, orientatsioon, valgus ja teised keskkonna näitajad. Neid näitajaid kasutatakse omakorda uute teenuste pakkumisel st rakenduste loomisel. Innovaatorid on turule toonud seadmed, mis sisaldavad palju sensoreid, mis kasutavad ära nutitelefonil eelist hõivata ja integreerida andmeid pilvest ja seadmest.

Võrgu- ja arvutusplatvorm pakub töötlust, lokaalset mälu ja ühenduvust. Asjade internet nõuab internetiühendust selleks, et kombineerida sensoritelt saadud info pilves resideeriva infoga. Näiteks HP laborite CeNSE projekt¹⁶⁹ kombineerib uudeid lahendusi materjalitehnoloogias, nanotehnoloogias ja elektromehaaniliste mikrosüsteemide alal selleks, et luua planeeti kattev sensorvõrk kasutades biljoneid või triljoneid ülipisikesi, odavaid ja vastupidavaid sensoreid. Projekti veebilehel öeldakse, et „*CeNSE consists of a highly intelligent network of billions of nanoscale sensors designed to feel, taste, smell, see, and hear what is going on in the world.*“ Projekti tulemuste potentsiaalsete kasutuskohtadena nähakse infrastruktuuri objekte nagu teed, sillad, ehitised. Lisaks saab selliseid sensorvõrke kasutada lennukites, tootmises, tervishoius, turvalisuse tagamisel jm. Paljudel kasutusjuhtudel piisab nendest kahest kihist. Uutele tõusvatele kasutusjuhtudele aga ei piisa ja neile on vaja lisada teenuste platvormi kiht.

Teenuste platvorm võib sisaldada vahekihti, analüütikat ja rakendustarkvara, mis kombineerib sensorandmeid muu kontekstis olulise infoga.

Ajakirja IoT World toimetaja Rich Quinell¹⁷⁰ ennustab asjade internetile ja masin-masin (*Machine-to-machine (M2M)*) kommunikatsioonile olulist kasvu 2014. aastaks. Ennustus põhineb Saksmaa telekommunikatsioonifirma Deutsche Telecom-i M2M tehnoloogiate arengu ennustustel aastaks 2014¹⁷¹. Kõigepealt on ette näha globaalsete alliansside laienemist, sest nii M2M kui asjade interneti areng ja kasutuselevõtt nõuavad piiriüleseid koostöö lahendusi. See tähendab, et andmete edasitoimetajad üle maailma saavad kokku leppida protokollides, andmete formaatides ja ligipääsutehnoloogiates. Teine trend on seotud suurandmete ja nende analüütika sidumisega asjade internetiga, mis peaks looma uusi teenuseid. Kuna asjade internet ja M2M lahendused genereerivad pidevalt suureneva andmehulga, siis kasvab ka nende andmete kaevandamise ja uuel innovatiivsel moel kasutamise võimalus. Siiski peavad andmete kogujad sel juhul otsustama, et kas teha andmed analüüsimiseks kättesaadavaks või arendada ise välja suurandmete analüütika võimekus.

Gartneri 2013. aasta tõusvate tehnoloogiatrendide ennustus¹⁷² keskendub inimese ja masina vahelise dialoogi tagamise tehnoloogiatele, millest erilist tähelepanu pälvivad targad masinad, asjade internet ja kognitiivne arvutus.

Tõusvate tehnoloogiate kasutamisel on Gartneri järgi kolm trendi: inimeste täiendamine tehnoloogiaga (näiteks kantava arvutusseadme kasutamine), inimeste asendamine masinatega (näiteks virtuaalne kognitiivne müügiassistent) ning inimeste ja masinate koostöö (näiteks inimese koostöö robotitega).

Gartner ennustab, et 2020 aastaks on Global 1000 firmade müügist 5% mõjutatud kantavate seadmete poolt kogutud tarbijaandmetest.

Kognitiivsete süsteemide komponentideks on neurosünaptilised kiibid (*Neurosynaptic chips*), mille arhitektuur on inspireeritud inimaju struktuurist. Selliseid kiipe, kus nii mälu, töötlus kui kommunikatsioon

¹⁶⁹ <http://www8.hp.com/us/en/hp-information/environment/cense.html#.Uue8oBD8LRY>

¹⁷⁰ Rich Quinell, Predicting 2014's IoT Trends, http://www.iotworld.com/author.asp?section_id=3150&doc_id=562274

¹⁷¹ <http://www.telecomsthechnews.com/news/2013/dec/19/what-are-deutsche-telekoms-m2m-predictions-2014/>

¹⁷² <http://www.gartner.com/newsroom/id/2575515>

on toodud üksteisele väga lähedale, demonstreeriti juba 2011. aastal IBM ja DARPA projekti SyNAPSE esitlusel¹⁷³. IBMi pikaajaline eesmärk on luua neurosünaptiliste kiipide süsteem, mis koosneb 10 biljonist neuronist ja 100 triljonist sünapstist (*synapses*), mis kõik kokku kulutavad ainult 1 kilovatt energiat ja võtavad enda alla vähem kui 2 liitrit.

Sellised neurosünaptilised kiibid (nn biokiibid) on mõeldud täiendamaks kognitiivseid süsteeme nagu Watson. Watson simuleeriks siis inimese aju vasaku poolkera tegevusi (keelest arusaamine ja analüütika) ja kognitiivsed kiibid tegeleksid parema poolkera tegevustega nagu tunnetus ja kujundite tuvastus. IBM-i teadlaste eesmärk on ühendada need kaks võimekust üheks nagu see on inimaju korral.¹⁷⁴

EL-is on kognitiivsete süsteemidega seotud Inimaju projekt (*Human Brain project*)¹⁷⁵, mille eesmärk on luua täiesti uus IKT infrastruktuur neuroteadusele ja ajuga seotud uuringutele nii meditsiini kui arvutuse valdkondades. Eesmärgiks on aru saada inimaju tegevusest ja haigustest ning püüda simuleerida aju arvutuslikku võimekust.

USA teatas aprillis 2013 üsna pea pärast EL *Inimaju projekti* algust USA ajuuuringute initsiatiivist BRAIN Initiative (*Brain Research through Advancing Innovative Neurotechnologies*)¹⁷⁶, mille eesmärgiks on kaardistada iga inimaju neuroni aktiivsus ja aru saada kuidas inimene mõtleb ja õpib. Projekti toetab suures mahus DARPA.

Euroopa ja DARPA SyNAPSE ajuprojektidel on erinev lähenemine, SyNAPSE projekt ei looda niipalju inimaju madalatasemeliste bioloogiliste mudelite simuleerimisele IKT eksperimentides kui EL projekt vaid modelleerib ajutegevust kasutades suurt hulka ühendatud neuroneid ja sünapse¹⁷⁷.

Gartneri järgi on biokiipide kasutuselevõtuni aega 5-10 aastat.

Kasutuselevõtt. Kaasajal kasutatakse asjade interneti tehnoloogial põhinevaid süsteeme näiteks targa kodu või linna rakendustes, transpordikorralduses, meditsiinis jm. Gartneri¹⁷⁸ arvates ei seisne asjade interneti väärtus mitte niivõrd seadmetes ega nende ühenduvuses, vaid pigem andmete agregeerimises, andmete töötlemises ja otsuste vastuvõtmises; võimaldades kasutada seadmetelt tulevat andmevoogu või seadet distantsilt mõjutada.

Autotööstus on üks põhilisi M2M (*machine-to-machine*) tehnoloogiate kasutuselevõtjaid. Deutsche Telecom ennustab, et sel aastal trend jätkub ja turule tuleb rohkem M2M lahendusi, mis pakuvad klientidele lisandväärtust omavaid teenuseid ja samal ajal ei ole need seotud kindla autotootjaga. Avalikus sektoris on asjade interneti arenevaks kasutusalaaks targa linna lahendused (näiteks tänavavalgustuse tark juhtimine), mille läbimurret nähakse ette aastal 2014.

M2M ja asjade internet on seni olnud põhiliselt B2B (*business-to-business*) turul. Nüüd toimub läbimurre tarbijaturule ja toodetakse üha rohkem tarku laiatarbetooteid. Personaalsete jälgimisvahendite ja kantava tehnoloogia turg peaks kasvama ülikiresti just sel aastal. Personaalsete jälgimisvahendite hinnad langevad ja erinevad targad kellad, tervise ja inimkeha seisukorra näitajate jälgimise vahendid jms tooted tulevad turule.

Asjade interneti laialdasel kasutuselevõtul on ka varjuküljed. Võrgustatud asjad võivad saada spioneerimise objektideks. Samas saab seda tehnoloogiat kasutada ka vastuluureks ja militaarsetel eesmärkidel. David Petraeus, endine USA luure keskagentuuri (CIA) juht ütleb luure kohta: „Huvipakkuvate objektide asukoht määratakse kindlaks, identifitseeritakse, jälgitakse ja kaugjuhitakse kasutades selliseid tehnoloogiaid nagu RFID (*radio-frequency identification*), sensorvõrgud, pisikesed sardserverid ja energiakorjajad, mis on kõik

¹⁷³ <http://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml#fbid=-pHjUq5w8zr>

¹⁷⁴ IBM. SyNAPSE: a cognitive computing project from IBM research. 2011; Available from:

http://www.ibm.com/smarterplanet/us/en/business_analytics/article/cognitive_computing.html

¹⁷⁵ <https://www.humanbrainproject.eu/>

¹⁷⁶ <http://www.whitehouse.gov/share/brain-initiative>

¹⁷⁷ <https://www.humanbrainproject.eu/documents/10180/17646/Vision+Document/8bb75845-8b1d-41e0-bcb9-d4de69eb6603>

¹⁷⁸ Gartner, Gartner Says Potential Size and Diversity of the Internet of Things Mask Immediate Opportunities, <http://www.gartner.com/newsroom/id/2564916>, Gartner 2013

ühendatud järgmise põlvkonna interneti kasutades odavat ja kõrgvõimsusega arvutust (praegu pilvearvutus, paljudel juhtudel superarvutus, pöördides kvantarvutusega).¹⁷⁹ Oma kõnes In-Q-Tel¹⁸⁰ juhtide kogunemisel 2012. aastal, toob David Petraeus välja CIA poolt kasutatavad ja neile huvipakkuvad tehnoloogiad, milleks on eelkõige asjade internet ja suurandmed.¹⁸¹

CIA on põhiliselt huvitatud tehnoloogiatest, mis on asjade interneti liikumapanevad jõud. Nendeks on järgmised tehnoloogiad:

- Objektide tuvastamine või seadmed, mis võimaldavad märgendada.
- Sensorid ja traadita sensorvõrgud – seadmed, mis tajuvad ja reageerivad.
- Sardüsteemid – süsteemid, mis mõtlevad ja annavad hinnanguid.
- Nanotehnoloogia, mis võimaldab kõiki neid seadmeid muuta piisavalt väikseks nii, et nad saaksid funktsioneerida igal pool.

CIA on selles kontekstis huvitatud ka suurandmetest (vt 2.1.3 Avaandmed).

Eesti hetkeseis ja järeldused. Asjade interneti alane kompetents Eestis on koondunud praegu põhiliselt TTÜ-sse. Seal tegeletakse sardüsteemidega, targa toluga jm. ELIKO-s on kompetents asjade interneti ja nutikate mobiilipõhiste tervishoiu lahenduste alal. Erasektoris pakutakse targa maja ja linna lahendusi. Arvestades üleilmset asjade interneti kasvutrendi peaks Eesti kindlasti seda kompetentsi edasi arendama. Riik võiks näiteks tellida tarku infrastruktuurilahendusi.

Linkandmete tehnoloogiatega seoses on asjade internet üks linkandmete allikatest ja samas tihti ka suurandmete allikatest. Eestis loodavad asjade interneti lahendused peaksid sellega arvestama ja olema võimalised nii kasutama linkandmeid kui ka väljastama linkandmeid.

Eesti riigikaitse peaks arvestama asjade interneti arengutrendidega, eriti tuleks jälgida biokiipide ja neurosünaptiliste võrkude uurimisprojektide tulemuste kasutuselevõttu luureagenteuride poolt.

2.1.8 Linkandmete seos teiste tehnoloogiatega

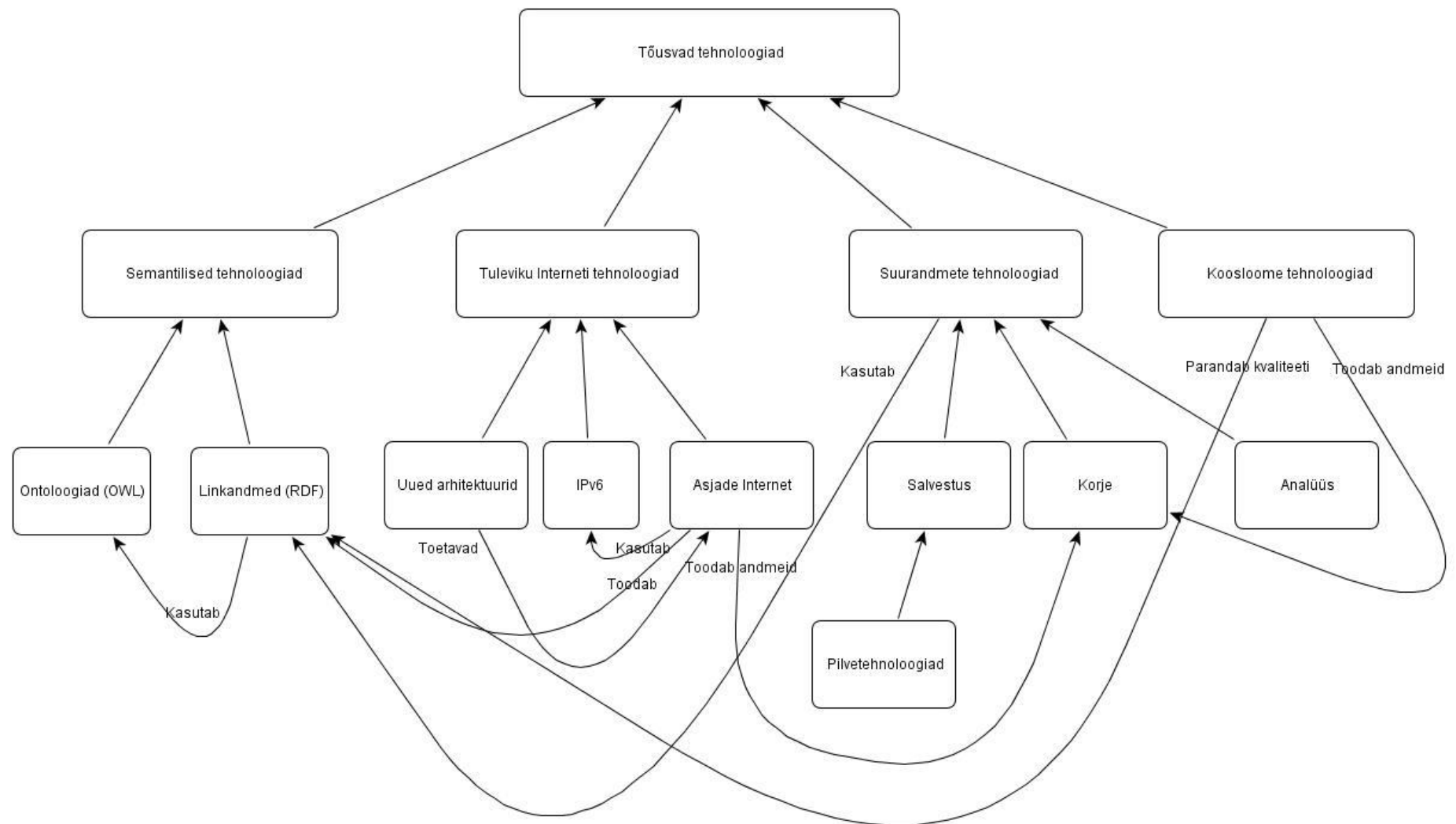
Tehnoloogiate seosed. Eelnevalt arutlesime projekti skoobis antud erinevate tõusvate tehnoloogiate, nende arengusuundade ja kasulikkuse üle. Iga tehnoloogiatrendi käsitle juures oli põgusalt mainitud ka seoseid teiste tehnoloogiate või suundumustega. Käesolevas jaotises esitame üldise pildi vaadeldud tehnoloogiate seostest. Peab märkima, et avaandmed ei kujuta endast eraldi tehnoloogiat, vaid andmeid, mida pakutakse ühel või teisel moel vabalt kasutamiseks.

Alltoodud joonisel (vt *Joonis 3*) on esitatud vaadeldud tehnoloogiate lihtsustatud mõisteline süsteem (ontoloogia). Ilma märgendita nooled joonisel tähistavad alam/ülemmõiste (ehk taksonoomilist) seost, näiteks linkandmete tehnoloogia on semantika tehnoloogiate alamtehnoloogia. Märgenditega nooled näitavad mittetaksonoomilisi seoseid tehnoloogiate vahel, kus märgend tähistab vastavat seost.

¹⁷⁹ <http://www.dailymail.co.uk/sciencetech/article-2115871/The-CIA-wants-spy-TV-Agency-director-says-net-connected-gadgets-transform-surveillance.html>

¹⁸⁰ <https://www.iqt.org/>

¹⁸¹ <https://www.cia.gov/news-information/speeches-testimony/2012-speeches-testimony/in-q-tel-summit-remarks.html>



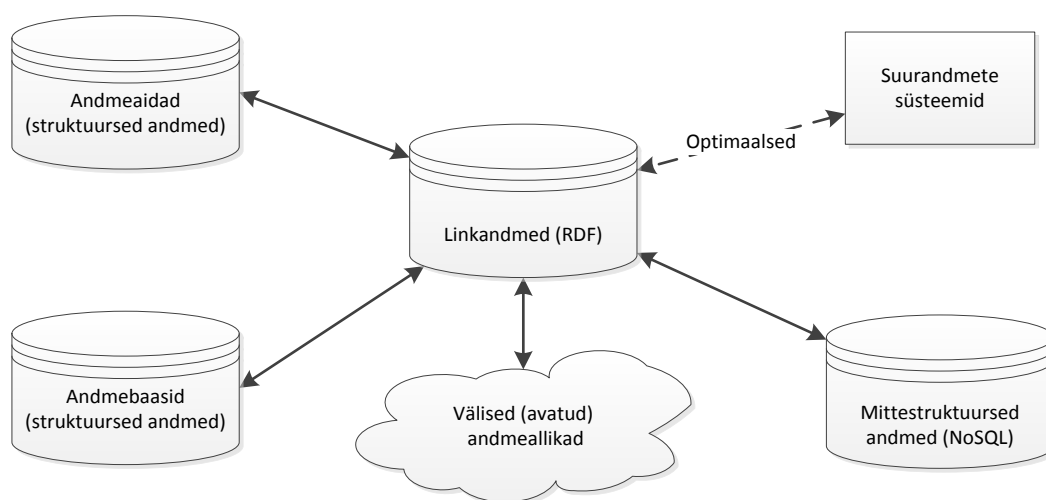
Joonis 3. Õhusvad tehnoloogiad ja nende seosed

Jooniselt võime näha, et linkandmed mängivad tehnoloogiate seostaja rolli. Neid kas toodavad või kasutavad nii asjade interneti, suurandmete kui koosloome tehnoloogiad. Pildil puuduvad avaandmed on seotud linkandmete tehnoloogiatega, sest neid võib esitada linkandmete kujul. Samas võivad ka näiteks suurandmed olla avaandmed.

Kuna Internet on muutumas üha enam sisukeskseks erinevalt varasemast ühenduste kesksusest, siis muutub eriti tähtsaks erinevate hajutatud andmete korje ja linkimine nende taaskasutamiseks ja tihti ka andmetele uue otstarbe andmiseks. Selles kontekstis on tähtis andmete algallika usaldusväärsus ja andmete kvaliteet.

Linkandmete tehnoloogiat saab tõhusalt kasutada Internetis erinevatesse allikatesse laialipaisatud reaaliaja-andmete vahendamiseks, kaardistamiseks, linkimiseks, indekseerimiseks ja tarbimiseks.

Linkandmed võimaldavad ka edukalt linkida igasuguseid andmeid, nii suurandmeid, traditsiooniliste andmebaaside andmeid kui ka mittestruktureeritud andmeid. Järgmisel joonisel (vt *Joonis 4*) on seda väidet illustreeritud.



Joonis 4. Linkandmete tehnoloogia kui erinevate andmete ja tehnoloogiate linkija

Järeldused Eestile. Järgnevalt vastame ülatoodud analüüsi põhjal järgmisele uurimisküsimusele:

Millised järgnevatest trendidest on Eestile esmaselt olulised linkandmete kasutuselevõtu plaanis: tulevikuinternet, semantiline veeb, asjade Internet, suurandmed, avaandmed, koosloome?

Vaadeldud tõusvate tehnoloogiate arengutrendide ja kasutuselevõtu taseme järgi otsustades on linkandmete kasutuselevõtu plaanis tehnoloogiliselt kõige tähtsam semantilise veebi ontoloogiate (sõnastike) kirjeldamise standardite areng ja kasutuselevõtt. Linkandmete laialdase kasutuselevõtu aspektist on eelkõige tähtsad avaandmed ja suurandmed (suur osa neist on toodetud asjade interneti kaudu).

2.2 Rahvusvaheliste initsiatiivide, tegevuste ja standardite analüüs

Käesolevas uuringus vaadeldavad linkandmete ning nendega seonduvate tehnoloogiate alased tegevused maailmas on klassifitseeritud organisatsioonide järgi. Erilist tähelepanu pööratakse Euroopa Komisjoni ja veebikonsortsiumi W3C tegemistele ning rahvusvahelistele standarditele.

Käesolev jaotis on seotud järgmiste uurimisküsimustega:

Millistes rahvusvahelistes initsiatiivides peaks Eesti olema aktiivne, sh riiklikul tasemel (prioriteetsuse järjekorras)? Millised kompetentsid ja ressursid peaksid iga sellise initsiatiivi lõikes Eestil osalemiseks olema?

Millised linkandmetega seotud avatud standardid on Eesti jaoks olulised (sh piisavalt küpsed kasutuselevõtuks)?

2.2.1 Majanduskoostöö ja arengu organisatsiooni tegevused (OECD)

Majanduskoostöö ja arengu organisatsiooni (OECD)¹⁸² üks missioone on pakkuda riikide valitsustele üle maailma võimalust jagada kogemusi ja leida koos lahendusi ühistele probleemidele. OECD-ga on käesoleval momendil on ühinenud 34 riiki, sh Eesti. OECD on eriti huvitatud koostööst valitsustega selleks, et mõista majanduslike, sotsiaalsete ja keskkonna muutuste põhjusi. Seetõttu viib OECD läbi statistilisi jm uuringuid, mõõtmaks erinevate valdkondade arengut erinevates riikides kui ka analüüsi tulevikutrendide ennustamiseks. OECD koostab ja publitseerib ka soovitusi erinevate valdkondade arengu kohta.

OECD teabevärava statistika saidil¹⁸³ võib leida erinevate valdkondade põhimõõdikute kohta andmed avalikuks kasutamiseks Exceli failidena, paljudel juhtudel on andmed saadaval ka reaalaajas. Eesti kohta seal andmeid pole, kuid riikide statistika rubriigis on andmed ka Eesti kohta.¹⁸⁴ OECD pakub ka riikide majandusülevaateid.¹⁸⁵

OECD avaliku sektori avaandmete (Open Government Data) projekti eesmärgiks on läbi erinevate riikide praktika soodustada avaandmete mõjude (nii majandusliku, sotsiaalse kui valitsemise mõju) hindamise meetodika väljatöötamist. Metodoloogia ja vastav uuring „Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives” avaldati 2013. a mais.¹⁸⁶ See uuring analüüsib avaliku sektori avaandmete initsiatiive, tuues välja nende initsiatiivide põhiprintsiibid, kasutatavad mõisted ja nende printsiipide rakendamise ees seisvad probleemid. Uuringus jõutakse järeldusele, et siiani on avaliku sektori avaandmete initsiatiivide mõju ja selle suurendamise võimalusi vähe uuritud. Väljapakutud metodoloogia on analüütiline raamistik nende initsiatiivide ja vastavalt kogutavate avaandmestike mõju uurimiseks OECD riikides. Metodoloogia loojad on arvamisel, et selle raamistiku rakendamine võimaldab leida piisava hulga tõendeid OECD maade avaandmete initsiatiivide tüpiseerimiseks (klassifitseerimiseks) ja mõju hindamise ühiste mõõdikute väljatöötamiseks.

OECD avaliku halduse direktoraat annab välja ülevaadet „Government at a Glance”. Novembris 2013 ilmus uus väljaanne¹⁸⁷, milles kajastuvad ka avaliku sektori avaandmete initsiatiive puudutavate mõõdikute väärtused, mis on leitud ülalmainitud meetodika järgi.

Näiteks on selles aruandes toodud OECD riikide tsentraalsetes portaalides publitseeritud avaandmehulkade arvud (30.10.2013 seisuga), mis on esitatud alljärgneval joonisel (vt Joonis 5).

¹⁸² <http://www.oecd.org>

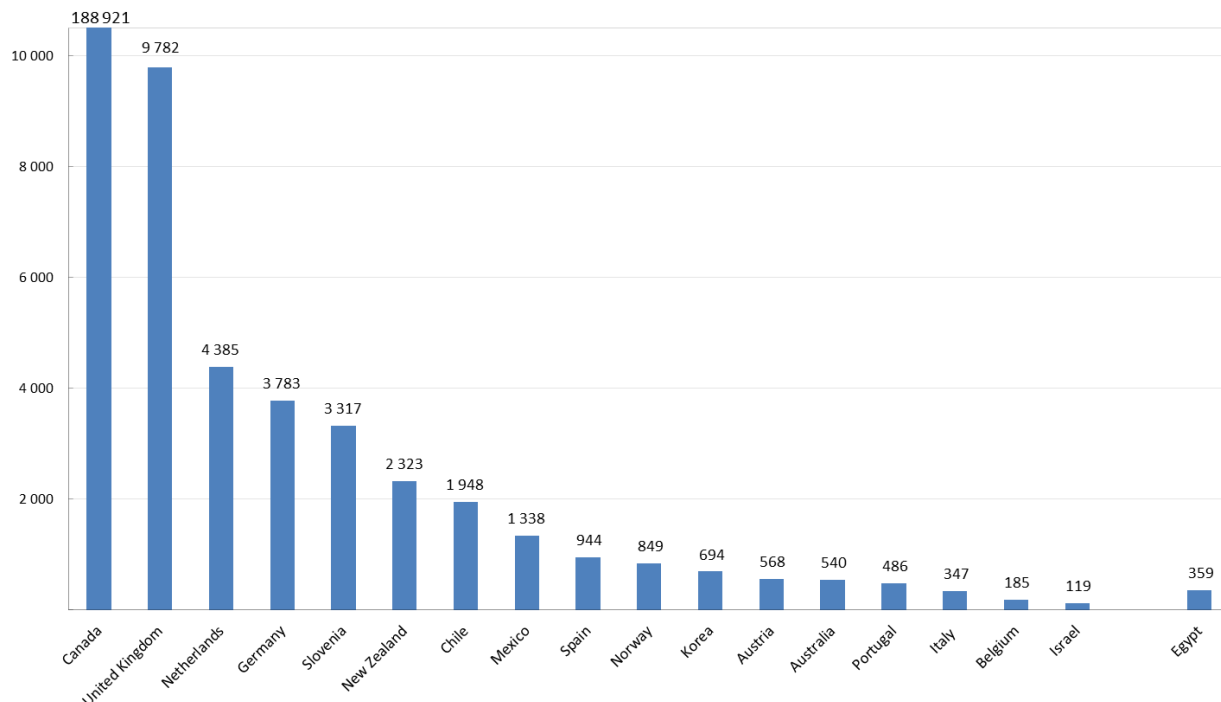
¹⁸³ <http://www.oecd.org/statistics/>

¹⁸⁴ <http://www.oecd.org/estonia/>

¹⁸⁵ <http://www.oecd.org/eco/surveys/>

¹⁸⁶ Ubaldi, B. (2013), “Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives”, OECD Working Papers on Public Governance, No. 22, OECD Publishing. <http://dx.doi.org/10.1787/5k46bj4f03s7-en>, <http://www.oecd-ilibrary.org/docserver/download/5k46bj4f03s7.pdf?expires=1385822932&id=id&accname=guest&checksum=62792E49ED3863A2CF9B8E9046597CB7>

¹⁸⁷ http://www.oecd-ilibrary.org/governance/government-at-a-glance-2013_gov_glance-2013-en



Joonis 5. OECD riikide tsentraalsetes portaalides publitseeritud avaandmehulkade arvud (2013)¹⁸⁸

Nimetatud ülevaates mainitakse, et 56% OECD liikmesriikidest on ühtne rahvuslik avaliku sektori avaandmete strateegia, 12%-l riikidest on olemas erinevad valdkondlikud (nt ministeeriumide) strateegiad ja 28%-l riikidest on olemas mõlemad strateegiad. Ainult 4 %-l OECD maadest pole üldse avaandmete strateegiat.

2.2.2 Euroopa Komisjoni ja tema peadirektoraatide algatused

Euroopa Komisjoni osakondi kutsutakse peadirektoraatideks.¹⁸⁹ Iga peadirektoraat tegeleb teatud valdkonna poliitikaga. Käesoleva projekti teemaga on enim seotud informaatika peadirektoraat (DIGIT), sidevõrkude, sisu ja tehnoloogia peadirektoraat (DG-CONNECT) ja Eurostat (ESTAT).

Informaatika peadirektoraadi initsiatiivid (*Directorate-General for Informatics – DIGIT*)

Informaatika peadirektoraadi ülesanneteks on välja töötada Euroopa Komisjoni IT strateegia, pakkuda Euroopa institutsioonidele IT infrastruktuuri lahendusi, e-teenuseid, tugiteenuseid ja telekommunikatsiooni vahendeid, pakkuda Ekle infosüsteeme e-komisjoni strateegia raames, propageerida e-riigi teenuste kasutuselevõttu Euroopa avalike haldusasutuste hulgas.

Järgnevalt on esitatud käesoleva projekti teemaga seotud DIGIT-i initsiatiivid:

Vabavara strateegiate väljatöötamine. DIGIT töötas välja vabavara strateegia perioodiks 2011-2013 ja on loonud ka varem kasutusel olnud vabavara alase praktika teabevärava <http://osor.eu>, mis nüüd on ühendatud portaaliga Joinup.¹⁹⁰

ISA. DIGIT haldab Euroopa Komisjoni avaliku halduse koosvõime programmi (ISA)¹⁹¹, mille eesmärgiks on toetada ja hõlbustada tõhusat ja tulemuslikku piiriülest elektroonilist koostööd Euroopa haldusasutuste

¹⁸⁸ Tsentraalseid portaale mitteomavad riigid puuduvad joonisel. Allikas: <http://dx.doi.org/10.1787/888932942868>

¹⁸⁹ Directorate-General –DG, <http://ec.europa.eu>

¹⁹⁰ <https://joinup.ec.europa.eu/community/osor/description>

vahel. Programmi põhieesmärgiks on avalike e-teenuste pakkumise võimaldamine ning ühiste lahenduste koostöö, taaskasutuse ja kättesaadavuse tagamine. ISA programm toetab paljusid tegevusi, mis on seotud semantilise jm koostöömega erinevate EL riikide haldusinfosüsteemide vahel.¹⁹² Käesoleva projektiga on tihedalt seotud järgmised ISA tegevusalad/projektid: haldusandmete taaskasutuse võimaldamine¹⁹³ ja INSPIRE.¹⁹⁴

ISA propageerib semantilist koostöömet läbi linkandmete. ISA liikmesmaade esindajad kiitsid heaks avaliku sektori linkandmete spetsifikatsiooni 2012. a. ISA töötab välja parimaid praktikaid linkandmete tehnoloogiate kasutuselevõtuks e-riigi valdkonnas, viib läbi avaliku sektori linkandmete pilootprojekte, koostab e-riigi valdkonna sõnastikke ja pakub IKT spetsifikatsioonide otsimise veebiteenust, mis baseerub linkandmete tehnoloogiate kasutamisel. Teenus Cesar¹⁹⁵ võimaldab otsida 1300 semantikavara jm spetsifikatsiooni hulgast, millesse on panustanud 15 organisatsiooni. See teenus kasutab Varakirjelduse Metaandmete Skeemi (ADMS)¹⁹⁶, mis on standarditud metaandmete sõnastik ühtseks semantiliste varade kirjeldamiseks.

Novembris 2013 ilmusid ka ISA programmi raames tehtud uuringu „*Study on Business Models for Linked Open Government Data – BM4LOGD*“ tulemused¹⁹⁷. Selle uuringu kokkuvõtte esitame uuringute osas (vt 2.4 Valdkonna juhtivate riikide parimate linkandmete alaste praktikate analüüs).

Epractice. EPractice on Euroopa komisjoni e-riigi lahenduste praktikate vahendamise foorum¹⁹⁸. Epractice.eu on EK poolt loodud portaal pakkumaks infot ja teenuseid e-riigi, e-kaasamise, ja e-tervise asjatundjatele. Portaali pakub suurt hulka juhtumiuuringuid, mis on esitatud erinevate EL riikide poolt. Samuti publitseeritakse portaalis Euroopa e-riigi ja e-kaasamise kohta teabelehti iga liikmesriigi jaoks.¹⁹⁹

Joinup (varem osor). Joinup²⁰⁰ on Euroopa Komisjoni avaliku halduse koostöömet lahenduste jagamise foorum. See on osalus- ja koostööplatvorm, mida EK finantseerib ISA programmi abil. Selle põhieesmärk on koostöömet lahenduste alase info ja kogemuste vahendamine e-riigi ekspertide hulgas. Põhirõhk on semantikavarade ja avatud lähtekoodiga tarkvara leidmisel, taaskasutamisel ja arendusel. Koostööplatvormiga on võimalik liituda ekspertidel nii EL riikidest kui väljastpoolt.

SEMIC. Semic²⁰¹ on e-riigi lahenduste semantilisele koostöömetele suunatud initsiatiiv, mille portaali on nüüd ühendatud joinup-iga pakkudes foorumit ning võrgustikku e-riigi projektidele ja initsiatiividele. Semic-u põhiliseks väärtuseks on semantilise koostöömet varade deponitoorium.²⁰²

Sidevõrkude, sisu ja tehnoloogia peadirektoraat (Communications Networks, Content and Technology – DG-CONNECT)

DG Connect²⁰³ haldab EL digitaalarengu tegevuskava²⁰⁴. Peale selle on nimetatud peadirektoraadi ülesandeks toetada kõrgetasemelist teadus ja arendustööd ning innovatsiooni; soodustada loomingulisust läbi Euroopa andmete väärtusahela; edendada digitaalsete toodete ja teenuste suuremat avalikku kasutamist ning tagada nende turvalisus. DG Connect üheks tähtsaks tegevuseks on ka toetada vaba Interneti üle kogu maailma.

¹⁹¹ The Interoperability Solutions for European Public Administrations, ISA, ISA teabevärv <http://ec.europa.eu/isa/>

¹⁹² http://ec.europa.eu/isa/actions/index_en.htm

¹⁹³ http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-15action_en.htm

¹⁹⁴ http://ec.europa.eu/isa/news/2013/the-benefits-of-geo-location-services-for-egovernment_en.htm

¹⁹⁵ <https://joinup.ec.europa.eu/community/cesar/description>

¹⁹⁶ Asset Description Metadata Schema, ADMS, <https://joinup.ec.europa.eu/asset/adms/release/all>

¹⁹⁷ <https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

¹⁹⁸ <http://epractice.eu>

¹⁹⁹ <http://www.epractice.eu/en/factsheets/>

²⁰⁰ <https://joinup.ec.europa.eu>

²⁰¹ <https://joinup.ec.europa.eu/community/semic/description>

²⁰² <https://joinup.ec.europa.eu/community/semic/highlights/asset/all>

²⁰³ <http://ec.europa.eu/dgs/connect/>

²⁰⁴ The Digital Agenda of the EU, <http://ec.europa.eu/digital-agenda/>

Euroopa Komisjoni digitaalarengu tegevuskava 2020²⁰⁵ lasti käiku 2010. a mais, see sisaldab 101 tegevust, mis on grupeeritud 7 prioriteetsesse valdkonda. Põhieesmärk on aidata Euroopa kodanikel ja firmadel maksimaalselt ära kasutada digitaalseid tehnoloogiaid.

Euroopa Komisjoni digitaalarengu tegevuskava 2020 on seotud käesoleva projekti teemaga, eelkõige avaandmete²⁰⁶ ja nende portaalide²⁰⁷ osas aga ka pilvearvutuse, turvalisuse jm seonduvate teemade osas.

Digitaalarengu tegevuskava raames anti 2003. a välja avaliku sektori teabe taaskasutamise direktiiv nn PSI direktiiv²⁰⁸, mis andis seni reguleerimata Euroopa avaliku sektori andmete turule ühise õigusliku raamistiku. Euroopa Parlament kinnitas 13. juunil 2013. a PSI direktiivi uuendatud versiooni.

DG Connect üheks ülesandeks on ka IKT valdkonna teadusuuringute ja innovatsiooniprojektide rahastamine. Projektide ülevaade on toodud eraldi jaotises (vt 2.2.3 *Euroopa Komisjoni raamprogrammid*).

Eurostatistika peadirektoraat (ESTAT)

Eurostat²⁰⁹ on EL statistikaamet, kelle ülesanne on pakkuda ELle Euroopa tasemel statistikat, mis oleks võrreldav eri riikide ja regioonide vahel.

Eurostat pakub avalikele andmetele juurdepääsu läbi Eurostati levitusveebiteenus²¹⁰. Teenus võimaldab nii juurdepääsu andmetikele kui ka nende struktuuri määratlustele (metaandmetele). Lisaks saab alla laadida mingi andmestiku alamhulga või andmestiku tervikuna. See tasuta veebiteenus väljastab andmed SDMX 2.0 ja 2.1 formaatides. SDMX (Statistical Data and Metadata eXchange, sdmx.org) standardib rahvusvaheliselt statistiliste andmete organiseerimise ja vahetamise viisid.

Tervishoiu ja tarbijate peadirektoraadi tegevused (DG Health & Consumers - DG SANCO)

Tervishoiu ja tarbijate peadirektoraat²¹¹ tegeleb rahvatervise kaitsmise ja parendamisega, Euroopa toiduohutusega ning põllumajandusloomade heaolu ja tervise ning põllukultuuride ja metsade kaitsega.

Tervishoiu ja tarbijate peadirektoraat pakub avalikke linkandmeid oma valdkonnas, nt pestitsiidide ja toidulisandite andmehulgad jm. Praeguseks on avatud 12 andmehulka, mis on ka juurdepääsetavad läbi EL avaandmete portaali²¹², mida haldab EL väljaannete talitus²¹³. Tervishoiu ja tarbijate peadirektoraat võttis osa ka ISA linkandmete pilootprogrammist, mis kasutas linkandmete tehnoloogiat 8 EL liikmesriigi taimekaitsevahendite andmete integreerimiseks²¹⁴ (vt 2.4 *Valdkonna juhtivate riikide parimate linkandmete alaste praktikate analüüs*).

²⁰⁵ <https://ec.europa.eu/digital-agenda>

²⁰⁶ <https://ec.europa.eu/digital-agenda/en/public-sector-information-raw-data-new-services-and-products>

²⁰⁷ <https://ec.europa.eu/digital-agenda/en/open-data-portals>

²⁰⁸ The Directive on the re-use of public sector information (Directive 2003/98/EC tuntud kui 'PSI Directive') <https://ec.europa.eu/digital-agenda/en/legislative-measures>

²⁰⁹ <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>

²¹⁰ http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

²¹¹ DG Health & Consumers is a DG of the European Commission, http://ec.europa.eu/dgs/health_consumer/about_us/who_we_are_en.htm

²¹² <http://open-data.europa.eu>

²¹³ <http://publications.europa.eu/>

²¹⁴ Linking data about applications and decisions for authorisation of plant protection products, <http://health.testproject.eu/PPP/>

2.2.3 Euroopa Komisjoni raamprogrammid

FP7 IKT tööprogramm

Euroopa Komisjon on toetanud tervet rida teadus- ja arendusprojekte oma FP7 IKT tööprogrammi raames²¹⁵ selleks, et vähendada nii linkandmete loojate kui kasutajate sisenemistõkkeid valdkonda.

Nende projektide hulgas on suurema mõjuga LOD2²¹⁶, LATC²¹⁷ ja PlanetData²¹⁸ projektid.

LOD2 on laiaulatuslik integreeritud projekt, mis ühendab juhtivaid lingitud avaandmete tehnoloogiate uurijaid, firmasid ja teenustepakkujaid 11 EL riigist selleks, et parendada veebis publitseeritud andmete sidusust ja kvaliteeti, luua usaldus lingitud andmete veebi vastu, muuta RDF andmete haldus efektiivsemaks ning kõige sellega vähendada tõkkeid linkandmete loomisel ja kasutamisel. LOD2 projekti üheks ülesandeks oli luua EL avaandmete portaali prototüüp (<http://publicdata.eu>). Seoses sellega viidi projekti raames läbi sidusrühmade uuring²¹⁹ saamaks ettekujutus erinevate sihtrühmade nõuetest EL tsentraalsele avaandmete portaale. Eesmärk oli jõuda nii kodanike, avaliku halduse, poliitikute kui erasektori esindajateni kõigis 27 EL riigis. Küsitlusest võttis osa 329 inimest vastates 19le küsimusele perioodil 8.11.2009 - 15.12.2011. Küsitluse tulemused on publitseeritud ka linkandmetena veebis.²²⁰

Küsitluse analüüsi põhitulemused olid järgmised:

- Andmeformaatide eelistuste suundumus on RDF/XML ja APIde suunas, ka JSON leidis eraldi märkimist.
- Andmekvaliteedi osas on tähtsamad andmete allikate info, andmete formaat ja metaandmete täielikkus.
- Potentsiaalsed kasutajad sooviksid avaandmeid kasutada põhiliselt uuringuks/analüüsiks, visualiseerimiseks ja lihtsalt kasutamiseks.
- Potentsiaalsed kasutajad soovivad, et portaali pakuks eelkõige toorandmeid, informatsiooni andmehulkade versiooni kohta ja võimalust andmehulki otsida, grupeerida.
- Kasutajad soovivad, et avaandmete kataloogis oleks vajadusel võimalik andmehulki salvestada (säilitada). Küsimus portaalis ainult andmehulkade metaandmete ja linkide avaldamisest ei saanud selget vastust.

LATC on spetsiaalne FP7 toetusmeede, mille missioon on toetada linkandmete loomist ja kasutamist inimeste ja organisatsioonide poolt. Põhiülesanneteks on luua katsetander andmeid intensiivselt kasutavate rakenduste jaoks publitseerides andmestikke, mis on loodud Euroopa institutsioonide poolt linkandmetena tekitades sel moel EL andmete pilve. Pilve ühendatakse järgmiste asutuste andmed: Euroopa Komisjoni finantslääbipaistvuse süsteem, CORDIS, EUR-Lex EL seadustele juurdepääsu süsteem, Eurostat ja Euroopa Keskpank jt.

Planetdata projekti põhieesmärk oli luua Euroopa Komisjoni ava- ja linkandmete portaali prototüüp, mida vaatleme portaali jaotises (vt 2.3 *Andme- ja teabevaravate analüüs*).

Optique projekt²²¹ väärrib äramainimist, sest on pühendatud suurandmete lõppkasutajatele andmeanalüüsi tarbeks skaleeruva juurdepääsu loomisele. Loodav lahendus peaks pakkuma semantilist ühendust kasutajate

²¹⁵ FP7 Information and Communication Technologies Work Programme, <http://cordis.europa.eu/fp7/ict/>

²¹⁶ <http://lod2.eu/>

²¹⁷ Linked Open Data Around-The-Clock, <http://latc-project.eu/>

²¹⁸ <http://planet-data.eu/>

²¹⁹ <http://wiki.lod2.eu/download/attachments/3473453/AnalysisSurvey.pdf?version=1&modificationDate=1329941495000>

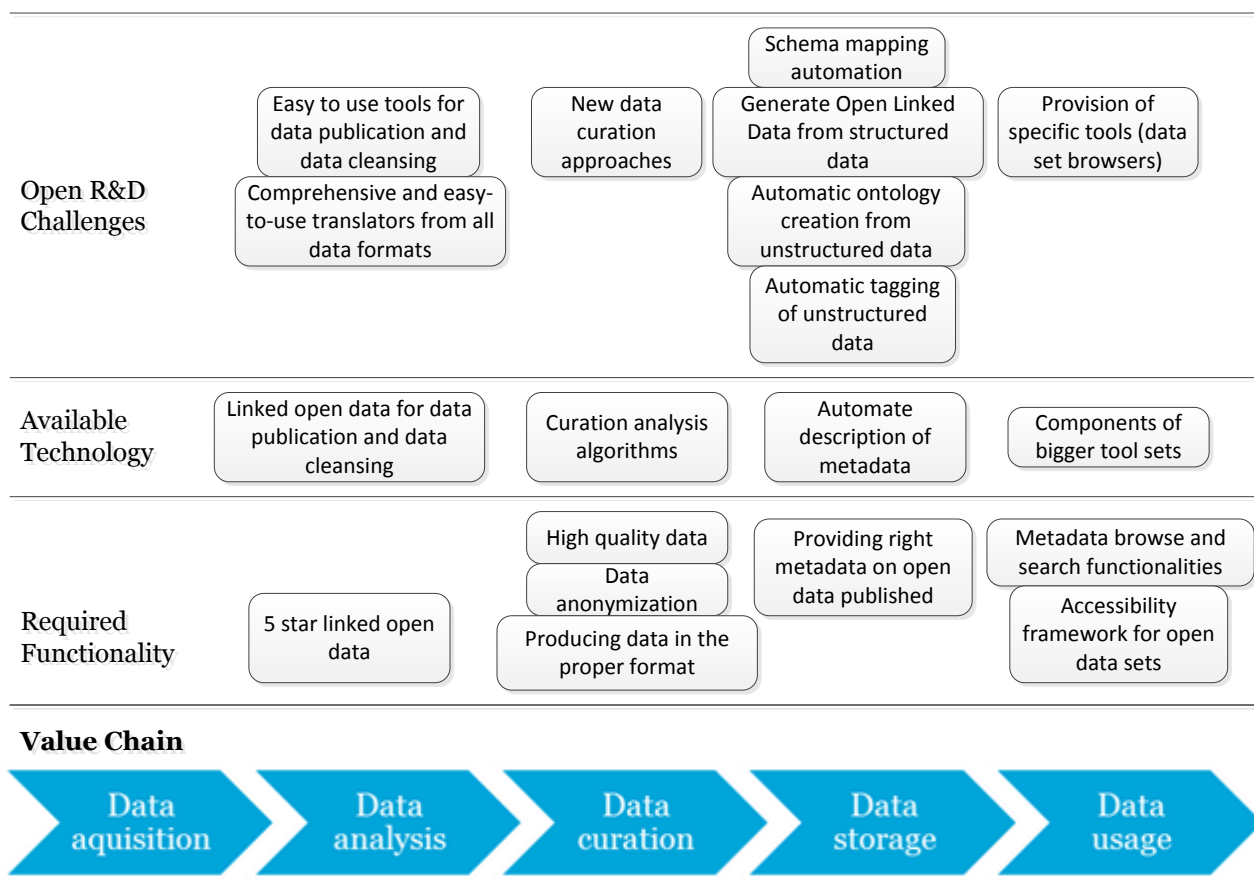
²²⁰ <http://data.lod2.eu/2010/ogd-survey/>

²²¹ <http://www.optique-project.eu/>

ja andmeallikate vahel võimaldades esitada kiireid intuiitviseid päringuid kasutades hästituntud sõnastikke ja kontseptualisatsioone. See lahendus kasutab laialdaselt ontoloogiaid.

Suurandmete avaliku foorumi projekt BIG²²² kujutab endast suurandmete tehnoloogiate ümber koondunud Euroopa majandustegevusharude kogukonda selleks, et propageerida suurandmete tehnoloogiate kasutuselevõtu esimesi tulemusi ja mõjutada otsustajaid ning võidelda poliitiliste ja õiguslike barjääridega. Projekti BIG üheks juba avaldatud tulemiks on suurandmete tehnoloogiline teekaart/tegevuskava²²³, mis sisaldab ka järgmist avaliku sektori avaandmete tehnoloogilist teekaarti/tegevuskava (vt allolev joonis).

Technical Roadmap Open Government Data



Joonis 6. Avalike andmete tehnoloogiline teekaart²²⁴

Suurandmete tehnoloogiate rolli avalike andmete juures nähakse mitte toorandmete publitseerimisel, vaid need on kasulikud avaandmete publitseerimisprotsessi harmoneerimisel ja aitavad uuendada dünaamiliselt üleslaetavaid avaandmeid ning luua kättesaadavate andmete avaldamiseks ühtset juurdepääsupunkti.

Envision projekt²²⁵ on seotud keskkonnateenuste infrastruktuuri loomisega. Tulem on mõeldud tavakasutajatele semantiliselt keskkonnateenuste avastamiseks, sidumiseks ja kompositsiooniks.

²²² <http://big-project.eu/>

²²³ http://big-project.eu/sites/default/files/D2.4.1_FINAL_vo_8.pdf

²²⁴ *Ibid.*

²²⁵ <http://www.envision-project.eu/>

TELEIOS projekt²²⁶ vaatleb maa seire andmeid, mis on suurandmed, mille suurust iseloomustab igapäevane mitme terabaidi suurune andmehulk saaduna satelliitidelt. Selline andmete hulk on väljakutse ESA, NASA ja Euroopa rahvuslikele kosmoseagentuuridele. TELEIOS projekt arendas välja tehnoloogiad nn virtuaalsete maa observatooriumide loomiseks. Need võimaldavad satelliitide andmeid kombineerida teiste väliste andmetega (nt maakaartide või muu veebis leiduva infoga) ja siis eraldada teadmisi, mida saab kasutada loomaks rakendusi maa seire teadlaste, otsustajate või üldsuse jaoks. Virtuaalsete observatooriumide loomiseks kasutatakse selles järgmisi tehnoloogiaid: ontoloogiaid, lingitud georuumilised ja ajalised andmed, spetsiaalseid massiivandmete või rasterandmete andmebaasisüsteeme ja andmekeldri (*data vault*) tehnoloogiad, satelliitide andmetest teadmiste tuvastamise/avastamise tehnikad, visuaalsed kasutajaliidesed. TELEIOS projekti raames loodi reaalselt töötav metsatulekahjusid jälgiv süsteem, mis kasutab semantilise veebi ja linkandmete tehnoloogiaid.²²⁷

EUCLID²²⁸ on projekt, mille tulemusena loodi linkandmete õppekava koos õppematerjalidega, et koolitada linkandmete praktikas kasutajaid. Õppematerjalid on avalikult kättesaadavad projekti kodulehel.

IKS²²⁹ on vabavara kogukond, mille projektid on suunatud semantiliselt rikastatud sisuhaldussüsteemide jaoks vaba ja paindliku tehnoloogilise platvormi loomisele, mis hõlmab ka linkandmeid. Pikemaks eesmärgiks on semantiliste tehnoloogiate arengule ja kasutuselevõtule kaasaaitamine, lootusega asendada nendega praegu kasutusel olevad LAMP (Linux, Apache, mySQL, PHP) ja Java-põhised sisuhaldussüsteemid. IKS-i projektide tulemusena loodud tarkvara on vabavara ja IKS kogukonnaga ühinemine on samuti vaba. IKS-i projektid on eriti suunatud keskmistele ja väikestele sisuhaldusettevõtetele. Üks selle kogukonna põhiline projekt on Apache Stanbol²³⁰, mis on vabavaraline modulaarne tarkvara ja hulk taaskasutatavaid komponente semantiliseks sisuhalduseks.

Horizon 2020

Horizon 2020 on uus EL teaduse ja innovatsiooni raamprogramm perioodiks 2014-2020²³¹. Vastavad raamprogrammi dokumendid on saadaval Horizon2020 kodulehel.²³²

Järgnevalt toome esile käesoleva uuringu teemaga enimseonduvad uurimistoetuste teemad 2014-2015 programmist „ICT LEIT Work Programme 2014-15“.²³³

Tulevikuinterneti väljakutse all pakutakse rahastamiseks järgmisi teemasid: uued mitte põhiarvuti/peaarvuti kesksed internetiarhitektuurid, uued optilised ja traadita ühendused, uued kõrgjõudlusega pilvelahendused (sh turvalisus) ja teenused (sh automaatne teenuste kompositsioon ja suurandmete tugi), avaliku sektori produktiivsuse ja innovaatilisuse tõstmine pilvearvutuse teenuste kasutuselevõtuga, kollektiivse teadvuse platvormid jätkusuutliku ja sotsiaalse innovatsiooni jaoks, veebiettevõtlus, 5G integreeritud, üldlevinud ja ultra-suure võimsusega võrgud (seotud asjade internetiga).

Sisutehnoloogiate ja infohalduse väljakutse all on järgmised neli võtmevaldkonda: suurandmete alane innovatsioon (nii tooted kui teenused) ja suurandmete kasutuselevõtt, masintõlge, vahendid loomemajanduse, teadmus- ja haridussektori innovatsiooni toetuseks, multimodaalne ja loomulik suhtlus arvutiga.

Suurandmete teema on selles programmis tihedalt seotud avaandmetega pakkudes rahastust väike- ja keskmistele ettevõtetele Euroopa avaandmete integratsiooni ja taaskasutuse inkubaatorite tekitamiseks eesmärgiga kiirendada avaandmete tarneahelate arengut.

²²⁶ <http://www.earthobservatory.eu/>

²²⁷ Koubarakis, M., et al.: Real-time wildfire monitoring using scientific database and linked data technologies. In: In the 16th International Conference on Extending Database Technology (EDBT 2013). Genoa, Italy (March 18-22 2013)

²²⁸ <http://euclid-project.eu/>

²²⁹ Interactive Knowledge Stack, <http://www.iks-project.eu/>

²³⁰ <http://www.iks-project.eu/projects/apache-stanbol>

²³¹ http://ec.europa.eu/research/horizon2020/index_en.cfm

²³² http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=h2020-documents

²³³ <https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/H2020%20LEIT-ICT%20WP%202014-15%20-%202031%2010%202013.pdf>

Horisontaalsed tegevused on eelkõige seotud asjade interneti ja nutikate objektide ühendamise platvormidega. Näiteks on aktuaalsed järgmised teemad: nutikate süsteemide integratsioon, nutikad võrgud, suurandmed jt. Nimetatud tehnoloogiate kasutusvaldkondadena nähakse e-tervist, toiduahelat, intelligentset transporti ja logistikat.

Asjade internetti, mis on üks nutika keskkonna ja nutikate platvormide osa, nähakse kui ühte järgmistest tähtsatest mõistetest, mis toetab sotsiaalseid väljakutseid ja majanduslikku kasvu hinnanguliselt 20% aastas.

2.2.4 Standardid

W3C standardid ja soovitusel

Rahvusvaheline veebikonsortium W3C²³⁴ on välja arendanud ja soovitanud terve rea standardeid²³⁵, millest osa on käesoleva projekti teema ja uurimisküsimustega otseselt seotud. Peale selle on W3C ellu kutsunud e-riigi huvigrupi²³⁶, kes korraldab internetis koosolekuid erinevatel e-riigiga seotud teemadel. Näiteks toimus 2013. a aprillis avaliku sektori avaandmete alane koosolek, millel esitatud ettekannetest nii mõnedki oleks Eestile õpetlikud. Näiteks, ettekanne „Open Government Data W3C Discussion Summary“²³⁷ võtab kokku 20 avaliku sektori avaandmete juhtumit, mis olid esitatud varasematel koosolekutel ja analüüsib neid W3C e-riigi huvigrupi loodud teekaardi valguses.

W3C lõi 2011. a spetsiaalse avalike linkandmete töögrupi.²³⁸ Töögrupp andis novembris 2013 välja linkandmete parimate publitseerimispraktikate teadaande²³⁹, mis sisaldab peale nõuannete ka linkandmete elutsükli mudeli. See on järg 2011. a avaldatud dokumendile „W3C's Cookbook for Open Government Linked Data“²⁴⁰. 2013. a avaldati ka linkandmete terminite sõnastik²⁴¹. W3C poolt pakutud parimad linkandmete publitseerimise praktikad võib kokku võtta järgmiselt:

- Modelleeri andmed;
- Taaskasuta sõnastikke niipalju kui võimalik;
- Nimeta asju püsivate URI-dega;
- Avalda andmete kohta nii inim- kui masinloetavad kirjeldused (metaandmed);
- Teisenda andmed RDF-i;
- Määra sobiv litsents;
- Avalikusta linkandmed ja kuuluta nad välja.

Töögrupp haldab ka avalike linkandmetega seotud organisatsioonide ja projektide kataloogi²⁴². Eraldi väärib märkimist W3C Biomeditsiiniliste andmete alamtöögrupp²⁴³, kes tegeleb ka biolinkandmetega.

Käesoleva projekti teemaga seotud W3C standardid ja soovitusel on järgmised:

- RDF, <http://www.w3.org/TR/rdf-concepts/>

²³⁴ <http://www.w3.org/>

²³⁵ <http://www.w3.org/standards/>

²³⁶ <http://www.w3.org/egov>

²³⁷ Tomasz Janowski, Jeanne Holm, Elsa Estevez, Open Government Data W3C Discussion Summary, http://www.w3.org/egov/wiki/images/5/5b/Open_government_data_5.pdf

²³⁸ The Government Linked Data (GLD) Working Group, <http://www.w3.org/2011/gld>

²³⁹ <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

²⁴⁰ http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

²⁴¹ <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html>

²⁴² <http://dir.w3.org>

²⁴³ http://www.w3.org/wiki/HCLSIG_BioRDF_Subgroup

- Hyper Text Markup Language (HTML), <http://www.w3.org/html/>
- Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- Web Ontology Language (OWL), <http://www.w3.org/TR/owl2-overview/>
- Extensible Markup Language (XML), <http://www.w3.org/XML/>
- Linked data principles , <http://www.w3.org/standards/semanticweb/data;>
<http://www.w3.org/DesignIssues/LinkedData>
- Cool URIs for the Semantic Web, <http://www.w3.org/TR/cooluris/>
- Publishing Open Governmental Data, <http://www.w3.org/TR/gov-data/>
- Linked Data Platform (LDP), Linked Data Platform Best Practices and Guidelines, <https://dvc.w3.org/hg/ldpwg/raw-file/default/ldp-bp/ldp-bp.html#bib-LDP-PRIMER>
- Linked Data Platform (LDP) 1.0 Primer, <https://dvc.w3.org/hg/ldpwg/raw-file/default/ldp-primer/ldp-primer.html>
- Data catalog vocabulary (DCAT), <http://www.w3.org/TR/vocab-dcat/>
- Data Cube Vocabulary, <http://www.w3.org/TR/vocab-data-cube>
- R2RML: RDB to RDF Mapping Language, <http://www.w3.org/TR/r2rml/>
- XGR-geo, <http://www.w3.org/2005/Incubator/geo/XGR-geo/>

Sõnastikud ja ontoloogiad

Järgmised sõnastikud ja ontoloogiad on seotud käesoleva projektiga:

- Lihtne teadmuse organiseerimise süsteem (*Simple Knowledge Organization System – SKOS*), <http://www.w3.org/2004/02/skos/>
- Dublincore metaandmete skeem, <http://dublincore.org/>
- *Friend of a Friend (FOAF)* sõnastik, <http://xmlns.com/foaf/spec/>
- Linkandmetega seotud terminite sõnastik, <https://dvc.w3.org/hg/gld/raw-file/default/glossary/index.html>
- Organisatsiooniontoloogia (*Organisation ontology, 2013*), <http://www.w3.org/TR/vocab-org/>
- Varakirjelduse Metaandmete Skeem (*Asset Description Metadata Schema, ADMS*) <https://joinup.ec.europa.eu/asset/adms/release/all>
- Isikute kirjeldamise sõnastik, <http://www.w3.org/TR/vocab-people/>

Statistika standardid

Statistikaametid kasutavad rahvusvahelist standardit SDMX²⁴⁴, mis kujutab endast statistiliste andmete kirjeldamise loogilist mudelit ja tagab arvutitevahelise automaatse kommunikatsiooni andmete vahetusel ja

²⁴⁴ <http://www.sdmx.org>

töötusel. SDMX-iga kaasneb ka vastav tehnoloogia, mis toetab standardiseeritud IT vahendeid, mida statistiliste andmete vahetuses ja töötustes osalevad pooled peavad kasutama.

Georuumiliste andmete standardid

Avatud georuumiline konsortsium (*Open Geospatial Consortium – OGC*) lasi hiljuti (2012. a) välja standardi staatiliste georuumiliste andmete jaoks, s.o geograafilise päringukeele GeoSPARQL RDF andmetele.²⁴⁵

OGC standardi baasil on loodud stRDF ja stSPARQL (SPARQL 1.1 laiendus), mis on mõeldud päringute esitamiseks ajas muutuvatele georuumilistele linkandmetele²⁴⁶. Seda kasutatakse süsteemis Strabon, mis on geolinkandmete andmebaas²⁴⁷.

Lisaks kasutatakse georuumilises valdkonnas vektor- ja rasterformaate nagu KML, GeoJSON, GeoTIFF jt. LinkedGeoData projekt kasutab OpenStreetMap projekti poolt kogutud andmeid selleks, et teha need avatud linkandmetena kättesaadavaks ja linkida siis teiste avatud linkandmetega.²⁴⁸

2.2.5 Soovitused Eestile

Rahvusvaheliste initsiatiivide ja standardite analüüsi tulemusena esitame soovitused järgnevate uurimisküsimuste lõikes:

Millistes rahvusvahelistes initsiatiivides peaks Eesti olema aktiivne, sh riiklikul tasemel (prioriteetsuse järjekorras)? Millised kompetentsid ja ressursid peaksid iga sellise initsiatiivi lõikes Eestil osalemiseks olemas olema?

Eesti osaleb juba mitmetes rahvusvahelistes initsiatiivides, mis tegelevad muuhulgas ka ava- ja linkandmete valdkonna edendamise. Nendest initsiatiividest võiks nimetada OECD, Avatud riigi partnerlusprogramm²⁴⁹, EK digitaalne tegevuskava ja e-valitsemise tegevuskava 2011-2015²⁵⁰, EK koosvõime programm ja Euroopa koosvõime raamistik²⁵¹ ning Euroopa avaliku sektori info tribüün (ePSIplaform).

Et ODI laieneb praegu erinevatesse riikidesse (sh Venemaa), siis võiksid ka Eesti äriettevõtted, idufirmad ja organisatsioonid astuda ODI liikmeks. Näiteks võiks Eesti ITL astuda ODI liikmeks ODI toetajana (liikmemaks 720 Inglise naela aastas). ODI partneri liikmemaks on ilmselt liiga suur Eesti firmadele (70000 Inglise naela aastas).

Eesti riigiasutused, teadusasutused ja erasektor võiksid aktiivselt osaleda H2020 projektikonkurssidel ja kasutada FP7 projektide tulemusi.

Millised linkandmetega seotud avatud standardid on Eesti jaoks olulised (sh piisavalt küpsed kasutuselevõtuks)?

Linkandmete standardid on rahvusvahelised ja eelkõige tuleks lähtuda järgmistest W3C ja teiste organisatsioonide rahvusvahelistest standarditest: URI, HTTP, RDF, RDFS, OWL, SKOS, SPARQL 1.1, W3C isiku ja organisatsiooni ontoloogiad, Dublin Core, Varakirjelduse Metaandmete Skeem.

Lisaks peaks lähtuma statistika ja ruumiandmete standarditest.

²⁴⁵ Open Geospatial Consortium: OGC GeoSPARQL - A geographic query language for RDF data. OGCR Implementation Standard (2012) GeoSPARQL <http://www.geosparql.org/>

²⁴⁶ <http://www.strabon.di.uoa.gr/stSPARQL>

²⁴⁷ K. Bereta, P. Smeros and M. Koubarakis. Representing and Querying the Valid Time of Triples for Linked Geospatial Data. In the 10th Extended Semantic Web Conference (ESWC 2013). Montpellier, France. May 26-30, 2013.

²⁴⁸ <http://linkedgeo.org/>

²⁴⁹ <http://www.opengovpartnership.org/>

²⁵⁰ <https://ec.europa.eu/digital-agenda/en/european-egovernment-action-plan-2011-2015>

²⁵¹ http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

2.3 Andme- ja teabevärvate analüüs

Andme- ja teabevärvate analüüs võimaldab näha, milline osa on ava- ja linkandmete portaalidel linkandmete publitseerimisel ja kasutuselevõtus. Lisaks on teabevärvate analüüs seotud uurimisküsimusega *sisutööluse ümberkorraldamisest veebisaitides ja –liidestest*.

2.3.1 Ava- ja linkandmete portaalid

Viimastel aastatel on plahvatuslikult suurenenud huvi avaliku sektori informatsiooni avamiseks selle taaskasutuse eesmärgil. Seoses sellega on paljudes riikides üle maailma loodud tsentraalsed avaandmete portaalid selleks, et lihtsustada avalikkuse juurdepääsu riigi avaandmetele. Näiteks võib tuua UK data.gov.uk portaali, Prantsusmaa data.gouv.fr portaali või USA portaali www.data.gov. Avaliku sektori informatsiooni publitseerimine veebis vastavalt avaandmete printsiipidele on esimene samm, mis veel ei tähenda, et andmed on tehtud inimestele kergesti leitavaks. Avaandmete portaalid kujutavad endast kesket kohta andmete lehitsemiseks ja allalaadimiseks, paljudel juhtudel ka andmete säilitamiseks, tehes andmete leidmise taaskasutuse eesmärgil lihtsamaks. Taoliste portaalide loomist võib julgelt ennustada ka järgnevatel aastatel. Vastavalt Euroopa avaandmete kataloogi (<http://datacatalogs.org>) andmetele oli oktoobriks 2013. a loodud 363 avaandmete portaali nii rahvuslikul kui lokaalsel tasemel. EL liikmesriikides oli oktoobriks 2013. a registreeritud 117 EL liikmesriikide kataloogi (vt <http://datacatalogs.org/group/eu-official>).

Käesolevas jaotises vaatlemegi nii mõningate tähtsamate riikide kui ka EL üleseid avaandmete katalooge ja nende võimalusi avaandmete taaskasutuseks.

Tavaliselt koondavad tsentraalsed avaandmete kataloogid automaatselt (*ingl harvest*) palju erinevaid andmehulki erinevatest hajutatud avaliku sektori asutustest ühte tsentraalsesse portaali. Kuna paljud avaandmete initsiatiivid omavad oma avaandmete katalooge kasutades oma spetsiifilist tarkvara, siis sellistes riikides, kus tsentraalne kataloog puudub, on võimalik avaandmete andmehulkade nimekiri paigutada mõnda üldisesse kataloogi nagu www.datahub.io või www.datamarket.com selle asemel, et investeerida oma kataloogi pidamise infrastruktuuri.²⁵²

Arvukate kataloogide tekkimisega kerkib kataloogidevaheline andmete vahetuse ja koosvõime probleem. Seda probleemi on püütud lahendada initsiatiividega, mis nõuavad standardiseeritud andmete ja andmevahetusprotokollide kasutamist. Eesmärgiks on võimaldada ühendatud otsingut üle mitmete kataloogide (indeksite) ja lihtsat metaandmete ning andmete vahetust.

Paljud avaandmete portaalid, eriti aga riiklikud (nt USA, UK jt) avaandmete portaalid on teinud kas oma metaandmed või osa andmeid kättesaadavaks ka linkandmete kujul RDF formaadis. Tavaliselt on sel juhul linkandmete portaal vaadeldav kui avaandmete portaali osa. Teisalt aga leidub ka riiklikke linkandmete portaale nagu näiteks Soomes.²⁵³ Tihti on loodud ka vastav SPARQL lüüs. Seetõttu vaatleme allpool ava- ja linkandmete portaale koos. Samas vaatleme ka spetsiaalseid (ka vadlkondlikke) linkandmete portaale.

2.3.2 Portaalide ja kataloogide hetkeseis

Riiklikud portaalid ja kataloogid

Esimene riiklik avaandmete portaal Euroopas oli Inglismaa avaandmete portaal (<http://data.gov.uk/>), mis loodi 2009. a. Olles esimene Euroopas, andis see eeskuju teistele riikidele oma portaalide loomisel. 2012. aastal portaali taaskäivitati selle täiustamise eesmärgil ja praegu on see Euroopa kõige kuulsam ava- ja linkandmete portaal (<http://data.gov.uk/linked-data>). Portaali on loodud kasutades CKAN-i (<http://ckan.org/>) avaandmete publitseerimise platvormi.

²⁵² <http://www.epsiplatform.eu/sites/default/files/Final%20TR%20Europes%20Data%20Catalogs.pdf>

²⁵³ <http://www.ldf.fi/index.html>

Käesoleval ajal on paljudel riikidel oma avaandmete portaalid – näiteks EL-i avaandmete portaalis (<http://open-data.europa.eu/>) on lingid 16 EL riigi (sh Eesti) avaandmete portaalidele. Sealt puuduvad aga Soome (<http://www.julkinendata.fi/>), Norra (<http://data.norge.no/>) jt riikide portaalide lingid.

Ameerika Ühendriikides loodi föderaalvalitsuse avaandmete portaal (<http://www.data.gov/>) 2009. aastal. Portaali pakub 18.11.2013 seisuga 87787 andmestikku, põhiliselt georuumilisest, keskkonna ja transpordi valdkondadest (kokku 55% andmestikest). Populaarsemad andmeformaadid on HTML, XML, ZIP, JSON, KML, CSV, Excel. Praeguseks on portaali osaliselt ümbertehtud kasutades CKAN platvormi.

Rahvusvahelised portaalid ja kataloogid

EK avaandmete portaali. Euroopa Komisjoni avaandmete portaali (<http://open-data.europa.eu/>) on ühine juurdepääsupunkt EL institutsioonide avaandmetele, mis on vabaks kasutamiseks ja taaskasutamiseks nii äri- kui mitteäri-eesmärkidel. Portaali eesmärgiks on aidata kaasa EL institutsioonide ja muude asutuste tegevuse läbipaistvuse suurendamisele. Praegu on tegemist portaali beeta versiooniga, mis avati kasutamiseks 2012. a detsembris ja milles oli 23.10.2013 seisuga 6138 allalaaditavat avatud andmehulka ja 7 visuaalset rakendust. Põhiline andmete publitseerija on Eurostat (5970). Rakendustest on IKT alal huvitav DG-Connect poolt tellitud *The Digital Agenda Scoreboard*, mis võimaldab analüüsida Euroopa infoühiskonna indikaatoreid nii maade kui teemade järgi. Portaali on teinud oma kataloogi metaandmed kättesaadavaks linkandmete kujul (<http://open-data.europa.eu/en/linked-data>). Lisaks kogutakse ka linkandmeid. On olemas SPARQL päringupunkt metaandmete otsinguks. Portaali sisaldab linke EL liikmesriikide avaandmete saitidele ja võimaldab oma andmestikke portaali lisada.

See portaali on loodud ainult vabavara kasutades: Drupal sisuhaldussüsteemi (<https://drupal.org/>), CKAN avaandmete publitseerimise platvormi ja LAMP tarkvarapaketti ning OKF poolt loodud tarkvara. Portaali metaandmete kataloog on loodud Dublin Core metaandmete skeemi, andmekataloogi sõnastikku DCAT ja varakirjelduse metaandmete skeemi ADMS kasutades.

Euroopa Komisjoni avaandmete portaali eelkäija ja prototüüp oli portaali publicdata.eu, mis loodi 7nda raamprogrammi projekti LOD2 tulemusena OKF-i poolt 2011. a. Portaali on ka praegu avatud. Sinna saab lisada oma andmekataloogi linki. Portaali on loodud kasutades CKAN-i (<http://ckan.org/>) avaandmete publitseerimise platvormi. Seal oli 23.10.2013 seisuga 19568 andmehulka. Kõige rohkem oli andmehulki UK-st, s.o 11063. Vormingutest kasutatakse kõige rohkem CSV (3865), XLS, RDF ainult 136 faili. Portaal on viited 79 rakendusele.

Muud rahvusvahelised avaandmete portaalid. Paljudel rahvusvahelistel organisatsioonidel on oma avaandmete portaalid. Allpool esitatud tabelis on toodud nimekiri mõnedest tähtsamatest portaalidest:

Tabel 1. Tähtsamad rahvusvaheliste organisatsioonide avaandmete portaalid

Nimetus	Link
EUR-Lex Access to European Union law	http://eur-lex.publicdata.eu/
Eurostat	http://eurostat.linked-statistics.org/
European Central Bank (ECB)	http://ecb.publicdata.eu
European Data Cloud	http://latc-project.eu/datasets
United Nations	http://data.un.org
World Bank	http://data.worldbank.org
Linkandmete kogukonna sait	http://linkeddata.org/
Datamarket (Andmed kui teenus (DaaS) firma)	http://datamarket.com/
Datacatalogs (peetakse kõige ulatuslikumaks avaandmete kataloogiks)	http://datacatalogs.org/
Veebikogukonna avaandmete portaali Datahub	http://datahub.io/

Veebiindeksi veebi mõju uurimise kataloog

<http://thewebindex.org/>

Semantilise veebi indeks

<http://sindice.com/>

Valdkondlikud portaalid

Georuumiliste andmete portaalid. Viimastel aastatel on andmete veebi tekkinud hulgaliselt georuumilisi andmeid. Ordnance Survey, UK rahvuslik kaardistamisamet, oli esimene, kes tegi UK georuumilised andmed avalikuks kui linkandmed²⁵⁴. Teine näide selles valdkonnas on LinkedGeoData²⁵⁵, kus on kättesaadavaks tehtud OpenStreetMap (OSM) andmed RDF kujul ja päringuid saab neile esitada SPARQL abil. Võib veel lisada Hispaania georuumiliste linkandmete avalikustamise saidi GeoLinked Data²⁵⁶ ja Kreeka avaandmete portaali²⁵⁷, kus on publitseeritud osa Kreeka georuumilisi andmeid linkandmetena.²⁵⁸

Arhiivide ja raamatukogude portaalid. Paljudel raamatukogudel ja mäluasutustel on oma linkandmete ja avaandmete portaalid. Järgnevalt vaatleme neist tähtsamaid.

Europeana. Europeana (<http://www.europeana.eu>) on Euroopa muuseumide, raamatukogude, arhiivide ja audio-visuaalsete kollektsioonide digiressursside varamu, mida rahastavad Euroopa Komisjon ja 21 liikmesriigi kultuuriministeriumid.

Europeana linkandmete pilootprojekt (*Europeana Linked Open Data Pilot*²⁵⁹) sisaldab käesoleval ajal (20.11.2013) avatud metaandmeid 25 miljoni Europeana säilitusühiku (teksti, pildi, video ja helisalvestise) kohta. Põhiline juurdepääs lingitud metaandmetele on läbi veebiteenuse liidese Europeana API (<http://pro.europeana.eu/api>). Rakendusliides kasutab standardseid tehnoloogiaid nagu REST üle HTTP. Vastused tagastatakse JSON-i vormingus. Selle pilootprojekti SPARQL päringupunkt on <http://europeana.ontotext.com/>.

USA Kongressi Raamatukogu. Kongressi raamatukogu pakub linkandmete teenust²⁶⁰, mis võimaldab nii inimloetavat kui masinloetavat juurdepääsu autoriandmetele.

Saksamaa Rahvusraamatukogu. Saksamaa rahvusraamatukogu pakub avatud litsentsiga linkandmete teenust bibliograafiliste andmete osas²⁶¹. Alates 2010. a pakutakse autorite andmeid ja nüüd ka lisaks pealkirjade andmeid.²⁶²

Muud valdkondlikud portaalid. Peale ülaltoodud valdkondlike portaalide on loodud palju teisi valdkondlikke avaandmete portaale, nii riiklikke kui rahvusvahelisi. Näiteks USA tervishoiu andmete portaal (<http://healthdata.gov/>), kus andmed on saadaval põhiliselt CSV, XLS ja RDF vormingus.

2.3.3 Soovitused Eestile

Portaalide ja andmevõrivate analüüs näitas, et kesksed avaandmete portaalid on vajalikud selleks, et tagada parem juurdepääs avalikele avaandmetele sh avalikele linkandmetele. Linkandmetele on sellistes portaalides tavaliselt avatud juurdepääs SPARQL lüüsi kaudu. Lisaks luuakse palju linkandmete lüüse otse andmete

²⁵⁴ <http://www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html>

²⁵⁵ <http://linkedgeodata.org/>

²⁵⁶ <http://geo.linkeddata.es/>

²⁵⁷ <http://linkedopendata.gr/>

²⁵⁸ C. Nikolaou, K. Dogani, K. Kyzirakos and M. Koubarakis. Sextant: Browsing and Mapping the Ocean of Linked Geospatial Data. In the 10th Extended Semantic Web Conference (ESWC 2013). Montpellier, France. May 26-30, 2013. Demo paper.

²⁵⁹ <http://data.europeana.eu>

²⁶⁰ Library of Congress Linked Data Service <http://id.loc.gov/>

²⁶¹ Linked Data service of the DNB, <http://www.dnb.de/EN/lids>

²⁶² <http://www.epsiplatform.eu/sites/default/files/Final%20TR%20Linked%20Data.pdf>

tootja portaalis vastava valdkondliku teenusena (näiteks raamatukogud, arhiivid jt) või on juurdepääs linkandmetele organiseeritud vastava API abil.

Eesti võiks sellist rahvusvahelist praktikat järgida oma avaandmete portaali(de) loomisel ja avalikele linkandmetele juurdepääsu tagamisel. Rahvusraamatukogu ja rahvusarhiiv võiksid võtta eeskujuks USA, UK või Saksamaa raamatukogude ja arhiivide kogemused. Eriti peaks pöörama tähelepanu valdkonnas rahvusvaheliselt kasutatavate metaandmete standardite ja ontoloogiade kasutuselevõtmisele selleks, et tagada koosvõimelisus teiste riikide linkandmetega (st et oleks võimalik linkida näiteks meie ja teiste riikide raamatukogude ja arhiivide andmeid).

Eesti avalikud avaandmed peaksid olema kättesaadavad ka EL keskse avaandmete portaali kaudu.

2.4 Valdonna juhtivate riikide parimate linkandmete alaste praktikate analüüs

Riiklikul tasandil on linkandmete loomine seotud avaandmete poliitikatega. Nagu eespool (vt 2.1.1 *Linkandmete tehnoloogia*) mainitud, on linkandmed oma olemuselt avatud andmed, sest nende kasutegur avaldub just siis kui kriitiline mass andmeid on omavahel lingitud ja teistele kasutatavaks tehtud. See ei takista linkandmeid ka kinniste süsteemide puhul kasutamast nagu ka eespool vaatlesime.

Seega, käsitledes juhtivate riikide poliitikaid linkandmete valdkonnas, näeme, et need on otseselt seotud avaandmete poliitikatega, mis omakorda on seotud avatud valitsemise poliitikaga. Linkandmeid vaadeldakse nendes kui avaandmete ühte võimalikku esitusviisi.

Nii linkandmete kui seonduvate töusvate tehnoloogiade alal on maailmas esikohal USA ja UK. Avaandmete ja linkandmete alal on eriti eesrindlik UK. Näiteks, avaandmete baromeetri (*Open Data Barometer*) hiljutise globaalse uuringu järgi, mis avaldati oktoobris 2013, on avaandmete valdkonnas üldkokkuvõttes esikohal UK, USA, Rootsi, Uus-Meremaa, Norra ja Taani. Seega on analüüsitavateks riikideks valitud UK ja USA.

Allpool toodud analüüs on abiks järgmisele uurimisküsimusele vastamisel:

Kuidas korraldada Lingitud Eesti suunas viivate ettepanekute elluviimine? Kas ja milliseid organisatsioonilisi (ümber)korraldusi selle realiseerimiseks on vaja teha, sh riiklikul tasandil?

2.4.1 Suurbritannia (UK)

UKs on linkandmete kasutuselevõtt just viimastel aastatel hoo sisse saanud tänu UK valitsuse poolt 2012. aastal vastu võetud mitmetele järgmistele riigi IKT strateegiadokumentidele²⁶³:

- UK valitsuse IKT strateegia – *UK government ICT strategy 2011*²⁶⁴
- UK valitsuse referentsarhitektuur – *UK government reference architecture (UKRA) 27 April 2012*²⁶⁵;
- UK avaliku sektori infoprintsiibid – *Information principles for the UK public sector 30 April 2012*²⁶⁶;

²⁶³ <https://www.gov.uk/government/collections/ict-strategy-resources>

²⁶⁴ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/85968/uk-government-government-ict-strategy_o.pdf

²⁶⁵ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266257/UK-Reference-Architecture-V1-O-HMG-Branded.pdf

- UK avaliku sektori kontseptuaalne mudel – *Concept model for the UK public sector 30 April 2012*²⁶⁷;
- Avaandmete valge raamat – *Open data white paper June 2012*²⁶⁸.

UK valitsuse IKT strateegia aastast 2011 ütleb, et valitsus loob ühtse ja turvalise IKT infrastruktuuri, mis baseerub ühistel avatud standarditel, mis muudavad IKT lahendused täiesti koosvõimelisteks. Selle eesmärgi täitmiseks lubati publitseerida referentsarhitektuur, avada andmed ja rakendusliidesed selleks, et ettevõtteid ja kodanikke julgustada arendama uusi turuvõimalusi. UK valitsus lubas luua uue organisatsiooni *Public Data Corporation*.

UK valitsuse referentsarhitektuur (UKRA) on loodud tagamaks valitsuse info taaskasutuse ja koosvõime. See sisaldab täielikku teenuste taksonoomiat. Referentsarhitektuur on suur ettevõtmine ja selle tegemine pidev protsess. UKRA sisaldab muuhulgas ka informatsiooni referentsmudelit (IRM). See kujutab endast standardseid vahendeid info kirjeldamiseks, kategoriseerimiseks ja jagamiseks. Selles osas juhindutakse omakorda UK avaliku sektori informatsiooni printsiipidest ja kontseptuaalsest mudelist.

Loodi spetsiaalne referentsarhitektuuriga tegelev töögrupp, kelle üheks ülesandeks oli leida konsensus erinevate valitsusametite ja organisatsioonide vahel. Referentsarhitektuuri dokumendis nenditi, et konsensuse leidmine oli väljakutse.

Avaliku sektori informatsiooni printsiipide dokument toob välja 7 printsiipi, millest tuleb lähtuda UK avaliku sektori informatsiooni töötlemise elutsükli vältel. Nendest neljas printsiip ütleb, et informatsioon peab olema standardiseeritud ja lingitav. Seal sätestatakse avatud standarditele pühendumust ja ühiste organisatsiooni standardite kehtestamist, informatsiooni linkimise raamistikku ja informatsiooni standardiseeritud linkandmeteks üleviimise pragmaatilise lähenemisviisi kehtestamist. Lisaks on linkandmetega seotud ka avaandmete publitseerimise ja info taaskasutamise printsiibid.

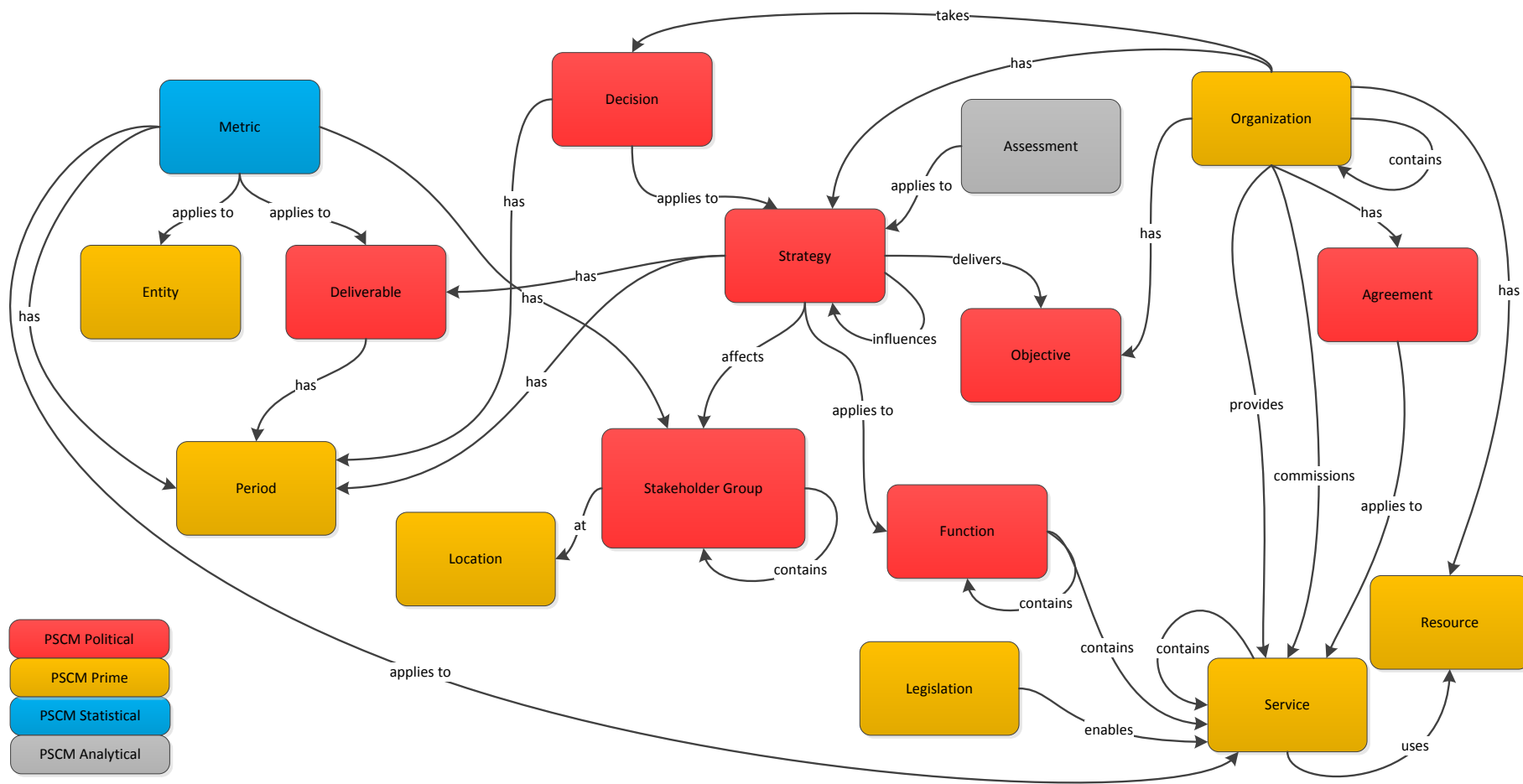
UK avaliku sektori kontseptuaalne mudel on väga huvitav ainulaadne dokument, mis esitab kontseptuaalsed mudelid 4 avaliku sektori informatsiooni konteksti kohta: operatsiooniline, statistiline, analüütiline ja poliitiline. Neid kontseptuaalseid mudeleid saab kasutada jagatavatest ja taaskasutatavatest andmetest ühtmoodi aru saamiseks, avaandmete kataloogide tegemisel, samade andmete mitmekordse kogumise vältimiseks, avaliku sektori lahenduste ressursside kokkuhoidmiseks jpm.

Iga konteksti jaoks on dokumendis esitatud vastav mudel lihtsa ontoloogia nii graafilisel kujul kui RDFS formaadis. Näiteks järgmisel joonisel (vt *Joonis 7*) on toodud poliitilise informatsiooni kontseptuaalne mudel teenuste kasutuselevõtuks.

²⁶⁶https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266284/Information_Principles_UK_Public_Sector_final.pdf

²⁶⁷ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266279/A-Concept-Model-for-the-UK-Public-Sector.pdf

²⁶⁸ Open Data White Paper – Unleashing the Potential, HM UK Government, June 2012, http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf



Joonis 7. UK avalike teenuste kasutuselevõtu kontseptuaalne mudel poliitilise konteksti jaoks²⁶⁹

²⁶⁹ Joonis on reprodutseeritud vastavalt UK Avatud Valitsuse Litsentsile; <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/>

Joonisel toodud mudelilt on näha, kuidas liigub info avalike teenuste kasutuselevõtus ja kuidas selleks ühendatakse erinevatest allikatest tulev info.

UK avaandmete valge raamat esitab UK avaandmete strateegia, mis on üles ehitatud järgmisele skeemile (vt *Joonis 8*).



Joonis 8. UK avaandmete strateegia

UK on panustanud avaandmete potentsiaalile toetada innovatsiooni ja majanduskasvu. ODI kaudu toetatakse avaandmete ärist kasutust. Selleks, et stimuleerida valitsusasutuste panustamist UK avaandmete portaali, nõutakse igalt osakonnalt avaandmete strateegiat ja tegevuskava. Viimaste täitmist kontrollitakse regulaarselt ministeeriumide tasemel.

UK valitsuse avaandmete valges raamatus väljendas valitsus üksmeelt luua uus valitsuse tasandi linkandmete töögrupp²⁷⁰. Selle töögrupi põhirolliks on töötada koos andmete valdajatega, kasutajatega ja organisatsioonidega nagu W3C linkandmete töögrupp, et kehtestada standardid ühistele valitsusetasandi URI-dele ja propageerida nende omaksvõtmist kogu valitsuse tasandil. Samas dokumendis peetakse tähtsaks andmete ühendamist võimaldavate põhiliste autoridatiivsete identifikaatorite loomist sellistes valdkondades nagu äri, lepingud, postikoodid, georuumilised objektid nagu teed ja bussipeatused. Valges raamatus väljendati ka soovi kasutada Sir Tim Berners-Lee poolt pakutud 5 tärni skeemi²⁷¹ avaandmete taaskasutatavuse hindamiseks.

UK on avalike linkandmete publitseerimise liider. Seda on poliitiliselt mõjutanud lisaks avaandmete valgele raamatule ka Euroopa Liidu direktiiv avaliku sektori informatsiooni taaskasutuse alal²⁷², mida UK-s rakendati juba 2005. aastal. UK kesk- ja omavalitsuste avaandmete portaali (<http://data.gov.uk>) loodi 2009. a, olles Euroopas esimene ja USA järel maailmas teine. UK-s on hästi arenenud lingitud andmete loomine ja neile juurdepääsu tagamine (vt <http://data.gov.uk/linked-data>). Peale selle avas Suurbritannia kaardistamise amet Ordnance Survey osa oma andmeid 2010. aastal, luues vastava portaali.²⁷³ Käesoleval ajal on erinevate valdkondade linkandmeid võimalik tarbida ka vastavatelt valdkondlikelt saitidelt nagu näiteks environment.data.gov.uk, location.data.gov.uk, legislation.data.gov.uk, education.data.gov.uk, transport.data.gov.uk, jt.

2012. aastal toimus muutus UK avaandmete portaali linkandmete tehnilise lahenduse osas. Andmestikud ja vastavad rakendusliidesed viidi üle uude vabavara abil loodud lahendusse, mis töötab pilvearvutuse infrastruktuuri baasil. Kasutatud vabavara oli Apache Jena, RDF halduseks kasutati Jena TDB ja linkandmete rakendusliideste tegemiseks kasutati süsteemi Elda²⁷⁴.

Veebikonsortsiumi (W3C) linkandmete töögrupi aastaaruanne 2012. a linkandmete alal²⁷⁵ tõi välja järgmised UK tugevused linkandmete publitseerimise vallas:

- UK on linkandmete publitseerimise liider (vt <http://data.gov.uk>), samas kui USA teeb edusamme 4-5 tärni²⁷⁶ avaandmete loomisel ja EK finantseerib projekte andmehulkade linkimiseks ja vastavate tööriistade loomiseks. UK keskkonnaamet (<http://environment.data.gov.uk>) publitseerib 5 tärni andmeid veekvaliteedi kohta. Suplusvee kvaliteedi andmed (<http://environment.data.gov.uk/bwq/>)

²⁷⁰ <http://data.gov.uk/linked-data/UKGovLD>

²⁷¹ www.w3.org/DesignIssues/LinkedData.html

²⁷² PSI-directive 2003. DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information,

http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf

²⁷³ www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html

²⁷⁴ <http://data.gov.uk/blog/linked-data-pilot-implementations-update>

²⁷⁵ (<http://www.slideshare.net/3roundstones/progress-update-on-government-linked-data-world>)

²⁷⁶ www.w3.org/DesignIssues/LinkedData.html

publitseeriti avalike linkandmetena 2011. a. Veenäite võetakse üle 500 asukoha Suurbritannia rannikult. Andmemudel on suhteliselt lihtne võimaldades identifitseerida asukohti, reostusohu ja vaatluste aegridu. Andmed on kättesaadavad läbi SPARQL päringupunkti ja linkandmete rakendusliidese.²⁷⁷

- UK University of Southampton avaandmete teenus (<http://data.southampton.ac.uk/datasets.html>) pakub valiku 5 täрни andmetest ülikooli kohta.
- Reegle (<http://data.reegle.info>) pakub kõrge kvaliteediga 5 täрни linkandmeid puhta ja taastuenergia ning kliima kohta maailmapangale ja UK avaandmete portaale.

Lisaks on UK Rahvusaarhiivi linkandmete initsiatiiv²⁷⁸ üks eesrindlikumaid avalike linkandmete publitseerimise näiteid nii pakutava andmekvaliteedi kui teostuse poolest. Iga UK, Šotimaa, Wales-i ja Põhja Iirimaa parlamendi poolt vastuvõetud õigusakt või selle muutus alates 1267. a on saadaval nii veebilehena kui XML-ina ja on kirjeldatud RDF-is.

Heaks näiteks on ka BBC, kes on linkandmeid kasutanud juba 2009. aastast oma arhiivimaterjalide esitamiseks²⁷⁹. Näiteks BBC Wildlife sait²⁸⁰ on üks nende varasemaid näiteid linkandmete kasutamisest. BBC kasutas linkandmeid ka 2012. a olümpiamängude programmi esitamisel²⁸¹. Motivatsioon selleks ja tehniline ülevaade firmasisesest linkandmete kasutamisest BBC sisuhaldussüsteemis on esitatud Jem Rayfiled-i blogis.²⁸²

UK-s loodi 2012. a avaandmete kasutajate grupid, kes toimivad kui nõudluse indikaatorid ja annavad valitsusele nõu, milliseid andmeid oleks vaja avada. Ka kohalikud omavalitsused on loonud oma avaandmete portaale. Ava- ja linkandmete alane koolitus on kättesaadav ja sageli tehakse temaatilisi häkkimispäevi, seminare, võistlusi ning valitsuse poolt rahastatakse innovatsiooni, mis on mõeldud ärisektori kaasamiseks avaandmete valdkonda.

Avaandmete baromeetri uuringu andmetel on UK nõrgim ala seotud avaandmete sotsiaalse mõju tagamisega, sest hiljutine valitsuspoliitika on põhiliselt rõhutanud andmete taaskasutamise majanduslikku mõju. Seega pole praegu prioriteediks avaandmete ümber olev kogukonna praktika ja selle kasutamine avatuma poliitika tegemisel.

Baromeetri raportis avaldati ka kahtlust, et UK poolt allakirjutatud G8 avaandmete harta printsiip „vaikimisi avaandmed“ ei pruugi rakenduda selliste uute tähtsate andmetike suhtes nagu firmade tegelike tulusaajate register, mis loodi maksupettuste ja korruptsiooni vastu võitlemiseks.

Olemasolevate laialdaste avaandmete vallas on näha jõupingutusi nende andmete kvaliteedi ja usaldusväärsuse tõstmiseks, standardimiseks ja nende linkimisvõimekuse tõstmiseks. Paljud andmehulgad, nagu näiteks kaardiandmed, maaregister, ärireister ja seadusaktid, on olemas 5 täрни linkandmetena.

UK avaandmete portaali kogemusi linkandmete alal võtavad selle autorid kokku järgmiselt²⁸³: „Avalike linkandmete väljalaskmine ei ole ainult tehniline probleem, kuigi sel on ka tehnilised väljakutsed. Avalikud linkandmed ei ole jäik riigi IT spetsifikatsioon, vaid nõuab viljakat dialoogi nii andmete valdajate, kasutajate kui arendajate vahel. Integreerides avalikud linkandmed üldisesse linkandmete maailma kahtlemata laiendab seda, kuid teisalt pakub see maailm ka täiendavaid ressursse avalike linkandmete rikastamiseks.“

²⁷⁷ Linked Data API, <http://code.google.com/p/linked-data-api>

²⁷⁸ <http://www.legislation.gov.uk>

²⁷⁹ <http://lists.w3.org/Archives/Public/public-lod/2009Apr/0162.html>

²⁸⁰ <http://www.bbc.co.uk/nature/wildlife>

²⁸¹ <http://www.bbc.co.uk/2012/>

²⁸² http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html

²⁸³ Shadbolt, Nigel, O'Hara, Kieron, Berners-Lee, Tim, Gibbins, Nicholas, Glaser, Hugh, Hall, Wendy and Schraefel, m.c. (2012) Linked open government data: lessons from Data.gov.uk. IEEE Intelligent Systems, 27, (3), Spring Issue, 16-24. (doi:10.1109/MIS.2012.23).

Listpoint (www.listpoint.co.uk) viis läbi avaandmete uuringu UK riigiametnike hulgas 2012. a.²⁸⁴

Sellele küsitlusele vastasid 1017 UK keskvalitsuse ja kohalike valitsuste, tervishoiu-, riigikaitse- ja politseiasutuste töötajat. Uuringu põhitulemus oli, et riigiametnikel puudub piisav teadlikkus avaandmetest ja riigi avaandmetega seotud plaanidest. 78% vastajatest ei tea riigi avaandmete plaanist ja selle kasulikkusest. Nad (66%) ei saa aru, mis on nende individuaalne roll riigi avaandmete vallas. Eksperdid (50%) tunnetavad, et erasektori juurdepääs andmestandarditele ja riigi andmekogumitele on vajalik uute töökohtade ja paremate teenuste pakkumiseks. Samas ei näe suur osa riigiametnikke, et andmete avamine erasektorile oleks prioriteet.

Eraldi väärrib märkimist Avaandmete Instituut (ODI)²⁸⁵, mis loodi 2012. a kui MTÜ, mida juhivad Sir Tim Berners-Lee ja prof Nigel Shadbolt Southampton-i ülikoolist. Selle instituudi eesmärgiks on propageerida avaandmete kultuuri selleks, et luua majanduslikku, keskkonna ja sotsiaalset väärtust. Instituut aitab kaasa teadmiste levitamisele ja nõudluse loomisele nii lokaalsel kui globaalsel tasandil. Avaandmete Instituut pakub ka professionaalset mentorlust ja veebikursusi avaandmete teemadel. Instituut viib läbi uuringuid ja osaleb teadus- ja arendusprojektides. Avaandmete Instituudil on suur liikmete võrgustik. Avaandmete Instituut toetab firmade innovatiivseid avaandmete projekte. Instituudi esimese tööaasta raportis väljendatakse rahulolu instituudi tööga ja finantseerimisega ja sellega, et võrgustikku lülitus üle 40 firma. Eriti ollakse rahul, et 2013. aastal toimusid järgmised tähtsad avaandmete ja nende masinloetavuse tähtsust rõhutavad sündmused: Inglismaa oli Avatud Valitsuse Partnerluse eesistujaks, kirjutati alla G8 avatud andmete harta ja anti välja USA Presidendi korraldus (9. mai 2013) selle kohta, et valitsuse info peaks olema vaikimisi avatud ja masinloetav.²⁸⁶

Avaandmete Instituut publitseerib edulugusid avaandmete kasutamisest, näiteks muuhulgas Opencorporates juhtum²⁸⁷ ja King's Colledge London edulugu.²⁸⁸

Avaandmete Instituut kuulutas välja oma initsiatiivi laienemise 28. oktoobril 2013. a. Laienemine tähendab avaandmete võrgustiku loomist ja initsiatiivi esialgset laienemist Prantsusmaale, Dubaisse, Kanadasse, USA-sse, Argentiinasse ja Venemaale.²⁸⁹

Avaandmete Instituut ja selle laiendamine võrgustikuks teisi riike kaasates on heaks näiteks linkandmete jõulisest initsiatiivist UKs. Ka UK ülikoolides on linkandmetega seonduvad tehnoloogiad au sees, näiteks üheks juhtivaks ülikooliks selles vallas on Southampton-i ülikool.

2.4.2 Ameerika Ühendriigid (USA)

USAs on linkandmete kasutuselevõttu avalikus sektoris mõjutanud 3 järgmist poliitilist initsiatiivi ja nendega seotud korraldused ning memorandumid:

- Avatud valitsemise initsiatiiv (2009)²⁹⁰ ja avatud valitsemise direktiiv.²⁹¹
- E-riigi ja infotehnoloogia ameti ²⁹² poolt väljaantud strateegiad.
- Riigivalitsemise uus digitaalstrateegia (2013) ja eelnev digitaalse valitsemise direktiiv (2012)²⁹³

²⁸⁴ Listpoint 2012. Listpoint open data survey, <http://www.publictechnology.net/news/open-datas-potential-remains-closed-public-sector-staff/37572>

²⁸⁵ The Open Data Institute (ODI) <http://www.theodi.org/>

²⁸⁶ <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>

²⁸⁷ <http://theodi.org/case-studies/opencorporates-case-study>

²⁸⁸ <http://theodi.org/news/odi-startup-demand-logic-saves-king-s-college-london-390000-year-energy-costs>

²⁸⁹ Bill Goodwin, US and Canada follow UK initiative on Open Data, 28.10.2013, Computer Weekly, <http://www.computerweekly.com/news/2240207982/US-and-Canada-follow-UK-initiative-on-Open-Data>

²⁹⁰ <http://www.whitehouse.gov/open>

²⁹¹ <http://www.whitehouse.gov/open/documents/open-government-directive>

²⁹² <http://www.whitehouse.gov/omb/e-gov>

Peale selle seonduvad nimetatud teemaga veel eespool käsitletud suurandmete initsiatiiv aastast 2012²⁹⁴ ja BRAIN initsiatiiv²⁹⁵.

Lisaks võib nimetada veel Nutika avalikustamise initsiatiivi (Smart Disclosure Initiative) ²⁹⁶ 2012. aastast. Nutikas avalikustamine viitab õigeaegsele toodete jm kohta käivate andmete ja info avaldamisele masinloetavas vormingus, et tarbijad saaksid võtta vastu informeeritud otsuseid. Ettevõtjad saavad siis seda avalikku infot kasutada uute teenuste ja firmade loomisel. Näiteks võib tuua sellised teenused nagu *Castlight* ja *LowerMyBills*.²⁹⁷

Kõigi nende initsiatiivide rakendamise tulemusena on USA järk-järgult üle minemas linkandmete kasutusele oma tsentraalses avaandmete portaalis ja teistes avalikes portaalides ning kinnistes infosüsteemides. Üheks põhiliseks motivaatoriks oli riigi infosüsteemi IT kulude oluline kärpimine ja majanduse elavdamine avaandmete tekitamisega.

Avatud valitsemise initsiatiiv. Avatud valitsemise initsiatiivi raames loodi USA-s 2009. aastal maailma esimene avaandmete portaali (<http://data.gov>). Eraldi portaali on loodud valitsuse kulutuste avalikustamiseks (<http://recovery.gov>, <http://USAspending.gov>), lisaks on veel teisigi andmete avamise initsiatiive nii energeetika, hariduse, avaliku julgeoleku kui tervishoiu (<http://healthdata.gov>) valdkondades²⁹⁸.

Avatud andmete baromeetri uuringu andmetel on USA-s ajalooliselt valitsuse andmed olnud avatult kättesaadavad, sest autoriõigused ei rakendu föderaalvalitsuse andmehulkadele. Praegune avaandmete poliitika aga löi teatud vastuolu, sest nõuab, et andmed oleksid litsentseeritud.

USA programmi „*Presidential Innovation Fellows*“ raames võeti valitsusasutustesse lühikeseks ajaks tööle eksperte äri sektorist, et toetada ja hoogustada avaandmete alast tegevust.

USA-s on ka palju kodanikeühendusi, kes soovivad töötada koos valitsusega avaandmete valdkondades tagades valitsemise läbipaistvuse. USA avaandmete portaali tehti spetsiaalne kogukonna sektsioon soodustamaks koostööd kodanikeühendustega.

USA põhiline avalike linkandmete projekt on *The Tetherless World Constellation (TWC) Linking Open Government Data (LOGD)* projekt (TWC LOGD)²⁹⁹. TWC LOGD uurib ja teostab andmete linkimist, kasutades semantilise veebi tehnoloogiaid. Projekti raames teisendatakse avalikke andmeid RDF-i ja lingitakse teiste linkandmetega. Peale selle viiakse läbi linkandmete loomise ja kasutamise alaseid koolitusi ja demonstratsioone. See projekt on seotud ka linkandmete portaali loomisega USA avaandmete portaali osana.³⁰⁰ Juba 2010. aastal tehti 400 data.gov andmestikku kättesaadavaks kui linkandmed, kokku 6,4 biljonit kolmikut³⁰¹, tekitades seega ka kohe lingitud suurandmed.

TWC LOGD projekt on aidanud teha ka linkandmete lahendust USA tsentraalsele tervishoiuportaalile www.healthdata.gov/jpt.

W3C linkandmete töögrupi 2012 . a aruanne³⁰² toob USA kohta välja järgmised saavutused:

- USA keskkonnakaitseamet (EPA) andis 2012. a välja 5 tärni linkandmeid.

²⁹³ http://www.whitehouse.gov/sites/default/files/uploads/2012digital_mem_rel.pdf

²⁹⁴ Whitehouse 2012, Big Data Press Release, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf

²⁹⁵ <http://www.whitehouse.gov/share/brain-initiative>

²⁹⁶ <http://www.whitehouse.gov/sites/default/files/omb/inforeg/for-agencies/informing-consumers-through-smart-disclosure.pdf>

²⁹⁷ <http://www.whitehouse.gov/sites/default/files/omb/inforeg/for-agencies/informing-consumers-through-smart-disclosure.pdf>

²⁹⁸ Obama 2011. B. Obama, Technology, <http://www.whitehouse.gov/issues/technology>

²⁹⁹ <http://logd.tw.rpi.edu/>

³⁰⁰ Li Ding, et al (2011). TWC LOGD: A Portal for Linked Open Government Data Ecosystems. Journal of Web Semantics: Science, Services and Agents on the World Wide Web (in press). <http://dx.doi.org/10.1016/j.websem.2011.06.002>

³⁰¹ <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

³⁰² <http://www.slideshare.net/3groundstones/progress-update-on-government-linked-data-world>

- USA uuendas terviseandmete portaali HealthData.gov.
- SEC register anti välja kui linkandmed (<http://edgarwrap.ontologycentral.com>) kasutades andmekuubi sõnastikku.
- USA Kongressi Raamatukogu esitas linkandmed (<http://id.loc.gov>).

Huvitav on jälgida näiteks USA tsentraalse avaandmete portaali arengut, mis teeb läbi kõik etapid ja õppetunnid üleminekul linkandmetele³⁰³. Portaali data.gov loomise alguses (2009) publitseeriti andmestikud seal avatud formaatides, nagu näiteks CSV, ja nende kasutamiseks oli vaja andmeid täismahus alla laadida. See tekitas aga probleeme andmekvaliteediga ja osa andmeid olid vananenud juba enne kui nad portaali üles pandi. Siis otsustati, et selleks, et andmed oleksid paremini kasutatavad, peaks need konverteerima RDF-i ja esitama linkandmetena. Eelpool mainitud TWC initsiatiivi (2010) raames loodigi raamistik andmete konverteerimiseks RDF, nende linkimiseks teiste andmetega ja JSON-i ning SPARQL-i kasutamiseks neil andmetel. See võimaldas erinevaid data.gov andmestikke linkida veebirakenduste tarbeks. Alates sellest ajast on USA valitsus publitseerinud üha rohkem oma andmeid RDFina, siiski on probleeme kvaliteediga ja ikka tuleb neid andmestikke rakenduste jaoks täismahus alla laadida. 2012 aastal vastu võetud Digitaalse valitsemise strateegia dokumendis püüti seda kitsaskohta lahendada andmete eraldamise teel rakendustest. Riigiametilt nõuti nn infokeskset lähenemist, mis selle dokumendi järgi tähendas nõuet, et kogu pakutavat sisu peab vaatlema kui andmeid, muutes mittestruktureeritud sisu struktureeritud andmeteks ja varustades selle kehtivate metaandmetega. Andmetele juurdepääs oleks sel juhul organiseeritud läbi ühiste veebipõhiste rakendusliidest (web API). Selle strateegia põhiidee oli teha data.gov portaalist publitseerimissaidi asemel arendajate ressurss ning koondada niimoodi rakendused selle avaandmete portaali ümber.

Arendajate ressursid RESTful rakendusliidestena, mis väljastavad JSON vormingus andmeid on saadaval saidilt <https://www.data.gov/developers/apis> (lisaks võib vaadata ka <http://project-open-data.github.io/>).

E-riigi strateegiad. USA e-riigi ja infotehnoloogia amet juhindus kuni 2013. a maikuuni strateegiast, mis on esitatud dokumendis *Digitaalse valitsemise strateegia (Digital Government Strategy (23. mai 2012))*³⁰⁴. Umbes aasta hiljem andis USA president B. Obama välja oma otsekorralduse uue digitaalstrateegia kohta, mida vaatame eraldi allpool. Digitaalse valitsemise strateegia on seotud ka Föderaalsete jagatud teenuste strateegiaga (*Federal Information Technology Shared Services Strategy (2. mai 2012)*)³⁰⁵ ja Föderaalsete pilvearvutuse strateegiaga (*Federal Cloud Computing Strategy (8. veebruar 2011)*)³⁰⁶. Pilvearvutuse strateegia sisaldab muuhulgas ka teenuste pilvearvutusele üleviimise ja selle riskijuhtimise raamistikke.

Jagatud teenuste strateegia üldised eesmärgid on järgmised:

- Vähendada riigiametite IT kulusid.
- Suurendada riigiametite produktiivsust rakendades integreeritud valitsemise protsesse ja innovatiivseid IT-teenuste lahendusi.
- Suurendada kommunikatsiooni kõigi osapoolte vahel selleks et jagatud teenuste haldajad, pakkujad ja tarbijad töötaksid koos kogu jagatud teenusega seotud tegevuste elutsükli vältel. Selleks on loodud vastavad koostööressursid nagu CIO.gov, ITDashboard.gov, Performance.gov, and BusinessUSA.gov.

Jagatud teenuste strateegia on seotud varasema reformiga 2010. aastal, mis oli 25 tegevusega tegevuskava USA riigi infosüsteemi kulutuste vähendamiseks ja efektiivsuse suurendamiseks.³⁰⁷ Põhiline selles tegevuskavas oli suund üleminekuks pilvearvutusele ja jagatud teenustele, föderaalsete andmekeskuste arvu

³⁰³ <http://arstechnica.com/business/2012/05/open-government-reboot-focuses-on-apis-instead-of-data/>

³⁰⁴ <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>

³⁰⁵ http://www.cio.gov/documents/Shared_Services_Strategy.pdf

³⁰⁶ http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/federal-cloud-computing-strategy.pdf

³⁰⁷ http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/25-point-implementation-plan-to-reform-federal-it.pdf

vähendamine vähemal 800 võrra (s.o umbes 40%) aastaks 2015 ning riigiametite ning ettevõtete veebipõhise koostööplatvormi loomine.

Digitaalse valitsemise strateegia 2012 seadis peaesmärgiks mobiiliseadmeid omavate kodanike varustamise digitaalse valitsemise info ja teenustega toetades printsiipi „*anywhere, anytime, on any device*”. Selleks nõuti riigiametilt oma sisupakkumise mudelite moderniseerimist ja masinloetavate ühenduste tekitamist riigi avaandmetele ja teenustele. Eesmärgiks oli anda ettevõtetele ja kodanikele võimalus kasutada saadavat infot nii ärilistel kui isiklikel eesmärkidel.

Riigivalitsemise uus digitaalstrateegia. USA President andis välja otsekorralduse riigivalitsemise uue digitaalstrateegia kohta 9. mail 2013³⁰⁸, mis sätestab, et avaliku sektori andmed peavad olema avatud ja masinloetavad vaikimisi. Uus korraldus on õiguslikult rohkem siduv ja spetsiifilisem kui eelmised avaandmetega seotud korraldused ning sätestab selge ja kindla õiguskeskkonna avaliku sektori avaandmetele. Samas dokumendis esitatakse ka uute printsiipide läbiviimise kava. Näiteks nõutakse, et iga valitsusasutus peab tagama, et mingi uue arvuti või seadme poolt kogutud andmed peavad kohe ilma eelneva digitaliseerimiseta olema eksporditavad. Valitsusasutustelt nõutakse oma andmete kataloogiseerimist. Uue avaandmete poliitika üks põhilisi aspekte on avaliku valitsusasutuste andmevarade listi loomine pärast vastava auditi läbiviimist. Põhjalik ümberkorraldus toimub ka USA valitsuse avaandmete portaalis, mis läheb üle CKAN platvormile ja mille sisu laiendatakse metaandmetega.

Uus USA riigivalitsemise digitaalstrateegia on ühtlasi ka tegevusteekaart (*roadmap*) konkreetsete, 12-kuulise tähtajaga ülesannetega ja tegevustega valitsusasutustele määrates ära ka konkreetsed vastutajad iga tegevuse jaoks. Allpool on toodud mõned näited tegevusteekaardi ülesannetest:

- Elektroonilise teabe ning teenuste pakkumine ja vastuvõtmine ükskõik kus, millal ja mis seadmel.
- Riigi kui teabe tootja ja publitseerija teabe publitseerimismudeli täiustamine nii, et oleks tagatud sisu ja andmete esitamine seadmetest sõltumatul viisil ning läbi paljude kanalite.
- Nutika, turvalise ja kulutõhusa teenuste haldamise struktuuri loomine.
- Avaliku sektori andmetes peituva innovatsioonipotentsiaali avamine.
- Dubleerivate süsteemide ja seadmete välja selgitamine valitsusasutustes eesmärgiga neid koondada ühisteenustesse.

Nimetatud otsekorraldusele vastavalt anti välja memorandum *Open Data Policy – Managing Information as an Asset*³⁰⁹, mis loob raamistiku efektiivse infohalduse printsiipide institutsionaliseerumisele kaasaaitamiseks, arvestades iga infohalduse elutsükli etappi ning edendades koosvõimet ja avatust. Memorandumis öeldakse, et vaatamata sellele, kas mingit infot saab avada või mitte, on selline raamistik kasutatav riigiasutuste poolt kõigi nende haldusalas olevate inforessursside kohta.

Memorandum nõuab valitsusasutustelt kõigi uute andmekogude või infosüsteemide algatamisel andmete masinloetavate ja avatud formaatide ning ühiste tuumikmetaandmete (peavad olema laiendatavad) kasutamist.

Memorandum sisaldab oma lisades palju informatsiooni elutsükli järgivaid poliitilisi nõudeid uutele loodavatele riiklikele andmekogudele. Põhiline nõue on *koosvõime*, millest lähtuvad teised nõuded. Allpool toome mõned põhilised nõuded:

³⁰⁸ Executive Order – Making Open and Machine Readable the New Default for Government Information, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government> ja <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-defaultgovernment>, [http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government.html](http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html)

³⁰⁹ <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

- Riigiasutustel tuleb kasutada masinloetavaid avatud andmete formaate.
- Riigiasutustel tuleb kasutada andmete koosvõime ja avatuse standardeid.
- Riigiasutustel tuleb kasutada avatud litsentse nii info kogumisel kui loomisel nii, et andmete avamisel poleks enam kitsendusi kopeerimisele, avaldamisele, edastamisele, levitamisele, kasutamisele jm.
- Riigiasutustel tuleb kirjeldada info kasutades ühiseid metaandmeid juba andmete kogumisel ja loomisel. Metaandmed peavad sisaldama infot algallika kohta, linkandmeid, geoasukohta, ajaseeriade järgnevusi, andmekvaliteedi näitajaid ja teisi asjakohaseid indekseid, mis näitavad seoseid andmestike vahel.
- Riigiasutustel on kohustus luua infosüsteeme, mis on koosvõimelised ja toetavad info kättesaadavust.

Memorandum viitab muuhulgas varasemale dokumendile *A Common Approach to Federal Enterprise Architecture* (2. mai 2012),³¹⁰, mis esitab USA riigiinfosüsteemi arhitektuuri. Käesoleva projekti raames on huvipakkuv, selle arhitektuuri andmete referentsmudel *The Data Reference Model* (DRM), mis peaks võimaldama efektiivselt jagada infot riigiasutuste vahel, suurendada integratsiooni, info taaskasutust, toetada semantilist koosvõimet tagades samas infokasutuse turvalisuse ja privaatsuse.

DRM hõlmab 3 standardiseerimise valdkonda.

- Andmete kirjeldus võimaldab andmeid ühtselt kirjeldada, andes edasi andmete tähenduse toetades seeläbi andmete avastamist ja jagamist.
- Andmete kontekst soodustab andmete avastamise sellist lähenemisviisi, mis kasutab andmete taksonoomiate põhist kategoriseerimist. Lisaks võimaldab andmete kontekst defineerida eriti mõjukaid andmevarasid.
- Andmete jagamine toetab andmetele juurdepääsu ja andmevahetust. Juurdepääs tähendab päringute esitamist andmetele ja andmevahetus tähendab siin fikseeritud andmete transaktsiooni partnerite vahel.

Kuigi üheski ülalnimetatud dokumentidest ei viidata konkreetsetele standarditele ega tehnoloogiatele selleks, et rakendada dokumentides toodud nõudeid, võib siiski tegelikkuse analüüsimisel märgata, et *fraasi* „*machine-readable by default*“ all mõeldakse praegu RDF-i ehk siis linkandmeid. Sama kehtib ka taksonoomiate (sõnastike) kohta, selles vallas kasutatakse nii RDFS, SKOS kui OWL ja teisi vajalikke standardeid.

Ilmselt ei soovi USA administratsioon piirata poliitikadokumentides ennast mingi kindla tehnoloogia mainimisega, sest tehnoloogia areneb selles vallas ülikiiressti ja seetõttu pole konkretiseeringud otstarbekad. Poliitikadokumendid esitavad ainult nõudeid tehnoloogiatele.

Samas on viiteid USA ja Iirimaa instituudi DERI (*Digital Enterprise Research Institute*) koostööle linkandmete ja semantiliste tehnoloogiate kasutuselevõtul USA valitsusametite poolt tsentraalse avaandmete portaali data.gov ja healthdata.gov raames juba aastal 2012³¹¹. Kasutatud tehnoloogia oli Neologism ja Google Refine RDF laiendus. Neologism³¹² põhineb sisuhaldussüsteemil Drupal ja võimaldab lihtsalt luua RDF andmete linkimiseks vajalikke sõnastikke (ontoloogiaid). Üks tähtsatest sõnastikest on *Vocabulary of Interlinked Datasets* (VOID), mille lõi DERI teadlased ja mis asub data.gov sõnastike saidil vocab.data.gov.

³¹⁰ http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf

³¹¹ <http://www.nuigalway.ie/about-us/news-and-events/news-archive/2012/april2012/us-government-joins-the-dots-with-irish-linked-data-technologies-1.html>

³¹² <https://drupal.org/project/neologism>

VOID on RDF(S) sõnastik kirjeldades 4 klassi ja 27 seost.³¹³ On loodud ka RDFi laiendus Google Refine-le, mis on graafiline kasutajaliides linkandmete eksportimiseks Google Refine andmetest.

2.4.3 Juhtivate riikide edutegurid linkandmete kasutuselevõtul

Võrreldes 2 maailma juhtivat riiki UK ja USA linkandmete kasutuselevõtus võime eelneva analüüsi põhjal välja tuua rea edutegureid, mis on esitatud alltoodud tabelis koos vastavate näidetega.

Tabel 2. Linkandmete kasutuselevõtu edutegurid

Edutegur	UK näited	USA näited
Suunatud riiklik strateegia	IKT strateegia, referentsarhitektuuri, kontseptuaalse mudeli ja infoprintsiipide dokumendid ja Avaandmete valge raamat (2012)	Avatud valitsemise initsiatiiv, e-riigi ja IT ameti strateegiad, riigivalitsemise digitaalstrateegiad, suurandmete initsiatiiv, BRAIN initsiatiiv jm
Rahvusvahelise avalikkuse ja alliansside mõjutamine	G8, ODI laienemine teistesse riikidesse	G8
Strateegiat toetav riiklik tegevuskava	IKT strateegia tegevuskava, Avaandmete valge raamat	Riigivalitsemise digitaalstrateegia (2013), memorandum <i>Open Data Policy-Managing Information as an Asset</i> (2013) jt
Pidev strateegia täiendamine	x	x
Tegevuskava viimine ametnikeni ja järelvalve	Selleks, et stimuleerida valitsusasutuste panustamist UK avaandmete portaali nõutakse igalt osakonnalt avaandmete strateegiat ja tegevuskava. Viimaste täitmist kontrollitakse regulaarselt ministeeriumide tasemel. Valitsuse tasandi linkandmete töögrupi loomine. Valitsuse tasandi referentsarhitektuuri töögrupi loomine.	Riigivalitsemise digitaalstrateegia (2013) <i>Memorandum Open Data Policy-Managing Information as an Asset</i> (2013) ³¹⁴ <i>A Common Approach to Federal Enterprise Architecture</i> (2012) Jagatud teenuste strateegia (2012) <i>Presidential Innovation Fellows</i>
Ametnike äriprotsesside ja töökultuuri muutus	Pole teada	Tegevuskava (2010) USA riigi infosüsteemi kulutuste vähendamiseks ja efektiivsuse suurendamiseks Jagatud teenuste strateegia (2012)
Riigisektori IT kulude joondamine	On üks IKT strateegia eesmäärke	Tegevuskava (2010) USA riigi

³¹³ <http://vocab.data.gov/VoID>

³¹⁴ <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>

Edutegur	UK näited	USA näited
		infosüsteemi kulutuste vähendamiseks ja efektiivsuse suurendamiseks
		Jagatud teenuste strateegia (2012)
Erasektori ja elanikkonna kaasamine parendades kommunikatsiooni e-teenuste haldajate, pakujate ja tarbijate vahel	IKT strateegia	Jagatud teenuste strateegia (2012)
Juhtfiguuride toetus	Tim Bernes-Lee	B. Obama
Teadlikkuse tõstmine	ODI ja selle laiendamine teistesse riikidesse	TWC avalike linkandmete projekt
Rahvusvaheline koostöö	DERI, ODI võrgustik, koostöö W3C-ga	ODI, DERI, koostöö W3C-ga
Rahvahulkade kaasamine	Avaandmete nõukogud kogukonna esindajatest	Avaandmete portaali kaudu võimalik mõjutada protsesse
Prioriteetsete linkandmete kasutusvaldkondade määramine	Arhiivid, raamatukogud, geoandmed, transport, haridus jt.	Missioonikriitilised rakendused, raamatukogud, tervishoid jt.

2.4.4 Kasulikke õppetunde välisriikide parimatest praktikatest Eestile

MKM avaldas ja suunas konsultatsiooniringile *Avaandmete roheline raamat* mustandi³¹⁵ 21. novembril 2013. aastal. Seda võib pidada esimeseks avaandmete strateegia ja tegevuskava dokumendiks Eestis. Dokument pole praeguseks veel vastu võetud, kuid on loota, et seda tehakse 2014. aasta jooksul. Avaandmete roheline raamat avab avaandmete valdkonna põhimõtted, olulisemad poliitikavalikud ning tegevuskava. Raamatus on öeldud, et „Rohelise raamatu heakskiitmisel valitsuse poolt järgneb rohelises raamatus ettenähtud tegevuste elluviimine (nt inim- ja raharessursside eraldamine, õigusaktide muutmine, konkreetsete projektide käivitamine)“.

Eesti võiks linkandmete loomisel ja kasutuselevõtul võtta arvesse UK ja USA praktikaid ning arvestada nende riikide edutegureid. Ülaltoodud tabelist on selgelt näha, et nii UK kui USA jaoks on suunatud strateegia ja konsensuslik tegevuskava esmatähtsad. Teiseks on tähtsuseti tegevuskava viimine ametnikeni ning selle täitmise kontroll.

Järgmises tabelis on toodud UK ja USA praktikate põhised soovitusel Eestile (muuhulgas on näidatud ka millise riigi praktikat soovitatakse eeskujuks võtta).

³¹⁵ Avaandmete roheline raamat, AARR, Avaliku teabe avamine taaskasutamiseks, versioon 0.4 2013-11-21, Eesti Vabariigi valitsus, <http://www.riso.ee/et/avaandmete-roheline-raamat>

Tabel 3. Soovitused Eestile teiste riikide edutegurite alusel (tähtsuse järjekorras)

Soovitus	Näidisriikide praktikad
Suunatud riikliku strateegia ja konsensuslik tegevuskava loomine	UK ja USA strateegiad ja tegevuskava
Tegevuskava viimine ametnikeni ja selle täitmise kontroll	UK ja eriti USA järelvalve
Ava- ja linkandmete töögrupi loomine valitsuse juurde	UK
ODI taolise instituudi või sarnase tegevusega osakonna loomine E-riigi Akadeemia juurde teadlikkuse tõstmise ja rahvusvahelise koostöö arendamise huvides	UK
Kuigi Eesti on Avaliku Teabe Seaduse järgi valinud avalike avaandmete vaikimisi pakkumise tee, võiks siiski avalike linkandmete pakkumise protsessi algstaadiumis määrata prioriteetsed avalike linkandmete pakkumise valdkonnad lähtudes nende rakendustes kasutatavusest	UK. Avaandmete pakkumisel on ka UK seoses G8 Hartaga liitumisele valinud avalike avaandmete vaikimisi pakkumise tee.
Riigisektori IT kulude joondamine	Nii UK kui USA loodavad riigisektori IT kulusid vaos hoida just ava- ja linkandmete tehnoloogiate kasutamisega andes teenuste loomise võimaluse erasektorile ja jättes endale põhiliselt selleks otstarbeks kvaliteetsete andmete pakkumise. Teine kokkuhoiu allikas on tark ja odav riigisektori andmete integratsioon kasutades linkandmete tehnoloogiaid.

3 Arhitektuurianalüüs

3.1 Infosüsteemide valim

Arhitektuurianalüüsi valim koosneb kolmest laiemat kasutust leidva lahenduse arhitektuurist. Analüüsitavad lahendused on valitud suuremamahulisi linkandmete komplekte (100+ miljonit kolmikut) kasutavate rakenduste seast, mille kohta oli piisavalt dokumentatsiooni. Kättesaadava dokumentatsiooniga lahenduste hulga kitsendamiseks kasutasime lisakriteeriumina linkandmete loomise ja rakendamise kasulikkust Eesti kontekstis. See tähendas nii hajusate kui tsentraliseeritud lahenduste, nii linkandmete repositooriumite kui nende esitluskihi rakenduste lahenduste esindamist valimis.³¹⁶

Tabel 4. Arhitektuurianalüüsi valim

Nimi	URL	Haldaja	Kolmikute arv
TWC LOGD	http://logd.tw.rpi.edu/	USA	10 miljardit ³¹⁷

³¹⁶ Valimi aluseks olnud dokumentatsiooni ülevaate annab järgnev metoodika ja visualiseerimine: Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) Systematic Mapping Studies in Software Engineering. In Proceedings of 12th International Conference on Evaluation and Assessment in Software Engineering.

³¹⁷ Seisuga 11.04.2011 (Allikas: <http://logd.tw.rpi.edu/>)

DBpedia Live	http://wiki.dbpedia.org/About	Kogukonna projekt, eestvedajateks OpenLink Software, University of Mannheim, Universität Leipzig	470 miljonit
Riiklik Ehitisregister	http://ehitisregister.ee/sparql	Eesti, Majandus- ja Kommunikatsiooniministeerium	156 miljonit

3.2 Referentssüsteem juhtumiuuringuks

W3C mustandis *Linked Data Platform 1.0*³¹⁸ on toodud välja nõuded linkandmete platvormidele. Käesolevas uuringus analüüsimise valimi lahenduste vastavust väljatoodud nõuetele. Lisaks platvormi arhitektuurile hindame ka platvormide elujõulisust, arvestades nende kasutajate ja arendajate kvantitatiivseid näitajaid, täienduste tegemise intervalle ning kolmandate osapoolte komponente, mida need platvormid toetavad.

Vaadeldavate nõuete nimekiri on järgnev:

- F1.1: The system shall provide the ability to create containers for composing resources, from UC1.
- F1.2: The system shall provide the ability to create nested containers, from UC1.
- F1.3: On deletion of a container, the system shall delete any contained resources and nested containers, from UC1.
- F2.1: The system shall provide the ability to create resources within a container, from UC2.
- F2.2: The system shall provide the ability to delete resources, from UC2.
- F3.1: The system shall provide the ability to retrieve resource descriptions, from UC3.
- F3.2: The system shall enable the client to retrieve the description of a hash URI, from UC3.
- F4.1: The system shall provide the ability to update an existing resource by substitution, from UC4.
- F4.2: The system shall provide the ability to perform a selective update of a resource, from UC4.
- F5.1: The system shall provide the ability to determine if a resource has changed, from UC5.
- F6.1: The system shall provide the ability to aggregate resources, from UC6.
- F6.2: The system shall support the addition of a resource to multiple aggregations, from UC6.
- F7.1: The system shall provide the ability to retrieve a collection-level description of a composition, from UC7.
- F7.2: The system shall provide the ability to retrieve an item-level description of a composition or aggregation, from UC7.
- F8.1: The system shall provide the ability to retrieve a paginated description of a composition or aggregation, from UC8.
- F9.1: The system shall provide the ability to store and access media resources, from UC9
- NF1.1: The system shall provide access guidance to resources, from UC1.

³¹⁸ Viide: <http://www.w3.org/TR/ldp-ucr/>

- NF2.1: The system shall encourage non-duplication of resources, from UC2.
- NF2.2: The system shall support distribution of resources, from UC2.
- NF2.3: The system shall support consistent, global naming, from UC2.
- NF3.1: The system shall support the use of standard vocabularies where appropriate, from UC3.
- NF3.2: The system shall provide a scalable linking model, from UC3.
- NF4.1: The system shall permit unrestricted vocabulary, from UC4.
- NF5.1: The LDP shall ensure consistent access in the case of multiple simultaneous attempts to access a resource, from UC5.
- NF6.1: The system shall allow resource descriptions that are a „mix of simple data and collections“, from UC6.
- NF6.2: The system shall support relative URIs enabling sharing of collections, from UC6.

Linkandmete esitluskäsi (veeb) ja Eesti spetsiifika paremaks esiletoomiseks on esitatud referentsarhitektuur (vt *Joonis 9*), mis toob välja valimi lahenduste arhitektuuride omavahel seotud olulisemad komponendid. Referentsarhitektuuri kasutame juhtumiuuringus valimi lahenduste arhitektuuride käsitlemiseks Eesti riigi infosüsteemi kontekstis.

Joonisel on kasutusel järgnevad värvikoodid:

- Kollaselt on välja toodud TWC LOGD lahenduse komponendid;
- Punaselt on esitatud DBpedia unikaalne osa;
- Siniselt on välja toodud Riikliku Ehitisregistri lahenduse komponendid;
- Roheliselt on välja toodud Riikliku Ehitisregistri ja TWC LOGD lahenduse ühisosa;
- Oranžilt on välja toodud DBpedia ja TWC LOGD lahenduse ühisosa.

Linkandmete tootmise keskkonna ja andmeid tarniva registri andmeteenuste andmevahetus üle X-tee toimub SOAP protokolliga kasutades. SPARQL rakendus, teeb SPARQL päringuid kas välisesse RDF andmebaasi või väiksema RDF andmehulga puhul lokaalsesse (mälu)baasi, rakendus ise suhtleb kasutajaga üle HTTP läbi veebirakenduse. RDF andmebaasi laetakse andmed kas RDF, CSV vms vormingut kasutades ning päringuid sinna tehakse SPARQL keeles, tulemused võivad olla RDF, CSV, JSON vms kujul sõltuvalt baasi lisavõimalustest. Ühendatud RDF andmebaas indekseerib andmekomplekte ning aitab tuvastada millises RDF andmebaasis konkreetse nimeruumi objektid asuvad. Samas osad andmekomplektid võivad olla kättesaadavad otse ühendatud baasist. Vajadusel suunab ühendatud RDF andmebaas päringud sinna edasi lüüsi kaudu. Semantiline portaali kasutab XSLT vms skripte SPARQL päringute vastuste serveerimiseks RSS, HTML (sh koos RDFa / microdata märgenditega) vms veebis aksepteeritud kujul. KOVide ja asutuste portaali sisust korjatakse semantiliselt portaali RDFa jms märgendid, mille serveerimiseks kasutatakse transformatsioone.

X-teega seotud peamised komponendid on registrid, andmeontoloogiad, andmete linkimise komponendid (sh integratsiooniserverid), RDF andmebaasid. RDF andmebaase kasutatakse linkandmete serveerimiseks. Ontoloogiaid kasutatakse ühelt poolt linkandmete metamudeli jaoks ning teiselt poolt andmete linkide tekkimise tagamiseks läbi ontoloogiatega ristkasutuse. Olemasolevate registre sisu linkandmetena eksponeerimiseks sobiks iseenesest ka DBpedia lahendus, kus linkandmed tehakse kättesaadavaks otse registri andmemudelist läbi teisendus- ja vastavusreeglite. Samas on selleks vajalik andmete integratsiooniserveri kasutuselevõtt, mis vastavad teisendused ära teeks. Integratsiooniserveri kasu tuleb

rohkem esile, kui organisatsioon kaalub ühendatud RDF andmebaasi kasutuselevõttu. Sel juhul toimuks andmebaaside ühendamine läbi integratsiooniserveri.

Linkandmete hajusa haldamise ja pakkumise toetamiseks on TWC LOGD lahenduses kasutusel SPARQL lüüs. Taolise konfiguratsiooni puhul hoitakse linkandmete metaandmeid (andmeontoloogiad + andmed linkandmete allikate kohta) eraldi repositooriumis lingitud kujul ning neid kasutatakse lüüsi poolt konkreetse linkandmete allika tuvastamiseks, kuhu konkreetset linkandmete päringud edastada. TWC LOGD juhtumi korral hoitakse lüüsis ka osasid linkandmeid, et neid kiiremini serveerida.

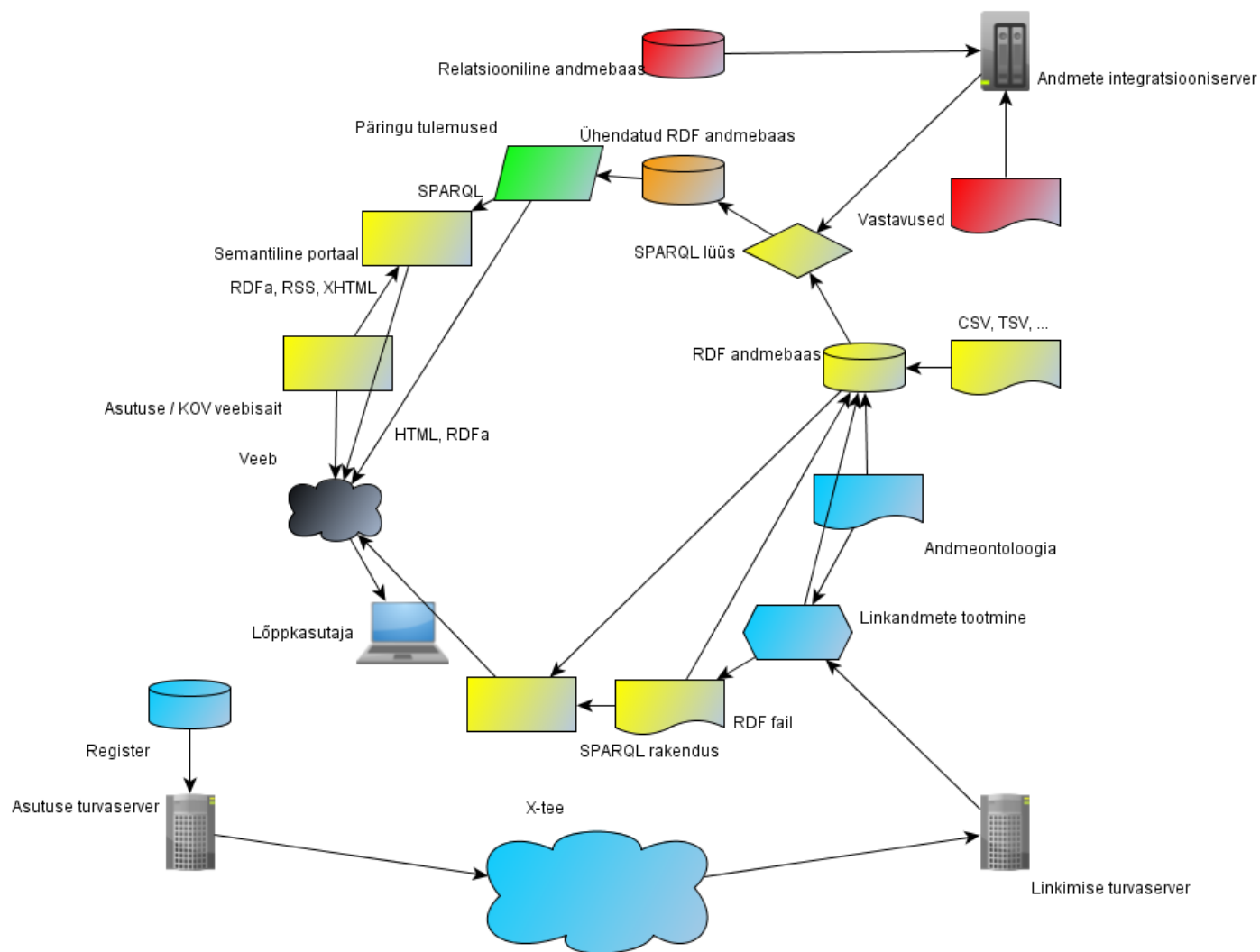
Linkandmete veebis kasutamiseks on kasutusel semantilised portaalid (TWC LOGD) ja spetsialiseeritud veebirakendused (Riiklik ehitisregister), mis kasutavad kas RDF faile või SPARQL päringuid andmete laadimiseks ja töötlemiseks. Lisaks on kasutusel lihtsakoeline esitluskiht RDF andmete esitamiseks (DBpedia), kus otsus andmete esitlusviisi (HTML, RDF, JSON jms vormingus) osas tehakse veebiserveri poolt. Semantilises portaalis agregeeritakse andmeid RDFa, RSS vormingus etteantud kanalitest (sh teised veebid), edastatakse RDF repositooriumite andmeid läbi sisuhaldussüsteemi semantikamoodulite (nt *Semantic Drupal*) ning esitatakse eelnevalt nimetatud XHTML kujul.

Riikliku Ehitisregistri avaandmete loomise lahenduse eripäraks on X-tee päringute kasutamine linkandmete loomisel. Linkandmete serveerimiseks luuakse esmalt ontoloogiates toodud metamudelit kasutades linkandmete komplekt RDF failina, mis seejärel laetakse RDF repositooriumisse SPARQL päringute toetamiseks ning tehakse samas kättesaadavaks ka RDF failina. Jooksvat SPARQL päringute teisendamist X-tee päringuteks ja vastuste tagasi RDF vormingusse teisendamist piisava jõudluse tagamiseks ei rakendata. X-tee päringute kasutamist õigustavad järgnevad argumendid:

- Andmete pärimiseks ei ole vaja teada registri seesmist andmemudelit, mis omakorda alandab linkandmete tootmise keerukust.
- Teenuste kaudu saab anda andmete linkimiseks ligipääsu vaid konkreetsele komplektile andmetele. Ideaalis oleks avaandmete jaoks eraldi andmeteenuseid.
- Teenused on juba olemas ning linkandmete tootmine ei eelda üldiselt täiendavaid muutusi registri enda lahenduses. Samas linkandmete efektiivsemaks loomiseks ja kaasajastamise automatiseerimiseks on vajalikud teenused andmete muutuste pärimiseks.
- Linkandmete loomine on lahus andmete haldamisest.

Nõrkused:

- X-tee andmeteenuste efektiivseks ja jätkusuutlikuks rakendamiseks on vaja andmeteenuseid uuenduste (kas siis muutunud objektide loetelu või konkreetsete muudatuste) pärimiseks ning avaandmete komplekti väljastamiseks registrisolevate objektide kohta.
- X-tee turvakiht poleks avaandmete iseloomu tõttu vajalik kui avaandmete tootmiseks oleks olemas andmeteenused, mis väljastaks vaid avaandmeid.



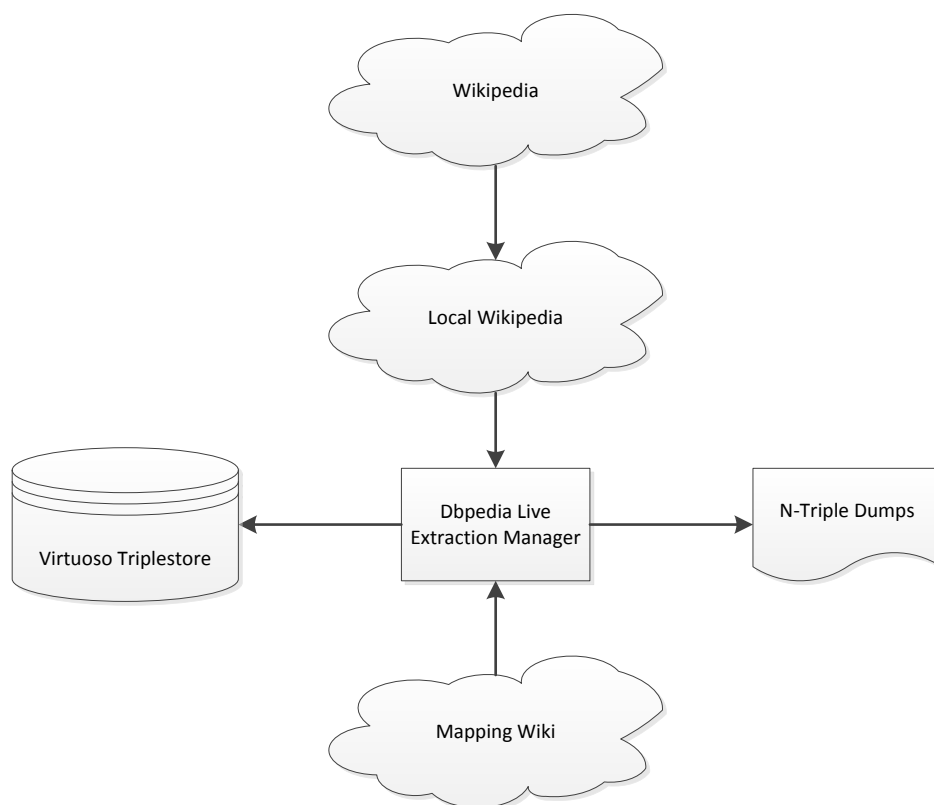
Joonis 9. Referentsarhitektuur valimi lahenduste arhitektuuridest

3.3 Juhtumiuuringud

3.3.1 DBpedia Live

DBpedia on koosloomel baseeruv kogukonnaprojekt, mille eesmärgiks on eraldada Wikipedia-st³¹⁹ struktureeritud andmeid ning teha need masinloetavalt kättesaadavaks ka teistele (veebi)rakendustele. DBpedia võimaldab esitada päringuid Wikipedia faktibaasile ja täiendada Wikipedia faktibaasi teiste andmetega läbi linkimise. Projekti ootuseks on Wikipedia andmete laiem kasutamine ning uudsete kasutamiskiiside leidmine, mis omakorda võiksid aidata kaasa Wikipedia enda platvormi arendamisele.

DBpedia inglisekeelses versioonis on esindatud ca 500 miljonit fakti RDF kolmikutena ning seetõttu on tegemist suhteliselt suure andmemahuga, mis siiski suurusjärgu võrra väiksem lingitud suurandmete mastaabist. Wikipedia andmete DBpedia andmekomplektiks teisendamise tehnilisel skeemil (vt *Joonis 10*) on näha, et DBpedia projekti Live lahenduses kasutatakse OpenLink Virtuoso Universal Server-it Wikipedia sisu lingitud kujul kättesaadavaks tegemisel.



Joonis 10. DBpedia Live System Architecture³²⁰

Virtuoso Universal Server lahendus võimaldab linkandmete tarbimist läbi veebibrauserite HTML vormingus, RDF andmekomplektina allalaadimiseks ning SPARQL klientrakenduste kaudu. Linkandmeid hoitakse RDF Quad Store RDF kolmikute repositooriumis.

³¹⁹ <http://www.wikipedia.org/>

³²⁰ <http://wiki.dbpedia.org/DBpediaLive>

3.3.2 Riiklik ehitisregister

Riikliku Ehitisregistri (EHR) linkandmete loomise lahenduses kasutatakse EHR X-tee andmeteenuseid linkandmete tootmiseks vajalike toorandmete pärimiseks. Nendest toorandmetest eemaldatakse konfidentsiaalne osa ning ülejäänud kodeeritakse RDF N-Triples vormingus kasutades EHR linkandmete ontoloogias esitatud sõnavara. Seejärel laetakse RDF andmekomplekt sarnaselt DBpedia Live lahendusele OpenLink Virtuoso RDF Quad Store'i. Andmete edasine serverimine toimub analoogselt DBpedia Live lahendusele.

EHR linkandmete SPARQL *endpoint* asub aadressil <http://ehitisregister.ee/sparql> ning lingitud avaandmete komplekti koopia ise RDF N-Triples vormingus aadressil http://ehitisregister.ee/ehr_rdf_dump.tar.gz. Linkandmete komplekti suuruseks 156 miljonit RDF kolmikut koos linkidega teistele andmekomplektidele. EHR linkandmed viitavad riikliku aadressandmete süsteemi (ADS) aadressiobjektidele ning katastriüksuse objektidele Maakatastris. Kuna ADS-i ja Maakatastri andmekomplekte endeid lingitud kujul pole, siis on vastaval lingid hetkel kohahoidjad tulevikus loodavate vastavate andmekomplektide objektidele. Taoline viis on sobilik ka näiteks mittevastavate linkandmetele või objektidele viitamisele. Kui andmeobjektide URI-de koostamise skeemid on kokkulepitud, need on üheselt määratud ning andmeobjekti URI koostamiseks piisab avalikest tunnustest (nt registripidaja nimeruum, registri lühinimi ja objekti avalik tunnus), siis saab tekitada linke objektidele ilma konkreetsete objektide täielikke kirjetele ligipääsu omamata või vastavate objektide kirjeldusi linkandmetena omamata. EHR-i lahendus detailsem kirjeldus on saadaval eraldi dokumendis (Riikliku ehitisregistri tehniline lahendus on avalikult kättesaadav järgmisest bakalaureusetööst - Hannes Mäehans, Registriandmete avalikustamisest Ehitisregistri näitel. Bakalaureusetöö. Informaatika Instituut, Tallinna Ülikool, 2012. http://www.cs.tlu.ee/teemaderegister/get_file.php?id=185).

EHR linkandmete komplekti ehitamise ja andmete linkimise lahendust saaks laiendada suvaliste andmekogude ja registrite jaoks järgnevatel eeldustel:

- Andmekogude/registrite jaoks on olemas andmeteenused andmeobjektide otsimiseks (sh uute/uuendatud kirjete pärimiseks) ning avaandmetesse minevate objektide kirjete pärimiseks.
- On olemas ontoloogiad vastavate linkandmete metamudeli jaoks.
- Vastavate X-tee andmeteenused on semantiliselt kirjeldatud kasutades vastavat metamudelit. Riigi infosüsteemi arhitektuuris on selleks ette nähtud SA-WSDL standard X-tee andmeteenuste WSDL kirjelduste jaoks.

EHR-i linkandmed on kasutusel kahes rakenduses: <http://www.ehitisregister.ee> ja <http://www.tehingud.ee/>. Esimeses rakendatakse korterite turuhinna automaatseks hindamiseks matemaatilist mudelit, mille sisendiks on vaja kortermajade ehitiste osade tehnilisi kirjeldusi. Viimased on pärit ehitisregistri avaandmetest ja tänu neile saab tehingud.ee-s rakendada keerulisi analüütilisi mudeleid. Teises rakenduses on lingitud EHR-i linkandmed ADS-i aadresside avaandmetega ning Tallinna miljööalade inventariseerimisaktide struktureerimata sisust saadud struktureeritud andmetega. Linkandmeid kasutatakse nn *citizen research* projektide koostoomes läbiviimiseks, a'la hooldamata kaevude ja mahajäetud ehitiste kaardistamisel ja ehitiste andmete kaasajastamisel. EHR-i andmete veebilahendus omakorda väljastab aadresside andmeid *microdata* vormingus (<http://schema.org/Place>) koos miljööväärtuslikkuse hinnanguga teistele rakendustele, a'la otsimootorid, tarbimiseks.

EHR-i andmete kasutusele võtmisel viidi läbi järgnevad tegevused:

1) Esmalt tuvastati, milline andmeteenuste komplekt väljastaks kõige efektiivsemalt EHR-i avaandmete komplekti ja võimaldaks automaatset andmete uuendamist. Selgus, et andmete uuenduste jaoks eraldi andmeteenust pole ning mõistlik on regulaarselt teha andmebaasist väljavõtte süsteemis leiduvatest ehitiste identifikaatoritest ning kasutada andmeteenust vastavate ehitiste andmete pärimiseks. Antud juhul piisas ühest andmeteenusest vajalike andmete kättesaamiseks.

2) Seejärel oli vaja tekitada juriidiline alus X-tee andmeteenuste kasutamiseks linkandmete ehitajale ja anda registripidaja poolt talle ligipääs päringute tegemiseks. Antud juhul oli aluseks pilootrakenduse arendusleping.

3) Järgmiseks sammuks oli andmete kogumine ja eeltöötlemine. Andmete pärimiseks kasutati X-tee andmeteenust, saadud andmetest eemaldati mitteavalikud objektid/atribuudid ning andmed talletati andmete linkimise serverisse. Kuigi konkreetseid andmete uuendamise ja kvaliteedi tagamise protseduure pole EHR-i andmete jaoks fikseeritud, on seni tehtud andmete uuendusi sagedusega kord aastas ja andmete parandusi andmete linkandmete kujule viimise käigus ei teostata.

4) Paralleelselt andmete kogumisega loodi ontoloogia EHR-i andmete metamudeli kirjeldamiseks ning lingiti see teiste olemasolevate ontoloogiatega. Aadresside metamudelina kasutati ADS-i ontoloogiat, koordinaatide jaoks ruumikujude ontoloogiat, katastriüksuste detailide jaoks loodi vajalikud elemendid uude ontoloogiasse, mis vajab edasiarendamist vastava valdkonna spetsialistide poolt. Loodud EHR-i andmete ontoloogia ei sisalda valdkondlikku teadmust ehitiste kohta.

5) Peale andmete kogumise lõpetamist ja ontoloogia realiseerimist kodeeriti EHR-i andmed linkandmete vormingus. Andmete publitseerimiseks loodi eraldi RDF fail N-Triples vormingus, mis tehti failina allalaetavaks ning lisaks laeti selle sisu OpenLink Virtuoso serverisse SPARQL päringute võimaldamiseks. Kuigi EHR-i andmete ontoloogia oli lingitud ADS-i ontoloogiaga, siis kuna ADS-i andmed ise polnud linkandmetena kättesaadavad, siis andmete otsese linkimise asemel genereeriti viited ADS-i objektidele, mida andmekomplektis kasutati. Vastavad viited tuleb vajadusel kaasajastada kui ADS-i andmed viiakse linkandmete kujule ning otsustatakse kasutada seal teistsugust viidete süsteemi kui EHR-i linkandmete komplektis.

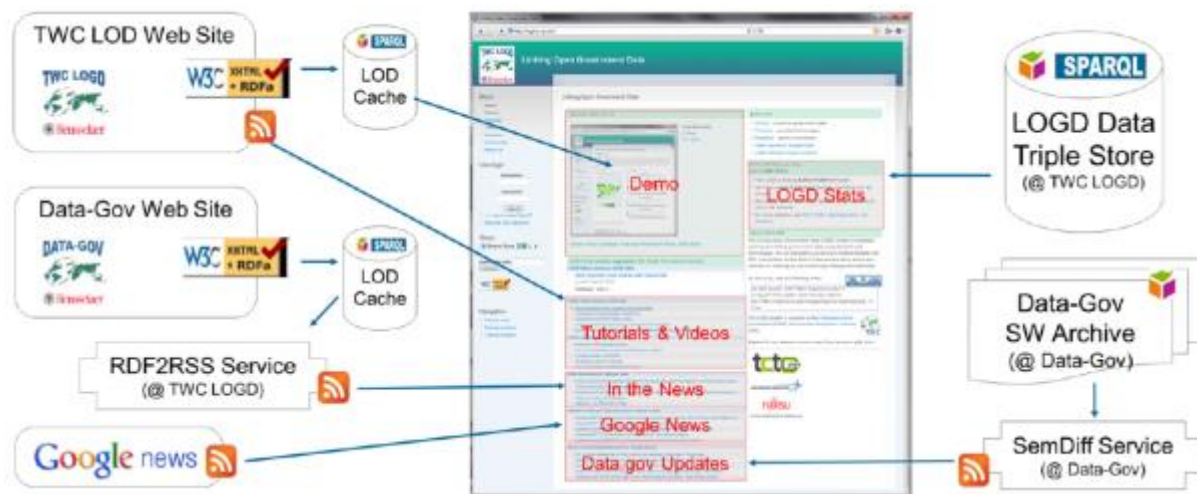
6) Ehitisregistri andmete rakendamiseks loodi ehitisregister.ee rakendus, mille esialgseks kontseptsiooniks on keskkond ehitistega seotud koosloome projektide läbiviimise lihtsustamiseks (a'la mahajäetud rajatiste või hooldamata kaevude kaardistamine). Rakenduse jaoks lingiti EHR-i linkandmed Tallinna miljöövääruslike alade inventeerimise aktidest saadud andmetega. Kuna vastavad andmed olid algselt kättesaadavad struktureerimata andmetena (tekstidokumendid PDF ja DOC vormingus), siis neile rakendati tekstitöötlemise algoritme andmete struktureeritud kujule viimiseks ning eraldati lisaks ehitiste pildid. Vastavad lingitud andmed tehti veebirakenduses kättesaadavaks.

7) Selleks, et huvitatud rakendused, nt otsimootorid, saaksid andmed ka otse veebirakendusest kätte, siis publitseeriti lehtede sisus olevad aadressid microdata vormingus (kasutades <http://schema.org/Place> definitsiooni).

8) ehitisregister.ee keskkonnas aadressidega seotud sündmuste agregeerimiseks on vajalik peale huvipakkuvate lehtede tuvastamist ja nende sisu regulaarse lugemise mehhanismi realiseerimist kodeerida neis sündmused samuti microdata vormingus (kasutades <http://data-vocabulary.org/Event> definitsiooni) ja määrata neis sündmuse asukoht. Antud rakenduse kontekstis pole seda võimalust veel realiseeritud, kuid erinevate koosloome initsiatiivide andmete agregeerimisel oleks sellest kasu.

3.3.3 TWC LOGD

TWC LOGD on ühendatud (*ingl federated*) linkandmete kataloog USA avaandmete jaoks, mis koondab kokku ca 10 miljardit RDF kolmikut, millest osa asuvad välistes repositooriumites (nt DBpedia) ning osa hoitakse kohapeal. Samas metaandmed andmekomplektide kohta on repositooriumis olemas tervikuna. Erinevate andmekomplektide jaoks on kasutusel potentsiaalsed erinevad SPARQL lüüsid (päringupunktid). 2011. a mai seisuga oli 6,4 miljardit kolmikut publitseeritud ka saidil <http://www.data.gov/semantic>, millest osad on kättesaadavad RDF failidena ja osad läbi SPARQL lüüsi (<http://services.data.gov/sparql>). Portaali iseseisev alus on baseerub Semantic Drupal'il (Drupal + selle RDF moodul) ja publitseerib andmeid RDFa-s annoteeritud XHTML lehtedena. Detailsemalt on kasutusel Drupali SPARQL laiendus veebilehe sisu koostamiseks SPARQL päringutega, mille tulemused muudetakse XSLT transformatsioonide kaudu HTMLiks. Näidise taolisest kasutamisest toob välja alljärgnev joonis (vt *Joonis 11*), kus erinevate sektsioonide täitmiseks on erinevad SPARQL päringud ja vastavad XSLT transformatsioonid.



Joonis 11. TWC LOGD portaali dünaamiliselt genereeritud esileht³²¹

Ühendatud repositoorium hoiab iga andmehulga kohta tema metaandmeid (nimi, kirjeldus, organisatsioon jne). SPARQL päringute tulemuste vormingute laiemaks valikuks on arendatud tööriist SparqlProxy ja RDFa-s annoteeritud veebilehtede sisu portaaliüleseks sünkroniseerimiseks on loodud komponent LOD cache. Lisaks on loodud tööriistad andmete importimiseks ja täiendamiseks. Päringute vahendamise SPARQL lüüs asub aadressil <http://logd.tw.rpi.edu/ws/sparqlproxy.php>. Linkide tekitamise erinevate andmekomplektide vahel tagavad URIdisainimise juhised/poliitikad.

³²¹ Allikas: <http://dx.doi.org/10.1016/j.websem.2011.06.002>

3.4 Vastavus Linked Data Platform 1.0 nõuetele

Kokkuvõtte valimi lahenduste vastavusest Linked Data Platform 1.0 mustandis toodud nõuetele on toodud alljärgnevas tabelis (vt Tabel 5). Veerud TWC LOGD DBpedia ja RE täistavad vastavalt TWC LOGD, DBpedia Live ja Riikliku Ehitisregistri andmete tarnimise tehnilisi lahendusi.

Tabel 5. Linked Data Platform 1.0 nõuete kaetus valimi lahenduste poolt

Linked Data Platform Use Cases and Requirements, W3C Working Draft 31 October 2013, http://www.w3.org/TR/ldp-ucr/	TWC LOGD	DBpedia	RE
F1.1: The system shall provide the ability to create containers for composing resources, from UC1.	Y	Y	Y
F1.2: The system shall provide the ability to create nested containers, from UC1.	N	N	N
F1.3: On deletion of a container, the system shall delete any contained resources and nested containers, from UC1.	Y	Y	Y
F2.1: The system shall provide the ability to create resources within a container, from UC2.	Y	Y	Y
F2.2: The system shall provide the ability to delete resources, from UC2.	Y	Y	Y
F3.1: The system shall provide the ability to retrieve resource descriptions, from UC3.	Y	Y	Y
F3.2: The system shall enable the client to retrieve the description of a hash URI, from UC3.	N	N	N
F4.1: The system shall provide the ability to update an existing resource by substitution, from UC4.	N	N	N
F4.2: The system shall provide the ability to perform a selective update of a resource, from UC4.	N	N	N
F5.1: The system shall provide the ability to determine if a resource has changed, from UC5.	N	N	N
F6.1: The system shall provide the ability to aggregate resources, from UC6.	Y	Y	Y
F6.2: The system shall support the addition of a resource to multiple aggregations, from UC6.	N	N	N
F7.1: The system shall provide the ability to retrieve a collection-level description of a composition, from UC7.	N	N	N
F7.2: The system shall provide the ability to retrieve an item-level description of a composition or aggregation, from UC7.	N	N	N
F8.1: The system shall provide the ability to retrieve a paginated description of a composition or aggregation, from UC8.	Y	Y	Y
F9.1: The system shall provide the ability to store and access media resources, from UC9	Y	Y	Y
NF1.1: The system shall provide access guidance to resources, from UC1.	Y	Y	N
NF2.1: The system shall encourage non-duplication of resources, from UC2.	N	N	N
NF2.2: The system shall support distribution of resources, from UC2.	Y	N	N
NF2.3: The system shall support consistent, global naming, from UC2.	Y	Y	Y
NF3.1: The system shall support the use of standard vocabularies where appropriate, from UC3.	Y	Y	N
NF3.2: The system shall provide a scalable linking model, from UC3.	Y	Y	Y
NF4.1: The system shall permit unrestricted vocabulary, from UC4.	Y	Y	Y
NF5.1: The LDP shall ensure consistent access in the case of multiple simultaneous attempts to access a resource, from UC5.	Y	Y	Y
NF6.1: The system shall allow resource descriptions that are a "mix of simple data and collections", from UC6.	Y	Y	Y
NF6.2: The system shall support relative URIs enabling sharing of collections, from UC6.	Y	Y	Y
Kokku Y	17	16	14

Peamised erinevused nimetatud lahendustes on nõuetes NF1.1, NF2.2 ja NF3.1. NF1.1-s tuleneb Riikliku Ehitisregistri lahenduse nõuetele mittevastavus SPARQL näidispäringute dokumentatsiooni mittekättesaadavusest. Samas on Riikliku Ehitisregistri andmete metamudel ontoloogiana kättesaadav opendata.riik.ee keskkonnas ja selle baasil saab RDF-ga kursisolev programmeerija vastavad näidispäringud koostada. NF2.2 näol on tegemist toote omadusega, mis on küll planeeritud Virtuoso 6-e, mis oli uuringu teostamise ajal DBpedia Live ja Riikliku Ehitisregistri linkandmete serverimiseks, kuid polnud seal realiseeritud (http://www.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSVirtuoso6FAQ#Does_Virtuoso_support_federated_triple_stores?). NF3.1 juures tuleneb nõudele mittevastavus asjaolust, et Riikliku Ehitisregistri andmete ontoloogia on standardiseeritud riigi ontoloogiate piires, nt ADSi ontoloogiast on pärit koordinaatide definitsioonid, kuid võiks kasutada globaalselt kasutuselolevaid ontoloogiaid.

3.5 Ettepanekud

Arhitektuurianalüüsis vaadeldi kolme lahendust, mis esindavad peamisi linkandmete loomise ja kasutamise viise. Arvestades suhteliselt suurt olemasolevate andmeteenuste mahtu, hästi juurdunud üleriiklikku teenustepõhist arhitektuuri avaliku sektori poolt andmete pakkumisel ja kohalike omavalitsuste veebide lahenduste standardiseerimist (KOVTP raames) on linkandmete puhul Eestis järgmise 5-10 aasta jooksul ratsionaalne laiendada andmeteenuste kasutamist registripõhiseks linkandmete tootmiseks ning veebe kohalike omavalitsuste linkandmete tootmiseks. Registripõhiseks linkandmete tootmiseks on vajalik andmeteenuste arendamise põhimõtete väljatöötamine, andmeteenuste semantiline kirjeldamine ja tehnilise lahenduse loomine avaandmeid väljastavate andmeteenuste rakendamiseks väljaspool X-teed. Veebide kasutamiseks linkandmete tootmiseks on vaja laiendada KOVTP lahenduse sisuhalduse mooduleid selliselt, et nad lisaks visuaalsele sisule väljastaks ka sisu RDFa/microdata märgenditega annoteeritult.

Esmajärjekorras annab suurima võidu vastavalt annoteeritud sisu kasutamise seisukohast kontaktisikute, asukohtade ja sündmuste annoteeritud kujul väljastamine.

Andmeobjektide vaheliste linkide tekkimise tagamiseks on vaja konkretiseerida poliitikad ressursside URIdes disainimiseks. Sama URIdes skeemi kasutamine tagab hajusalt hallatud ning linkandmete vahel linkide tekke. Praegu saaks URIdes disainimisel võtta aluseks RIHA (<https://riha.eesti.ee>) valdkonnasõnastike jaoks loodud URIdes disainimise skeemi, kuid see on mõeldud peamiselt metamudelite, mitte konkreetsete andmeobjektide jaoks. Seega, enne vastava skeemi kasutuselevõtmist peab seda kaasajastama linkandmete seisukohast. Samuti, plaanides andmeteenuste kasutamist linkandmete tootmiseks, vajavad kaasajastamist semantilise koosvõimega seotud juhised ja meetodikad tervikuna.

Kiirendamaks andmete linkimist ja lihtsustamaks linkandmete kasutamist koos linkimata andmetega tuleb luua nn linkandmete tuumik. Linkandmeid saab kasutada koos linkandmete vormingusse viimata andmetega kasutades integratsiooniservereid vms tehnoloogiaid, mis lihtsustavad suure variatiivsusega andmemudelite metamudelite ühitamist ja andmete koos kasutamist. Luues linkandmete komplektid peamistele andmeobjektidele, mis on esindatud enamikes andmekoosseisudes ja rakendustes (juriidilised ja füüsilised isikud, aadressid, geoinfo), tekib nn linkandmete tuumik, mis lihtsustab andmete linkimist või linkimata andmete kasutamist integratsiooniserverite jt tööriistade kaudu. Seda nn tuumikut võib tinglikult vaadelda kui riigi infosüsteemi andmete integratsioonikihi võimaldajat.

4 Õigusruumi analüüs

Õigusruumi analüüs käsitleb järgmisi küsimusi:

Millised on linkandmete kasutuselevõtuga kaasnevad võimalikud tugevused ja nõrkused, võimalused ja ohud, sh tavaelanike, ettevõtete ja riigi jaoks? Kuidas ohtusid ja riske saaks õigusaktide abil ennetada ja maandada?

Kuidas tagada linkandmete kasutuselevõtu raames isikuandmete kaitse, kas ja mida selleks on täiendavalt vajalik teha?

Kas ja kuidas peaks linkandmete kasutuselevõtuks muutma Eesti õigusruumi? Millised on hetkel olevad piirangud ning tulevikus reguleerimist vajada võivad teemad?

Kui eeldada, et on võimalik maandada ava- ja linkandmetega seotud riske, siis kuidas stimuleerida linkandmete avaldamist ja kasutamist?

4.1 Analüüsi eesmärgid, ulatus ja piirangud

4.1.1 Õigusruumi analüüsi üldised eesmärgid

Linkandmete kasutuselevõtt ning sellega seoses ka õigusruumi analüüs peaks soodustama järgmiste üldiste eesmärkide saavutamist:

- Inimeste parem elu Eestis.
- Ettevõtete konkurentsivõime kasv.
- Riigi poolt pakutavate teenuste efektiivsem arendamine.
- Paremate otsustuste tegemine.
- Jätkuv kinnipidamine Eesti Vabariigi põhiseaduses sätestatud põhiõigustest, vabadustest ja kohustustest, sealhulgas intellektuaalomandi ning inimeste turvalisuse, isikuandmete ja privaatsuse jätkuv kaitse.
- Linkandmetega seotud võimalike riskide maandamine.

Olulised on ülaltoodud üldised eesmärgid, mitte linkandmete kasutuselevõtt kui selline. Sellest järeldub, et kui linkandmete laiem kasutuselevõtt neid eesmärke ei toeta (või kui negatiivsed mõjud on suuremad kui positiivsed), peaks linkandmetest pigem loobuma või nende kasutamist kitsendama. Näiteid uute tehnoloogiate kitsendatud kasutamise kohta võib tuua erinevatest valdkondadest nagu geenitehnoloogia või tuumaenergeetika.

4.1.2 *Ava- ja linkandmete kasutuselevõtu võimalike tugevuste, nõrkuste, võimaluste ja ohtude analüüs*

Ava- ja linkandmete kasutuselevõtt tulenevad võimalikud positiivsed efektid, mida õigusruum peaks võimaldama ning mida on pakutud mitmetes allikates (sh^{322,323}), on järgmised:

- Riikidele, era- ja vabasektorile võimaluste loomine uut tüüpi teenuste loomiseks.
- Majanduse elavdamine, sisemajanduse kogutoodangu suurendamine.
- Avalikus sektoris toodetud andmete taaskasutuse edendamine ja kasutusse andmine, sh vastava taristu loomine.
- Loomemajanduse, IKT ja teiste võtmetehnoloogiate potentsiaali laiem kasutamine teiste sektorite lisandväärtuse tõstmisel.
- Läbipaistvuse suurendamine, teabenõuete arvu vähendamine ja sellega avaliku sektori koormuse vähendamine, avatud teenuste loomise ja haldamise loovutamine era- ja vabasektorile.

Riskid, mida õigusruumi arendamine ja muud meetmed peaksid aitama neutraliseerida, on järgmised:

- Oht isikule. Isikuandmete kaitse ja privaatsuse probleemid.
- Oht ettevõttele. Konfidentsiaalsete andmete kaitse probleemid. Intellektuaalomandi kaitse probleemid³²⁴.
- Oht efektiivsusele avaliku sektori asutustes. Asjatu lisatöö ja bürokraatia.
- Oht riigile, sealhulgas riigisaladusele või kriitilisele taristule. Linkandmete laiaulatuslik kasutamine võimaliku riikidevahelise konflikti korral. Konfidentsiaalsete andmete kaitse probleemid.
- Oht ühiskonnale. Otstarbekas on arvestada võimalust, et inimesed ei suuda säilitada kontrolli omavahel seotud tehnilike süsteemide üle, mis võivad töötada sõltumatuid eesmärgi seades ning linkida ja kasutada tehisseadmeid (sealhulgas droone jms).

Tehnilised efektid, millel võib olla nii negatiivne kui ka positiivne mõju, on muuhulgas järgmised:

- Seni hajutatult eksisteerivate andmebaaside, sotsiaalmeedia, dokumentide, faktide, seadmete, teadmiste jms integreerimine ühtseks tervikuks.
- Kõikide füüsiliste ja abstraktsete „asjade“ muutmine arvutite abil leitavateks, loetavateks, seostatavateks ja töödeldavateks.

³²² Avaandmete roheline raamat (kavand). Versioon 0.4. 2013-11-18

³²³ Eesti 2020 tegevuskava 2011-2015, <http://valitsus.ee/et/riigikantselei/eesti2020>

³²⁴ Legal barriers to linked open data <http://linkedup-project.eu/2013/05/23/legal-barriers-to-linked-open-data/>

Järgnevas tabelis on toodud linkandmete kasutuselevõtu võimalike tugevuste, nõrkuste, võimaluste ja ohtude analüüs (SWOT-analüüs) seoses püstitatud eesmärkidega:

Tabel 6. Linkandmete kasutuselevõtu SWOT-analüüs

	Kasulikud eesmärkide suhtes	Kahjulikud eesmärkide suhtes
Sisemised	<i>Tugevused</i> <ul style="list-style-type: none"> Avalikus sektoris toodetud andmete taaskasutuse edendamine ja kasutusse andmine. Halduskoormuse vähenemine. Võimalused uut tüüpi teenuste loomiseks. 	<i>Nõrkused</i> <ul style="list-style-type: none"> Tekivad lisatöö, -kulutused ja -bürokratia avaliku sektori (link)andmete loomisel ja haldamisel. Uued regulatsioonid osutuvad kitsendavaks uute tehnoloogiate suhtes.
Välised	<i>Võimalused</i> <ul style="list-style-type: none"> Inimeste parem elu. Ettevõtete parem konkurentsivõime. Paremad teenused ja otsustused. Sisemajanduse kogutoodangu (SKT) suurendamine. Parem maine rahvusvahelistes hinnangutes ning sellega kaasnevad efektid, nt investeeringud ja SKT kasv. 	<i>Ohud</i> <ul style="list-style-type: none"> Oht privaatsusele, isikuandmete leke seoses linkandmete kasutamisega. Intellektuaalomandi kaitsega seotud ohud. Ohud suurte konfliktide korral. Oht, et inimesed ei suuda säilitada kontrolli integreeritud tehnilike süsteemide üle.

4.1.3 Analüüsi piirangud: hõlmataavad tehnoloogiad ja õigusruum

Arvestades jaotises 7.1.2 toodud järeldust, et komplekt „avaandmed + semantiline veeb“ on linkandmete kasutuselevõtuks primaarne, lülitame õiguslikku analüüsi eelkõige avaandmete, semantilise veebi ja linkandmetega seotud väljakutsed ja probleemid. Kuna semantilist veebi võib käsitleda linkandmete tehnoloogia ühe komponendina, vaatame edasises põhiliselt ava- ja linkandmeid, arvestades seejuures ka teisi tehnoloogilisi suundumusi – asjade Internetti, semantilist veebi, tulevikuinternetti, suurandmeid ja koosloomet - mis võimendavad ava- ja linkandmete mõju ning tugevdavad tehtud järeldusi.

Andmed või sisu on avatud, kui see kõigile antud vabalt kasutada, taaskasutada ja jagada – kitsendatult maksimaalselt viitamise või samadel tingimustel taasjagamise nõudega³²⁵. Mõiste „avaandmed“ räägib õiguslikust koosvõimest, mõiste „linkandmed“ semantilistest ja tehnilisest koosvõimest³²⁶. On võimalikud kõik neli andmete avatuse ja lingituse kombinatsiooni, sealhulgas avaandmed, mis pole linkandmed ning linkandmed, mis pole avaandmed.

Hetkel olevad piirangud on lisaks tehnoloogiale eelkõige seotud Eesti õigusruumiga. See sisaldab Eesti õigusakte, Eestile kohaldatavaid rahvusvahelisi õigusakte ning muid õigusruumi puudutavaid materjale. Järgnevas tekstis on õigusaktid toodud allviidetena. Eelkõige tuleb arvestada järgmiste õigusaktide ning

³²⁵ <http://opendefinition.org/>, <http://www.w3.org/2005/Incubator/lld/wiki/Scope>

³²⁶ <http://www.w3.org/2005/Incubator/lld/wiki/Scope>

muude materjalidega (seaduste puhul on sulgudes toodud ametlikud lühendid³²⁷, mida tekstis on kohati kompaktsuse huvides kasutatud).

- Eesti Vabariigi põhiseadus (PS).
- Avaliku teabe seadus (AvTS).
- Isikuandmete kaitse seadus (IKS).
- Riikliku statistika seadus (RStS).
- Autoriõiguse seadus (AutÕS).
- Riigi infosüsteemi haldussüsteem. Vabariigi Valitsuse määrus. Vastu võetud 28.02.2008 nr 58.
- Infosüsteemide turvameetmete süsteem. Vabariigi Valitsuse määrus. Vastu võetud 20.12.2007 nr 252.
- Andmekogude asutamist ja toimimist käsitlevad õigusaktid.
- Muud seadusandlikud aktid vastavalt vajadusele, näiteks õigusaktid, milles sätestatakse tasud andmete väljastamise eest.
- Eesti infoühiskonna arengukava 2020 (heaks kiidetud Vabariigi Valitsuse poolt 14.11.2013). Majandus- ja Kommunikatsiooniministeerium, <http://263654.edicypages.com/eesti-infouhiskonna-arengukava-2020/infouhiskonna-arengukava-2020-loppversioon>.
- Konkurentsivõime kava „Eesti 2020“ (kinnitatud Vabariigi Valitsuse poolt 25.04.2013) <http://valitsus.ee/et/riigikantselei/eesti2020>.
- Avaandmete roheline raamat (kavand). Versioon 0.4. 2013-11-18.
- DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:ET:HTML>, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:en:NOT>.
- EUROOPA PARLAMENDI JA NÕUKOGU DIREKTIIV 2013/37/EL, 26. juuni 2013, millega muudetakse direktiivi 2003/98/EÜ avaliku sektori valduses oleva teabe taaskasutamise kohta, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:ET:PDF>.
- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=31995L0046&model=guichett.
- EL uus andmekaitse direktiiv. 2012. European new data protection law. http://www.mlawgroup.de/news/publications/detail.php?we_objectID=227.
- Berni kirjandus- ja kunstiteoste kaitse konventsiooniga ühinemise seadus, <https://www.riigiteataja.ee/akt/24662>. Berni kirjandus- ja kunstiteoste kaitse konventsioon, <https://www.riigiteataja.ee/akt/13101723>.

³²⁷ <https://www.riigiteataja.ee/lyhendid.html?sorteeri=pealkiri&kasvav=true>

4.2 Avaandmete/linkandmete publitseerimine

Et avaandmeid / linkandmeid kasutada, peaksid nad olema olemas. Analüüsime seepärast kõigepealt, kuidas on õigusruumis sätestatud avaandmete/linkandmete publitseerimine, nende masinloetavus, andmete allalaadimise viis ning mida oleks vaja õigusaktides sellega seoses muuta.

4.2.1 Avalikustatavad andmed ja avalikustamise tähtajad

AvTS § 58² reguleerib avaliku teabe taaskasutamist:

- (1) Teabevaldaja viib pärast käesoleva paragrahvi jõustumist loodavas andmekogus sisalduva käesoleva seaduse § 28 lõike 1 punktis 30 nimetatud teabe vastavusse § 29 lõikes 4 nimetatud nõudega.
- (2) Igäihele tagatakse tasuta juurdepääs pärast käesoleva paragrahvi jõustumist asutatavas andmekogus sisalduvatele täiendavalt töötlemata andmetele.
- (3) Igäihele tagatakse tasuta juurdepääs enne käesoleva paragrahvi jõustumist asutatud andmekogus sisalduvatele täiendavalt töötlemata andmetele hiljemalt 2015. aasta 1. jaanuariks.“

Kuna AvTS § 58² lg 1 sisaldab viiteid muudele paragrahvidele, toome selguse mõttes need viited toodud koos nurksulgudes kommentaaridega: „§ 58². Avaliku teabe taaskasutamist reguleerivate sätete rakendamine (1) Teabevaldaja viib pärast käesoleva paragrahvi jõustumist [29.12.2012] loodavas andmekogus sisalduva käesoleva seaduse § 28 lõike 1 punktis 30 nimetatud teabe [andmekogudes sisalduvad andmed, millele ei ole kehtestatud juurdepääsupiirangut] vastavusse § 29 lõikes 4 nimetatud nõudega [Käesoleva seaduse § 28 lõike 1 punktis 30 kirjeldatud teave peab olema avalikustatud masinloetaval kujul ning olema tervikliku andmekogumina allalaaditav teabevärava kaudu.]“.

Lisaks avaliku teabe a seaduse § 28 lõike 1 punktis 30 nimetatud teabele võib avaliku sektori infosüsteemides olla mitmesugust muud teavet, mida oleks otstarbekas taaskasutada. AvTS § 43³ lõike 1 põhjal asutatakse andmekogu seadusega või selle alusel antud õigusaktiga, seega nende infosüsteemide andmeid, mida pole sellisel moel asutatud, avalikustama ei pea. Ka Andmekaitse Inspeksiooni poolt 2013. a avaldatud andmekogude juhendis³²⁸ tehakse selget vahet andmekogude ja muude avaliku sektori infosüsteemide vahel. Lisaks andmekogude andmetele tuleks publitseerida ka muud avaliku sektori juurdepääsupiiranguteta andmed vastavalt AvTS § 58² ning direktiivide 2003/98/EL ja 2013/37/EL nõuetele.

Et andmete avalikustamise ja seega ka linkandmete kasutamise üks olulisemaid regulatsioone on Euroopa direktiiv 2013/37/EL, tuleks kehtestada selle järgimiseks vajalikud õigus- ja haldusnormid. Muuhulgas tuleks täiendada RIHA määrust avaandmete publitseerimisega seotud sätetega; see vajadus tuleneb ka üldisemast ava- ja linkandmete temaatikast. Samuti tuleks avaliku sektori asutustel vaadata läbi oma andmekogude aluseks olevad õigusaktid ning viia need vastavusse taaskasutuse printsiipidega. Näited võimalike täienduste kohta on järgnevatel alajaotistes.

Arvestades seda, et olemasolevate andmekogude avalikustamist pole tihti veel alustatud, et juba avalikustatud andmed pole paljudel juhtudel masinloetavad ning et asutused enamasti ei loe andmete avalikustamist esmase prioriteediga ülesandeks, võib tähtaeg 01.01.2015 olla ebarealistlik. Õiguslikus tähenduses juhul, kui andmete avaldamise tähtaeg lõpeb ning seaduses sätestatud andmetele ligipääsud ei ole ette valmistatud ning andmetele seaduses ettenähtud viisil ligi ei pääse, rikuvad teabevaldajad kehtivat seadust. Isikutel, kellel on õigus seaduses sätestatud andmetele ettenähtud viisil juurde pääseda ning kes seda teabevaldaja seadust rikkuva tähtaja järgimata jätmise tõttu teha ei saa, saavad pöörduda kaebuse esitamiseks halduskohtusse, kes võib teabevaldajat kohustada oma tegevus seadusega kooskõlla viia. Arvestades, et direktiiv 2013/37/EL kehtestab direktiivi sätete kehtestamiseks tähtaja 18.07.2015, oleks võimalik avaliku teabe seaduses nimetatud tähtaja ühtlustamine direktiivis nimetatud tähtajaga, võttes seejuures arvesse ka direktiivist tulenevaid lisanõudeid.

³²⁸ https://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/Andmekogude%20juhend%202013_o.rtf

AvTS § 58² lõiked 2 ja 3 ei viita § 28 lõike 1 punktis 30 nimetatud teabele, nõudes seega tasuta juurdepääsu tagamist kõikide peale 2012. a asutatud andmekogude andmetele ning (peale 01.01.2015) kõikide olemasolevate andmekogude andmetele. Siiski sätestab avaliku teabe seadus teabevaldaja kohustuse järgida andmetele kehtestatud juurdepääsupiiranguid (§ 9 lg 2 p 7).

4.2.2 Avalikustatavate andmete masinloetavuse tase

AvTS § 8 lg 3 sätestab, et teabele juurdepääs hõlmab ka õigust seda teavet taaskasutada, kusjuures taaskasutamise mõiste avab sama seaduse § 3¹. Seega võib selle seaduse põhjal avalikustatud andmeid enamasti kasutada kui ava- või linkandmeid (küsimusi võib tekkida seoses intellektuaalomandi õigustega, mida analüüsitakse allpool).

Kasutamiseks peaksid linkandmed olema kättesaadavad struktureeritud andmetena, mis on kirjeldatud ja avaldatud viisil, mis lubab neid automaatselt seostada. Avaliku teabe seadus seda ei kindlusta, sest kuigi seaduse § 29 lg 4 sätestab: „Käesoleva seaduse § 28 lõike 1 punktis 30 kirjeldatud teave [andmekogudes sisalduvad andmed, millele ei ole kehtestatud juurdepääsupiirangut] peab olema avalikustatud masinloetaval kujul...“, pole masinloetavuse mõiste seaduses määratletud. Seega ei kindlusta see formuleering andmete avaldamist linkimiseks sobival kujul. Esimene samm olukorra täpsustamisel võiks olla masinloetavuse mõiste täpsustamine avaliku teabe seaduses.

Kui tugev võiks masinloetavuse nõue olla? Linkandmete parema kasutusele võtmise seisukohast võiks eelistada masinloetavuse nõude täpsustamist andmekogude jaoks vähemalt nelja täрни tasemele. Selle lahenduse risk on, et asutuste kulutused andmete publitseerimiseks on liiga suured ja seadust ei hakata täitma. Võimalik lahendus oleks seega viia avaliku teabe seaduses andmete publitseerimise nõue vastavusse direktiiviga 2013/37/EL („masinloetav vorming” – failivorming, mis on struktureeritud selliselt, et tarkvararakendused suudavad spetsiifilisi andmeid, sealhulgas üksikuid faktiväiteid, ja nende sisemist struktuuri kergelt tuvastada, ära tunda ja välja lugeda“). Masinloetavust vähemalt nelja täрни tasemel võiks nõuda eelkõige olulisemate kesksete andmekogude puhul.

Märgime ka, et Eesti seadustik seab mõnevõrra tugevamad nõuded kui direktiiv 2013/37/EL. Viimane kohustab avaliku sektori asutusi tegema „dokumendid kättesaadavaks mis tahes olemasolevas vormis või keeles ning, kui see on võimalik ja asjakohane, avatud ja masinloetavas vormingus koos metaandmetega. Nii vorming kui ka metaandmed peaksid olema võimalikult suures ulatuses kooskõlas ametlike avatud standarditega“. Kuigi mainitud direktiivi nõue dokumentide kättesaadavaks tegemiseks masinloetaval kujul vaid võimaluse ja asjakohasuse korral on suurel määral orienteeritud paberdokumentidele ja raamatutele, on see direktiivis sõnastatud üldise printsiibina.

4.2.3 Andmete avalikustamise viis ja nõuded

Et andmete linkimine oleks lihtsam, on soovitatav sätestada andmete avalikustamise viis ja nõuded võimalikult arusaadavalt. Muuhulgas, AvTS § 29 lg 4 („Käesoleva seaduse § 28 lõike 1 punktis 30 kirjeldatud teave peab olema avalikustatud masinloetaval kujul ning olema tervikliku andmekogumina allalaaditav teabevärava kaudu“) jätab lahtiseks, millisest teabeväravast on jutt. Vastav nõue täpsustatakse vastavas rakendusjuhendis.

Toodud säte jätab samuti lahtiseks, kas teabeväravas piisab viitamisest konkreetse andmekogu andmetele või on vaja midagi rohkemat. Kuna ka rahvusvahelises praktikas kasutatakse sellist viitamist (nt UK data.gov.uk portaalis) ning see muudab andmete publitseerimise ja ajakohasena hoidmise oluliselt lihtsamaks, võiks rakendusaktides ja juhendites täpsustada, et teabeväravas võib kasutada nii viitamist konkreetse andmekogu andmetele kui ka andmete esitamist teabeväravas. Tuleks valmistada ette Eesti avaandmete varamut asutav ja korraldav õigusakt.

AvTS § 58² tekstis on mõiste „täiendavalt töötlemata andmed“, mida tuleks täpsustada, sest praegune tekst lubab mitmesuguseid interpretatsioone. Näiteks tuleb linkimiseks sobival kujul avalikustamiseks andmeid tihti täiendavalt töödelda, et rahuldada avaandmete viie tärni süsteemi kõrgemate tasemete nõudeid.

AvTS § 29 lg 4 sätestab: „Käesoleva seaduse § 28 lõike 1 punktis 30 kirjeldatud teave peab olema avalikustatud masinloetaval kujul ning olema tervikliku andmekogumina allalaaditav teabevärava kaudu“. Kuna reaalsed andmekogud on pidevas muutumises, võib sellist nõuet mitmeti tõlgendada ning selle täitmine võib olla raskendatud. Näiteks, kas mitme aasta vanune koopia andmekogust, millele viidatakse teabevärava kaudu, rahuldab avaliku teabe seaduse nõuet?

Idealis peaks avaliku teabe seaduse mõte andmete avalikustamise osas olema selline, et tagada igapäevasele ajakohaste (hetkel kehtivate, õigete) andmete kättesaadavus. Kahjuks ei ole see üldjuhul ei põhimõtteliselt ega tehniliselt reaalne, sest muutuvate andmete puhul kajastavad süsteemi andmed reaalsust alati teatava hiline misega. Koopiad, avalikustatavad andmed jm teisene informatsioon jääb omakorda süsteemi andmetest maha. Selliste ajaliste hiline misete minimeerimine võib nõuda väga kulukaid tehnilisi lahendusi. Küsimus pole seega, kas avalikustatavad andmed on reaalsusest taga, vaid kui palju nad seda on.

Teabe avalikustamise puhul kehtiv põhimõte, mille kohaselt vajalik teave peaks jõudma selle vajajani võimalikult kiiresti, sisaldub AvTS § 30 lõike 1 esimeses lauses: „Teabevaldaja on kohustatud teabe avalikustama viisil, mis tagab selle jõudmise võimalikult kiiresti igapäevasele, kes teavet vajab“. See sõnastus jätab siiski tegelikud nõuded raskesti defineeritavateks. Üks võimalus andmete allalaadimise nõuete täpsustamiseks on toodud direktiivi 2003/98/EÜ sättes (12), mis ütleb muuhulgas: „ Kui taaskasutamise taotlus on rahuldatud, peaksid avaliku sektori asutused dokumendid kättesaadavaks tegema sellise aja jooksul, mis võimaldab ära kasutada nende kogu majanduslikku potentsiaali. Eriti oluline on see dünaamiliste andmete (nt liiklusandmed) puhul, mille majanduslik väärtus oleneb andmete ja korrapäraste uuenduste viivitamatust kättesaadavusest. Litsentseerimise korral võivad litsentsi osaks olla sätted kiire kättesaadavuse kohta“.

Õigusaktis, millega andmekogu asutatakse, peaksid olema täpsustatud andmekogude andmete allalaadimise nõuded, muuhulgas näiteks sätestades allalaaditava versiooni maksimaalse ajalise mahajäämuse tegelikust jooksvast versioonist. Sellised nõuded võib kehtestada ka andmekogu registreerimisel riigi infosüsteemi haldussüsteemis RIHA.

4.2.4 Ettepanekud

1. AvTS § 58² räägib ainult andmekogudest, viidates § 28 lõike 1 punktile 30. Linkandmete kasutuselevõtu soodustamiseks tuleks kiiremas korras võimaldada ka AvTS §28 lg 1 kõigis ülejäänud alapunktides viidatud andmete taaskasutamist, mida soovivad direktiivid 2003/98/EL ja 2013/37/EL.
2. Et andmete avalikustamise ja seega ka linkandmete kasutamise üks olulisemaid regulatsioone on Euroopa direktiiv 2013/37/EL, tuleks kehtestada selle järgimiseks vajalikud õigus- ja haldusnormid, muuhulgas täiendades RIHA määrust avaandmete korrastamisega seotud sätetega.
3. Avaliku sektori asutustel tuleks vaadata läbi oma andmekogude aluseks olevad õigusaktid ning viia need vastavusse taaskasutuse printsiipidega, sealhulgas sätestades andmete avalikustamise nõude ja selle võimalikud piirangud.
4. Täpsustada avaliku teabe seaduses masinloetavuse mõistet, soovitatavalt kooskõlas direktiivis 2013/37/EL sätestatuga. Masinloetavust vähemalt nelja tärni tasemel võiks nõuda eelkõige olulisemate kesksete andmekogude puhul.
5. Täpsustada rakendusaktides ja juhendites seda, et teabeväravas võib kasutada nii viitamist konkreetse andmekogu andmetele kui ka andmete esitamist teabeväravas. Valmistada ette Eesti avaandmete varamut asutav ja korraldav õigusakt.

6. Sätestada nõue, et andmekogu registreerimisel riigi infosüsteemi haldussüsteemis RIHA tuleb näidata, kui kiiresti peab vajalik teave peaks jõudma selle vajajani (näiteks määrates ära allalaaditava versiooni maksimaalse ajalise mahajäämuse tegelikust jooksvast versioonist).

4.3 Avaliku teabe taaskasutamise tasustamine

AvTS sätestab §-s 4 lg 4: „Juurdepääs teabele võimaldatakse tasuta, välja arvatud juhul, kui teabe väljastamisega seotud otseste kulutuste eest maksmine on seadusega ette nähtud“.

Vastavalt AvTS § 58² lõikele 2 tuleb igapähele tagada tasuta juurdepääs pärast 29.12.2012 asutatavas andmekogus sisalduvatele täiendavalt töötlemata andmetele. Vastavalt sama paragrahvi lõikele 3 tuleb hiljemalt 2015. aasta 1. jaanuariks igapähele tagada tasuta juurdepääs enne 29.12.2012 asutatud andmekogus sisalduvatele täiendavalt töötlemata andmetele. Mittespetsialistidele võivad § 4 lg 4 ning § 58² lg 2 ja 3 nõuded näida vastuolulised, neid võiks rakendusaktides või juhendites täpsustada.

Näitena muudes õigusaktides sätestatud tasude kohta toome Eesti topograafia andmekogu (ETAK). ETAK on riigi infosüsteemi kuuluv andmekogu, kuhu kantakse üldist tähtsust omavate topograafiliste nähtuste andmed ning andmed, mis kirjeldavad nende nähtuste sisu, suhteid ja konteksti³²⁹. Ruumiandmete seaduse § 73 lg 2 toob ära tasuliste andme- ja võrguteenuste, millega võimaldatakse ETAK-i andmete kasutusse andmine, teenustasude alam- ja ülempiirid teenuste liikide kaupa.

AvTS § 58² lõigete 2 ja 3 nõuded pole kitsendatud andmekogudes sisalduvatele andmetele, millele on kehtestatud juurdepääsupiirangud. Seega peaksid need nõuded laienema ka näiteks rahvastikuregistrile, millest andmete väljastamine on praegu osaliselt tasuline³³⁰.

Äriregistri, mittetulundusühingute ja sihtasutuste registri ning kommertsandiregistri elektrooniliste andmete väljastamise eest võetava tasu suurus on kinnitatud justiitsministri määrusega³³¹.

Analoogilisi näiteid võib tuua muudest valdkondadest, näiteks on elektroonilise kinnistusraamatu andmete väljastamine tasuline³³², samuti on tasuline katastriandmetega tutvumine³³³ jne.

Direktiivides 2003/98/EL ja 2013/37/EL on sätestatud põhimõte, mille kohaselt tasu võtmine andmete avaldamise eest ei tohiks üldjuhul ületada andmete avaldamiseks tehtavaid otseseid kulutusi. Teatud juhtudel, näiteks avaliku sektori asutuste puhul, kes peavad teenima tulu, et katta oluline osa oma avalik-õiguslike ülesannete täitmisest tulenevatest kuludest, lisandub mainitud summadele mõistlik investeeringutulu (direktiivi 2013/37/EL artikkel 6).

Viimase direktiivi artikli 7 lg 1 sätestab, et avaliku sektori asutuse valduses olevate dokumentide taaskasutamise standardtasude puhul määratakse eelnevalt kindlaks ja avaldatakse, kui see on võimalik ja asjakohane, siis elektrooniliselt, tasu kohaldamise tingimused ja tasu tegelik summa, sealhulgas tasu arvutamise alus.

Direktiivi 2013/37/EL artikli 13 lg 2 ütleb: „Liikmesriigid esitavad iga kolme aasta järel komisjonile aruande avaliku sektori valduses oleva taaskasutatava teabe kättesaadavuse, kõnealuse teabe kättesaadavaks tegemise tingimuste ja õiguskaitsetavade kohta. Aruanne avalikustatakse ja liikmesriigid viivad selle põhjal läbi artikli 6 rakendamise läbivaatamise, eriti piirkulusid ületava tasu võtmise osas“.

³²⁹ Ruumiandmete seadus, <https://www.riigiteataja.ee/akt/128022011002>

³³⁰ Rahvastikuregistrist andmete väljastamise hindade kinnitamine. Regionaalministri määrus, RT I, 10.11.2010, 15, <https://www.riigiteataja.ee/akt/110112010015>

³³¹ Äriregistri, mittetulundusühingute ja sihtasutuste registri ning kommertsandiregistri elektrooniliste andmete väljastamise eest võetava tasu suuruse ja nende isikute loetelu, kellele väljastatakse äriregistri, mittetulundusühingute ja sihtasutuste registri ning kommertsandiregistri elektroonilisi andmeid tasuta, kinnitamine. Justiitsministri määrus, RT I, 08.11.2013, 12, <https://www.riigiteataja.ee/akt/108112013012>

³³² Kinnistusraamatuseadus, <https://www.riigiteataja.ee/akt/102052013002>

³³³ Riigi maakatastri pidamise korra kinnitamine, <https://www.riigiteataja.ee/akt/24854>

Kuna ingliskeelne sõna „free“ on mitmetähenduslik, siis avaandmete mõiste („...*anyone is free to use, reuse, and redistribute it...*“) tingimata tasuta kasutamist ei eelda, kuigi seda võiks soovi korral sealt välja lugeda. Küll aga on avaliku sektori teabe tasuta kasutamise põhimõte selgelt sõnastatud dokumendis „G8 Open Data Charter and Technical Annex“³³⁴: „20) We recognise that open data should be available free of charge in order to encourage their most widespread use“.

Avaliku sektori avaandmete eest tasu küsimise kui Eestis linkandmete loomist ja kasutamist pärssiva teguri mõju hindab keskmiseks, suureks või väga suureks 50% avaliku sektori ja 90% erasektori vastajatest. Ankeedivastustes on mainitud ka paljude eri andmetike linkimiseks hädavajalike alusandmete suhteliselt suurt maksumust ruumiandmete puhul. See võib muuta need alusandmed erasektorile praktiliselt kättesaadamatuks.

Otstarbekas on viia Eesti õigusaktid võimalust mööda kooskõlla rahvusvaheliste arengutega, rakendades järjekindlamalt avaliku sektori teabe tasuta kasutamise, tasu suuruse, tasustamise aluste avalikustamise ning nende aluste perioodilise läbivaatamise põhimõtteid.

4.3.1 Ettepanekud

1. Vaadata läbi ja vajadusel uuendada riigi andmekogude kohta käivad õigusaktid, kus sätestatakse andmete väljastamise tasumäärad, eriti piirkulusid ületava tasu võtmise osas.
2. Sätestada avaliku teabe seaduses direktiivi 2013/37/EL artikkel 7 lg 1 nõue, et kui rakendatakse avaliku sektori asutuse valduses olevate andmekogude andmete taaskasutamise tasustamist, siis määratakse standardtasude puhul eelnevalt kindlaks ja avaldatakse tasu kohaldamise tingimused ja tasu tegelik summa, sealhulgas tasu arvutamise alus.
3. Kuna tasustamise teema on andmete linkimise seisukohast oluline, võiks sõnastada selgemalt AvTS § 4 lg 4 ning § 58² lg 2 ja 3, et vältida nendevahelist näilist vastuolu tasude osas mittespetsialistide jaoks (näiteks täpsustades, et üks räägib üksikpäringutest, teine andmete allalaadimisest).

4.4 Isikuandmete kaitse ja privaatsuse tagamine linkandmete kasutuselevõtu raames

Isikuandmete kaitse ja privaatsuse tagamine on olulisemaid küsimusi ava- ja linkandmete kasutamisel. Inimesed on mures oma andmete turvalisuse pärast, eriti arvestades luureskandaale, massilisi isikuandmete avalikustamisi seoses küberrünnetega ning sotsiaalmeedia levikut. Ava- ja linkandmete levik võib neid probleeme võimendada.

Analüüsime allpool, mida oleks vaja teha Eesti õigusruumis seoses isikuandmete kaitse ja privaatsuse tagamisega.

4.4.1 Isikuandmete esitamine avaandmetena

Kas isikuandmeid võiks esitada avaandmetena? Analüüsime seda küsimust kahes osas:

1. kas isikuandmeid võib avalikustada ja
2. kas isikuandmeid võiks esitada avaandmetena.

Märgime kõigepealt, et isikuandmed võivad olla avalikustatud erinevatel, muuhulgas järgmistel alustel:

³³⁴ G8. Policy paper. G8 Open Data Charter and Technical Annex. Published 18 June 2013.
<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

- IKS § 11 ja 12 sätestavad, et andmesubjekt võib isikuandmed ise avalikustada või anda nõusoleku andmete töötlemiseks. Ta võib sellise siiski igal ajal tagasi võtta, see nõusolek võib olla osaline ning vaidluse korral eeldatakse, et andmesubjekt ei ole oma isikuandmete töötlemiseks nõusolekut andnud. Tekib ka põhimõtteline küsimus, kas andmesubjekt saab anda informeeritud nõusoleku kõigiks võimalikeks taaskasutusteks (näiteks, kas ta saab aru sellise nõusoleku kõigist tagajärgedest) ja kas sellist võimalust tuleks inimestele üldse anda. Andmesubjekti nõusoleku tõendamise kohustus on isikuandmete töötlejal.
- Andmekogudes salvestatavad isikuandmed võivad olla avalikustatud ka seaduse alusel. Nii hoitakse äriregistris mitmesuguseid isikuandmeid, sealhulgas Eesti isikukoodi ja aadressi (ärieadustik³³⁵, § 62). Vastavalt ärieadustiku § 28 lõikele 1 on äriregistri kanded avalikud ning igal inimesel on õigus tutvuda registrikaardi ja äritoimikuga ning saada registrikaardist ja äritoimikus olevast dokumendist ärakirju.
- Vastavalt isikuandmete kaitse seaduse § 11 lõikele 2 võib isikuandmeid ilma andmesubjekti nõusolekuta ajakirjanduslikul eesmärgil töödelda ja avalikustada meedias, kui selleks on ülekaalukas avalik huvi ning see on kooskõlas ajakirjanduseetika põhimõtetega. Andmete avalikustamine ei tohi ülemääraselt kahjustada andmesubjekti õigusi.

Enamikul juhtudel on andmekogudes sisalduvad isikuandmed siiski konfidentsiaalsed. Näiteks tervise infosüsteemis ei ole avalikustatavaid isikuandmeid³³⁶. Ka rahvastikuregistri andmed ei ole avalikustatud - juurdepääs neile on piiratud³³⁷. Riikliku statistika seadus³³⁸ sätestab otseselt või kaudselt tuvastatavate andmete konfidentsiaalsuse järgmiselt: „§ 34. Konfidentsiaalsed andmed

(1) Andmed, mis võimaldavad statistilise üksuse otsest või kaudset tuvastamist ja seeläbi üksikandmete avalikustamist, on konfidentsiaalsed andmed.

(2) Statistiline üksus käesoleva seaduse tähenduses on andmete alusel otseselt tuvastatav, kui andmed sisaldavad kas statistilise üksuse nime, aadressi või isiku- või registrikoodi.

(3) Statistiline üksus käesoleva seaduse tähenduses on andmete alusel kaudselt tuvastatav, kui otsest tuvastamist võimaldavate tunnuste puudumisel on võimalik statistilist üksust tuvastada muude andmete alusel. Et otsustada, kas statistiline üksus on tuvastatav, võetakse arvesse kõik võimalused, mida kolmas isik võib eeldatavasti kasutada nimetatud statistilise üksuse tuvastamiseks.“

Kas avalikustatavad isikuandmed võivad olla avaandmed? Andmed või sisu on avatud, kui see kõigile antud vabalt kasutada, taaskasutada ja jagada – kitsendatult maksimaalselt viitamise või samadel tingimustel taasjagamise nõudega. Vaatame erinevaid võimalusi andmete avalikustamiseks avaandmete seisukohast:

- Kui andmesubjekt on andnud ise nõusoleku andmete töötlemiseks, tuleb muuhulgas talle teatavaks teha isikuandmete töötleja või tema esindaja nime ning isikuandmete töötleja aadressi ja muud kontaktandmed (isikuandmete kaitse seaduse § 12 lg 3). Neid ja muid isikuandmete kaitse seaduse § 11 ja 12 sätteid arvestades oleks väga raske tõlgendada andmesubjekti luba andmete töötlemiseks kui tema luba andmeid vabalt kasutada, taaskasutada ja jagada.
- Kui isikuandmed andekogudes on avalikustatud seaduse alusel, siis võib nende kujutamine avaandmetena olla mitmeti tõlgendatav. Näiteks, äriregistris hoitavad isikuandmed, sealhulgas Eesti isikukood ja aadress on avalikud (Ärieadustik³³⁹, § 62 ja 28). Vastavalt isikuandmete kaitse seaduse § 11 lõikele 1, kui isikuandmed avalikustatakse seaduse alusel, siis ei kohaldata isikuandmete töötlemisele isikuandmete kaitse seaduse teisi paragrahve. Seega siit võiks teha järelduse, et andmeid võib avaandmetena kasutada. Samas on RIHA andmetel äriregistri konfidentsiaalsuse

³³⁵ Ärieadustik, <https://www.riigiteataja.ee/akt/102072013063>

³³⁶ Tervise infosüsteemi põhimäärus. Vabariigi Valitsuse määrus. Vastu võetud 14.08.2008 nr 131, <https://www.riigiteataja.ee/akt/13251011>

³³⁷ Rahvastikuregistri seadus, <https://www.riigiteataja.ee/akt/114032011005>

³³⁸ Riikliku statistika seadus, <https://www.riigiteataja.ee/akt/13338093>

³³⁹ Ärieadustik, <https://www.riigiteataja.ee/akt/102072013063#para70lg2>

turvaosaklass S1 (info asutusesiseseks kasutamiseks: juurdepääs teabele on lubatav juurdepääsu taotleva isiku õigustatud huvi korral), mis näib olevat vastuolus äriseadustiku paragrahviga 28. Samuti on nende andmete omandiõigused ebaselged.

- Kui isikuandmeid on avalikustatud meedias, siis on väljaannetel tavaliselt omad reeglid nende andmete kasutamise lubamiseks ning neid ei saa üldjuhul avaandmetena käsitleda.

Ülaltoodust nähtub, et isikuandmed võiksid olla avaandmed juhul, kui nad on avalikustatud seaduse alusel. Siiski tuleks sellised olukorrad õigusaktides täpsemalt sätestada – näiteks, kas ja millistel tingimustel neid andmeid võib vabalt kasutada, taaskasutada ja jagada. Sellisel täpsustamisel võib juhtuda, et käsitletavaid isikuandmeid siiski linkandmetena kasutada ei saa, aga olukord ja piirangud on vähemalt kindlaks tehtud.

IKS § 11 ja 12 sätestavad, et andmesubjekt võib isikuandmed ise avalikustada või anda nõusoleku andmete töötlemiseks. Tihti on selline võimalus siiski üsna piiratud. Näiteks ei ole ei rahvastikuregistri seaduses ega ka riigiportaalis eesti.ee inimesel võimalust muuta oma isikuandmed rahvastikuregistris avalikuks, sealhulgas täpsustades sellise nõusoleku tingimusi vastavalt isikuandmete kaitse seaduse §-le 12.

Sellised võimalused toovad lisaks andmete efektiivsemale kasutamisele kaasa ka mitmeid riske. Eelkõige on need seotud ebapiisavalt läbimõeldud isikuandmete avalikustamise ohtudega nii inimesele endale või tema lähikondlastele, suuremahulise avalikustamise või turvaprobleemide korral ka riigi toimimisele. Tuleks arvestada ka seda, et isikuandmete osalise avalikustamise võimalus, sealhulgas isiku võimalus seda määratleda, võib kaasa tuua potentsiaalselt suured andmemahud, süsteemi keerukuse ja lisanduvad turvaohud.

Seega on otstarbekas andmekogusid sätestatavates õigusaktides täielikumalt realiseerida IKS § 11 ja 12 pakutavad andmesubjekti poolse isikuandmete avalikustamise võimalused, arvestades samas sellega kaasnevaid isikuga seotud ning tehnoloogilisi riske.

4.4.2 Isikuandmete töötlemine linkandmetena

Linkandmete kasutamine annab tänu erinevast allikast saadud andmete seostamisele võimaluse teha üha laiemalt järeldusi ja otsuseid ilma andmesubjekti osaluseta. Sellistest otsuste kohta sätestab isikuandmete kaitse seadus: „§ 17. Automaatsed otsused

(1) Andmetöötlussüsteemi poolt ilma andmesubjekti osaluseta otsuse (edaspidi *automaatne otsus*) tegemine, millega hinnatakse tema iseloomuomadusi, võimeid või muid isiksuseomadusi, kui see toob andmesubjektile kaasa õiguslikke tagajärgi või mõjutab teda muul viisil märkimisväärselt, on keelatud, välja arvatud järgmistel juhtudel:

- 1) automaatne otsus andmesubjekti suhtes tehakse lepingu sõlmimise või täitmise käigus tingimusel, et andmesubjekti taotlus lepingu sõlmimiseks või täitmiseks rahuldatakse või talle on antud võimalus esitada tehtava otsuse suhtes vastuväide oma õigustatud huvi kaitseks;
- 2) automaatse otsuse tegemine on ette nähtud seadusega, kui seaduses on sätestatud meetmed andmesubjekti õigustatud huvide kaitseks.

(2) Andmesubjekti peab enne automaatse otsuse vastuvõtmist teavitama mõistataval viisil automaatse otsuse vastuvõtmise aluseks olevast andmete töötlemise protsessist ja tingimustest.“

Kuna seadus ei sätesta, mida tähendab märkimisväärne mõjutamine, on siin väga lai ruum tõlgendamiseks ning linkandmete erinevate kasutamiskiiside vaidlustamiseks. Vastavalt isikuandmete kaitse seaduse §-le 17 isikuandmeid üldjuhul linkandmetena kasutada ei saa, sest on väga raske näiteks ette näha, kas mingi konkreetne linkandmete kasutusviis võib tulevikus andmesubjekti märkimisväärselt mõjutada. Soovitav oleks õigusakte selles osas täpsustada. Üks võimalus selleks on tuua analoogiliselt IKS § 11 lõikele 6 välja valdkonnad, millede puhul automaatsed otsused on lubatud.

Toodud järelendus ei sõltu üldjuhul sellest, kas isikuandmed on avalikustatud või mitte.

Erandiks on olukord, kus isikuandmed on seaduse alusel avalikustatud – sellisel juhul neile isikuandmete kaitse seaduse § 17 ei rakendu. Seega võib sellisel juhul isikuandmeid töödelda kui linkandmeid.

4.4.3 Isikuandmete kaitse ja privaatsuse tagamine seoses linkandmetega

Mõistliku tasakaalu leidmine isikuandmete kaitse ja andmete kasutamise efektiivsuse vahel on olnud Eestis ja Euroopa Liidus pikka aega aktuaalne teema, sealhulgas näiteks seoses teadusuuringute ja andmekaitse vahekorraga³⁴⁰. Lisaks efektiivsuse nõuetele peab olema tagatud inimeste põhivabaduste ja õiguste kaitse, isikuandmete ja identiteedi kaitse. Inimene on oma andmete omanik ning tal peab olema võimalus jälgida oma isikuandmete kasutamist³⁴¹.

Järgnevas vaatame ainult linkandmetega seoses tekkivaid lisaküsimusi, püüdmata analüüsida seniseid olemasolevaid probleeme.

Linkandmete kasutamine süvendab polaarsusi, võimaldades efektiivsemat töötlust, samas aga tekitades uusi ohte privaatsusele. Uued ohud võivad võimenduda või avalduda mitmeti:

- Näiliselt ebaolulised isikulised detailid võivad kombineeritult anda isikust põhjaliku pildi.
- Andmete linkimine võib laiendada võimalusi statistilise üksuse kaudseks tuvastamiseks.
- Isikustatud andmed kombineeritult isikustamata andmetega võivad kaasa tuua isikuandmete kaitse nõuete riive.
- Kõik ohud süvenevad seetõttu, et linkandmed võimaldavad seostada isikuandmeid avaliku sektori andmete ja muude allikate (meedia, sotsiaalmeedia jm) kombineerimisel.
- Kuna kombineeritakse andmeid kontrollitud ja kontrollimata allikatest, on suurem oht väärarusaamade, põhjendamatu hinnangute, inimesele suunatud vaenulike kampaaniate jms tekkeks.

Seetõttu tuleb isikuandmete kaitseks tagada, et ükski avalik lingitud andmete komplekt ei sisaldaks isikuandmeid. Eelmistes jaotistes toodud analüüs näitab siiski, et olemasolevas õigusruumis on isikuandmete kaitse ja privaatsus tagatud ka seoses linkandmete seadusliku kasutuselevõtuga, eelkõige seepärast, et isikuandmete kasutamine linkandmetena on äärmiselt raskendatud.

Erandjuhtudel, kui see on võimalik (nt äriregister), tuleks riskid hoolikalt läbi analüüsida. Näiteks, kui vaadata koos äriseadustiku § 62 ja § 28 lg 1 (äriregistris hoitakse mitmesuguseid isikuandmeid, sealhulgas Eesti isikukoodi ja aadressi ning äriregistri kanded on avalikud), AvTS § 8 lõiget 3 (teabele juurdepääs hõlmab ka õigust seda teavet taaskasutada) ning isikuandmete kaitse seaduse § 11 lõiget 1 (kui isikuandmed avalikustatakse seaduse alusel, siis ei kohaldata isikuandmete töötlemisele isikuandmete kaitse seaduse teisi paragrahve) tekib küsitav olukord, kus äriseadustiku alusel saab isikukoodi avalikustada, AvTS alusel saab seda teavet taaskasutada ning tulemusele ei kohaldata isikuandmete seadusega võimaldatavat kaitset.

IKS § 16 lg 1 sätestab: „Andmesubjekti nõusolekuta võib teadusuuringu või riikliku statistika vajadusteks töödelda andmesubjekti kohta käivaid andmeid üksnes kodeeritud kujul“. Sama paragrahvi lg 2 näib rääkivat

³⁴⁰ Andmekaitse ja teadustöö. Sirp, 25.01.2013, http://www.sirp.ee/index.php?option=com_content&view=article&id=17020:andmekaitse-ja-teadustoe&catid=9:sotsiaalia&Itemid=13&issue=3426

³⁴¹ Eesti infoühiskonna arengukava 2020 (heaks kiidetud Vabariigi Valitsuse poolt 14.11.2013). Majandus- ja Kommunikatsiooniministeerium, <http://263654.edicypages.com/eesti-infohiskonna-arengukava-2020/infohiskonna-arengukava-2020-loppversioon>

esimesele vastu: „Andmesubjekti tuvastamist võimaldaval kujul on teadusuuringu või riikliku statistika vajadusteks andmesubjekti nõusolekuta tema kohta käivate andmete töötlemine lubatud üksnes juhul, kui pärast tuvastamist võimaldavate andmete eemaldamist ei oleks andmetöötlemise eesmärgid enam saavutatavad või oleks nende saavutamine ebamõistlikult raske“. Esimesele lõikele vasturääkivana näib ka lg 4, lubades piiranguteta: „Kogutud isikuandmeid on lubatud töödelda teadusuuringu või riikliku statistika vajadusteks, olenemata sellest, millisel eesmärgil neid isikuandmeid algselt koguti“. Võimalusel on soovitatav esitada IKS § 16 lõiked selliselt, et nende omavaheline kooskõla oleks paremini jälgitav.

4.4.4 Ettepanekud

1. Reguleerida või täpsustada juhendmaterjalides, rakendusaktides või õigusaktides selliseid isikuandmete kaitse aspektist küsitavaid olukordi, kus ühe õigusakti alusel saab isikukode või muid isikuandmeid avalikustada, AvTS § 8 lg 3 alusel saab seda teavet taaskasutada ning isikuandmete kaitse seaduse § 11 lg 1 alusel ei kohaldata tulemusele isikuandmete seadusega võimaldatavat kaitset. Üks võimalus selleks oleks kitsendada AvTS § 8 lg 3, kitsendades selles isikuandmete taaskasutamist. Teine võimalus oleks täpsustada õigusaktides, mis lubavad isikuandmete avalikustamist, kas ja millistel tingimustel neid andmeid võib kasutada avaandmetena (sh vabalt kasutada, taaskasutada ja jagada). Samuti võib täpsustada isikuandmete kaitse seaduse § 11 lõiget 1 (kui isikuandmed avalikustatakse seaduse alusel, siis ei kohaldata isikuandmete töötlemisele isikuandmete kaitse seaduse teisi paragrahve). Võimalik on kasutada ka juhendmaterjale ja soovitusi.
2. Realiseerida täielikumalt andmekogusid sätestatavates õigusaktides IKS § 11 ja 12 pakutavad andmesubjekti poolse isikuandmete avalikustamise võimalused. Näitena võib tuua andmesubjekti võimaluse muuta oma isikuandmed rahvastikuregistris avalikuks (niihästi õigusaktide kui ka tehnilise teostuse tasemel), sh täpsustades sellise nõusoleku tingimusi vastavalt isikuandmete kaitse seaduse §-le 12. Seejuures tuleks püüda maandada sellega kaasnevaid isikuga seotud ning tehnoloogilisi riske, muuhulgas jälgides IKS § 12 toodud kitsendusi isikuandmete töötlemise nõusoleku andmiseks. Märkus: võimalik siiski, et kodanikku peab kaitsma ka iseenda eest; näitena võib tuua kiirraenude reguleerimise teemalised vaidlused.
3. Täpsustada juhendmaterjalides või rakendusaktides, millal rakendub isikuandmete kaitse seaduse § 17 (automaatsed otsused). Sealhulgas, kuna andmete linkimine, mille võimalikud kasutusviisid on ette määramata ja mis seega võib märkimisväärselt mõjutada andmesubjekti, on lubatud vaid §-s 17 märgitud erijuhtudel, võiks kas täpsemalt välja tuua, mida võib lugeda märkimisväärses andmesubjekti mõjutamiseks, või märkida analoogiliselt IKS § 11 lõikele 6 valdkonnad, millede puhul automaatsed otsused on lubatud.
4. Nendel juhtudel, kui isikuandmete kasutamine linkandmetena on võimalik (nt äriregister), tuleks kaasnevad riskid hoolikalt läbi analüüsida.
5. Selgitada IKS § 16 lõikeid juhendmaterjalides, et nende omavaheline suhe oleks paremini arusaadav.

4.5 Õigused, vastutused ja kohustused

Linkandmete seisukohast on oluline, et vastavalt autoriõiguse seadusele on teosed, millele tekib autoriõigus, muuhulgas arvutiprogrammid ning teoste kogumikud ja informatsiooni kogumikud (sealhulgas andmebaasid). Autoriõiguse seadust ei kohaldata näiteks ideedele ja teooriatele, aga ka faktidele ja andmetele.

Võivad tekkida küsimused, kas autoriõiguse seaduse põhjal kaitstavas teoses sisalduvaid üksikuid fragmente võib vabalt linkida, kes vastutavad linkandmete uuendamise ja nende andmete alusel pakutavate teenuste kvaliteedi eest jne.

4.5.1 Intellektuaalomandi õiguste kaitse tagamine

Eestis rakendatav intellektuaalomandi õiguste kaitse on sätestatud mitmete õigusaktidega. Siia kuuluvad Eesti Vabariigi põhiseadus (§ 39), autoriõiguse seadus³⁴², Berni kirjandus- ja kunstiteoste kaitse konventsioon³⁴³, viimasel põhinev WIPO (Maailma Intellektuaalse Omandi Organisatsiooni) autoriõiguse leping³⁴⁴, intellektuaalomandi õiguste kaubandusaspektide leping³⁴⁵ (lisa protokollile Eesti ühinemise kohta Maailma Kaubandusorganisatsiooni Asutamislepinguga³⁴⁶), Euroopa Liidu vastavad direktiivid jne.

Intellektuaalomandi õiguste kaubandusaspektide leping määratleb muuhulgas II jao osades 1–7 intellektuaalomandi kategooriad: autoriõigus ja sellega kaasnevad õigused, kaubamärgid, geograafilised tähised, tööstusdisainilahendused, patendid, mikrolülituste topograafia ja avalikustamata teabe kaitse.

Tööstusomandi õiguskorralduse aluste seadus määratleb tööstusomandi esemeteks leiutised, mida registreeritakse patendiseaduse, kasuliku mudeli seaduse või Euroopa patentide väljaandmise konventsiooni kohaldamise seaduse alusel; mikrolülituse topoloogiad; kauba- ja teenindusmärgid; tööstusdisainilahendused³⁴⁷.

AvTS § 8 lg 3 sätestab, et teabele juurdepääs hõlmab ka õigust seda teavet taaskasutada, kusjuures taaskasutamise mõiste avab sama seaduse § 3¹. Viimane ei sätesta siiski kõiki avaandmete kasutamise erisusi, sealhulgas omandiõiguste temaatikat. Näiteks, kas isik võib kasutada avaliku sektori poolt avalikustatud andmeid, täiendada neid ja lugeda siis need andmed enda omandiks?

Ei avaliku teabe seadus ega ka jaotise alguses mainitud õigusaktid ja lepingud ei sätesta, millised intellektuaalomandi õigused seonduvad avaliku teabega. Vastavalt autoriõiguse seaduse §-le 6 tekib autoriõigus teosele teose loomisega ning autoriõiguse tekkimiseks ning teostamiseks ei nõuta teose registreerimist, deponeerimist või muude formaalsuste täitmist. Seega võib avaliku teabe seaduse kohaselt avaldatud teave olla kaitstud autoriõigustega. Vastavalt AvTS § 4 lõikele 3 peab teabele juurdepääsu võimaldamisel olema tagatud autoriõiguste kaitse.

Et sellega seonduvaid probleeme vältida, tuleks avaliku teabe seaduses sätestada avalikustatava teabe litsentseerimise kord (näiteks, kelle pädevuses on litsentse anda) ja litsentsitingimused või litsentsitüübid. Kasutada võiks seejuures sobivat litsentsi *creative commons* litsentsiperest³⁴⁸.

4.5.2 Vastutus linkandmete ja nende kaudu esitatavate teenuste kvaliteedi eest

Kui linkandmete põhjal tehakse igapäevaseid või riigi juhtimise otsuseid, on oluline nii toorandmete kui ka töödeldud andmete kvaliteet.

Avaliku teabe seaduse § 9 lg 8 sätestab üldise vastutuse publitseeritavate andmete kvaliteedi eest. Samas „absoluutset“ kvaliteeti ei ole, näiteks konkreetsete andmekogude puhul on avalikustatavad andmed tehnoloogilistel ja muudel põhjustel enamasti reaalsete andmete koopiad, mis muuhulgas võivad olla mitte täiesti kaasaegsed. Selliseid objektiivseid piiranguid tuleks arvestada, seejuures vältides olukorda, kus

³⁴² Autoriõiguse seadus, <https://www.riigiteataja.ee/akt/114062013005>

³⁴³ Berni kirjandus- ja kunstiteoste kaitse konventsiooniga ühinemise seadus, <https://www.riigiteataja.ee/akt/24662>. Berni kirjandus- ja kunstiteoste kaitse konventsioon, <https://www.riigiteataja.ee/akt/13101723>

³⁴⁴ Välisministeeriumi teadaanne, <https://www.riigiteataja.ee/akt/13267527>. WIPO (Maailma Intellektuaalse Omandi Organisatsiooni) autoriõiguse leping, [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=DD:11:33:22000A0411\(01\):ET:PDF](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=DD:11:33:22000A0411(01):ET:PDF)

³⁴⁵ Intellektuaalomandi õiguste kaubandusaspektide leping, <https://www.riigiteataja.ee/akt/79299>

³⁴⁶ Protokoll Eesti ühinemise kohta Maailma Kaubandusorganisatsiooni Asutamislepinguga (Marrakeši lepinguga), <https://www.riigiteataja.ee/akt/79272>

³⁴⁷ Tööstusomandi õiguskorralduse aluste seadus, <https://www.riigiteataja.ee/akt/128122011046>. Vt ka Tööstusomandi kaitse Pariisi konventsioon, <https://www.riigiteataja.ee/akt/13088459>.

³⁴⁸ <http://creativecommons.org>

andmed tehakse linkandmetena ühekordselt avalikuks ning edasi neid andmeid enam ei hallata. Seepärast tuleks nõuded avalikustatavate andmete kvaliteedile ilmutatult sätestada ning teha kättesaadavaks, soovitatavalt RIHA kaudu.

Linkandmete kasutamisel tuleks ka arvestada võimalust, et andmete algne avaldaja muutub avalikkuse silmis (põhjendatult või mitte) automaatselt vastutajaks vahendatud andmete kvaliteedi eest. Näiteks võib riigi- või kohaliku omavalitsuse asutus avaldada mingid andmed avaandmetena ning teine osapool luua rakenduse nende andmete põhjal teenuse osutamiseks, kusjuures avalikkus näeb andmete allikana esimesena mainitud asutust. Kui see rakendus muutub populaarseks, kuid teine osapool lakkab mingil põhjusel teenust osutamast või osutab ebakvaliteetset teenust, pöördub kasutajate pahameel asutuse vastu. Selliste olukordade leevendamiseks saab kasutada asjakohaseid litsentse, paremat teavitust, viitamist asjakohastele õigusaktidele (nt tarbijakaitseseadus) jne.

4.5.3 Ettepanekud

1. Sätestada avalikustatava teabe litsentseerimise kord ja litsentsitingimused või litsentsitüübid avaliku teabe seaduses, võimalusel soovitades seejuures kasutada sobivat litsentsi *creative commons* litsentsiperest.
2. Kehtestada andmekogude põhimäärustes ja juhendmaterjalides avalikustatavate andmete kvaliteedi nõuded ning publitseerida nende nõuete ja nende vastavuse taseme hinnangud vastavas andmekogus või teha nad kättesaadavaks RIHA kaudu.

4.6 Omandi, kriitilise taristu ja riigi kaitse

4.6.1 Omandi andmete kaitse

Praegune õigusruum käsitleb eelkõige isikuandmete kaitset, kuid tehnoloogia arenedes suureneb vajadus käsitleda ka omandi andmete kaitsmist, eriti kui see on seotud konkreetse isikuga. Eelkõige võib olla oluline hinnalisema vara kohta käiva info kaitsmine (sh kinnisvara, maaomand, autod), kuid ka arvutite, telefonide jm info võib olla kergesti kättesaadav ning nad võivad sisaldada olulist infot³⁴⁹.

Linkandmed võivad suurendada riske muuhulgas tänu sellele, et integreeritakse mitmest allikast pärit infot - näiteks äriregistrist (saades kätte nimele vastava isikukoodi), kinnistusregistrist (saades kätte kinnistute andmed), maakatastrist (saades kätte katastriüksuste omanike nimed) jne.

Lisanduv riskide allikas on see, et kui üksikute päringute info tehakse lingituna internetis kõigile kättesaadavaks, võib seda kasutada selliste päringute tegemiseks, mida seni polnud võimalik teha vastavate registrite seadustes määratletud kitsenduste tõttu. Teisisõnu, võivad tekkida vähem või rohkem täielikud koopiad piiratud ligipääsuga registritest ja andmebaasidest.

Näiteks on maakatastriseaduse³⁵⁰ §-s 6 sätestatud, et igaüks võib tutvuda katastriandmetega ja saada nendest väljavõtteid, kui seadusega ei ole sätestatud teisiti, kusjuures katastriandmetele juurdepääsu sooviv isik võib korraka tutvuda ühe katastriüksuse ja sellega piirnevate katastriüksuste andmetega. Kui linkandmete loomine ja publitseerimine on lihtne tegevus, võivad aja jooksul tehtud päringute tulemused olla linkandmetena säilitatud ning nendest võib moodustuda koopia märkimisväärselt maakatastri osast. Sellele osale ei pruugi enam rakenduda maakatastri päringute arvu piirang.

³⁴⁹ http://www.itworld.com/slideshow/129283/how-cops-and-robbers-are-using-google-earth-383397?source=ITWNLE_nlt_tonight_2013-11-19

³⁵⁰ Maakatastriseadus, <https://www.riigiteataja.ee/akt/13335388>

Selliste probleemide ennetamiseks tuleks sätestada avaliku teabe seaduses või konkreetsete andmekogude kohta käivates õigusaktides, millised on õigused päringute tulemusena saadavate andmete publitseerimiseks.

4.6.2 Kriitiline taristu ja riigi kaitse

Linkandmed võimaldavad rohkem infot saada ka mitte alati heasoovlikel osapooltel, näiteks välisluureteenistustel. Muuhulgas võib asjade Interneti tehnoloogia arenedes tekkida vajadus käsitleda ka seadmete andmete kaitsmist, eriti kui nad on seotud konkreetsete isikute, asukohtade, ajahetkede või riigi kriitilise taristuga.

Võiks arvata, et seadmete info internetis on kaitstud, kuid tegelikkuses on palju näiteid vastupidise kohta: ruuterites on üldsusele teadaolevaid tagauksi³⁵¹; internetis on nimekirju seadmetest, mille kontrolleri kasutajanimed ja paroolid on jäetud tehases seadistatuks^{352, 353}; jne. Kui kõik need andmed panna linkandmetena kokku eelmises jaotises mainitud isikute ja omandi andmetega, võib tekkida üsna detailne pilt riigi toimimisest, sealhulgas ka elutähtsast infrastruktuurist. Seda pilti saavad erinevatel eesmärkidel kasutada mitmesuguste riikide teenistused³⁵⁴.

Sellist pilti saab kasutada üksikute isiku- või varavastaste rünnete teostamiseks, pannes kokku mitmesugust tüüpi andmed. Kui otsest infot ei ole võimalik saada, on võimalik rakendada kaudseid meetodeid. Muuhulgas on vastavalt riikliku statistika seaduse³⁵⁵ § 34 lõikele 3 statistiline üksus andmete alusel kaudselt tuvastatav, kui otsest tuvastamist võimaldavate tunnuste puudumisel on võimalik statistilist üksust tuvastada muude andmete alusel. Et otsustada, kas statistiline üksus on tuvastatav, võetakse arvesse kõik võimalused, mida kolmas isik võib eeldatavasti kasutada nimetatud statistilise üksuse tuvastamiseks.

Linkandmed annavad uusi suuri võimalusi rämpskirjade individualiseerimiseks ning seoses sellega isiklikult kohandatud rünnete teostamiseks. Kasutades avalikke, sotsiaalmeedia ja muid (link)andmeid on võimalik koostada väga usutavaid kirju, lisada neile näiteks äriregistrist saadud isikukoodi kasutades saaja nimele krüpteeritud manuseid jne. Sellised kirjad soodustavad klikkamist ohtlikele linkidele, nakatatud manuste avamist jms.

Potentsiaalselt väga ohtlikud on linkandmete kasutamise võimalused suurte rahvusvaheliste konfliktide puhul. Rahu ajal kogutud ja mittekasutatavaid suuri linkandmete andmebaase võib kasutada elutähtsate seadmete üle kontrolli saamiseks, seadmete töö halvamiseks, paanika tekitamiseks, muude rünnete mõju tugevdamiseks, desinformeerimiseks jne.

Otstarbekas on arvestada võimalust, et inimesed ei suuda säilitada kontrolli omavahel seotud tehnilike süsteemide üle, eriti kui need valdavad ja suudavad üheaegselt töödelda informatsiooni, mis inimestele endile kunagi integreeritult hõlmata ei ole. Kui sellistes süsteemides hakkavad arenema sellised omadused nagu näiteks sõltumatute eesmärkide püstitamine, mis tehniliselt on realiseeritav ka praegu ning mida võivad rakendada nii entusiastlikud, rumalad kui ka pahatahtlikud osapooled, siis on tulemused prognoosimatud. Arvestades lisaks seda, et lingitakse seadmete infot ning võttes arvesse robotika (sealhulgas droonide jm sõjalise eesmärgiga autonoomsete väikeste ja suuremate seadmete) arengut, saab isikuandmete ning ressurse käsitlevate andmete kaitse hoopis uue tähenduse – linkandmed koos muu tehnoloogia arenguga võivad tõepoolest olla „killer application“. Areng ilmselt linkandmete suunas läheb ja tuleks mõelda, kas selliseid riske on võimalik kuidagi leevendada. Muuhulgas tuleks silmas pidada võimalust, et praegune andmete maksimaalse avalikustamise suund võib mõnekümne aasta pärast asenduda vastupidise tendentsiga, vähendamaks tehissüsteemide kasvava võimsusega seotud riske – tendents, mille esimesi ilminguid võib jälgida mõnede riikide küberkaitse programmides.

³⁵¹ <http://www.devttys0.com/2013/10/reverse-engineering-a-d-link-backdoor/>

³⁵² <http://www.technologyreview.com/news/514066/what-happened-when-one-man-pinged-the-whole-internet/>

³⁵³ <http://money.cnn.com/2013/04/08/technology/security/shodan/>

³⁵⁴ <https://www.cia.gov/news-information/speeches-testimony/2012-speeches-testimony/in-q-tel-summit-remarks.html>

³⁵⁵ Riikliku statistika seadus, <https://www.riigiteataja.ee/akt/13338093>

Mainitud probleemide mõju vähendamiseks tuleks andmete (linkandmetena) publitseerimisel arvesse võtta riske, mis tekivad nende andmete kombineerimisel juba olemasolevate publitseeritud andmetega. Võimalik andmete publitseerimise kitsendamine selliste riskide tõttu peaks siiski olema pigem osaline (enamikes andmekogudes on andmeid, mida saab suurema riskita avalikustada), põhjalikult läbi analüüsitud ning detailselt põhjendatud.

4.6.3 Ettepanekud

- Sätestada konkreetsete andmekogude kohta käivates õigusaktides, võimalusel ka avaliku teabe seaduses (nt AvTS § 3¹, § 4 või § 8), millised on õigused päringute tulemusena saadavate andmete publitseerimiseks.
- Andmete (linkandmetena) publitseerimisel tuleks arvesse võtta riske (sh isikutele, omandile, seadmetele, kriitilisele taristule ja riigi julgeolekule), mis tekivad nende andmete kombineerimisel juba olemasolevate publitseeritud andmetega. Võimalik andmete publitseerimise kitsendamine selliste riskide tõttu peaks siiski olema pigem osaline (enamikes andmekogudes on andmeid, mida saab suurema riskita avalikustada), põhjalikult läbi analüüsitud ning detailselt põhjendatud.

5 Kvalitatiivuuringu tulemused

Kvalitatiivuuringu tulemus on koond kaheksast individuaalintervjuust ning kahest fookusgruupiintervjuust valdkonna ekspertidega, kellel on teadmised ja praktilised kogemused mitmetes linkandmetega seonduvates tehnoloogiavaldkondades. Tulemused on esitatud vastavalt intervjuu kavale ning ei ole seatud prioriteetsuse järjekorda.

5.1 Linkandmete ja seonduvate tehnoloogiate kasutuselevõtu võimalused ja takistajad

Kas ja mil määral on nimetatud tehnoloogiad vajalikud Eesti riigi IS arenguks? Millised neist on Eestile esmaselt olulised linkandmete kasutuselevõtu plaanis?

Peamiselt peetakse riigi infosüsteemide seisukohalt oluliseks linkandmete ja semantilise veebi tehnoloogiaid. Neid kahte tehnoloogiat nimetatakse sageli paaris, kuna semantika loob linkandmete kasutuselevõtuks vajalikud eeldused. Linkandmete kasutamise eelisena tuuakse välja riigi paindlikumat reageerimist muutustele ning kvaliteetsemat poliitikakujundamist, mis ilma linkandmeteta ebaõnnestub. Informatsiooni kättesaadavuses ja kasutamises ei nähta konkurentsieelist mitte ainult erasektorile, vaid ka riigile.

Avaandmete osas arvamused lahknevad. Osalt peetakse olemasolevat X-tee platvormi piisavaks, et andmeid avada, teisalt aga ei usuta, et riigil oleks piisavalt motivatsiooni ja ressursse andmete avamisega tegeleda, sest peamine kasusaaja avatud andmetest ei ole riik, vaid erasektor.

Asjade interneti, suurandmete ja koosloome osas arvamused hajuvad, kuna kogemus ja teadmised nende tehnoloogiate kasutamisel ei ole ekspertidel sarnased.

Kõige tähtsamaks linkandmete kasutuselevõtus peetakse andmete avamist ja semantiliste tehnoloogiate kasutuselevõttu, milleta ei ole linkandmete kasutuselevõtt võimalik. Ülejäänud tehnoloogiate järjestuses tekivad eriarvamused ning ei joonistu välja selgeid prioriteete.

Milliseid võimalusi pakub avaandmete ja linkandmete kasutamine riigile, erasektorile, elanikkonnale tervikuna?

Nutikas ja läbipaistev riik. Ava- ja linkandmete osas nähakse avalikku sektorit kui võimaldajat, erasektorit ja kodanikke kui tarbijat ja teenuste loojat. Andmete avamise üheks olulise vajadusena nimetatakse riigi ja kodaniku vahelise suhtluse paranemist, tagasiside andmist kodanike poolt riigile, et riik saaks teha paremat poliitikat.

Avaandmed – kõigile vabalt ja avalikult kasutamiseks antud masinloetavas formaadis andmed, millel puuduvad kasutamist ning levitamist takistavad piirangud.

Linkandmed – hajusad, seotud ja koosvõimelisi andmehulki, mis on esitatud ühtses formaadis (nt RDF), on publitseeritud veebis ja omavad juurdepääsu läbi päringupunktide (nt SPARQL).

Semantilised tehnoloogiad võimaldavad lisada andmetele tähendust ning esitada andmeid koos nende tähendustega masinloetaval kujul.

Asjade internet – füüsiliste objektide (näiteks esemete, seadmete) varustamine sensoritega, mis on lingitud traadiga või traadita võrgu abil.

Suurandmed on mahukad, muutuvad, mitmekesised ja keerukad andmed.

Koosloome – praktika, kus teenuste, ideede ja andmete saamiseks kasutatakse suurt hulka tavaliselt online-režiimis ühendatud inimesi. Koosloome protsess eeldab, et teenused, andmed, seadmed jne. kasutaks ühist mõistete süsteemi, ontoloogiat, semantikat ja viitaks nendele.

Tulevikuinternet – tehnilisest aspektist lähtudes tähendab tulevikuinternet uute internetiarhitektuuride ja internetiprotokollide väljatöötlust ja kasutuselevõttu (näiteks IPv6).

LEO³⁵⁶ (Läbipaistev Eesti Omavalitsus) on hea näide riigi avamisest kodanikele.

Monitooringu võimaldamine. Elutähtsate teenuste osutajatel on võimalik saada oluliselt kiiremini infot, kui elutähtis teenus on häiritud. Avalike teenuste osutamise kvaliteedi tõstmine muutub võimalikuks ja vajadused arusaadavaks.

Kiirem reaktsiooniaeg sündmustele. Andmete kiirem analüüs võimaldab kiiremini ja selgemalt aru saada, milline on otsuste mõju. Valedale otsustele on võimalik kiiremini reageerida ning uusi otsuseid langetada.

Avaliku sektori kulude vähenemine. Kui avalik sektor kaotaks ära vajaduse vastata kodanike ja erasektori teabenõuetele, päringutele, ning avaks andmed, hoitaks selle arvelt kokku pikas perspektiivis hulk ressursi. Koosloome tehnoloogiad kasutades saaks osa avaliku sektori teenustest korraldada ka kodanike kaasamisega (nt tänavaaукude kaardistamine mobiilirakendusega). Lisaväärtus on andmete avamisest võimaldatud majanduse elavnemine.

Elanikkonna elukvaliteedi tõstmine ja tervishoiu valdkonna kulude vähendamine. Tervishoius annaks andmete pikaajaline kogumine ja analüüsimine võimaluse hoida tulevikus kokku ravikuludid. Lisaväärtus võiks tekkida nt geenivaramu andmete sidumisel e-Tervise andmetega.

Uued teenused erasektoris, avaliku sektori taandumine andmete tootjaks. Teenuste loomist nähakse eelkõige erasektoris ning sel juhul taandatakse avalik sektor teenuseosutaja rollist. Selline stsenaarium võimaldaks paremat konkurentsi teenuseosutajate vahel.

Nähtamatud teenused, vähem e-teenuseid. Avalike teenuste paljususe asemel nähakse andmete linkimises võimalust, et riik teeks osa teenuseid kodanikule „nähtamatult“ ära. Nähtamatute teenuste all mõeldakse eelkõige proaktiivseid teenuseid, millele inimesele on seadusandlik õigus ning mida ta peab saama sõltumatult sellest, kas ta tellib teenust või mitte. Sagedase näitena tuuakse siinkohal lapse riikliku sünnitoetuse teenust, mida ei peaks spetsiaalselt taotlema, kuna riigil on piisav hulk informatsiooni olemas, et teenus nähtamatult osutada. Kodanik saab teada teenuse saamisest läbi raha laekumise pangakontole. Teise sarnase näitena saab tuua riikliku vanaduspension, mille puhul inimene ei peaks pensioni taotlema. Tulemuseks on vähem e-teenuseid.

Millised on Eesti tänased võimekused nimetatud tehnoloogiate raames, millised on arenguvajadused?

Kuigi linkandmete tehnoloogia arendamist on toodud välja pea kõige olulisemana, ei nähta Eesti arenguvajadust siin kriitilisena, kuna ka maailmas on mahajäämus linkandmete kasutuselevõtus ühtlaselt suur. Seevastu avaandmete osas nähakse suurt mahajäämust maailma keskmisest. Üksikud initsiatiivid avaandmete osas on olnud, kuid riigi tasandil puudub koordineeritus. Andmete avamise kontekstis räägitakse võimekusest kahes erinevas võtmes – ühelt poolt tahtmises/tahtmatutes avaliku sektori poolt andmeid avada ning teisalt võimetusest andmeid avada, kuna ressursse selleks napib. Tahtmatuse põhjuseks on tõenäoliselt andmete ebahütlane kvaliteet ning väärsti mõistmise võimalus, mis võib tuua kaasa ettenägematud probleeme ja täiendavaid kulusid avaliku sektori asutustele.

Võimekust andmeid omavahel linkida on tajutud juba kümmekond aastat tagasi läbiviidud pilootprojektides, kuid saadud tulemusi ei ole kas osatud või on kardetud neid kasutada. Avaliku sektori teadmatus võimalike tagajärgede eest, mida andmete avamine ja linkimine endaga kaasa toob, on takistanud tehnoloogilist arengut.

Laialdase linkandmete kasutuselevõtu osas on arengut pidurdavaks teguriks ka standardite puudumine. Ühe olulise arenguvajadusena nimetataksegi mõistete, ontoloogiate ja standardite loomist, et saaksime mõistetest ühtviisi aru.

³⁵⁶ Läbipaistev Eesti Omavalitsus

<http://www.riigipilv.ee/QvAJAXZfc/pendoc.htm?document=LEO.qvw&host=local&anonymous=true>

Semantiliste tehnoloogiate kasutuselevõtt on valmidus tunnetatav, ontoloogiaid on välja töötatud X-tee raames. Samas ei ole semantiliste tehnoloogiate kasutuselevõttu saanud edu, kuna riigi poolt pole loodud selleks nõuet ning erasektori teadlikkus ja vajadus on vähene. Peamise arenguvajadusena nähakse siin just selle tehnoloogia olulisuse tajumist.

Asjade interneti kontekstis on nimetatud oluliseks puuduseks tehnoloogia kasutuselevõtul IPv6 internetiprotokolli³⁵⁷ puudumist, millest tingituna väheneb Eesti konkurentsivõime maailmas.

Suurandmete osas kaheldakse, kas suurandmeid Eestis üldse olemas on. On levinud aramus, et suurandmeid Eesti praktikas ei esine ning see on vaid teadlaste pärusmaa. Võimekust suurandmeid töödelda nähakse üksikutel organisatsioonidel, kes tegutsevad riikliku julgeoleku, keskkonnaseire vms valdkonnas. Vajadust suurandmete töötlemise järele nähakse linkandmetega seondult, kuna linkandmete laialdase kasutuselevõtu kasvavad hüppeliselt ka andmemahud.

Koosloome võimekust illustreerib kõige paremini eesti.ee portaali testimine, millesse on kaasatud sadu vabatahtlikke asjast huvitatud kodanikke. Samuti ei oleks ilma koosloometa võimalik olnud „Teeme ära“ kampaaniad. Seega võib väita, et silmapaistvad näited on olemas, kuid pigem ei ole koosloome laialt kasutatav ega teadlikkus koosloome võimalustest piisav.

Millistel koosvõime tasemetel on Eestis barjäärid kõige suuremad? Millised takistused pärsivad avatud ja linkandmete (efektiivset) loomist ja kasutamist Eestis?

Ekspertide arvamused barjääride osas lähevad lahku. Küll aga paigutub suurimaks õiguslik barjäär. Poliitiliste ja organisatoorsete barjääride osas jagunevad arvamused kahte äärmusesse ning tehnilisi takistusi peetakse kõige väiksemaks.

Andmekaitse. Juriidilisel tasemel seab andmekaitse tugevad piirangud andmete kasutamisele. Osalt on see mõistatav, kuna Eesti väiksuse tõttu on oht mosaiikandmete tekkeks, mille puhul mitmest andmekogust saadud isikustamata andmeid on võimalik taasisikustada. Isiku privaatsus on tugevalt kaitstud. Andmete kogumine on registripõhine ja see ei võimalda andmekomplektidega fantaseerimist, uute andmete avastamist.

Visiooni, strateegia ja tegevusplaani puudumine. Avaliku sektori üleselt puudub reeglistik infosüsteemide korraldamiseks. Iga ministeerium korraldab oma haldusalas infosüsteeme oma parema äranägemise järgi, kuid ministeeriumide vahel on ebakõlad.

Ei tunnetata vajadust. Ametnik osutab teenust, sest arvatakse, et kodanik niikuinii ise ei oska. Kardetakse andmete interpreteerimist.

Ei usalda innovatsiooni. Ei usuta, et uued tehnoloogiad toovad suurt edu. Innovatsiooni on tehtud, kuid see on enne pooleli jäetud, kui tulemus tekib. Vajab poliitilist survet, et tekitada innovatsiooni sundusaldamine.

Kehv andmekvaliteet. Kartus, et andmetes on vead. Ametniku ja poliitikute hirm, mis juhtub meedias, kui avastatakse hulgaliselt vigu andmetes. Ametnik tunnetab töömahu kasvu, et andmeid parandada.

Tarkade tellijate puudus. Tellija ei tea, mida ja milleks ta vajab. Kui riigi poolt ettekirjutust ei tehta, tellitakse odavaid ja halbu lahendusi. Oluline on, et riigihanked muutuksid – nõudmised arendajatele peavad suurenema, tellijale peab jääma võimalus läbirääkimisteks.

Andmed ei ole masinloetavad. Andmete loojad ei ole teadlikud standarditest, andmed ei ole loodud ja avalikustatud standardeid arvestades. Ei eristata teabe avalikustamist ja taaskasutamist. Ühtsed mõisted, ontoloogiad, standardiseeritud identifikaatorid puuduvad. Kompetentsid standardite kasutuselevõtuks puuduvad.

³⁵⁷ IPv6 (Internetiprotokolli versioon 6) ehk "uue põlvkonna" Internetiprotokoll (inglise keeles IPng, *Internet Protocol Next Generation*) on andmesideprotokoll, mis on loodud praegusel ajal üldkasutatava Internetiprotokolli IPv4 asendamiseks

Andmed ei ole sisuliselt lingitavad. Eesti ja rahvusvahelised klassifikaatorid ei ole alati ühtivad. Isegi, kui kasutame samu mõisteid, siis asutuste vahel andmeid linkides võib selguda, et mõistete taga olev sisu ei ühti täielikult. Samuti tekib probleem ajalooliste andmekogudega, mille puhul võib ühe mõiste tähendus olla aja jooksul moondunud, meetodika andmete kogumiseks muutunud.

Linkandmed on riigile kallid. Vajadus täiendava projekteerimise ja arendamise järele, mida tänases arenduses pole ette nähtud.

Inimesele kasulikke andmeid on vähe. Riik korjab andmeid, mida tal on lihtne korjata ning kodanikule on sellest tagasipeegelduvalt vähe kasu. Nt on toodud välja terviseinfo andmestik, millega kodanikul ei ole palju peale hakata. Küll aga on vaja infot meditsiinasutustel, et kodanikku paremini teenindada. Inimese kasu oleks näiteks sportimise andmete ja meditsiiniandmete omavahelises kombineerimises, kuid siis võib tekkida meditsiinisektoril oht kaotada kontroll inimese ravimise üle.

Millised ohud võivad tekkida seoses potentsiaalse linkandmete laialdase kasutuselevõetuga?

Privaatsuse riive oht. Masinad teavad inimese kohta liiga palju. Info on liiga kergelt kättesaadav. Meil puudub ülevaade, kuhu ja kelleni info ulatub. Nt ei pruugi ekstreemspordilasele sobida see, kui info tema harrastuste kohta lekib kindlustusseltsidele või Haigekassale, kuna otsused võivad muutuda seeläbi inimesele kahjulikumaks.

Manipulatsioonide oht. Meedial võib tekkida võimalus teha suunatud negatiivseid kampaaniaid kellegi või millegi vastu. Privaatsus ja ärisaladus on kaitsmata.

Oht riigi julgeolekule. Andmete avalikustamine ei tähenda vaid avalikustamist Eestis. Kui andmed on avalikud, on nad kättesaadavad ka piirideüleselt. Infot saavad enda jaoks ära kasutada ka need riigid, kes kujutavad endast ohtu Eesti riigi julgeolekule.

Ebapiisav analüüs andmete avalikustamisel. Oht, et andmed, mida poleks tohtinud mingite asjaolude kokkulangemisel avalikustada, saavad avalikuks. Juba avalikustatud andmeid ei ole võimalik isegi sunniviisiliselt tagasi võtta.

Mosaiikandmete tekke oht. Andmed, millelt on eemaldatud isikustamist võimaldavad tunnused, ent mitmest andmekogust pärinevate andmete linkimisel ja koosseisitamisel tekib võimalus andmeid taasisikustada.

Andmete kvaliteet ja usaldusväärsus on madalad. Avatud andmed väljuvad algallika kontrolli alt. Probleemid teenusega võivad olla teenuseosutaja, mitte aga andmete tootja probleemid, teenuse tarbija ei pruugi omada ülevaadet. Isegi kui usaldusväärsed andmed on olemas, ei ole teenusepakkujatel sundust, oskust või teadlikkust nende andmete kasutamiseks teenuse osutamisel.

Andmete omanik ei arvesta andmekasutajatega. Kuna andmete omanik ei tea, kes andmeid kasutavad, ei saa ta muudatuste elluviimisel arvestada andmekasutajate huvidega. Ei ole võimalik tuvastada, keda muutused mõjutavad.

Andmete väärlinkimine või osaline linkimine. Näiliselt sarnaste, ent sisu poolest erinevate andmete linkimine võib põhjustada probleeme ja segadust. Samuti võib olukorras, kus osa andmestikke on lingitud ja osa mitte, tekkida oht valede järelduste tegemiseks. Inimene ei pruugi vaevuda kontrollima, kas masin võttis kõik vajaliku arvesse või mitte.

Millistes valdkondades ava- ja linkandmete projektide käivitamisel võiks olla Eestile suurim mõju?

Alljärgnevalt on olulisuse järjekorras välja toodud valdkonnad, kus võiks ava- ja linkandmete initsiatiivi elluviimisel olla suurim mõju projektide käivitamisest:

- linkandmete turvalisus, privaatsus ja kvaliteet;

- metaandmete (sh ontoloogiad) loomine;
- linkandmete loomine/tootmine, haldus (koos metaandmetega/ontoloogiatega) ja publitseerimine;
- avaandmete loomine/tootmine, haldus ja publitseerimine;
- linkandmete loomise metoodika ja juhendite koostamine ning nende publitseerimine;
- koosloome kasutamine linkandmete mahu suurendamisel, nende semantilise tähenduse tõstmisel või kvaliteedi määramisel;
- linkandmete rakendused (sh valdkonnaspetsiifilised liidesed);
- avaandmete rakendused (näiteks teenustes);
- temaatilised rakendusuringud ja tarkvaraarendusprogrammid;
- linkandmete alase teadlikkuse suurendamine, sh kasutuselevõtu propageerimine;
- linkandmete ja teiste seonduvate tehnoloogiate alaste koolituste pakkumine/laiendamine.

Täiendavalt toodi välja eelnevalt nimetamata valdkondi, mille tähtsust hinnati väga oluliseks:

- terviklike andmekaitse ja infoturbe lahenduste väljatöötamine;
- osalemine avatud standardite väljatöötamisel riiklikul tasandil;
- õigusloome väljatöötamine.

Milliste valdkondade põhiandmehulkade avamine ja linkimine on kõige kõrgema kasulikkuse määraga Eestile?

Alljärgnevalt on olulisuse järjekorras välja toodud valdkonnad, kus võiks ava- ja linkandmete initsiatiivi elluviimisel saavutada suurimat kasu:

- tervishoid;
- georuumiline;
- haridus;
- transport ja infrastruktuur;
- teadus ja uurimistöö;
- äritegevus (nt äriregister);
- energeetika ja keskkond;
- maa seire (nt meteoroloogia, metsandus, kalandus, jahindus, põllumajandus);
- finants ja lepingud (nt eelarve ja kulutused, lepingud, pakkumused);
- valitsuse aruandlus ja demokraatia;
- statistika;

- sotsiaalne mobiilsus ja heaolu (nt tervisekindlustus, töötushüvitised, elamispind);
- kuritegevus ja õigus;
- globaalne areng (nt abi, toiduturvalisus).

Lisaks toodi välja Eesti mõistes oluline valdkond:

- kultuuripärand.

Mida oleks vaja teha ning kelle toetus on oluline ava- ja linkandmete initsiatiivi õnnestumiseks Eestis?

Ava- ja linkandmete initsiatiivi õnnestumiseks peetakse võrdselt oluliseks poliitilist toetust ning juhtfiguuride/eeskõnelejate olemasolu. Neile järgnevad huvigrupid ning oluliselt viimasele kohale paigutatakse kodanikud.

Kõige olulisemate tegevustena tuuakse välja riigi poolset eestvedamist ning vastutuse võtmist. Kui riigi poolt astutaks esimene samm, oleks teiste huvigruppide tegutsemine lihtsam. Samuti nimetatakse edulugude vajadust, mis võimaldaks näidata innovatiivset probleemilahendust ning mida Euroopasse skaleerida.

Oluline on riigi poolt vajalik õigusruumi korrastamine, eelkõige andmekaitsepoliitika väljatöötamine, standardite, kokkulepete ja koosloome algatamine et luua valdkonna arenguks vajalikud eeldused.

Riigi toetusmeetmed võiksid olla suunatud rakendusuuringute läbiviimiseks ning erasektoris ava- ja linkandmetega seotud projektide käivitamise finantseerimiseks.

5.2 Lingitud Eesti koosvõime arhitektuur

Milliseid loodava riikliku avaandmete varamu omadustest on olulised varamu kasutavuse seisukohast?

Alljärgnevalt on olulisuse järjekorras välja toodud omadused, mis on olulised avaandmete varamu kasutavuse seisukohast:

- kvaliteedi tagamise mehhanismid;
- andmehulkade (toorandmete) kättesaadavus masinloetavas formaadis;
- andmehulkade registreerimine (metaandmete esitamine);
- andmehulkade versioonide teave;
- analüüsi ja visualiseerimisvahendite olemasolu;
- andmehulkade (toorandmete) kättesaadavus;
- andmete teisendajad publitseerijatele (näiteks csv formaadist rdf formaati);
- andmehulkade otsimine, sirvimine, grupeerimine;
- sotsiaaltarkvara vahendid;
- koosloome võimalused (näiteks andmete parandamine jms);

- andmehulkade kommenteerimine, vigadest teatamine, märgendamine.

Samuti nimetati:

- andmete korje, hõive, integratsioon;
- metaandmete esitamine;
- töö kiirus;
- intuitiivne otsing ja kasutusmugavus;
- piisav jõudlus 24/7 suurte andmehulkadega toimetulemiseks;
- autentimise võimalused;
- andmete omanikele garantiid;
- kvaliteedi ja usaldusväärsuse hindamise meetrikad;
- ontoloogiliste teisenduste tööriistad.

Milleks võiks kasutada avaandmeid?

Alljärgnevalt on loetletud avaandmete kasutusviisid alustades olulisemaist:

- turunduseks või müümiseks;
- integreerimiseks teiste andmetega;
- uute andmehulkade avastamiseks;
- visualiseerimiseks;
- uuringuteks ja analüüsiks;
- uute andmete loomiseks;
- uute teenuste loomiseks;
- linkimiseks;
- andmete puhastamiseks;
- lihtsalt kasutaksin andmeid nagu nad on;
- publitseerimiseks.

Seejuures on nimetatud uue ärivõimalusena ka ava- ja linkandmetel põhineva faktidepõhise teadmuse tagasimüümist riigile.

Avaliku sektori vaatenurgast annavad ava- ja linkandmed paremad võimalused protsesside efektiivsemaks muutmiseks, kuna annavad võimaluse välja arvutada reaalseid tegevuskulusid.

Millistes formaatides oleks oluline avaandmeid publitseerida?

Enim sobivate formaatidena on välja toodud RDFa (RDF veebilehtedes), CSV, RDF (*Resource Description Framework*) ja JSON (JavaScript Object Notation). Pigem äärmuslikumaks peetakse XML ja PDF

formaatide kasutamist ning pigem taunitakse KML (Keyhole Markup Language (geoandmed)), Excel, HTML ja SQL Database Dump formaatides andmete esitamist.

Üldjoontes ei soovi ka eksperdid väga kindlat suunda võtta ühe formaadi poole, vaid lähtuvad masinloetavuse põhimõttest. Kui andmed on masinloetavad, on ka formaat sobiv.

5.3 Organisatsiooniline koosvõime

Millised võiksid olla avaliku sektori ülesanded infosüsteemide tellijana seoses ava- ja linkandmete tehnoloogiate rakendamisega?

Täiendavad reeglid infosüsteemide tellimisel. Linkandmete tehnoloogiate rakendamiseks tuleb riigihangetele lisada ava- ja linkandmete reeglid, mis on kohustuslikud iga IS tellimisel. Samuti tuleb teostada järelevalvet nõuete täitmise osas ning sanktsioneerida toetussummade väljamaksmisel nõuete mittetäitmise korral. Uute reeglite kehtestamiseks on vaja ka kompetentse, kes tehnoloogiaid valdavad.

Kuidas kaasata linkandmete tehnoloogia arendamisse/kasutamisse Eesti erasektor, vabasektor ja kodanikkond?

Innovatsioonitoetused. Toetused projektidele, mis on seotud ava- ja linkandmete kasutuselevõtu initsiatiividega. Finantseerimise abil saab riik survestada innovatsiooni.

Garage48-tüüpi ettevõtmised. Huvigruppide kaasamine rakenduste väljatöötamisel. Motivatsioon ja kasu tekib huvigruppidele läbi avaliku sektori säästetud raha, millest avalik sektor saaks selliseid projektettevõtmisi finantseerida.

Teadlikkuse tõstmine. Mida suuremaks muutub teadlikkus uutest tehnoloogiatest, seda paremini tajub kodanikkond oma võimalusi. Kodaniku huvi on eelkõige isiklik huvi.

6 Kvantitatiivuuringu tulemused

Kvantitatiivuuringu jaoks töötasime välja veebipõhise ankeedi. Ankeet on toodud lisas (vt 9.1 Kvantitatiivküsitluse ankeet).

Ankeedi levitamiseks koostasime juhtrühma kontaktibaasi abil asjakohase sihtrühmade nimekirja, mille koosseis on toodud allolevas tabelis. Lisaks levitasime ankeedi URL-i sotsiaalmeedias (Facebook, LinkedIn).

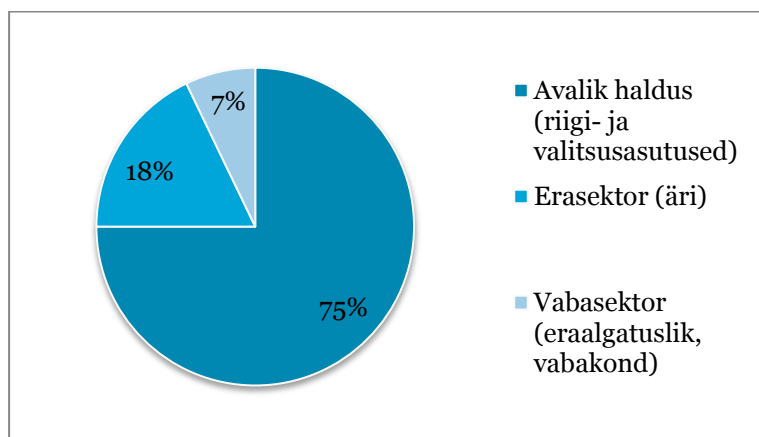
Tabel 7. Ankeetküsitluse sihtrühmade koosseis

Sihtrühm	Saadetud kutseid	Vigased kutsed
Vabakond	49	4
Erasektor	148	12
Avalik sektor	388	15
Kokku	585	31

Ankeetküsitlusele tulemusena laekus 77 täidetud ankeeti, millest 56 andmed olid uuringu kontekstis kasutatavad. Ülejäänud 21 vastanust ei omanud küsitud tehnoloogiate kohta teadmisi ja sisulisi vastuseid seetõttu ei andnud.

6.1 Vastanute profiil

Vastanute jaotust sihtrühmade lõikes iseloomustab alljärgnev joonis.



Joonis 12. Vastanute profiil: sihtrühmade jaotus

Vastanute teadmisi erinevatest tehnoloogiatest iseloomustab alljärgnev tabel (vt Tabel 8). Valdavalt on teadmised üldisel või keskmiselt tasemel. Põhjalike või ekspertteadmistega isikuid leidub vastanute seas üksikuid.

Tabel 8. Vastanute profiil: teadmised tehnoloogiatest

	Sektor	Kokku sektoris	Ei tea sellest tehnoloogiast midagi	Üldisel tasemel teadmised valdkonnast	Keskmiised teadmised valdkonnast	Põhjalikud teadmised valdkonnast	Olen valdkonna ekspert
Avaandmed	Avalik sektor	42	4	20	13	5	0
	Erasektor	10	1	3	3	3	0
	Vabasektor	4	0	1	2	0	1
Linkandmed	Avalik sektor	42	8	24	8	2	0
	Erasektor	10	0	4	2	4	0
	Vabasektor	4	1	2	1	0	0
Semantilised tehnoloogiad	Avalik sektor	42	8	24	9	1	0
	Erasektor	10	3	3	2	2	0
	Vabasektor	4	1	2	1	0	0
Asjade internet	Avalik sektor	42	5	24	10	3	0
	Erasektor	10	1	4	3	0	2
	Vabasektor	4	1	0	1	1	1
Suurandmete töötlemise ja hõive tehnoloogiad	Avalik sektor	42	11	15	14	2	0
	Erasektor	10	0	6	2	0	2
	Vabasektor	4	1	0	1	1	1
Koosloome tehnoloogiad	Avalik sektor	42	14	17	11	0	0
	Erasektor	10	0	6	3	1	0
	Vabasektor	4	0	1	1	0	2

Vastanute praktilisi kogemusi iseloomustab alljärgnev tabel:

Tabel 9. Vastanute profiil: praktilised kogemused

	Sektor	Kokku sektoris	Kogemused puuduvad	Kogemused 1-2 projektis	Kogemused 3 või enam projektis
Avaandmed	Avalik sektor	42	18	22	2
	Erasektor	10	4	4	2
	Vabasektor	4	1	1	2
Linkandmed	Avalik sektor	42	27	14	1
	Erasektor	10	3	4	3
	Vabasektor	4	2	1	1

	Sektor	Kokku sektoris	Kogemused puuduvad	Kogemused 1-2 projektis	Kogemused 3 või enamas projektis
Semantilised tehnoloogiad	Avalik sektor	42	28	13	1
	Erasektor	10	6	2	2
	Vabasektor	4	3	1	0
Asjade internet	Avalik sektor	42	30	11	1
	Erasektor	10	7	1	2
	Vabasektor	4	1	2	1
Suurandmete töötlemise ja hõive tehnoloogiad	Avalik sektor	42	27	14	1
	Erasektor	10	6	2	2
	Vabasektor	4	1	1	2
Koosloome tehnoloogiad	Avalik sektor	42	26	16	0
	Erasektor	10	6	4	0
	Vabasektor	4	1	0	3

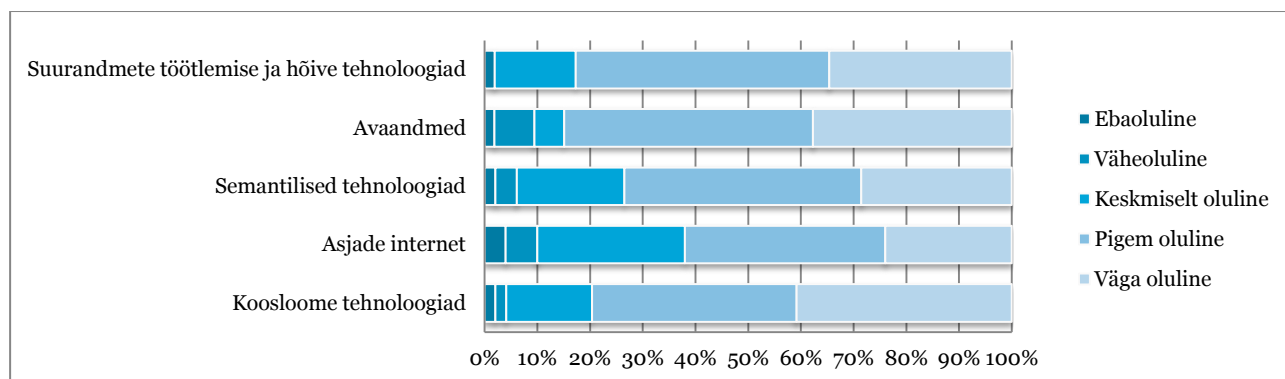
6.2 Vastused küsimustele

Kokkuvõttes saab kvantitatiivuuringu raames teha järgmised järeldused:

- Kõiki käsitletud tehnoloogiaid (suurandmed, avaandmed, semantilised tehnoloogiad, asjade internet, koosloome tehnoloogiad) peetakse linkandmete kontekstis suures ulatuses oluliseks (60-85%).
- Teadlikkus valitsuse plaanidest linkandmete vallas on valdavalt väike. Eristub erasektor, kus teadlikkus on suurem.
- Eesti võimekust ja kompetentsi eelnimetatud tehnoloogiate osas peetakse valdavalt halvaks (65-80%), eriti semantiliste tehnoloogiate osas (90%).
- Linkandmete pärssivate tegurite mõju on valdavalt suur. Väikseimaks peetakse õiguslikke takistusi (35%) ja rahastamise puudulikkust (36%). Suurim pärssiv tegur on vähene linkandmete alane teadlikkus (77%), järgnevad riikliku visiooni, strateegia ja tegevusplaani puudumine (64%), organisatoorsed takistused (62%) ja eestvedaja puudumine (62%).
- Peamiste ohuna linkandmete kasutuselevõtul nähakse andmete madalat kvaliteeti ja usaldusväarsust (62%). Teiste ohtude osatähtsus jääb alla 50%.
- Linkandmete potentsiaalset kasu nähakse valdavalt kõikides valdkondades, eriti tervishoius (94%), statistikas (80%) ja geourimises (78%).
- Valdavalt suurt kasulikku mõju linkandmete kasutuselevõtu arendamiseks nähakse kõikides väljapakutud tegevustes ja projektides (48-68%). Suurima kasuga tegevusteks hinnati praktilise väärtusega tegevusi nagu linkandmete ja avaandmete pilootrakenduste väljatöötamist. Olulisel kohal on linkandmete turvalisus, privaatsus ja kvaliteet ning linkandmete loomise meetoodika ja juhendite koostamine.
- Lingitud avaandmete kontekstis on kõige atraktiivsemad kasutusvõimalused uuringute ja analüüside tegemine (75%) ning integreerimisvõimalused teiste andmetega (78%). Väiksema atraktiivsusega on turundus ja müügitegevus (32%) ning andmete publitseerimine (32%).

Küsimuste detailsemas kokkuvõttes toome erisused sihtrühmade lõikes välja olukordades, kus arvamused on lahknevad. Kui arvamuste lahknevus on marginaalne, siis esitame sihtrühmade ülese kokkuvõtte.

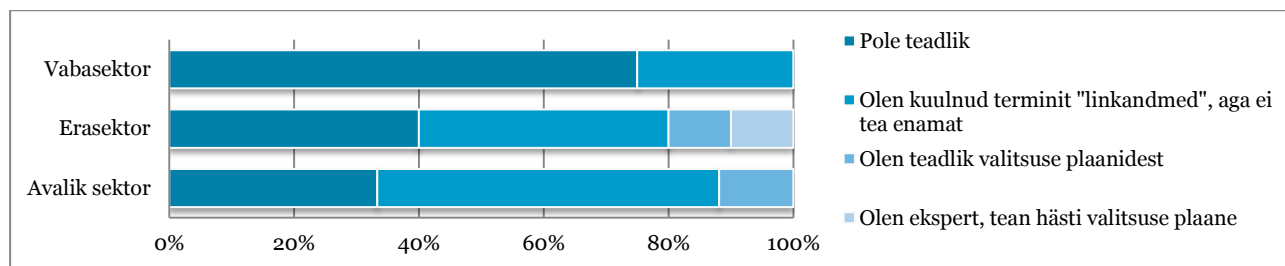
Palun hinnake tehnoloogiatega olulisust Eestile linkandmete kasutuselevõtu kontekstis.



Joonis 13. Tehnoloogiatega olulisus Eestile linkandmete kontekstis

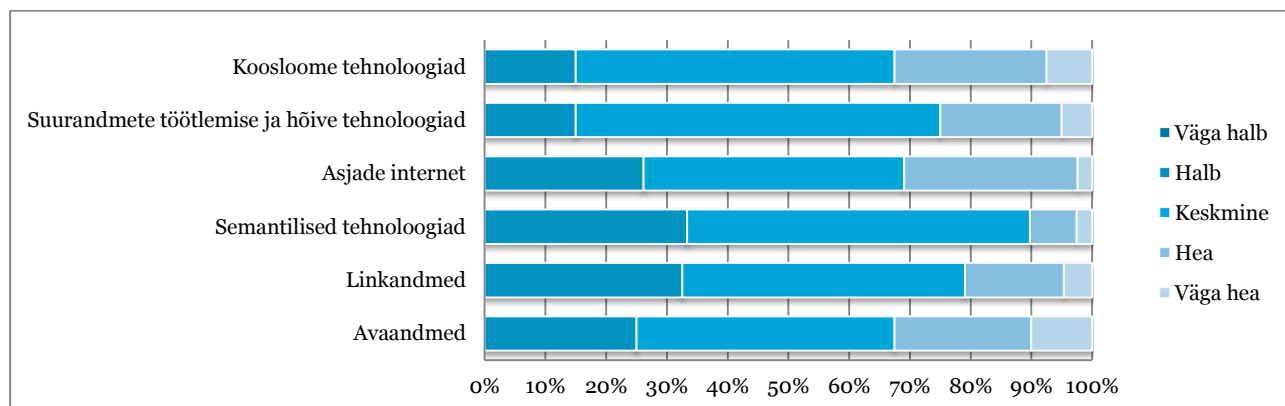
Kuivõrd olete teadlik valitsuse plaanidest linkandmete vallas?

Teadlikkusega valitsuse plaanidest linkandmete vallas eristub erasektor. Avalikus ja vabasektoris on teadlikkus väike.



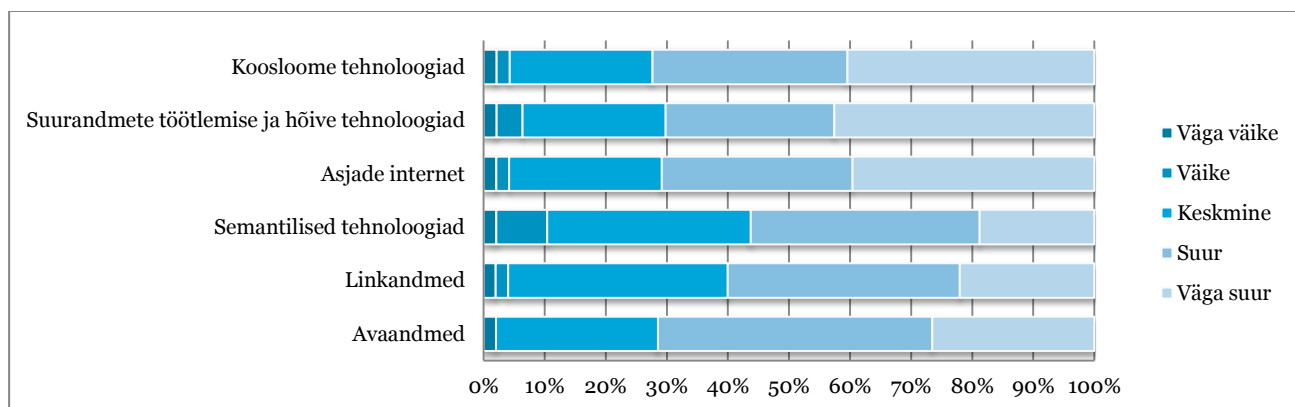
Joonis 14. Sihtrühmade teadlikkus valitsuse plaanidest linkandmete vallas.

Milline on Teie arvates Eesti tänane tehnoloogiline võimekus/kompetents nimetatud tehnoloogiarendide raames?



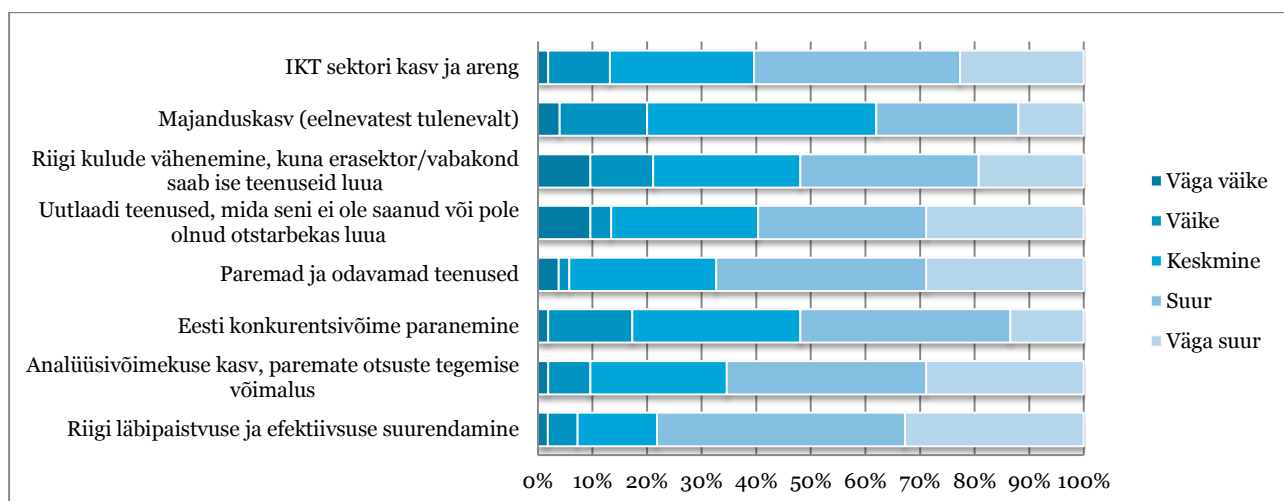
Joonis 15. Eesti võimekus tehnoloogiarendide raames

Milline on Teie arvates Eesti arenguvajadus järgmiste tehnoloogiarendide raames?



Joonis 16. Eesti arenguvajadus tehnoloogiarendide raames

Palun hinnake potentsiaalseid kasusid linkandmete kasutamisest riigile, erasektorile ja elanikkonnale tervikuna.



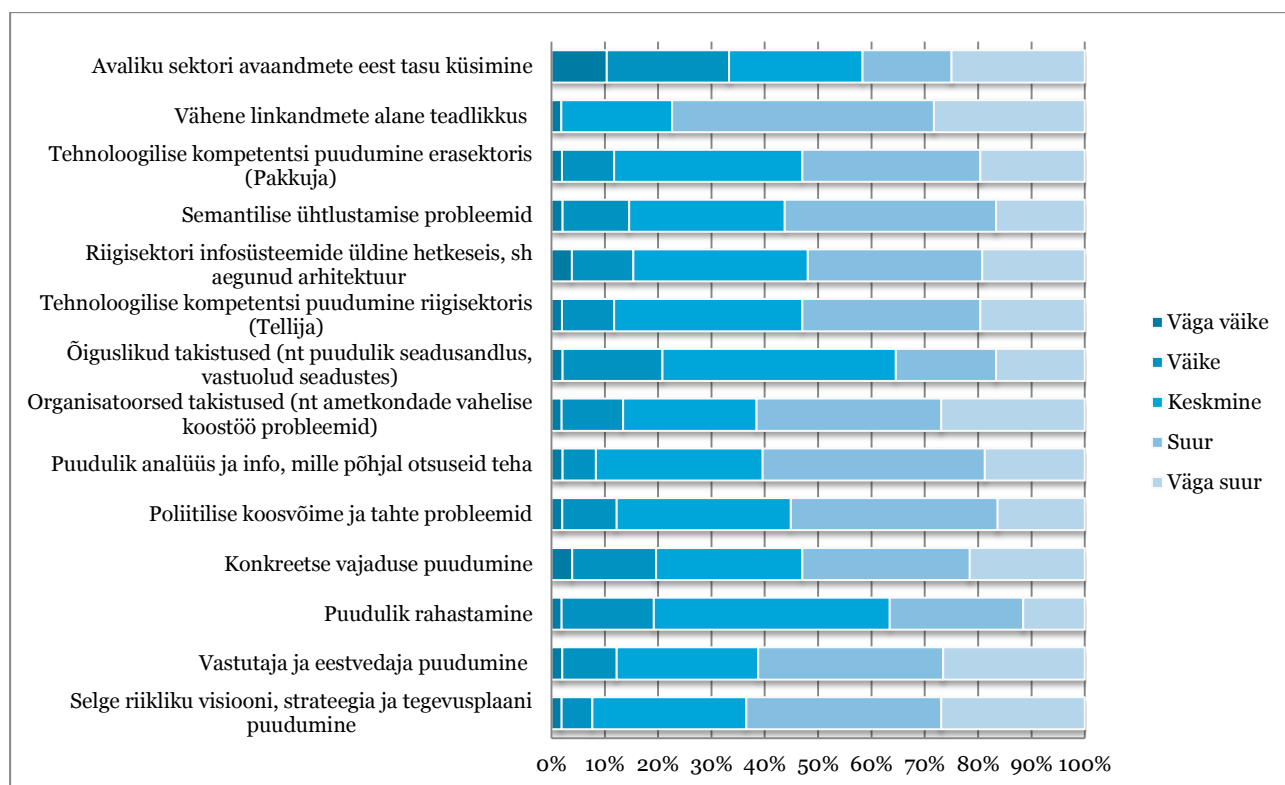
Joonis 17. Potentsiaalsed kasud linkandmete kasutamisest

Täiendusena pakuti kasuna välja veel ettevõtte ärieeelse saavutamist ning inimeste infootsingu oskuse paranemist.

Palun hinnake linkandmete loomist ja kasutamist pärssivaid tegureid Eestis.

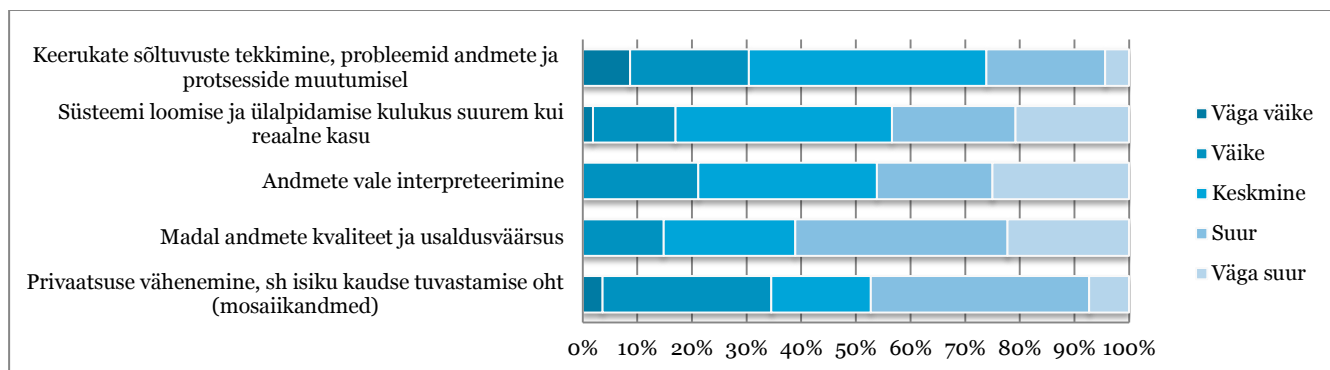
Pärssivate tegurite osas läheb era- ja avaliku sektori arvamus lahku avaandmete eest tasu küsimise osas. Erasektor hindab, et avaandmete tasu on pigem suur pärssiv tegur.

Ühe kommentaarina andmetasudele toodi konkreetne näide: suur probleem tuleneb olukorrast, kus riigisektor (konkreetselt Maa-amet) küsib paljude eri andmetike linkimiseks hädavajalike alusandmete nagu näiteks Eesti Topoloogiline Andmekogu (ETAK) või katastripiirid, eest raha. Ruumiandmete seaduses sätestatud hinnakiri paistab ühikuna väike, kuid andmekogus on miljoneid objekte, mis teeb summadeks kümneid tuhandeid eurosid. Mõnel andmekogul (kõrgusinfo, aeropildid) lähevad hinnad sadadesse tuhandetesse eurodesse andmetiku kohta. See muudab andmed praktiliselt kättesaadamatuks, v.a riigisektori enda sees, kes saab andmed kätte tasuta. On palju andmetikke, mis on lingitavad just nende ruumiomaduste tõttu. Paljudes teistes riikides (nt Soome, Taani) on samad ruumiandmed tehtud tasuta kättesaadavateks avaandmeteks.



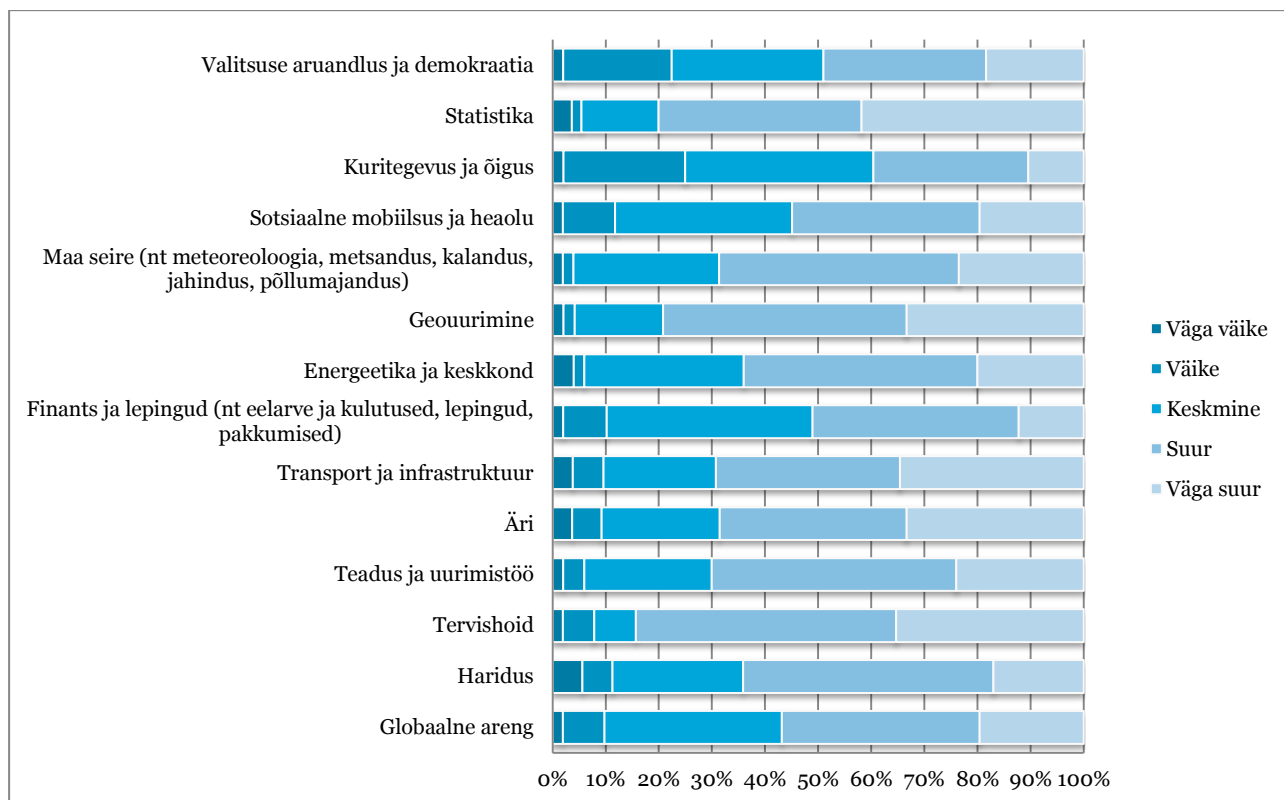
Joonis 18. Linkandmete loomist ja kasutamist pärssivad tegurid

Palun hinnake ohtude suurust seoses potentsiaalse linkandmete laialdase kasutuselevõttuga Eestis.



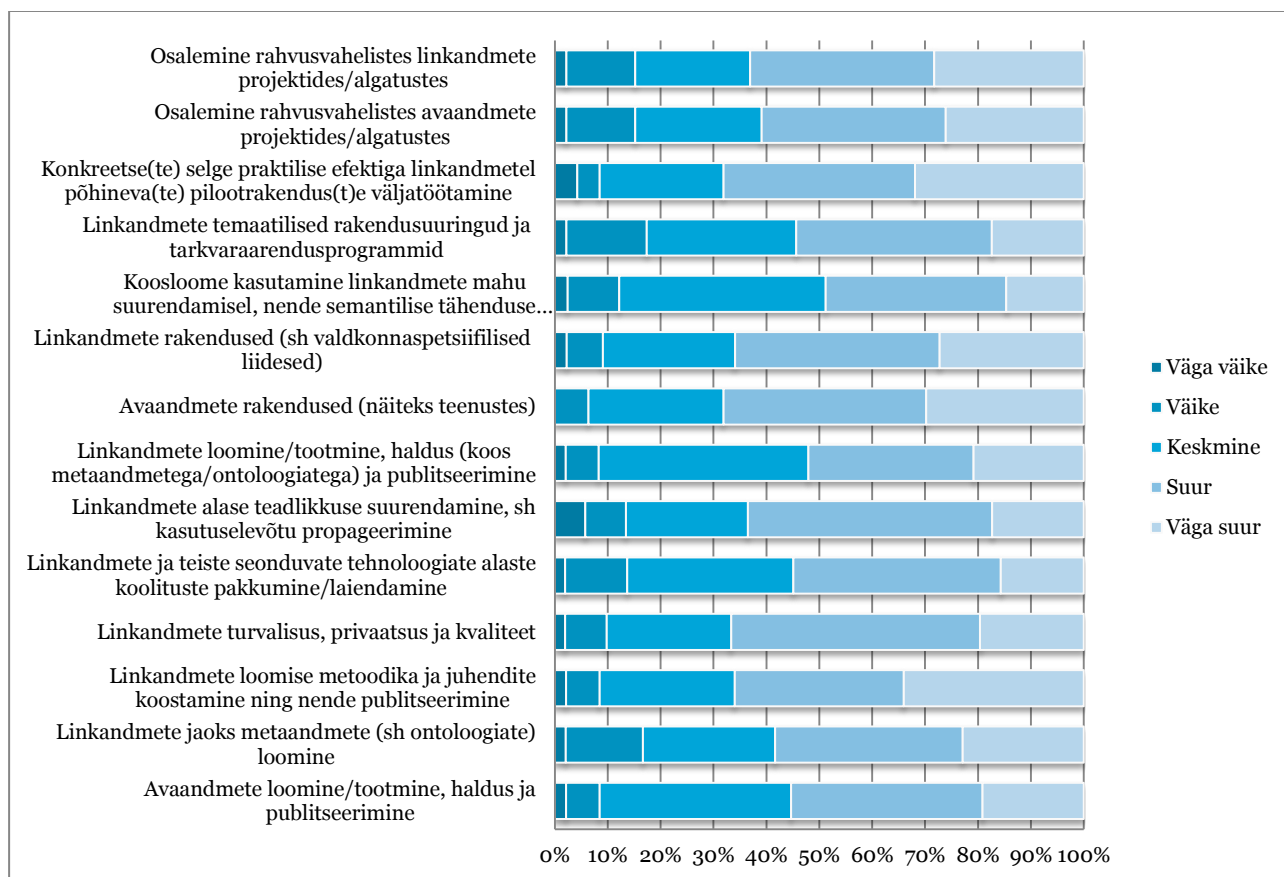
Joonis 19. Ohtude suurus seoses linkandmete laialdase kasutuselevõttuga

Palun hinnake valdkondade põhiandmehulkade linkandmetena avamise potentsiaalset kasu Eestile tervikuna.



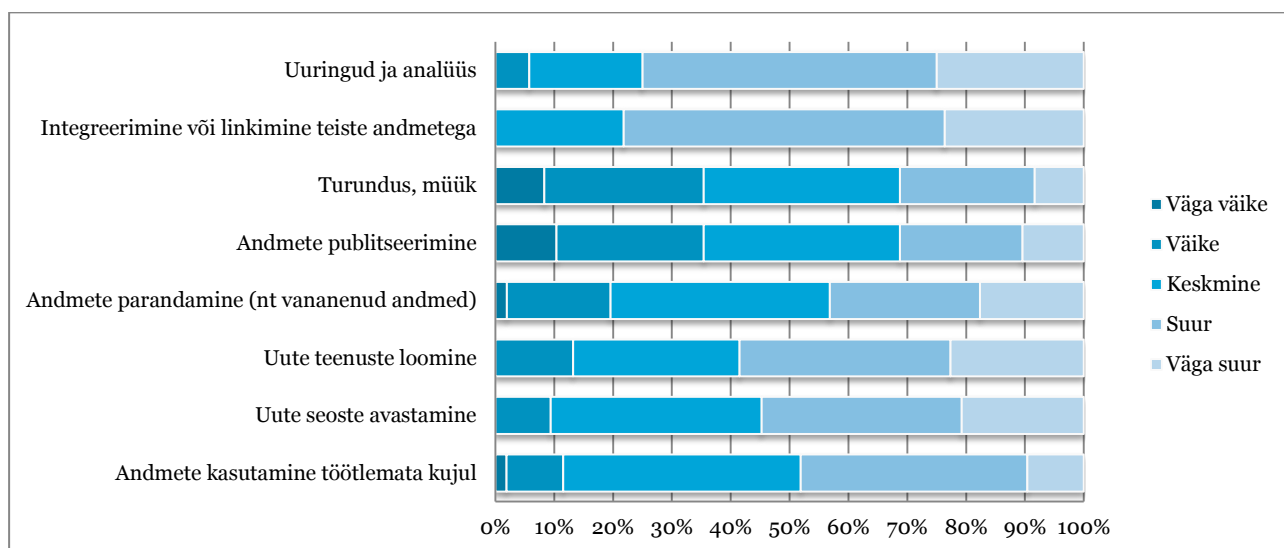
Joonis 20. Linkandmete avamise potentsiaalne kasu valdkonniti

Palun hinnake, milline mõju on erinevate projektide/tegevuste käivitamisel Eestile.



Joonis 21. Projektide ja tegevuste mõju Eestile

Linkandmete kontekstis peetakse üha enam oluliseks esitada avaandmeid linkandmetena. Palun hinnake avaandmete kasutusvõimaluste atraktiivsust enda jaoks (kuivõrd kasutaksite avaandmeid loetletud viisidel).



Joonis 22. Lingitud avaandmete kasutusvõimaluste atraktiivsus

7 Vastused uurimisküsimustele, ettepanekud ja poliitikasoovitused

Jaotises pakutakse esialgsed vastused uuringuga seoses tõstatatud uurimisküsimustele. Iga uurimisküsimusega seoses pakutakse ettepanekud ja poliitikasoovitused.

7.1 Linkandmete kasutuselevõtu üldised ettepanekud

Alltoodud üldised ettepanekud käsitlevad linkandmete kasutuselevõtu eesmäärke, trende, initsiatiive, pilootprojekte, ohtusid, riske, meetmeid ja ettepanekute elluviimist ning neid võib vaadata kui Lingitud Eesti esmast arengustrateegiat. Jaotis keskendub eelkõige koosvõime poliitilisele kontekstile ning õiguslikule ja organisatsioonilisele koosvõimele.

7.1.1 Linkandmete tehnoloogiate kasutamise eesmärgid ja tulemused

Milliste eesmärkide ja ülesannete jaoks oleks Eesti tingimustes mõistlik linkandmete tehnoloogiaid enim kasutada – nii avalikus kui ka era- ja vabasektoris? Millist kasu saaksid sellisest linkandmete rakendamisest Eesti tavaelanikud ja ettevõtted?

Linkandmete tehnoloogiad võimaldavad seostada omavahel isoleeritud andmeid ja seeläbi luua uusi teenuseid ja teadmisi ning edendada innovatsiooni. Näiteks on üleeuroopalise küsitluse aruandes „*Analyses of the LOD2 (Creating Knowledge out of Interlinked Data) – OGD Survey*“³⁵⁸ toodud positiivsete võimalustena välja valitsemise läbipaistvus ja efektiivsus, innovatsioon, 21. sajandi infrastruktuur. Valitsustele annavad linkandmete tehnoloogiad võimaluse nutikamate ja efektiivsemate avaliku sektori teenuste ja rakenduste väljatöötamiseks³⁵⁹. Neid rahvusvahelisi tulemusi toetab ja konkretiseerib käesoleva analüüsi käigus läbi viidud kvalitatiivuuring, millest selgusid järgmised olulisemad linkandmete kasutamise võimalikud eesmärgid ja tulemused.

- Nutikas ja läbipaistev riik, riigi ja kodaniku vahelise suhtluse paranemine. Näide: LEO³⁶⁰ (Läbipaistev Eesti Omavalitsus).
- Andmete linkimine võimaldaks paremini pakkuda „nähtamatuid“ teenuseid, millele inimesele on seadusandlik õigus ning mida ta peab saama sõltumatult sellest, kas ta tellib teenust või mitte. Näited: lapse riikliku sünnitoetuse teenus, mida ei peaks spetsiaalselt taotlema, kuna riigil on piisav hulk informatsiooni olemas, et teenus nähtamatult osutada; riiklik vanaduspension, mille puhul inimene ei peaks pensioni taotlema. Tulemuseks on vähem e-teenuseid.
- Andmete kiirem analüüs võimaldab kiiremini ja selgemalt aru saada, milline on otsuste mõju. Valedale otsustele on võimalik kiiremini reageerida ning uusi otsuseid langetada.

³⁵⁸

<http://wiki.lod2.eu/download/attachments/3473453/AnalysisSurvey.pdf?version=1&modificationDate=1329941495000>

³⁵⁹ https://joinup.ec.europa.eu/sites/default/files/D4.3.2_Case_Study_Data_Integration_vo.15.pdf

³⁶⁰ Läbipaistev Eesti Omavalitsus

<http://www.riigipilv.ee/QvAJAXZfc/opendoc.htm?document=LEO.qvw&host=local&anonymous=true>

- Monitooringu võimaldamine elutähtsate teenuste osutajate jaoks, mis lubab saada oluliselt kiiremini infot, kui elutähtis teenus on häiritud.
- Avaliku sektori kulude vähenemine, näiteks seoses teabenõuete vastamise vajaduse vähenemisega ja kodanike kaasamisega. Näide: tänavaaukude kaardistamine mobiilirakendusega.
- Elanikkonna elukvaliteedi tõstmine ja tervishoiu valdkonna kulude vähendamine – tervishoiu andmete pikaajaline kogumine ja analüüsimine annaks võimaluse hoida tulevikus kokku ravikulusid. Näide: lisaväärtus võiks tekkida geenivaramu andmete sidumisel e-Tervise andmetega.
- Uute teenuste osutamine erasektoris, avaliku sektori taandumine andmete tootjaks. Selline stsenaarium võimaldaks taandada avalik sektor teenuseosutaja rollist, pakuks paremat konkurentsi teenuseosutajate vahel ning vähendaks avaliku sektori kulutusi.

Ankeetides hinnati kõige olulisemateks linkandmete potentsiaalseteks kasudeks analüüsivõimekuse kasvu ja paremate otsuste tegemise võimalust, uutlaadi teenuseid ning riigi läbipaistvuse ja efektiivsuse suurenemist.

Riigi infosüsteemi koosvõime raamistik ning mõned muud raamistiku dokumendid (sh semantilise koosvõime raamistik, veebide raamistik) sisaldavad linkandmete ning nendega seotud tehnoloogiate elemente, kuid nende tehnoloogiate käsitus vajaks ajakohastamist ja täiendamist.

Järgnevates alajaotistes analüüsitakse lähemalt nende eesmärkidega seotud võimalikke initsiatiive, pilootprojekte, tehnoloogiaid jne.

Märgime, et kõiki ülalmainitud rakendusi ja teenuseid saab luua ka ilma linkandmeteta. Linkandmete kasutamine (eeldusel et neid on piisavalt olemas) võimaldab luua rakendusi kiiremini, lihtsamalt ja laiaulatuslikumalt. Analooiliselt võimaldab näiteks assemblerkeel põhimõtteliselt kõike programmeerida, kuid kõrgema taseme keelte kasutamine on tunduvalt tootlikum.

Tuleb siiski arvestada, et linkandmete tehnoloogiate rakendamise puhul on tegemist tüüpilise innovatsiooniprotsessiga, mille puhul uue tehnoloogia kuluefektiivne rakendamine eeldab eelnevat mahukat investeerimist toimiva taristu loomisesse, linkandmete eelnevasse publitseerimisse, kvalifitseeritud arendajate koolitamisega jne. Jaotises 7.3.1 on toodud avaliku sektori andmete avalikustamisest saadava võimaliku kasu suuruse kaudsed hinnangud. See kasu hakkab ilmuma siiski alles siis, kui linkandmete tehnoloogiad saavutanud piisava küpsuse. Jaotises 2.1.1 on toodud linkandmete ning nendega seotud tehnoloogiate küpsuse hinnangud.

Ettepanekud ja poliitikasoovitused

- Linkandmete tehnoloogiate kasutuselevõttu tuleks igati soodustada, sest nad võimaldavad seostada omavahel isoleeritud andmeid ja seeläbi luua uusi teenuseid ja teadmisi, edendada innovatsiooni ning radikaalselt parendada avalike teenuste pakkumist.
- Linkandmete kasutamise eesmärkide ja tulemuste hulka kuuluvad riigi ja kodaniku vahelise suhtluse paranemine (nt LEO, Läbipaistev Eesti Omavalitsus); „nähtamatud“ teenused, millele inimesele on seadusandlik õigus ning mida ta peab saama sõltumatult sellest, kas ta tellib teenust või mitte (nt lapse riikliku sünnitoetuse teenus või riiklik vanaduspension); andmete kiirem analüüs (valedetele otsustele on võimalik kiiremini reageerida ning uusi otsuseid langetada); monitooringu võimaldamine elutähtsate teenuste osutajate jaoks (lubab saada oluliselt kiiremini infot, kui elutähtis teenus on häiritud); avaliku sektori kulude vähenemine, sh seoses teabenõuete vastamise vajaduse vähenemisega ja kodanike kaasamisega (nt tänavaaukude kaardistamine mobiilirakendusega); elanikkonna elukvaliteedi tõstmine ja tervishoiu valdkonna kulude vähendamine (nt geenivaramu andmete sidumisel e-Tervise andmetega).
- Linkandmete ning nendega seotud tehnoloogiate temaatika tuleks ajakohastatult ja täiendatult lülitada riigi infosüsteemi koosvõime raamistikku ja sellega seotud dokumentidesse.

- Arvestada, et linkandmete puhul on tegemist on tüüpilise innovatsiooniprotsessiga, mille puhul uue tehnoloogia kuluefektiivne rakendamine eeldab eelnevat mahukat investeerimist toimiva taristu loomisesse, linkandmete eelnevasse publitseerimisse, kvalifitseeritud arendajate koolitamisega jne. Pikemas perspektiivis tuleks tutvustada linkandmete kasutuselevõtu edulugusid ning alustada linkandmete kasutamise projektidest tulenevate kulude vähenemise jälgimist asutustes.

7.1.2 Olulisemad trendid, Eesti võimekused ja võimalused

Millised järgnevatest trendidest on Eestile esmaselt olulised linkandmete kasutuselevõtu plaanis: tulevikuinternet, semantiline veeb, asjade Internet, suurandmed, avaandmed, koosloome? Millised on Eesti tänased võimekused ja arenguvõimalused oluliste trendide raames?

Trendide olulisuse analüüsil lähtume rahvusvahelisest kogemusest ja ülaltoodud ava- ja linkandmete valdkonna analüüsist, käesoleva uuringu raames läbiviidud küsitlustest ning uute tehnoloogiate sobivusest olemasolevasse riigi infosüsteemi arhitektuuri oma kasu tõestanud tehnoloogiate kõrvale. Üldisi seoseid nimetatud tehnoloogiate ja lingitud andmete vahel võib näha järgmiselt. Avaandmete initsiatiivist tulevad andmed, mida omakorda saab esitada linkandmetena. Semantiline veeb on üheks linkandmete allikaks ja esitlusvormiks, sh ontoloogiad annavad linkandmete (semantilise) meta-info ja läbi nende kasutamise tagatakse andmete lingitavus. Kogudes üleriigiliselt kokku linkandmed tekib vajadus suurandmete kogumise, haldamise ja töötlemise tehnoloogiate järele. Koosloome võib anda lisaväärtust andmete linkimisele ning seostub asjade Internetiga kui koosloomes osalevad ka asjad. Samas omab asjade Internet potentsiaali saada üheks olulisemaks linkandmete allikaks väljaspool avalikku sektorit. Tulevikuinternet on pigem tehniline ja linkandmete rakendustele nähtamatu kiht linkandmete automaatsel tekitamisel, levitamisel ja kasutamisel.

Praeguseks momendiks on nimetatud tehnoloogiad ja nende kasutuselevõtt erinevas arengujärgus ja erinevas küpsuse astmes. Gartneri poolt esitatud graafik nende tehnoloogiate arengu kohta³⁶¹ pakub, et semantiline veeb jõuab stabiilse kasutamise platoole enam kui kümne aasta pärast, avaliku sektori avaandmed kahe kuni viie aasta pärast, pilvepõhised tehnoloogiad suurandmete töötlemiseks viie kuni kümne aasta pärast, asjade internet enam kui kümne aasta pärast, kusjuures nende kõigi puhul on oodata vahepealset pettumuse perioodi peale esialgset vaimustust. Kui see prognoos vastab tõele, võiks 2020. aasta perspektiivis kõige suuremat rõhku panna avaandmetele, mille platoo peaks eeldatavasti saabuma kõige varem ja enne 2020. a.

Käesolevas uuringus läbi viidud intervjuudes peetakse riigi infosüsteemide seisukohalt oluliseks peamiselt linkandmete ja semantilise veebi tehnoloogiaid. Neid kahte tehnoloogiat nimetatakse sageli paaris, kuna semantika loob linkandmete kasutuselevõtuks vajalikud eeldused. Linkandmete kasutamise eelisena tuuakse välja riigi paindlikumat reageerimist muutustele ning kvaliteetsemat poliitikakujundamist. Informatsiooni kättesaadavuses ja kasutamises ei nähta konkurentsieelist mitte ainult erasektorile, vaid ka riigile. Asjade interneti, suurandmete ja koosloome osas arvamused hajuvad, kuna kogemus ja teadmised nende tehnoloogiate kasutamisel ei ole ekspertidel sarnased. Kokkuvõttes, kõige tähtsamaks linkandmete kasutuselevõttu peetakse andmete avamist ja semantiliste tehnoloogiatele kasutuselevõttu, millela ei ole linkandmete kasutuselevõtt võimalik. Samuti tundub avaandmete osas olevat mahajäämus suurim võrreldes teiste linkandmetega seotud tehnoloogiatega. Ülejäänud tehnoloogiate järjestuses tekivad eriarvamused ning ei joonistu välja selgeid prioriteete.

Ankeetide vastustes hinnati koosloome, suurandmete ja avaandmete tehnoloogiate olulisust üsna sarnaseks, keskmiselt vähem oluliseks peeti asjade Interneti tehnoloogiat. Eesti võimekust hinnati kõige kõrgemaks avaandmete valdkonnas, kõige madalamaks hinnati võimekust asjade Interneti ja semantiliste tehnoloogiate osas.

³⁶¹ Gartner, Hype Cycle for Big Data, 2012, Published: 31 July 2012

Ei tohiks alahinnata ka muid suundumusi, sh tulevikuinterneti vajalikkust, suurandmete teema kasvavat populaarsust ning asjade Interneti standardimisele asunud suurfirmade algatusi³⁶². Linkandmete kasutuselevõtu seisukohast ei saa neid siiski hetkel hinnata Eesti jaoks esmaselt olulistena.

Trendide analüüsil tuleb arvestada ka olemasolevat riigi infosüsteemi arhitektuuri. Selles domineerivad X-tee andmeteenused andmete loomise, haldamise ja levitamise ühe olulise mehhanismina. On läbi viidud pilootprojekte avaandmete tootmise osas, kasutades X-tee andmeteenuseid. Saab esitada andmete ja teenuste semantilist metainfot (nt ADS-i andmeteenuste semantilised kirjeldused RIHA-s).

Lähtudes ülaltoodust võib järeldada, et linkandmete kasutuselevõtuks on esmajoonel olulised avaandmed ja semantiline veeb. Eesti tänased võimekused ja arendusvõimalused on esmajoonel seotud teenustepõhise mudeli rakendamisega linkandmete tootmisel (taas)kasutades X-tee jaoks realiseeritud andmeteenuseid. See on ka pragmaatiline valik, arvestades olemasolevat riigi infosüsteemi arhitektuuri.

Ettepanekud ja poliitikasoovitused

- Linkandmetega seotud trendidest (tulevikuinternet, semantiline veeb, asjade Internet, suurandmed, avaandmed, koosloome) tuleks Eestile esmaselt oluliseks pidada avaandmeid ja semantilist veebi. Nendes valdkondades on Eestil täna esialgne võimekus olemas ning arenguvõimalused väga head, kui arvestada olemasolevat riigi infosüsteemi arhitektuuri, läbiviidud avaandmete tootmise pilootprojekte ning andmete ja teenuste semantilise metainfo esitamise juba olemasolevaid kogemusi.

7.1.3 Aktiivsus rahvusvahelistes initsiatiivides

Millistes rahvusvahelistes initsiatiivides peaks Eesti olema aktiivne, sh riiklikul tasemel (prioriteetsuse järjekorras)? Millised kompetentsid ja ressursid peaksid iga sellise initsiatiivi lõikes Eestil osalemiseks olemas olema?

Aktiivsus rahvusvahelistes initsiatiivides võib anda uut kogemust, aidata levitada oma tulemusi, parandada riigi mainet ja tuua kaasa lisaressursse. Samas nõuab see omapoolset kõrge kvalifikatsiooniga inimeste ajaressursside ja rahaliste vahendite panustamist. Eesti riigi väiksuse ja ressursside piiratuse tõttu tuleks initsiatiive hoolikalt valida. Ankeetides hinnati Eesti osalemise mõju rahvusvahelistes linkandmete ja avaandmete projektides/algatustes keskmiseks, võrreldes teiste uuritud valdkonna alaste tegevustega.

Vastavalt ava- ja linkandmete valdkonna analüüsile on selles valdkonnas olulised OECD, Euroopa Komisjoni, standardimise ja muud initsiatiivid. Euroopa Komisjoni initsiatiivid võiksid siinkohal olla kõrgeima prioriteediga, kuna osavõtt nendes võib anda reaalselt tulemust, nad mõjutavad Eestit kõige vahetumalt ning võivad kaasa tuua lisaressursse. Muuhulgas, Eesti võtab osa Euroopa Komisjoni tegevusest ning saaks seal rõhutada linkandmete osatähtsust. See on eriti oluline strateegiliste dokumentide puhul, mis võivad määrata suuremahuliste programmide käivitamist. Selline osavõtt vajab üldist link- ja avaandmete teemalist kompetentsi ning eeldab orienteeruvalt 1-2 inimkuu ulatuses lisaressurssi aastas, eriti juhul, kui Eesti esindajad niikuinii osalevad vastavate dokumentide loomises või kooskõlastamises.

Oluline on osavõtt Euroopa Komisjoni programmidest, nagu avaliku halduse koostöö programm (ISA)³⁶³, mille eesmärgiks on toetada ja hõlbustada tõhusat ja tulemuslikku piiriülest elektroonilist koostööd Euroopa haldusasutuste vahel. Muuhulgas, Joinup³⁶⁴ on Euroopa Komisjoni avaliku halduse koostöö lahenduste jagamise foorum, mida EK finantseerib ISA programmi abil ja mille põhieesmärk on koostöö lahenduste alase info ja kogemuste vahendamine e-riigi ekspertide hulgas. Euroopa Parlamendi ja Nõukogu otsuse nr 922/2009/EÜ³⁶⁵ artiklil 11 sätestab: "Ühiste raamistike ja baasvahendite loomist ja täiustamist rahastatakse täielikult ISA programmi vahenditest. Kõnealuste raamistike ja vahendite kasutamist

³⁶² http://bits.blogs.nytimes.com/2014/03/27/consortium-wants-standards-for-internet-of-things/?_php=true&_type=blogs&_php=true&_type=blogs&hp&r=1

³⁶³ The Interoperability Solutions for European Public Administrations, ISA, ISA teabevärv <http://ec.europa.eu/isa/>

³⁶⁴ <https://joinup.ec.europa.eu>

³⁶⁵ <http://eur-lex.europa.eu/legal-content/ET/TXT/PDF/?uri=CELEX:32009D0922&qid=1401088969847&from=EN>

rahastavad kasutajad". Analoogiliselt sätestatakse ühisteenuste loomist, industrialiseerimist ja täiustamist. Seega eeldab ISA tegevustest osavõtt eelkõige linkandmete alal pädeva tööjõu panustamist huvitatud asutuste poolt ning lähetuste rahastamist, kusjuures märgatava mõju avaldamiseks mingi tegevuse raames tuleks sellest tegevusest osa võtta vähemalt poole töökoha mahu ulatuses. Loodud vahendeid saab kasutada käivitavate pilootprojektide raames ning see ei nõua lisaressurssi.

EL uus teaduse ja innovatsiooni raamprogramm Horizon 2020 perioodiks 2014-2020. a rahastab aastatel 2014-2015 mitmeid ava- ja linkandmetega, tulevikuinternetiga, suurandmetega, asjade Internetiga ja koosloomega seonduvaid teadus- ja arendusprojekte. Sellistest programmidest osavõtt eeldab põhjalikku tehnilist koostööd, linkandmete ja teadustöö alast kompetentsi ning võib nõuda suurt eeltööd osavõtjate hulka pääsemiseks, kuid pakub vastukaaluks teadmisi ja kontakte, detailset tehnilist kogemust ning paljudel juhtudel ka lisarahastust. Suur osa Horizon 2020 uurimissuunalisi tegevusi on rahastatud 100% ulatuses, samas valdavalt innovatsioonile suunatud tegevuste rahastamise määr on 70%³⁶⁶. EL 7. raamprogrammi puhul pärssis väikese ja keskmise suurusega ettevõtete osaluse suurendamist see, et selliste ettevõtete puhul oli raske leida omafinantseeringut³⁶⁷. Seepärast võiks kaaluda rahastamisskeemi, mis toetab väikese ja keskmise suurusega ettevõtete osalemist Horizon 2020 programmi linkandmetega seotud projektides. Omafinantseeringu kogumahtu võib kaudsete meetoditega prognoosida, kasutades EL 7. raamprogrammi IKT programmi projekti Eesti poolse osaluse keskmist mahtu (suurusjärg 250 tuhat eurot) ja eeldatavat Horizon 2020 omafinantseeringu määra (30%). Vajalik rahastamisskeemi maht sõltub eeldatavast projektide arvust ja sellest, kui suur osa omafinantseeringust kompenseeritakse. Tööjõu vajadus sõltub Horizon 2020 puhul konkreetsest projektist, see on projekti raames rahastatud, see võib olla väga suur ning seda tuleb iga projekti taotluses eraldi ette planeerida.

Ava- ja linkandmete rakendamisel on tähtis roll rahvusvahelistel standarditel ja soovitel. Uuring „Standardiseerimise majanduslik mõju“³⁶⁸ väidab, et ligi 25% iga-aastasest SKP kasvust tuleneb standardiseerimisest. Tegemist on patenteerimise järel teise suurima mõjuga teguritest, mis mõjutab majandust. Ava- ja linkandmete standardimise valdkonnas on olulisim rahvusvaheline veebikonsortsium W3C (sh *W3C Government Linked Data Working Group*), kuid ka ülalmainitud Euroopa Komisjoni programmid püüavad saadud tulemusi fikseerida standardimise kaudu. Rahvusvaheline standardiorganisatsioon ISO on välja töötanud mitmeid linkandmetega otseselt või kaudselt seotud standardeid, näiteks geograafilise informatsiooni metaandmete standardi ISO 19115:2003³⁶⁹. Eesti saab rahvusvaheliste standardite väljatöötamises kõige kuluefektiivsemalt osaleda läbi olemasolevate struktuuride. Näiteks on Eesti Standardikeskus ISO täisliige, mis võimaldab osaleda ava- ja linkandmetega seotud standardimises; samuti on otstarbekas ühendada Euroopa Komisjoni projektidest osavõtt standardistegevustega. Võimalik on ka otsene osavõtt W3C tööst, seda nii era- või avaliku sektori organisatsioonide kui ka projektide kaudu³⁷⁰. Mainitud tegevused eeldavad väga head kompetentsi standarditavas valdkonnas, teadmisi standardimise protsessist, lisatööjõudu soovitatavalt vähemalt poole töökoha mahu ulatuses ning lähetuste rahastamist.

OECD avaliku sektori avaandmete (*Open Government Data*) projekti eesmärgiks on läbi erinevate riikide praktika soodustada avaandmete mõjude (nii majandusliku, sotsiaalse kui valitsemise mõju) hindamise meetodika väljatöötamist. Tõenäoliselt nõuaks selle meetodika mõjutamine suurt ajaressurssi, kui see üldse on võimalik. Eesti saab kasutada OECD poolt levitatavat kogemust, jälgida OECD meetodikaid ning kui see osutub otstarbekaks, siis sättida oma tegevusi nii, et mitte olla OECD ülevaadetes viimastel kohtadel. Seda saab teha muude tegevuste raames ning ei nõua suurt rahalist või ajalist ressurssi.

Kasulik võib olla ka osavõtt muudest initsiatiividest, nagu Euroopa andmete foorumi³⁷¹ konverentside seeriast, Euroopa avaliku sektori infotribüünist ePSIplatform³⁷², avatud valitsemise partnerlusprogrammist³⁷³ jne. Avatud valitsuse partnerlusprogramm ist osavõtuks tuleb valitsusel tasuda

³⁶⁶ http://ec.europa.eu/research/participants/docs/h2020-funding-guide/grants/applying-for-funding/find-a-call/what-you-need-to-know_en.htm

³⁶⁷ <http://www.etag.ee/wp-content/uploads/2012/05/7RP-vahekokkuvote.pdf>

³⁶⁸ http://www.sis.se/pdf/Economic_impact_of_standardization_France.pdf

³⁶⁹ http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

³⁷⁰ <http://www.w3.org/Consortium/membership-faq#project>

³⁷¹ <http://www.data-forum.eu/>

³⁷² <http://epsiplatform.eu/>

³⁷³ <http://www.opengovpartnership.org/>

iga-aastast liikmemaksu, mille suurus sõltub elanike sissetulekust ja on Eesti puhul 200,000 USD³⁷⁴. Osavõtt muudest eelpoolnimetatud initsiatiividest eeldab ekspertide tööaega ning lähetuste rahastamist.

Eesti erasektori esindajad (näiteks ITL aga ka erafirmad, eriti idufirmad) võiksid astuda ODI toetajaliikmeks (liikmemaks 720 Inglise naela aastas). ODI partnerliikmeks saamine ja olemine on suhteliselt kallis (70000 Inglise naela aastas), mis pole ilmselt jõukohane ja ei vasta ka ootustele.

Ettepanekud ja poliitikasoovitused

- Eesti peaks olema aktiivne järgmistes linkandmetega seotud initsiatiivides (prioriteetsuse järjekorras): Euroopa Komisjoni initsiatiivid ja programmid (sh strateegiliste dokumentide loomine, ISA programm, Joinup foorum, Horizon raamprogramm, eSENS projekt jt), standardimistegevused, OECD avaliku sektori avaandmete (*Open Government Data*) projekt, Euroopa andmete foorumi konverentside seeria, Euroopa avaliku sektori infotribüün ePSIplatform, avatud valitsemise partnerlusprogramm jne.
- Osavõtt Euroopa Komisjoni tegevusest vajab üldist link- ja avaandmete teemalist kompetentsi ning tööaega. Osavõtt EL raamprogrammist Horizon 2020 eeldab põhjalikku tehnilist koostööd, linkandmete ja teadustöö alast kompetentsi; kuna EL 7. raamprogrammi puhul pärssis väikese ja keskmise suurusega ettevõtete osaluse suurendamist omafinantseeringu vajadus, võiks kaaluda rahastamisskeemi, mis toetab väikese ja keskmise suurusega ettevõtete osalemist Horizon 2020 programmi linkandmetega seotud projektides. Osavõtt standardimisest eeldab väga head kompetentsi standarditavas valdkonnas, teadmisi standardimise protsessist, üsna palju tööaega ning lähetuste rahastamist. Arvestades Eesti väiksust ning ressursside nappust tuleks linkandmetega seotud tegevus võimalusel ühendada muude tegevustega vastavate initsiatiivide raames.
- Eesti erasektori esindajad (näiteks ITL aga ka erafirmad, eriti idufirmad) võiksid astuda ODI toetajaliikmeks (liikmemaks 720 Inglise naela aastas). ODI partnerliikmeks saamine ja olemine on suhteliselt kallis (70000 Inglise naela aastas), mis pole ilmselt jõukohane ja ei vasta ka ootustele.

7.1.4 Pilootprojektide käivitamine

Millised pilootprojekte peaks Eesti käivitama siseriiklikult? Millises järjekorras neid algatada ja milliseid ressursse nad vajaksid (sh projektide hinnanguline maksumus), millist kasu riigile ja/või ühiskonnale toodsid?

Linkandmete kasutuselevõtt eeldab järgmisi etappe: andmete avamist (kättesaadavaks tegemist), nende puhastamist ja korrastamist (vajadusel), viimist vähemalt neljanda või viienda järjekorras, linkimist, taaskasutamist (nt päringute võimaldamist) ja haldamist³⁷⁵. Kõigil neil etappidel on vaja lahendada suur hulk uurimisküsimusi ning näiteks EL uus teaduse ja innovatsiooni raamprogramm Horizon 2020 perioodiks 2014-2020. a rahastab aastatel 2014-2015 kokku enam kui poolteise miljardi euro maksumuses ITK alaseid initsiatiive, sealhulgas palju ava- ja linkandmetega seonduvaid teadus- ja arendusprojekte.

Eesti siseriiklikke linkandmete kasutuselevõtu pilootprojekte on soovitatav kavandada nii, et nad (1) arvestaksid Eesti eripära (nt X-tee ja RIHA olemasolu), (2) ei dubleeriks mõjuva vajaduseta rahvusvahelisi suuremahulisi uurimisprojekte, (3) looksid võimaluse kolmandatele osapooltele linkandmete rakenduste loomiseks, (4) tekitaksid lõppkasutajale või üldsusele demonstreeritava kasuliku näitetulemuse („*proof of concept*“), (5) annaksid asjaosalistele uusi kogemusi ja (6) oleksid jätkusuutlikud. Näiteks võiks suhtuda ettevaatlikult linkandmete publitseerimise pilootprojektide algatamisse, mis ei demonstreeri selgelt saadavat kasu ning mis ei kindlusta selle kasu jätkusuutlikkust, sealhulgas loodud linkandmete uuendamist ja haldust.

³⁷⁴ <http://www.opengovpartnership.org/node/1330>

³⁷⁵ <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>

Üheks pilootprojektide käivitamise eesmärgiks on kogemuste andmine uute tehnoloogiatega töötamisel. Anketeerimise tulemustest selgus, et kogemused avaandmetega puudusid 41% vastajatest, kogemused ülejäänud vaadeldavate tehnoloogiatega – 57-68% vastajatest.

Kui analüüsida võimalikke pilootprojekte ülaltoodud linkandmete kasutuselevõtu etappidel, siis andmete avamise vajadus peaks saama rahuldatud, sest avaliku teabe seaduse § 29 lg 4 nõuab andmekogudes sisalduvate andmete avalikustamist masinloetaval kujul hiljemalt 1.01.2015. Kuna masinloetavuse mõiste pole seaduses määratletud, võib tekkida vajadus käivitada pilootprojekte andmete viimiseks linkimiseks sobivale vähemalt neljanda või viienda järni tasemele. Samuti tuleks käivitada pilootprojektid toetamaks linkandmete kasutuselevõttu nii avaliku sektori veebisaitides kui erasektorile mõeldud lahendustes, mille järele on tuvastatud teistes uuringutes (nt koosvõimeliste e-äri/e-kaubanduse lahenduste komplekt firmadele).

Vastava kogemuse saamiseks ja üldistamiseks on otstarbekas käivitada orienteeruvalt kuni kuus tehnilist pilootprojekti alternatiivsete linkandmete tehnoloogiate järeleproovimiseks praktikas. Need pilootprojektid peaksid:

- arvestama Eesti eripära, eriti X-tee kasutamist;
- demonstreerima infoühiskonna arengukava täidesaatmiseks vajalike tegevuste toetamist ja mõõtmist (a'la veebide seire ja linkimine WCAG nõuete täitmisepidevaks jälgimiseks);
- looma üldise mudeli ja pakkuma konkreetseid lahendusi edasisteks suuremateks linkandmete alasteks projektideks, nt avatud linkandmete platvormi loomine lingitud andmetel baseeruvate e-äri lahenduste jaoks;
- arvestama arengutega lähema 5 aasta jooksul - näiteks tuvastama, milline platvorm sobiks linkandmete jaoks suurandmete kontekstis või kuidas haakida koosloomet linkandmete loomisesse, haldamisse ja kasutamisesse;
- analüüsima, milliseid W3C mustandis „Linked Data Platform Use Cases and Requirements“³⁷⁶ toodud nõudeid linkandmete platvormidele koos kasutuslugudega oleks mõistlik Eesti linkandmete arhitektuuri realisatsioonis ära katta;
- sisaldama metaandmete, sealhulgas ontoloogiate loomist ja täiendamist;
- taotlema realselt kasuliku tulemuse saamist ning jätkusuutlikkust, andes kolmandatele osapooltele võimaluse linkandmete rakenduste loomiseks, tekitades lõppkasutajale või üldsusele demonstreeritava kasuliku näitetulemuse ning luues andmete uuendamise ja haldamise mehhanismi;
- soodustama linkandmete varamu määratlemist ja kasutuselevõttu, võimalusel kasutades või täiendades selleks juba olemas olevaid lahendusi;
- vajadusel kasutama koosloome (nt linkandmete tootmine teeaukude, avatud kaevude või muude kriitiliste objektide kaardistamine) ja suurandmete (nt linkandmete analüütika reaalajas koos välisriikide linkandmetega) tehnoloogiaid.

Projektide alustamisel saab kasutada mitmete riikide, sh Soome näiteid³⁷⁷.

Kui selliste projektide maksumuse hindamisel lähtuda juba läbi viidud avaandmete projektidest, sealhulgas Ehitisregistri pilootprojektist³⁷⁸, lisades ka lihtsa reaalset probleemi lahendava laiemale üldsusele arusaadava kasuliku rakenduse loomise, teise andmekomplektiga linkimise, valdkondlike ontoloogiate täiendamise ning mitme erineva baastehnoloogia rakendamise (a'la koosloome linkandmetega koosneb kahest

³⁷⁶ <http://www.w3.org/TR/2013/WD-ldp-ucr-20130131/#scope-and-motivation>

³⁷⁷ <http://www.seco.tkk.fi/events/2014/2014-01-24-ldf/>

³⁷⁸ Ehitisregistri avaandmed - <http://opendata.riik.ee:8080/dataset/ehitisregistri-avaandmed>

baastehnoloogias) maksumuste hinnangud, siis võib ühe pilootprojekti maksumuse suurusjärguks pakkuda mõnikümme kuni sada tuhat eurot sõltuvalt mil määral väljatoodud aspekte arvestatakse. Üldsuse teavitamise seisukohast on võimalik siduda linkandmete loomist riigi poolt tellitud uuringute läbiviimisega, kus uuringu raames tehakse sisendandmed ja tulemused lingitud kujul kättesaadavaks.

Lisaks ülalmainitud väiksema mahuga pilootprojektide läbiviimisele tuleks arvestada, et andmete linkimisel annab suurima efekti enamkasutatavate andmete linkandmete kujule viimine. Nii näitab X-tee andmeteenuste analüüs³⁷⁹, et Äriregistri, Rahvastikuregistri, Maakatastri ja ADS-i metaandmed võimaldavad kirjeldada umbes 80% riigi andmeteenuste andmemudeleist. Andmete linkimise kontekstis tähendab see, et vastavate registrite ja andmekogude andmed on lingitavad praktiliselt kõikide teiste andmekogude ja registrite andmetega. Järgmine suurem programm oleks seega enamkasutatavate andmete põhjal linkandmete tuumiku loomine, kuhu ülejäänud andmeid saab vastavalt vajadusele juurde linkida. See programm peaks algama õiguslike aspektide analüüsi sisaldava pilootprojektiga ning sisaldama mitmeid järjestikuseid projekte oluliste andmekogude linkimiseks. Programmi mahtu saab täpsemalt hinnata peale tehniliste pilootprojektide läbiviimist.

Eestis on olemas väga head koosloome näited („Teeme ära“) ja vajadusel saab koosloome elemente lülitada algatatavatesse projektidesse. Nii võib lisaks ülalmainitud tehnilistele pilootprojektidele kasutada linkandmete platvormi samaliigiliste andmete erinevatest allikatest kokkukogumiseks. Näitena võib tuua Belgia üleriigilise ja regionaalsete aadressiregistrite andmete ühendamine³⁸⁰. Sellised pilootprojektid sobivad koosloomega seotud aspektide katmiseks, kuna koosloome eripäraks on andmete hajus tekkimine ja haldamine väiksemate andmekomplektide kaupa. Eesti puhul võiks ühendada kohalike omavalitsuste andmed, üldistades kohalike omavalitsuste raamatupidamise avaandmete rakendust³⁸¹. Selliste projektide maksumus sõltub oluliselt püstitatud ülesannetest ja lingitavate andmete valikust, jäädes tehniliste pilootprojektide maksumuse ja linkandmete tuumiku programmi maksumuse vahepeale.

Vajadus reaalajas andmetepõhiselt otsustada võib kaasa tuua vajaduse suurandmete järele. Samas kahtlevad kvaliteetivõime osalejad, kas Eestis on suurandmeid. Ilmselt on see tingitud suurandmete osas mahu aspekti ületähtsustamisest ning kiiruse ja variatiivsuse alahindamisest. Suurandmete tehnoloogiaid tuleks lülitada ülaloodud projektidesse vastavalt vajadusele, kuid eraldi spetsiifiliste suurandmete projektide vajadus ei ole ilmne. Küll võib olla otstarbekas suurandmete ja koosloome alane koolitus.

Asjade Interneti ja tulevikuinterneti arendamiseks oleks esmajoones vaja pöörata tähelepanu internetiprotokolli 6. versiooni kasutusele võtmisele Eestis, sest ilma selleta ei ole võimalik näiteks asjade Internet. Kui teenusepakkujad lähevad IPv6-le üle vastavalt nõudlusele ja võimalustele³⁸², siis riigi poolne initsiatiiv võiks seisneda ülemineku käivitamises IPv6-le avalikus sektoris. Vastavalt USA Kaubandusministeeriumi uuringule³⁸³ jagunevad suurte kasutajate puhul kulud riistvarale, tarkvarale ja tööjõule proportsiooniliselt 10/20/70, ehk põhiosa on tööjõukuludel. Kulud võib jaotada pikale ajavahemikule, kui varakult alustada. APNIC, üks viiest Interneti regionaalsest registrist maailmas, hindab oma uurimuses „IPv6 for decision makers“³⁸⁴, et üleminek IPv6-le koos paralleelse IPv4 kasutamisega võib maksta 40 USD-d ühe kasutaja kohta ühes aastas viieaastase perioodi jooksul. Need arvud on väga hinnangulised ja võivad tegelikkuses olla erinevates riikides erinevad, kuid pakutud suurusjärgud näitavad, et näiteks üle tuhande töötajaga ministeeriumi valitsemisalas võivad rahalised ja tööjõu kulutused IPv6-le üleminekuks olla märkimisväärsed ning need tuleks eelarvesse sisse planeerida või rahastada eraldi projektidena.

Ettepanekud ja poliitikasoovitused

- Otstarbekas on käivitada orienteeruvalt kuni kuus tehnilist pilootprojekti alternatiivsete linkandmete tehnoloogiate järeleproovimiseks praktikas erinevate seotud tehnoloogiate võtmes valdkonnaülestes infoühiskonna arengukavas väljatoodud oluliste mõõdikute parendamise võtmes. Ühe pilootprojekti

³⁷⁹ X-tee andmeteenuste analüüs (Haav, Hele-Mai; Küngas, Peep (2013). Semantic data interoperability: the key problem of big data. In: Big Data Computing: (Toim.) Akerkar, Rajendra. Chapman and Hall / CRC Press, 2013. [ilmumas]

³⁸⁰ https://joinup.ec.europa.eu/sites/default/files/D4.3.2_Case_Study_Data_Integration_v0.15.pdf

³⁸¹ <http://www.riigipilv.ee/>

³⁸² <https://www.zone.ee/et/2013/06/06/zone-ee-kaivitas-ipv6-toe/>

³⁸³ <http://www.ntia.doc.gov/legacy/ntiahome/ntiageneral/ipv6/final/IPv6final2.htm>

³⁸⁴ http://www.apnic.net/community/ipv6-program/ipv6-for-decision-makers/ipv6_for_decision_makers.pdf

maksumuse suurusjärguks võib pakkuda kuni sada tuhat eurot. Pilootprojektid võiksid olla soovitatavalt tervishoiu, georuumilises, hariduse või transpordi, e-äri ja infrastruktuuri valdkondades, kusjuures väga oluline on vastavate organisatsioonide valmisolek projektide algatamiseks. Pilootprojekte on soovitatav kavandada nii, et nad arvestaksid Eesti eripära, ei dubleeriks mõjuva vajaduseta rahvusvahelisi suuremahulisi uurimisprojekte, looksid võimaluse kolmandatele osapooltele linkandmete rakenduste loomiseks, tekitaksid lõppkasutajale või üldsusele demonstreeritava kasuliku näitetulemuse, annaksid asjaosalistele uusi kogemusi ja oleksid jätkusuutlikud.

- Järgmine suurem programm oleks enamkasutatavate andmete põhjal linkandmete tuumiku loomine, kuhu ülejäänud andmeid saab vastavalt vajadusele juurde linkida. See programm peaks algama õiguslike aspektide analüüsi sisaldava pilootprojektiga ning sisaldama mitmeid järjestikuseid projekte oluliste andmekogude linkimiseks. Programmi mahtu saab täpsemalt hinnata peale tehniliste pilootprojektide läbiviimist.
- Lisaks ülalmainitud tehnilistele pilootprojektidele tuleks kasutada linkandmete platvormi samaliigiliste andmete erinevatest allikatest kokkukogumiseks, näiteks kohalike omavalitsuste andmete ühendamiseks, üldistades kohalike omavalitsuste raamatupidamise avaandmete rakendust. Selliste projektide maksumus jääb tehniliste pilootprojektide maksumuse ja linkandmete tuumiku programmi maksumuse vahepeale.
- Asjade Interneti ja tulevikuinterneti arendamiseks oleks esmajoonel vaja pöörata tähelepanu internetiprotokoll 6. versiooni kasutusele võtmisele Eestis. Riigi poolne initsiatiiv võiks seisneda ülemineku käivitamises IPv6-le avalikus sektoris. Tuleb arvestada, et rahalised ja tööjõu kulutused IPv6-le üleminekuks üle tuhande töötajaga ministeeriumi valitsemisalas võivad olla märkimisväärsed ning need tuleks eelarvesse sisse planeerida või rahastada eraldi projektidena.

7.1.5 Linkandmete alase võimekuse tõstmine

Kuidas tõsta Eestis linkandmete ja nendega seotud trendide alast võimekust avalikus, era- ja vabasektoris?

Võimekuse tõstmine on kompleksne teema. Nii näitab UK kogemus³⁸⁵, et linkandmete RDF kujul kättesaadavaks tegemisest ei piisa nende laialdaseks tarbimiseks rakenduste kaudu. Seetõttu loodi Ühendkuningriikides vahevara linkandmete kättesaadavaks tegemiseks ka JSON ja XML andmekirjelduskeeltes. Lisaks loodi valdkonnaspetsiifilised API-d vastavate andmete kasutamiseks. Samuti toob aruanne välja linkandmete tootjate vajaduse konkreetsete eeskirjade ja muustrite järele linkandmete tootmiseks ning vajaduse juhendamise järele linkandmete tootmisel ja haldamisel. Eestis näitavad esialgsed Ehitisregistri linkandmete kogemused RDF kasutamise osas nii õppetöös kui erasektori tarbijate poolt, et piisava huvi olemasolu korral ei saa linkandmete tehnilised detailid tarkvaraarendajatele rakenduste loomisel takistuseks. Pigem on küsimus teadlikkuse tõstmises ja ajendi tekitamises linkandmete loomise ja kasutamise motiveerimiseks vahendite piiratud arvestades.

Kuna võimekuse tõstmise erinevaid aspekte käsitlevad mitmed käesoleva töö uurimisküsimused, viitame allpool võimekuse tõstmise küsimuse analüüsil teistele peatükkidele. Kasutame seejuures metoodikat, mis põhineb eBCM ehk e-äri kogukonna mudeli (*The E-Business Community Model*) lähenemisviisil³⁸⁶, mis on välja töötatud projektis eBCM-RAP (finantseeritud Norra Innovatsioonikeskuse poolt) ning mida on kasutatud Riigikantselei tarkade otsuste fondi uurimisprojekti „Andmeaitade (teiseste andmekogude) loomise põhimõtete väljatöötamine“³⁸⁷. Metoodika jagab uuritava valdkonna teemaplokkideks, igas teemaplokkis on omad alamvaldkonnad. Mudeli loogika on järgmine.

- Põhivalmisoleku teemaplokk iseloomustab seda, kui võrd ühiskond on valmis uue tehnoloogia rakendamiseks. Rakendamine eeldab kokkuleppeid – seadusi ja regulatsioone, standardeid ning

³⁸⁵ http://events.linkedata.org/ldow2010/papers/ldow2010_paper14.pdf

³⁸⁶ <http://www.ebaltics.com/00904523>

³⁸⁷ http://valitsus.ee/UserFiles/valitsus/et/riigikantselei/strateegia/poliitika-analuusid-ja-uuringud/tarkade-otsuste-fondi-uuringute-kokkuvotted/Andmeaidad_1%C3%B5pparuanne.pdf

lepinguid, mis toetavad tehnoloogia juurutamist ja kasutamist. Samuti on hädavajalik kindlustunne suuremate riskide suhtes, mille annab infoturbe tagamine ja andmete usaldusväärsuse kindlustamine. Lõpuks on oluline vastav mõtteviis – tehnoloogia kohta käivad teadmised, inimeste motivatsioon ning asutuste vaheline koostöö.

- Ainult põhimõttelisest valmisolekust ei piisa, on vaja ka võimaldajaid – taristut, eriti IT osas, aga samuti kvalifitseeritud inimressurssi ning praktilisi kogemusi.
- Valmisolek ja võimaldajad pannakse tööle liikumapanevate jõudude poolt. Eelkõige loovad uued tehnoloogiad uusi võimalusi ning motiveerivad maandama kaasnevaid ohte. Kogu suuna on muutnud võimalikuks tehnoloogia ja innovatsioon. Turud ja firmadevaheline konkurents võimendab tehnoloogilist arengut. Ühiskond ja poliitilised jõud tunnetavad vajadust agregeeritud informatsiooni põhjal tehtavate otsuste järele, mis motiveerib linkandmete arengut.

Käesolevas uuringus on teemaplokid ja alamvaldkonnad kaetud järgmiselt:

- Põhivalmisolek: kokkulepped, sh õigusaktid ja standardid – käsitletud eelkõige punktides 7.1.8 ja 7.2.1.
- Põhivalmisolek: kindlus, sh infoturve ja usaldusväärsus – käsitletud eelkõige punktides 7.1.7 ja 7.1.8.
- Põhivalmisolek: mõtteviis (motivatsioon, teadmised, koostöö) – käsitletud eelkõige punktides 7.1.1, 7.1.3, 7.3.1, 7.3.6.
- Võimaldajad (IT taristu, kvalifitseeritud inimressursid, praktilised kogemused) – käsitletud eelkõige punktides 7.1.2, 7.2.2, 7.3.1.
- Liikumapanevad jõud (võimalused ja ohtude maandamine, tehnoloogia ja innovatsioon, turud ja konkurents, ühiskond ja poliitika) – käsitletud eelkõige punktides 7.1.1, 7.1.2, 7.2.2, □, 7.3.5.

Ettepanekud ja poliitikasoovitused

- Linkandmete alase võimekuse tõstmiseks tuleks eelkõige paralleelselt käivitada linkandmete teemalised kursused ja koolitused vastavalt ettepanekutele ja poliitikasoovitustele jaotisest 7.3.1, algatada pilootprojektid vastavalt soovitudele jaotisest 7.1.4 ning luua tehnilise infrastruktuuri lahendused vastavalt soovitudele jaotisest 7.2.2.
- Lisaks tuleks osaleda rahvusvahelistes initsiatiivides vastavalt soovitudele jaotises 6.1.3, maandada riske vastavalt soovitudele jaotises 7.1.7 ning korrastada õigusruumi vastavalt soovitudele jaotises 7.1.8.

7.1.6 Linkandmete alase teadlikkuse tõstmine

Millised on võimalused linkandmete alase teadlikkuse tõstmiseks nii avalikus, era- ja vabasektoris kui ka tavaelanike hulgas?

Tavaelanikud võivad tänu linkandmete kasutamisele saada paremaid teenuseid, kuid nad pole eeldatavalt linkandmete nn toorel kujul kasutajate peamine sihtgrupp. Seda arvestades peaks peamine teavitustöö linkandmete olemuse osas toimuma sihtgruppides, kes loovad, haldavad või kasutavad otseselt linkandmeid (sh andmete omanikud ja rakenduste loojad).

Anketeerimine näitas, et teadlikkuse tõstmine on Eestis oluline – 77-st laekunud täidetud ankeedist vastati kahekümne ühes (27%), et vastajal ei olnud teadmisi ühestki küsitatud tehnoloogiast. Nende seas, kes olid mõnest tehnoloogiast teadlikud, oli jaotus tehnoloogiate lõikes küllaltki ühtlane – 9-25% vastajatest ei teadnud vastavast tehnoloogiast midagi, 38-53% vastajatest olid üldisel tasemel teadmised, 20-32% vastajaid hindas oma teadmisi keskmiseks.

Artiklis „*Impediments, challenges and recommendations for using open government data*“³⁸⁸ soovitatakse avaandmete kasutuse laiendamiseks ning parendamiseks 1) poliitikajuhiste väljatöötamist avaandmete tsentraliseerimiseks, 2) tehnoloogiliste ja õiguslike ligipääsumehhanismide väljatöötamist andmete avamiseks, 3) metaandmete lisamist andmetele sobivate andmekomplektide leidmise hõlbustamiseks ning 4) infrastruktuuri arendamist andmete (taas)kasutamiseks. Andmekomplektide kasutamise jaoks tsentraalsete ligipääsupunktide loomine, koos teabe levitamise selle olemasolust, aitab teadvustada uusi võimalusi. Metaandmete olemasolu teeb andmete kasutamise operatiivsemaks.³⁸⁹ Andmete levitamise ning kasutamise infrastruktuuri (nt analüütiline keskkond) loomine loob soodsa pinnase uute ideede tekkeks andmete rakendamise osas³⁹⁰. Analoogilised soovitused võiksid kehtida ka avaliku sektori linkandmetele.

Riigi linkandmete kõrval tuleks soodustada ka koosloomest tekkivate linkandmete kasutamist. Seda on mainitud ka Ühendkuningriikide kogemust kirjeldavas kokkuvõttes³⁹¹.

Intervjuudes toodi välja vajadus teadmuse koondamiseks ja parimate praktikate levitamiseks töövahendite näol. Teavitustöö linkandmetega seotud võimalustest ja ohtudest võiks jälgida häid praktikaid, mis on tekkinud muuhulgas Eestis läbi viidud ISKE turvameetmete süsteemi või riigi IT koosvõime raamistiku alase teadlikkuse tõstmise käigus. Vastava teavitustöö korraldamisel tuleks tugineda linkandmete kasutamist hõlbustavate tehnoloogiate ja rakenduste tekkimise ja arengu näidetele ja parimatele praktikatele.

Teadlikkuse tõstmise üks oluline teema on andmete linkimiseks suurema potentsiaaliga andmekogude väljavalimine ning nende omanike aitamine ja motiveerimine andmete avamiseks ja linkandmete ettevalmistamiseks. Selleks peaks avamise ja linkimise mõju olema ära kirjeldatud ja tasuvus piisav. Seega tuleks edukate linkandmete taotluste esitamiseks pakkuda asutustele metoodikad ja argumendid loodavate linkandmete mõju hindamiseks ja esiletoomiseks. Üks võimalus selleks on anda hinnang linkandmete kasutatavuse efektile läbi andmekoosseisude kirjelduste kattuvuse analüüsi – mida rohkem on sama andmeobjekti erinevates andmekogudes kasutatud, seda suurema tõenäosusega saaks seda teiste andmetega linkida ja mida rohkem on andmekogus teiste andmekogude objektidega lingitavaid objekte, seda suurem mõju on vastava andmekogu avamisel või linkimisel.

Samuti on suurema mõjuga nende andmete avamine või linkimine, mis on ka teistes riikides avatud või lingitud – siis saaks Eesti andmetele tehtud rakendust ka eksportida eeldusel, et kasutatud on sama metamodelit kui teiste riikide andmetes. Samuti saaks Eesti andmetele rakendada välisriikides loodud lahendusi. See viib omakorda suurandmete teemani, mis pole Eesti andmeid üksikuna vaadates veel aktuaalne, kuid võib muutuda aktuaalseks koos Skandinaavia või teiste laiema piirkonna andmetega.

Analoogsed soovitused kehtivad teiste uuringuga seotud tehnoloogiate kohta.

Ettepanekud ja poliitikasoovitused

- Linkandmete alase teadlikkuse tõstmiseks tuleks alustada infopoliitika ja selle rakendusplaanide uuendamisest, koosvõime raamistiku täiendamisest ja juhendite koostamisest. Tuleks käivitada esmased pilootprojektid (vt jaotis 7.1.4), mis demonstreerivad linkandmete kasulikkust ja mida saab koolitustel näidetena kasutada. Koolituste esimene etapp peaks keskenduma üldistele (orienteeruva mahuga üks päev) ja süvendatud (tehniline lisakoolitus orienteeruva mahuga kaks päeva) koolitustele. Teavitustöö linkandmetega seotud võimalustest ja ohtudest võiks jälgida häid praktikaid, mis on tekkinud ISKE turvameetmete süsteemi või riigi IT koosvõime raamistiku alase teadlikkuse tõstmise kampaaniate käigus.
- Tuleks välja valida andmete linkimiseks suurema potentsiaaliga andmekogud ning aidata ja motiveerida nende omanikke andmete avamisel ja linkandmete ettevalmistamisel. Samuti on

³⁸⁸ http://www.w3.org/2012/06/pmod/pmod2012_submission_7.pdf

³⁸⁹ A. M. G. Zuiderwijk, et al., "The necessity of metadata for open linked data and its contribution to policy analyses," presented at the Conference on E-Democracy and Open Government (CeDEM12), Krems, Austria, 2012

³⁹⁰ A. Zuiderwijk, et al., "Open Data Policies: Impediments and Challenges" presented at the 12th European Conference on eGovernment – ECEG 2012, Barcelona, Spain, Forthcoming

³⁹¹ http://events.linkedata.org/ldow2010/papers/ldow2010_paper14.pdf

suurema mõjuga selliste andmete avamine või linkimine, mis on ka teistes riikides avatud või lingitud.

7.1.7 Linkandmete kasutuselevõttuga kaasnevad võimalikud ohud ja riskid

Millised on linkandmete kasutuselevõttuga kaasnevad võimalikud ohud ja riskid, sh tavaelanike ja ettevõtete jaoks? Kuidas neid ennetada ja juhtida? Sealhulgas:

- kuidas tagada linkandmete kasutuselevõtu raames isikuandmete kaitse, kas ja mida selleks on täiendavalt vajalik teha?
- kuidas vähendada otsingu tulemustes rämpsandmete hulka (ebavajalikke, dubleerivaid, vananenud ja valesid andmeid)?

Linkandmete kasutuselevõttuga kaasnevaid võimalikke ohtusid ja riske on käsitletud käesoleva uuringu jaotistes 4.1.2 ja 4.4. Muuhulgas, linkandmete puhul võib piisata isikuandmetest ühe lingi küljes selleks, et tuvastada ülejäänud andmete seotus vastava isikuga; näiliselt ebaolulised isikulised detailid võivad kombineeritult anda isikust põhjaliku pildi; andmete linkimine võib laiendada võimalusi statistilise üksuse kaudselt tuvastamiseks; isikustatud andmed kombineeritult isikustamata andmetega võivad kaasa tuua isikuandmete kaitse nõuete riive; linkandmed võimaldavad seostada isikuandmeid avaliku sektori andmete ja muude allikate (meedia, sotsiaalmeedia jm) kombineerimisel; oht väärarusaamade, põhjendamatute hinnangute, inimesele suunatud vaenulike kampaaniate jms tekkeks on suurem. Seetõttu tuleb isikuandmete kaitseks tagada, et ükski avalik lingitud andmete komplekt ei sisaldaks isikuandmeid.

Ankeetides hinnati kõige suuremateks andmete vale interpreteerimise, privaatsuse vähenemise ja keerukate sõltuvuste ning andmete muutumise ohtusid. Ankeetküsitlus näitas ka, et avaliku sektori töötajad on rohkem mures selle pärast, et süsteemi loomise ja ülalpidamise kulukus on suurem kui loodav kasu; samuti hindavad nad kõrgemaks andmete ja protsesside muutumisest tulenevaid probleeme (erinevused avaliku ja erasektori gruppide vastuste keskmistes on statistiliselt olulised olulisusnivool $\alpha = 0.05$).

Õigusliku analüüsi jaotistes 4.4.3 ja 4.4.4 pakutud ettepanekutest isikuandmete kaitseks ja privaatsuse tagamiseks seoses linkandmetega on jaotise lõpus ettepanekute ja poliitikasoovitustena välja toodud olulisemad punktid.

Rämpsandmete (ebakorreksete andmete) vähendamiseks tulemustes saab kõige rohkem ära teha andmete haldaja ja avaldaja. Õigusliku analüüsi jaotises 4.5 pakutud meetmetest intellektuaalomandi kaitse tagamiseks, vastutuse omistamise sätestamiseks linkandmete kvaliteedi eest ning linkandmete uuendamise tagamiseks on jaotise lõpus välja toodud olulisemad punktid.

Lisaks õigusruumi täiendamisele nähtub eelnevalt linkandmeid rakendanud riikide kogemusest, et linkandmete loojad vajavad kvaliteetsete andmete loomiseks ja rämpsandmete osakaalu vähendamiseks konkreetseid eeskirjasid ja juhendamist³⁹². Nt Ühendkuningriikide ja Belgia kogemus (üleriigilise ja regionaalsete aadressiregistrite andmete ühendamine), FP7 (LOD) ja teised rahvusvahelised linkandmete projektid annavad hea sisendi parimate praktikate näol kvaliteetsete linkandmete loomise, haldamise (sh versioneerimise) jm eeskirjade jaoks. Üks võimalus linkandmete kvaliteedi parendamiseks on koosloome meetodite kasutamine, nagu näitavad EU 7 raamprogrammi projektide, sh LOD2 toel käivitatud eksperimendid³⁹³.

Ettepanekud ja poliitikasoovitused

- Täpsustada sellised isikuandmete kaitse aspektist küsitavad olukorrad, kus ühe õigusakti alusel saab isikukode või muid isikuandmeid avalikustada, AvTS alusel saab seda teavet taaskasutada ning

³⁹² http://events.linkedata.org/ldow2010/papers/ldow2010_paper14.pdf

³⁹³ http://jens-lehmann.org/files/2013/iswc_crowd_qa.pdf

isikuandmete kaitse seaduse alusel ei kohaldata tulemusele isikuandmete seadusega võimaldatavat kaitset (vt jaotis 4.4.4). Realiseerida andmekogusid sätestatavates õigusaktides täielikumalt IKS § 11 ja 12 pakutavad andmesubjekti poolse isikuandmete avalikustamise võimalused. Täpsustada, millal rakendub isikuandmete kaitse seaduse § 17 (automaatsed otsused).

- Rääpsandmete vähendamiseks kehtestada andmekogude põhimäärustes ja juhendmaterjalides avalikustatavate andmete kvaliteedi nõuded ja määratleda vastutus andmete uuendamise eest. Publitseerida nende nõuete ja nende vastavuse taseme hinnangud vastavas andmekogus või eha nad kättesaadavaks RIHA kaudu. Lisada avaliku teabe seadusesse säte, mis selgelt määratleb vastutuse väliste osapoolte töödeldud andmete ja nende osutatud teenuste kvaliteedi eest.
- Töötada välja eeskirjad ja juhendid abistamaks linkandmete loojaid kvaliteetsete andmete avaldamisel ja rääpsandmete osakaalu vähendamisel ning kasutada linkandmete kvaliteedi parendamiseks koosloome meetodeid.

7.1.8 Vajalikud muudatused õigusruumis

Kas ja kuidas peaks linkandmete kasutuselevõtuks muutma Eesti õigusruumi (sh eeltoodud küsimuste vastuseid arvestades)? Millised on hetkel olevad piirangud ning tulevikus reguleerimist vajada võivad teemad?

Linkandmete kasutuselevõtt hõlmab endas kokkuleppeid koosvõime neljal tasemel - tehnilisel, semantilisel, organisatoorsel ja õiguslikul tasemel. Muudatused õiguslikul tasemel peavad toetama ka ülejäänud tasemeid ja vastupidi. Näiteks uuring „Open data: an international comparison of strategies“³⁹⁴ toob välja nn suletud valitsemiskultuuri kui ühe barjäärdest andmete avalikuks tegemisel - nimelt ametnikul organisatsioonis on lihtsam andmeid mitte avada kui riskida sellest tulenevate sanktsioonidega. Taoliste barjäärde mahavõtmine eeldab muutusi nii organisatsioonides kui õigusruumis.

Ankeetide vastustes paigutusid linkandmete loomise ja kasutamise õiguslikud takistused pigem vähem olulisse pärssivate tegurite rühma.

Vajalikud muudatused õigusruumis, sealhulgas piirangud ja reguleerimist vajavad teemad, on toodud käesoleva uuringu jaotises 4 ning kokku võetud allpool.

Ettepanekud ja poliitikasoovitused

- Viia AvTS ning andmekogude aluseks olevad vastavusse taaskasutuse printsiipidega ja direktiividega 2003/98/EL ning 2013/37/EL. Kehtestada direktiivi 2013/37/EL järgimiseks vajalikud õigus- ja haldusnormid, muuhulgas täiendades RIHA määrust avaandmete korrastamisega seotud sätetega (vt jaotis 6).
- Vaadata läbi ja vajadusel uuendada riigi andmekogude kohta käivad õigusaktid, kus sätestatakse andmete väljastamise tasumäärad, eriti piirkulusid ületava tasu võtmise osas (vt jaotis 4.3.1).
- Sätestada avalikustatava teabe litsentseerimise kord ja litsentsitingimused või litsentsitüübid avaliku teabe seaduses, võimalusel soovitades seejuures kasutada sobivat litsentsi *creative commons* litsentsiperest. Kehtestada andmekogude põhimäärustes ja juhendmaterjalides avalikustatavate andmete kvaliteedi nõuded ning publitseerida nende nõuete ja nende vastavuse taseme hinnangud vastavas andmekogus või teha nad kättesaadavaks RIHA kaudu (vt jaotis 4.5.3).
- Sätestada konkreetsete andmekogude kohta käivates õigusaktides, võimalusel ka avaliku teabe seaduses (nt AvTS § 31, § 4 või § 8), millised on õigused päringute tulemusena saadavate andmete publitseerimiseks. Andmete (linkandmetena) publitseerimisel tuleks arvesse võtta riske (sh isikutele, omandile, seadmetele, kriitilisele taristule ja riigi julgeolekule), mis tekivad nende andmete

³⁹⁴ http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

kombineerimisel juba olemasolevate publitseeritud andmetega. Võimalik andmete publitseerimise kitsendamine selliste riskide tõttu peaks siiski olema pigem osaline (enamikes andmekogudes on andmeid, mida saab suurema riskita avalikustada), põhjalikult läbi analüüsitud ning detailsetl põhjendatud (vt jaotis 4.6.3).

7.1.9 Ettepanekute elluviimise korraldamine

Kuidas korraldada Lingitud Eesti suunas viivate ettepanekute elluviimine? Kas ja milliseid organisatsioonilisi (ümber)korraldusi selle realiseerimiseks on vaja teha, sh riiklikul tasandil? Kuidas leida vajalikud rahalised vahendid, sh kaasata erasektori ja vabakonna panust?

Uuringud toovad välja konkreetse poliitilise juhtfiguuri olulisuse kriitilise tegurina linkandmete ja avaandmete initsiatiivi õnnestumiseks³⁹⁵. Positiivseteks näideteks on USA-s president Obama poliitiline tugi avaandmetele ja eelmise Ühendkuningriikide peaministri Gordon Browne osalemine andmete avamise poliitikate loomisel. Hispaanias olid avaandmete poliitikate tšempioniteks regionaalsed poliitikutud. Mainitud aruanne, nagu ka UK avaliku sektori andmete linkimise kogemuste analüüs³⁹⁶ ei sisalda rohkem organisatsiooniliste (ümber)korralduste tüüpi soovitusi. Põhjus võib olla selles, et andmete avamine ja lingitaval kujul esitamine ei ole eesmärgid omaette, vaid pigem muude ülesannetega kaasnevad tegevused.

USA-s on asutatud e-teenuste innovatsioonikeskus (*Digital Services Innovation Center*), milles töötab 2014. a jaanuari seisuga kuus põhitöötajat³⁹⁷. Arvestades riikide elanike arvude suhet ei pea käesolev uuring otstarbekaks sellise struktuuri loomist Eesti valitsusasutuste juurde vaid selleks, et ellu viia Lingitud Eesti suunas viivaid ettepanekuid, välistamata sellise keskuse võimalikkust laiemate ülesannete täitmiseks. Välisriikide parimate praktikate õppetundidest lähtudes tuleks arendada suunatud strateegiat ja konsensuslikku tegevuskava, viia tegevuskava ametnikeni ning kontrollida selle täitmist, luua valitsuse juurde ava- ja linkandmete töögrupp ning teha rahvusvahelist koostööd ODI-ga (vt jaotis 2.4.4). Neid tegevusi saaks teha RISO, RIA, MKM infoühiskonna teenuste arendamise osakonna (ITAO) ja ministeeriumide vahenditega.

Muud Lingitud Eestit toetava organisatsioonilise koosvõime arendamise ettepanekud riiklikul, erasektori ja vabakonna tasemel koos rahaliste vahendite ja nende allikate hinnangutega on toodud käesoleva uuringu jaotistes 7.1.4 ja 7.3 ning võetud valikuliselt kokku allpool. Rahalised vahendid tuleks planeerida jooksvalt koos toodud ettepanekute elluviimisega.

Ettepanekud ja poliitikasoovitused

- Kuna linkandmete puhul on tegemist on tüüpilise innovatsiooniprotsessiga, mille puhul uue tehnoloogia kuluefektiivne rakendamine eeldab eelnevat mahukat investeerimist (jaotis 7.1.1), siis tuleks leida poliitiline tugi linkandmete ja avaandmete tehnoloogiate soodustamiseks. Sellise toe motiveerimiseks võib tugineda rahvusvahelisele praktikale (nt USA-s president Obama poliitiline tugi avaandmetele ja eelmise Ühendkuningriikide peaministri Gordon Browne osalemine andmete avamise poliitikate loomisel) ning jaotises 7.3.1 toodud avaliku sektori andmete avalikustamisest saadava võimaliku kasu suuruse kaudsetele hinnangutele.
- Eraldi struktuuri loomist Eesti valitsusasutuste juurde vaid selleks, et ellu viia Lingitud Eesti suunas viivaid ettepanekuid, ei pea käesolev uuring otstarbekaks, välistamata sellise keskuse võimalikkust laiemate ülesannete täitmiseks. Samas tuleks tähtsustada suunatud strateegiat ja konsensuslikku tegevuskava, viia tegevuskava ametnikeni ning kontrollida selle täitmist, luua valitsuse juurde ava- ja linkandmete töögrupp ning teha rahvusvahelist koostööd ODI-ga.
- Andmete avamisel ja lingitaval kujul esitamisel tuleks eristada kolme erinevat tüüpi tegevusi. Spetsiifilised tegevused ava- ja linkandmete loomiseks tuleks läbi viia nende asutuste organisatsiooni

³⁹⁵ http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

³⁹⁶ http://events.linkedata.org/ldow2010/papers/ldow2010_paper14.pdf

³⁹⁷ <http://gsablogs.gsa.gov/dsic/about/>

ja vahenditega, kes andmeid toodavad ja haldavad, vastavalt õigusaktidele. Eelnevaid tegevusi toetavad ettevõtmised, näiteks koolitused, tuleks finantseerida ja tellida keskselt, nt RIA, RISO või Riigikantselei poolt. Konkreetseid arendusprojekte võib läbi viia nii asutuste vahenditega kui ka keskselt; neid tuleks võimalusel teha laiemate projektide kontekstis, mis viivad asutuste või kodanike jaoks nähtavate kasulike tulemusteni.

- Era- ja vabasektori kaasamiseks tuleks kehtestada kõikidele tellitavatele infosüsteemidele koosvõimelisuse, avaandmete ja linkandmete osa kohustuslikkuse ning semantiliste klassifikaatorite ja URI-de olemasolu nõuded vastavalt jaotisele 7.2.4, propageerida edukaid linkandmete projekte nii era- kui ka avalikus sektoris ning arendada tarkvara linkandmete esitamiseks ja päringute tegemiseks vastavalt jaotisele 7.3.4 ning pakkuda Eesti vabasektori ja kodanikkonna kaasamiseks linkandmete alast teavitamist ja koolitamist, huvipakkuvaid rakendusi, suhteliselt lihtsaid tööriistu ja kättesaadavaid linkandmeid vastavalt jaotisele 7.3.5. Nende jaotiste, samuti intervjuude ja ankeetide põhjal ei saa prognoosida olulist era- ja vabasektori rahastuse kaasamist Lingitud Eesti suunas viivate ettepanekute elluviimisesse.

7.2 Lingitud Eesti koosvõime arhitektuuri arendamise ettepanekud

Jaotise põhiteema on semantiline ja tehniline koosvõime.

7.2.1 Linkandmetega seotud avatud standardid

Millised linkandmetega seotud avatud standardid on Eesti jaoks olulised (sh piisavalt küpsed kasutuselevõtuks)? Milline on iga soovitatava standardi puhul selle rakendamise saadav kasu ja rakendamiseks vajalikud ressursid (sh rahalised ressursid ehk maksumus)? Kuidas peaks standardite juurutamist (sh avalikus sektoris) korraldama, et nende kasutuselevõttu kiirendada?

Uuring „Standardiseerimise majanduslik mõju“³⁹⁸ toob välja, et ligi 25% iga-aastasest Prantsusmaa SKP kasvust tuleneb standardimisest. Uuringus viidatakse ka analoogilistele tulemustele teiste Euroopa tööstusriikide Saksamaa ning Ühendkuningriikide kohta. Lisaks otsestele tuludele suurendab standardimine ettevõtte väärtust, võimaldab kiiremat innovatsiooni, parandab vastavust regulatsioonidega, muudab tulemused rahvusvaheliselt kasutatavateks ning soodustab toodete ja teenuste kvaliteeti. (Rahvusvahelise) koostöö võimaldamine on standardimise üks olulisemaid tulemusi. Näiteks on Eestis kokku 24003 kehtivat standardit, neist vaid 275 ehk 1,1 % on algupärased Eesti standardid³⁹⁹ (8.01.2014 seisuga). Üaltoodud põhjustel on investeeringud standardite kasutuselevõttu suhteliselt suure tulususega.

Linkandmete ja semantilise veebi standardite kasutuselevõtmine ei tähendaks olulisi muutusi riigi infosüsteemi arhitektuuris – standardid RDF, RDFa, microdata, OWL, SA-WSDL, URI, URL, HTTP jt on riigi infosüsteemi koosvõime raamistik⁴⁰⁰ juba olemas ning piisavalt küpsed kasutuselevõtuks. Pigem on neist osade puhul probleemiks nende rakendamine praktikas, nagu ka linkandmete puhul üldiselt. Nendele probleemidele pakutakse lahendusi muude käesoleva analüüsi uurimisküsimuste vastustes, sh jaotistes 7.2.2, 7.2.5, 7.2.6.

Kuna linkandmete standardite rakendamine ei ole eesmärk omaette, tuleks seda korraldada seoses linkandmetega seotud projektide, koolituste ja muude algatustega. Ankeetide vastustes märgitakse, et avatud ja lingitud andmed ei ole väärtuslikud „asjana iseeneses“ ning mingi standardi eelistamine ei tähenda ilmtingimata rahalist säästu. Linkandmetega seotud projektide, sealhulgas ka standardi IPv6 rakendamise vajalikkust ja maksumuse hinnanguid käsitletakse jaotises 7.1.4.

Mitmed standardid sisaldavad asjade interneti, suurandmete, koosloome ja tulevikuinterneti teemalisi komponente. Näitena võib tuua Internet Engineering Task Force (IETF) poolt 1998. a avaldatud IPv6

³⁹⁸ <http://www.iso.org/iso/fr/home/standards/benefitsofstandards/benefits-detail.htm?emid=7>

³⁹⁹ <http://www.evs.ee/Standardimine/Eestistandarditearv/tabid/96/Default.aspx>

⁴⁰⁰ <http://www.riso.ee/et/koosvoime/raamistik>

standardi RFC 2460 ning ISO/IEC JTC 1/SC 17 töörühma dokumendi „*Collection of Standards Data related to Internet of Things*“, milles tuuakse ära 418 asjade Internetiga otseselt või kaudselt seotud standardit, soovitusi jne. Siiski pole nende teemadega seotud standardid praeguseks hetkeks piisavalt küpsed, et neid hakata Eestis ulatuslikult kasutusele võtma. Sellele viitab ka suurfirmade AT&T, Cisco, General Electric, IBM, Intel jt konsortsiumi hiljutine kava luua asjade interneti standardid⁴⁰¹.

Ettepanekud ja poliitikasoovitused

- Jätkata tööd juba riigi infosüsteemi koosvõime raamistikus esitatud ja kasutuselevõtuks piisavalt küpsede standardite juurutamisega, nagu RDF, RDFa, *microdata*, OWL, RDF (S), SA-WSDL, URI, URL, HTTP jt.
- Kasutada ava- ja linkandmete publitseerimisel W3C linkandmete publitseerimise parimaid praktikaid.
- Rakendada linkandmete standardeid seoses linkandmetega seotud projektide, koolituste ja muude algatustega.

7.2.2 Tehnilise infrastruktuuri lahendused

Kas Eesti peaks looma juurde või täiendama mingeid tehnilise infrastruktuuri lahendusi (nt avaandmete X-tee või avaandmete portaali arendus, andmeanalüüsi töövahendid vm), et toetada linkandmete kasutuselevõttu ja kogu elutsükli protsessi?

W3C konsortsium esitab ülevaate linkandmete elutsüklitest⁴⁰², sealhulgas 7nda raamprogrammi projekti LOD2 raames välja pakutud elutsükli LOD2 Stack⁴⁰³. Enamiku elutsükli sammude läbimiseks pole vaja statsionaarsete infrastruktuurilahenduste olemasolu. Samas osade sammude jaoks, nagu näiteks andmete kogumine, andmete sirvimine/otsimine ja päringute tegemine, on mõistlik luua toetavaid lahendusi. Vaatame kolme suurt teemat: linkandmete tootmine, linkandmete kasutamine ja linkandmete varamu.

Linkandmete tootmiseks on mõistlik võimalusel ära kasutada olemasoleva riigi infosüsteemi arhitektuuri poolt pakutavaid võimalusi. Riikliku Ehisregistri linkandmete pilootprojektis kasutati näiteks X-tee andmeteenuseid linkandmete komplekti ehitamiseks. Kuna X-tee kaudu on riigi infosüsteemide andmed kättesaadavad, on X-tee kasutamine linkandmete tootmiseks igati otstarbekas. Selline kasutamine võimaldab paremini korraldada ka linkandmete uuendamist.

Seejuures tekib küsimus, kui tihe peaks olema linkandmete tootmise keskkonna ja X-tee integratsioon. Vastuste skaala ühes otsas on X-tee oma senisel kujul, mida kasutatakse vaid andmete kättesaamiseks; viide X-teele sellise lahenduse võimalikus nimetuses (nt „avaandmete X-tee“) on pigem illustratiivne. Skaala teises otsas on ümberehitatud X-tee, mille erinevatel tasemetel on kasutajatel erinevad turvanõuded ja juurdepääs andmetele; viide X-teele sellise lahenduse võimalikus nimetuses on sisuline.

Esimese variandi eelisteks on süsteemide selge piiritlemine, arvatavalt väiksem uuenduste maht ja maksumus ning väiksemad turvariskid, eriti X-tee kasutamise osas. Puuduseks võib olla andmete publitseerija väiksem tegutsemise mugavus. Andmete kasutajate jaoks olulist vahet pole, samas lingitud andmeid luuakse eelkõige kasutajate, mitte niivõrd publitseerijate jaoks. Teise variandi eelised ja puudused on vastupidised. Kaalukas argument teise variandi kasuks võib olla see, kui linkandmetele on vaja lisada tõestusväärtust või garanteerida autoriseeritud ligipääs. Oleme arvamisel, et minimaalne vajalik integratsioon X-teega (esimene variandi variatsioonid, kus võib-olla saab kasutada X-tee komponente, kuid linkandmete taristu on X-teest lahus) minimeerib riske ja kulusid ning et seda ei kaalu üles teise variandi eelised. Detailne analüüs ja lõpliku arhitektuuri valik ei kuulu käesoleva uuringu raamidesse.

⁴⁰¹ http://bits.blogs.nytimes.com/2014/03/27/consortium-wants-standards-for-internet-of-things/?_php=true&_type=blogs&hp&r=0

⁴⁰² http://www.w3.org/2011/gld/wiki/GLD_Life_cycle

⁴⁰³ <http://stack.lod2.eu/>

Linkandmete kasutamise osas näitab UK kogemus, et linkandmeid kasutavate arendajate jaoks on RDF andmekirjelduskeel ja SPARQL päringud ebamugavad ning nad eelistavad lihtsamaid variante⁴⁰⁴. Analoožilisi seisukohti on toodud ka Eesti avaandmete portaali edasise arengu ettepanekutes⁴⁰⁵. Seda arvestades võib soovitada RESTful APIdel põhineva vahekihi loomist, mis kasutab SPARQL päringuid ning tagastab linkandmete JSON või XML vaated, võimaldab teha URI põhiseid otsinguid ning toetab andmete publitseerijaid valdkondlike APIde loomisel.

Kui uurida võimalusi piirdumaks vaid JSON keelega, tuleks silmas pidada, et XML on praeguseks arendajate kogukonna poolt hästi omaks võetud.

Tuleks edasi arendada Eesti avaandmete portaali opendata.riik.ee ning avatud linkandmete portaali opendata.riik.ee/en/linked-data, seostades neid teiste andmeallikatega, nagu Eesti geoportaal⁴⁰⁶, Maa-ameti geoportaal⁴⁰⁷, Keskkonnaagentuuri portaali⁴⁰⁸ ning Statistikaameti andmebaas⁴⁰⁹. Kuna paljud avaliku sektori asutuste andmed on portaalis registreerimata, tuleks käivitada projekt Eesti avaliku sektori avaandmete loetelu loomiseks ning publitseerimiseks.

Ankeetide vastustes paigutus riigi infosüsteemide üldine hetkeseis, sh aegunud arhitektuur, keskmise olulisusega pärssivate tegurite rühma.

Kokkuvõttes, lähtudes Eesti olemasolevast IT infrastruktuurist ning rahvusvahelisest kogemusest, on mõistlik kasutada X-tee linkandmete publitseerimisel toorandmete allikana, pakkuda andmete kasutajale andmeid harjumuslikus andmekirjelduskeeles üle kasutaja pöördumiste jaoks suhteliselt lihtsa vahekihi ning arendada edasi avatud linkandmete portaali.

Ettepanekud ja poliitikasoovitused

- Kasutada X-tee jaoks loodud andmeteenuseid linkandmete publitseerimisel toorandmete allikana, võimaldades andmete pakkujatel esitada andmeid harjumuslikus andmekirjelduskeeles ja taaskasutades olemasolevaid andmeteenuseid üle suhteliselt lihtsa vahekihi.
- Täpsustada andmeteenuste disainimise metoodikat ja teha selle jälgimine kohustuslikuks, läbi Riigi infosüsteemi haldussüsteemi määruse, et uute andmeteenuste loomisel tekiks linkandmete loomiseks vajalikud andmeteenused ning samal ajal pareneks andmeteenuste taaskasutus.
- Luua avatud linkandmete tootmiseks infrastruktuur, kus saaks avaandmeid pakkuvaid X-tee jaoks loodud andmeteenuseid rakendada, sh linkandmete tootmiseks, ilma turvaservereid kasutamata.
- Arendada edasi avatud linkandmete portaali andmete kogumise, linkimise ja haldamise automatiseerimiseks kasutades andmeteenuseid. Kuna paljud avaliku sektori asutuste andmed on portaalis registreerimata, tuleks käivitada projekt esmalt Eesti avaliku sektori avaandmete loetelu loomiseks, seejärel publitseerimiseks, ning seejärel nende linkimiseks.
- Tellida ideekonkursi kaudu avatud tehnilise pilveplatvormi (PaaS) loomine ja haldamine linkandmete kasutamise lihtsustamiseks tehnoloogilistes pilootides (nt linkandmete loomine ja haldamine koosloome kaudu, linkandmete reaalajaanalüütika infoühiskonna arengukava meetrikate pidevaks mõõtmiseks, valdkonnaülene linkandmete kasutamine e-äri lahendustes, ...) ja mis uuendaks linkandmeid linkandmete portaalist ning oleks avatud linkandmetele, funktsioonidele ja linkandmete rakendustele. Platvorm peab toetama valdkondlike REST teenuste loomist andmete publitseerijate, firmade või kogukonna poolt nii kommertseesmärkidel kui tasuta kasutamiseks.

⁴⁰⁴ http://ceur-ws.org/Vol-628/ldow2010_paper14.pdf

⁴⁰⁵ <http://opendata.riik.ee/juhendid/ettepanekud-edasisteks-arendusteks>

⁴⁰⁶ <http://inspire.maaamet.ee/>

⁴⁰⁷ <http://geoportaal.maaamet.ee>

⁴⁰⁸ <http://www.keskkonnainfo.ee/>

⁴⁰⁹ <http://pub.stat.ee/px-web.2001/dialog/statfile2.asp>

- Avatud tehnilise platvormi jätkusuutlikkuse tagamiseks on vaja eraldi eelarvelised vahendid selle organisatsiooni ülalpidamiseks. Organisatsiooni ülesandeks oleks linkandmete metamudelite loomise tehniline juhendamine, andmete linkimise tehnilise toe pakkumine, linkandmete kasutamise praktiliste juhiste väljatöötamine ja platvormi tehniline haldamine/töõshoidmine.

7.2.3 Koosloomel põhinevate lahenduste loomine

Kas ja kuidas korraldada Eestis linkandmete kasutuselevõttuga seotud koosloomel põhinevate lahenduste loomist?

Koosloomel on palju määratlusi. Lihtsaim neist ütleb, et koosloome on probleemilahenduse ja tootmise internetipõhine hajutatud mudel⁴¹⁰. Riigi seisukohast on koosloome oluline instrument kodanikuühiskonna kujundamisel, kuna see lihtsustab indiviidide panustamist suurte probleemide lahendamisse. Koosloomel on rakendusi nii üleriigilistes teenustes, kus vajadus teenuse järele on olemas laiemas plaanis, kui väikeste sihtgruppide kontekstis, kus vajadus tekib seismiste ajendite toimet. Eestis hästi tuntud koosloome näide on „Teeme ära!“ koristustalgud⁴¹¹, mis on tänapäevaks jõudnud enam kui sajasse riiki⁴¹².

Koosloome elluviimiseks on kriitilise tähtsusega mitmeotstarbelise koosloome platvormi (sh liitunud invidiidid) loomine, mis lihtsustab isikute kaasamist erinevatesse initsiatiividesse. Sellised platvormid eeldavad ühiste mõistete ja probleemihalduse kasutamist, mida saab korraldada nii linkandmete kasutamisega kui ka ilma selleta.

Riik saab läbi viia pilootprojektid koosloome platvormi ja selle näidisrakenduste loomiseks kas siis üldiste konkurssidena või konkreetsete teenuste moderniseerimise eesmärgil (nt seotult julgeoleku või keskkonna seirega). Selliste pilootprojektide raames tuleks üheks eesmärgiks seada sellise platvormi loomine, mida saaks kasutada ka väljaspool avaliku sektori teenuste pakkumist. See lihtsustaks tunduvalt uute koosloome lahenduste loomist.

Ankeetide vastustes hinnati koosloome kasutamise tegevuste mõju Eestile keskmiselt väiksemaks muude projektide ja algatuste mõjust.

Selliste konkursside algatamisel tuleks tugineda vähemalt ühele sisukale rakendusele, kus linkandmete kasutamine annaks koosloome protsessile lisaväärtust. Kuna koosloome eelduseks olevate ühiste mõistete ja probleemihalduse kasutamist saab paljudel juhtudel realiseerida ka ilma linkandmeteta, pole sellise rakenduse leidmine triviaalne. Näidetena võib tuua koosloome kasutamist linkandmete haldamisel^{413, 414}, sealhulgas linkandmete kvaliteedi parendamisel⁴¹⁵, aga ka linkandmete kasutamist hädaolukordade haldamisel⁴¹⁶ jne. Analoogilised rakendused võivad olla kasulikud ka Eestis.

Ettepanekud ja poliitikasoovitused

- Viia läbi pilootprojektid koosloome platvormi ja selle näidisrakenduste loomiseks kas siis üldiste konkurssidena või konkreetsete teenuste moderniseerimise eesmärgil (nt seotult julgeoleku või keskkonna seirega). Eesmärgiks tuleks seada sellise platvormi loomise, mida saaks kasutada ka väljaspool avaliku sektori teenuste pakkumist, tuginedes vähemalt ühele sisukale rakendusele, kus linkandmete kasutamine annaks koosloome protsessile lisaväärtust. Pilootprojektis tuleks eelistada suurema mõjuga koosloome projekte, millel on pikem ajalugu ja väljakujunenud koosloome mallid.
- Eesti valitsus võiks koondada ava- ja linkandmete kasutajate grupid, kes kasutades koosloome platvormi annaksid riigile nõu, milliseid andmeid avada ja linkida.

⁴¹⁰ http://www.clickadviser.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf

⁴¹¹ <http://www.teemeara.ee/et>

⁴¹² www.letsdoitworld.org

⁴¹³ http://ceur-ws.org/Vol-782/SimperlEtAl_COLD2011.pdf

⁴¹⁴ <http://www.aifb.kit.edu/web/Inproceedings3284>

⁴¹⁵ http://jens-lehmann.org/files/2013/iswc_crowd_qa.pdf

⁴¹⁶ http://ifgi.uni-muenster.de/~j_ortmo2/publications/OrtmannetalTC2011.pdf

7.2.4 Vajalikud semantikaressursid

Millised semantikaressursse Eesti vajab linkandmete tehnoloogia rakendamiseks?

Linkandmete efektiivne kasutamine eeldab selgeid andmete metamudeleid. Jaotises 2.1.2 on toodud põhilised linkandmetega seonduvad kasutusküpsed tehnoloogiad, mida on esmajoones vaja linkandmete rakendamiseks: ressursside identifikaatorid (*Uniform Resource Identifiers-URI*), ressursside kirjeldamiseks kasutatavad keeled (*The Resource Description Framework-RDF*, *RDF/XML* ja *RDF Schema-RDFS*) ning ontoloogiate semantilise kirjeldamise keel *Web Ontology Language-OWL*. RDF on graafipõhine andmete esitus, mis on defineeritud W3C poolt koos SPARQL⁴¹⁷ päringukeele ja ontoloogiate kirjelduskeele OWL-ga moodustamaks aluse semantilisele andmete veebile.

Riigi infosüsteemide semantilise koosvõime raamistikus on viimaste esitamiseks ette nähtud valdkonnasõnastikud, klassifikaatorid, ontoloogiad, nendega seotud veebiteenuste semantilised kirjeldused ja muud semantilised varad, mis tuleb publitseerida RIHA-s⁴¹⁸.

RIHA-s (kõik andmed on esitatud seisuga 24.01.2014) on 74 erinevat valdkonda, neist 14-l on olemas valdkonnasõnastikud. Klassifikaatoreid on 747; ühel klassifikaatoril võib olla arvukalt erinevaid versioone, näiteks on Eesti haldus- ja asustusjaotuse klassifikaatoril 16 erinevat versiooni. Arhiiviväärtust omavaid andmeobjekte on 4467, sealhulgas näiteks 10 andmeobjekti nimetusega „EESNIMI“ või „eesnimi“ ning üle 200 andmeobjekti nimetusega „ID“ või „id“.

Kvalitatiivuuringu tulemuses on märgitud, et olemasolevad andmed ei ole sisuliselt lingitavad, Eesti ja rahvusvahelised klassifikaatorid ei ole alati ühilduvad, tekivad probleemid ajalooliste andmekogudega ning samu mõisteid kasutades võib asutuste vahel andmeid linkides selguda, et mõistete taga olev sisu ei ühti täielikult. Analoogilisi probleeme on teistes riikides, näiteks Soomes⁴¹⁹.

Ankeetide vastustes hinnati Eesti tänast kompetentsi semantiliste tehnoloogiate vallas pigem madalamaks võrreldes teiste uuritavate tehnoloogiatega.

Eelnevast nähtub, et linkandmete juures kasutamiseks on vaja olemasolevaid semantilisi varasid täiendada, ühtlustada ja/või kooskõlastada. Eriti oluline on see valdkonnaülest rakenduste puhul. Edasine semantiliste varade loomine ja korrastamine võiks toimuda koos ava- ja linkandmete loomisega.

Linkandmete kontekstis on võimalik kasutada ka lingvistilisi ontoloogiaid nagu WordNet jt. WordNet on juba RDF/OWL kujul (ka Eesti WordNet). Siiski, üldise andmete lingitavuse saavutamiseks pole Eesti WordNeti kasutamine eriti otstarbekas, mis ei välista selle kasutamist mingiteks spetsiifilisteks linkandmete rakendusteks.

Lisaks olemasolevate semantiliste varade täiendamisele vastavalt kehtivale semantilise koosvõime raamistikule tuleks seda raamistikku linkandmete juures kasutamiseks edasi arendada. Lähtudes linkandmete loomise primatest praktikatest^{420,421} tuleks linkandmete publitseerimiseks muuhulgas:

- valida ja soovitada sobivad sõnastikud;
- luua metoodika sobivate linkandmete URI-de loomiseks^{422,423,424};
- pakkuda soovitusi andmete lähtekehtade (*provenance*) kirjeldamiseks⁴²⁵;

⁴¹⁷ <http://www.w3.org/TR/rdf-sparql-query/>

⁴¹⁸ http://www.riso.ee/sites/default/files/elfinder/article_files/RISsemantikaVo7-lopplik.pdf

⁴¹⁹ <http://www.seco.tkk.fi/events/2014/2014-01-24-ldf/ap/frosterus-koko.pdf>

⁴²⁰ <http://www.w3.org/2011/gld/charter>

⁴²¹ http://www.w3.org/2011/gld/wiki/Best_Practices_Discussion_Summary

⁴²² <http://www.w3.org/TR/cooluris/>

⁴²³ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-URI-sets-uk-public-sector.pdf

⁴²⁴ <http://data.gov.uk/resources/uris>

- kehtestada versioneerimise põhimõtted.

Seda on otstarbekas teha olemasoleva semantilise koosvõime raamistiku edasise arendamise raames. Seega tuleks luua semantilise koosvõime raamistiku järgmine, täiendatud ja uuendatud versioon seoses raamistiku rakendamisel saadud senise kogemusega ning uute tehnoloogiasuundade ja parimate praktikate arenguga.

Ettepanekud ja poliitikasoovitused

- Linkandmete kasutamiseks tuleks olemasolevaid semantilisi varasid täiendada, ühtlustada ja/või kooskõlastada, eriti valdkonnaüleste rakenduste puhul. Lisaks olemasolevate semantiliste varade täiendamisele vastavalt kehtivale semantilise koosvõime raamistikule tuleb kaasajastada semantilise koosvõime raamistiku järgmist versiooni linkandmete ja seotud tehnoloogiate vaatevinklist.
- Esmajoones tuleks keskenduda juriidilisi ja füüsilisi isikuid, aadresse, kinnistuid ning ruumiandmeid kirjeldavate ontoloogiate loomisele, sest need andmed on esindatud enim riigi infosüsteemides, neil on suurim ühisosa riigi infosüsteemide lõikes ja nad on eelduseks andmete linkimisele. Seoses linkandmete kasutamisega peaks nende üleilmse lingitavuse parandamiseks võtma kasutusele nii palju standardontoloogiaid kui võimalik. Ka Eesti riigi infosüsteemi ontoloogiaid tuleks linkandmete loomise aspektist üle vaadata, eriti just avalike linkandmete loomise nõuetest lähtuvalt. Lisaks tuleks luua rohkem valdkonnaüleseid ontoloogiaid.
- Kuigi RIHA määruses on öeldud, et kõik uued infosüsteemid tuleb varustada semantiliste kirjeldustega, pole selge, kas see on nii läinud ja kas toimub vastav järelevalve. Seda küsimust tuleks eraldi uurida näiteks auditi raames.

7.2.5 Linkandmete tehnoloogiad andmete avaandmetena publitseerimisel

Kuidas peaks registrites/ infosüsteemides/ andmekogudes (sh andmete avaandmetena publitseerimisel) linkandmete tehnoloogiaid kasutusele võtma?

Esitatud küsimusel on järgmised aspektid:

- õiguslikke nõudeid avaandmetele/linkandmetele, sealhulgas nende publitseerimisele ja haldamisele, on analüüsitud jaotises 0, eriti alajaotistes 4.2 ja 4.5;
- soovitatavaid linkandmetega seotud standardeid on käsitletud jaotises 7.2.1;
- soovitatavaid linkandmete kasutuselevõttu toetavaid tehnilise infrastruktuuri lahendusi on analüüsitud jaotises 7.2.2;
- linkandmete kasutuselevõtuks soovitatavaid semantikaressursse on pakutud jaotises 7.2.4;
- võimalikke arhitektuurilisi lahendusi on käsitletud jaotises 3.

Kuna linkandmete üks põhilisi ideid on võimaluse andmine kolmandatele osapooltele uute teenuste ja rakenduste loomiseks, siis suur osa linkandmete tehnoloogiate kasutuselevõtu raskusest langeb kolmandatele osapooltele. Ankeetides hinnati, et avaandmete rakenduste loomine kuulub projektide rühma, millel on Eestile kõige suurem mõju.

⁴²⁵ http://www.w3.org/2011/gld/wiki/228_Best_Practices_Pragmatic_Provenance

Konkreetsete registrite/infosüsteemide/andmekogude andmete haldaja tegevused sõltuvad sellest, kuidas üaltoodud soovitusel on rakendatud. Kindlasti tuleks jälgida 1. osas viidatud W3C linkandmete publitseerimise parimaid praktikaid⁴²⁶:

1. modelleeri andmed;
2. nimeta asju püsivate URI-dega;
3. taaskasuta sõnastikke niipalju kui võimalik;
4. avalda andmete kohta nii inim- kui masinloetavad kirjeldused (metaandmed);
5. teisenda andmed RDF-i;
6. määra sobiv litsents;
7. avalikusta linkandmed ja kuuluta nad välja.

Arvestades konkreetset Eesti olukorda, võiks linkandmete tehnoloogia kasutuselevõtt andmete publitseerimisel sisaldada minimaalselt järgnevat:

- Kasutatakse üaltoodud W3C linkandmete publitseerimise parimaid praktikaid.
- Andmed publitseeritakse vastavalt kehtivatele õigusaktidele, masinloetaval kujul ja avatud litsentsiga. Sätestatakse vastutused andmete kvaliteedi ning nende haldamise osas (vt jaotis 0, sh 4.2 ja 4.5).
- Järgitakse riigi infosüsteemi koosvõime raamistik⁴²⁷ toodud standardeid andmete ja metaandmete publitseerimiseks (vt jaotis 7.2.1).
- Kasutatakse valitud linkandmete kasutuselevõttu toetavaid tehnilise infrastruktuuri lahendusi, soovituslikult kasutades X-teed linkandmete publitseerimisel andmete allikana ning pakkudes andmete kasutajale andmeid harjumuslikus andmekirjelduskeeles üle kasutaja pöördumiste jaoks suhteliselt lihtsa vahekihi (vt jaotis 7.2.2).
- Kasutatakse sobivaid semantikaressursse andmete metamudelite kirjeldamiseks vastavalt semantilise koosvõime raamistikule, vajadusel olemasolevaid kirjeldusi täiendades ja korrastades (vt jaotis 7.2.4).

Ettepanekud ja poliitikasoovitused

- Registrites, infosüsteemides ja andmekogudes (sh andmete avaandmetena publitseerimisel) tuleks linkandmete tehnoloogiatega kasutusele võtmiseks jälgida W3C linkandmete publitseerimise parimaid praktikaid: modelleerida andmed, nimetada asju püsivate URI-dega, taaskasutada sõnastikke niipalju kui võimalik, avaldada avaandmete kohta nii inim- kui masinloetavad kirjeldused (metaandmed), teisendada andmed RDF-i, määrata sobiv litsents, avalikustada avaandmed linkandmetena ja kuulutada nad välja. Mitteavalike andmete puhul on mõistlik eelkõige lingitavana hoida see osa andmetest, mille juures on näha vajadust kombineerida neid andmetega välistest allikatest.
- Andmed tuleks publitseerida vastavalt kehtivatele õigusaktidele, masinloetaval kujul ja avatud litsentsiga, sätestades vastutused andmete kvaliteedi ning nende haldamise osas (vt jaotised 7.1.7 ja 7.1.8), järgides riigi infosüsteemi koosvõime raamistik⁴²⁷ toodud standardeid andmete ja metaandmete publitseerimiseks (vt jaotis 7.2.1), kasutades valitud linkandmete kasutuselevõttu toetavaid tehnilise infrastruktuuri lahendusi ning soovituslikult pakkudes andmete kasutajale

⁴²⁶ http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

⁴²⁷ <http://www.riso.ee/et/koosvoime/raamistik>

andmeid harjumuslikus andmekirjelduskeeles üle kasutaja pöördumiste jaoks suhteliselt lihtsa vahekihi (vt jaotis 7.2.2), kasutades sobivaid semantikaressursse andmete metamudelite kirjeldamiseks vastavalt semantilise koosvõime raamistikule, vajadusel olemasolevaid kirjeldusi täiendades ja korrastades (vt jaotis 7.2.4).

7.2.6 Sisutöötlus veebisaitides ja -liidestest

Kuidas korraldada ümber sisutöötlus veebisaitides ja -liidestest, et toetada linkandmete kasutuselevõttu?

Linkandmete kasutuselevõtt veebisaitides ja -liidestest peaks algama uute vajaduste selgitamisest, mida linkandmete kasutuselevõtt rahuldab, eelistest, mida linkandmed seejuures pakuvad ning linkandmete kasutuselevõtu tulu-kulu analüüsist. Eriti tuleks olemasolevate veebisaitide puhul hinnata, kas nende ümberkorraldamise ja linkandmete kasutuselevõtu positiivsed efektid ületavad tehtavaid kulutusi.

Olemasolevat sisu saab lingitud kujule viia erinevate vahendite abil. Näiteks saab metaandmete lisamiseks kasutada RDFa⁴²⁸ (*Resource Description Framework in Attributes*), mikroandmete⁴²⁹ (*Microdata*), *microformats*⁴³⁰ ja teisi laiendusi HTML dokumentidele. Valik vahendi vahel sõltub paljudest teguritest; ühe kriteeriumina on otstarbekas kasutada esitust, mille puhul mitmete valdkondade jaoks on loodud vastavad metamärgendite komplektid. Nii on mikroandmete jaoks loodud üldised märgendite komplektid, nagu schema.org⁴³¹, mida laiendatakse uute valdkondadega nagu näiteks e-äri komplekt GoodRelations⁴³². Olemasolevast sisust linkandmete tootmise kohta leidub kirjanduses mitmeid praktilisi juhtumiuuringuid⁴³³. Sisutöötluse ümberkorraldamine saitides tähendaks vastavate märgendite kasutuselevõttu ning sobivate agregaatrite kasutuselevõttu portaalides, nt eesti.ee-s, kus seda sisu kokku kogutakse.

Vastavalt semantilise koosvõime raamistikule kirjeldatakse andmeteenuste liideseid RIHA-s juba praegu semantiliselt. Linkandmete osas tuleb nende semantilise kirjeldamise skeemid üle vaadata, et andmeteenused saaks linkida andmetega. Linkandmete kasutuselevõtul tuleks joondada sisutöötlus LOD2 Stack poolt pakutud välja linkandmete elutsükliga. Tuleks järgida veebide koosvõime raamistiku soovitusi ja jaotises 7.2.5 toodud W3C linkandmete publitseerimise parimaid praktikaid.

Veebis koos inimloetava sisuga publitseeritud linkandmetest on otsene kasu mitmetes rakendustes, mida seni nii pragmaatilistel kui organisatoorsest kaalutlustest pole täide viidud. Üheks linkandmete kasutamise rakenduseks sisutöötluse vaatevinklist oleks KOV-ide ja asutuste veebidest sisu, nt avalike kontaktide loendi ja uudiste, automaatne kokkukogumine ja esitamine suuremates portaalides, näiteks eesti.ee. Sellisel juhul on vaja tagada, et suuremate portaalide agregaatrite poolt aktsepteeritud märgendeid ka sisuliselt kasutataks sisu tootjate juures, mis omakorda eeldab vastavaid kokkuleppeid osapoolte vahel.

Ülaltoodut tuleks silmas pidada ka uute veebisaitide loomisel. Kui lisakulutused linkandmete kasutuselevõtuks on väikesed võrreldes kogu arenduse maksumusega, on uute rakenduste korral lisaargumendiks ka võimalikud tulevased linkandmete rakendused. Lihtsamatest taolistest rakendustest tuleks alustada võimalikult vara, luues nt eesti.ee keskkonda avaliku sektori kontaktide linkandmetena kokkukogumise funktsionaalsuse.

Ettepanekud ja poliitikasoovitused

Linkandmete kasutuselevõtt veebisaitides ja -liidestest peaks algama RDFa, mikroandmete või *microformats* märgendite lisamisest HTML dokumentidele. Konkreetne märgendite keel ja sõnavara valitakse pragmaatilistest kaalutlustest lähtuvalt. Avaliku sektori veebides on esmajärjekorras oluline märgendada organisatsiooni töötajate kontaktid ja sündmused. Suuremate organisatsioonide puhul on mõistlik selleks laiendada olemasoleva sisuhaldussüsteemi mooduleid, mis töötajate kontakte ja

⁴²⁸ <http://www.w3.org/TR/xhtml-rdfa-primer/>

⁴²⁹ <http://www.whatwg.org/specs/web-apps/current-work/multipage/microdata.html>

⁴³⁰ <http://microformats.org/about>

⁴³¹ <http://schema.org/docs/gs.html>

⁴³² <http://blog.schema.org/2012/11/good-relations-and-schemaorg.html>

⁴³³ http://events.linkedata.org/ldow2010/papers/ldow2010_paper01.pdf

uudiseid veebi väljastavad, selliselt, et väljastatud andmetele lisatakse märgendid automaatselt. KOVide puhul annaks suurima mõju KOVTP lahendusele analoogse toe lisamine. Otstarbekas on kasutada esitusi, mille puhul mitmete valdkondade jaoks on loodud vastavad metamärgendite komplektid, nt mikroandmete jaoks loodud <http://schema.org>, <http://purl.org/goodrelations> jt. Linkandmete kasutuselevõtul tuleks joondada sisutöötlus LOD2 Stack poolt pakutud linkandmete elutsükliga ning järgida veebide koosvõime raamistiku soovitusi ja jaotises 7.2.5 toodud W3C linkandmete publitseerimise parimaid praktikaid. Üheks linkandmete kasutamise rakenduseks sisutöötluse vaatevinklist oleks KOV-ide ja asutuste veebidest sisu automaatne kokkukogumine ja esitamine suuremates portaalides, näiteks eesti.ee.

- Kirjeldatud nõuete ja tehnoloogiate rakendamiseks tuleks olemasolevate veebisaitide puhul käivitada linkandmete kasutuselevõttu toetavad projektid, mis võivad olla iseseisvad või seotud linkandmete kasutuselevõtu pilootprojektidega (vt jaotis 7.1.4). Uute veebisaitide projektide hangetesse tuleks lülitada linkandmete kasutuselevõttu toetavad sisutöötluse nõuded (vt jaotis □), muuhulgas seoses metaandmete lisamise, semantilise kirjeldamise ja andmete agregeerimisega.

7.3 Lingitud Eestit toetava organisatsioonilise koosvõime arendamise ettepanekud

Järgnevas käsitletakse eelkõige organisatsioonilist koosvõimet. Tuuakse ära ka esialgne tulu-kulu analüüs, iseloomustamaks soovitude elluviimise jätkusuutlikkust. Ettepanekute tegemisel lähtutakse võimaluse piires juba olemasolevatest Eesti ja rahvusvahelistest organisatsioonidest ja protsessidest.

7.3.1 Linkandmete tehnoloogiate monitoorimine, õpetamine ja propageerimine

Mida peaksid tegema Eesti akadeemilised ringkonnad ja koolituste korraldajad linkandmete tehnoloogiate monitoorimiseks, õpetamiseks ja propageerimiseks?

Linkandmeid on seni Eesti ülikoolides käsitletud väheste kursuste ja tudengiprojektide raames õppejõudude initsiatiivil. Seda initsiatiivi tuleks võimendada läbi temaatiliste õppe-, koolitus- ja teadusprojektide, mille raames kaasataks nii eksperte välismaalt kui kohalikke tudengeid. Muuhulgas on vaja ülikoolide õppekavasid uute trendide valguses regulaarselt kaasajastada. Sel juhul peaks hoolitsemise vahendite eest parimate praktikute importimiseks Eestisse, olemasoleva tööjõu koolitamise ja õppematerjalide väljatöötamiseks. Sama kehtib ka muude käesolevas uuringus käsitletud suundumuste kohta, nt semantiline veeb ja suurandmed.

Ankeedivastustes keskmiselt hinnati linkandmete alast vähest teadlikkust kõige olulisema linkandmete loomist ja kasutamist pärssiva tegurina Eestis. Linkandmete ja teiste seonduvate tehnoloogiate alaste materjalide loomise, teadlikkuse suurendamise ja koolituste pakkumise/laiendamise mõju hinnangud olid positiivsed, kuid kuulusid pigem keskmiste või väiksemate hinnangute rühma. Erasektori töötajad hindasid ankeetides linkandmete alast vähest teadlikkust olulisema pärssiva tegurina kui avaliku sektori vastajad (erinevused avaliku ja erasektori gruppide vastuste keskmistes on statistiliselt olulised olulisusnivool $\alpha = 0.05$).

Linkandmete jaoks on pakutud arvukalt juhend- ja õppematerjale, sealhulgas nii W3C konsortsiumi⁴³⁴, Euroopa Komisjoni projektide (nt FP7 projekt EUCLID⁴³⁵) kui ka muude allikate poolt⁴³⁶. See muudab lihtsamaks esmaste põhimõtete õpetamise, kuid nõuab lisaväärtuse pakkumiseks konkreetsete kogemuste ja rakenduste ning praktiliste näidete ja tehniliste vahendite käsitlemist. Seda arvestades võib linkandmete teemalise keskmise mahuga ülikoolikursuse (kursuse maht 5–6 EAP, nädalas 2 loengu- ja 2 harjutustundi,

⁴³⁴ http://www.slideshare.net/fabien_gandon/w3c-tutorial-on-semantic-web-and-linked-data-at-www-2013

⁴³⁵ <http://euclid-project.eu/>

⁴³⁶ <http://linkeddata.org/guides-and-tutorials>

eeldatav üliõpilase töömaht 120-150 tundi) väljaarendamise töömahuks lugeda orienteeruvalt 3-6 inimkuud. Sõltuvalt kursuse iseloomust, tehtavate tööde mahust ja üliõpilaste arvust võib iga-aastast koormust hinnata samasse suurusjärku.

Kuna koolitused on väiksema mahuga, on nende ettevalmistuse ja läbiviimise töömahud väiksemad. Siiski ei vähene koolituse töömaht proportsionaalselt tundide arvuga; võib hinnata, et ühepäevase töötoa ettevalmistuse maht on orienteeruvalt paar nädalat. Arvestades koolitusfirmade praktilist orienteeritust võib eeldada, et uusi kursusi hakatakse välja töötama siis, kui nendel on olemas potentsiaalne tellija.

Monitoorimise, õpetamise ja propageerimise töömahu hindamiseks võib kasutada ka andmeid seni läbiviidud programmide kohta. Nii oli Euroopa Liidu struktuurifondide poolt rahastatud programmi „Infoühiskonna teadlikkuse tõstmine“ kogumaht 50 miljonit krooni ning programmi tegevused viidi läbi aastatel 2007-2013⁴³⁷. Alates 2009. aastast on selle programmi raames toimunud koolituste sarjas „Tark e-riik“ toimunud kokku 185 koolitust, seminari ja infopäeva, millest on osa võtnud ligikaudu 6000 inimest⁴³⁸.

Läbiviidud koolituste kogemust arvesse võttes võib linkandmete alase koolituse jagada üldiseks, süvendatud ja erikoolituseks. Üldist koolitust orienteeruva mahuga üks koolituspäev tuleks anda kõikidele keskmiste või suuremate andmekogude haldajatele. Seisuga 7.02.2014 oli RIHA-s registreeritud kokku 690 andmekogu, neist 268 kasutusel, 31 lõpetatud või lõpetamisel ning ülejäänud mitmesugustes asutamise/kasutuselevõtmise etappides. Osa andmekogudest on väga väikesed ja mõned haldajad haldavad mitut andmekogu. Siit võib eeldada, et koolitatavate suurusjärk on viissada inimest ehk orienteeruvalt 20 koolituspäeva.

Hinnanguliselt vajaksid süvendatud kahepäevalist tehnilist lisakoolitust eelkõige nende suuremate andmekogude haldajad, kes avaldavad oma andmeid ava- ja linkandmetena. Kui eeldada nende arvuks 200 (10 rühma, tehnilisemas rühmas on vähem inimesi), saame kokku 20 süvendatud koolituse päeva. Linkandmete mitmesugustele kitsamatele aspektidele suunatud eriotstarbeliste koolituste vajaduse selgub peale üldiste ja süvendatud koolituste läbiviimist. Peale esmaste koolituste läbiviimist tuleks korraldada jätkukoolitusi eeldatava mahuga orienteeruvalt 20% esialgsest koolituste mahust aastas. Ühe koolituspäeva maksumus võib Rahandusministeeriumi hinnangul olla 1250 eurot koos lektoritasu ja ruumide rendiga⁴³⁹.

Eelnevad koolitused olid mõeldud eelkõige avaliku sektori töötajatele. Muuhulgas on riigikaitse seisukohast soovitatav parandada riigiametnike teadlikkust linkandmetest ja nendega seotud tehnoloogiast. Erasektori spetsialistid võiksid koolitustest soovi korral osa võtta, kuid nagu üks erasektori intervjuueeritav ütles – pole probleemi, kui vaja, õpime, mida vaja ja teeme ära. Kui on nõudlus, siis tuleb ka pakkumine. Sellist järeldust toetab ka koolituste sarja „Tark e-riik“ kogemus, kus põhilised osavõtjad olid avalikust sektorist.

Linkandmete rakendamise tulu-kulu analüüsiks oleks vaja arvestada ka sellise rakendamise tulusid. Avaandmete puhul on sellist analüüsi korduvalt tehtud või teha püütud. Ühe Euroopa Komisjoni poolt tellitud ja Graham Vickery poolt läbi viidud analüüsi väide on, et avaliku sektori informatsiooni taaskasutamise otsesed kasud (*gains*) Euroopa Liidus on 40 mlrd eurot aastas; otsesed ja kaudsed kasud kokku on 140 mlrd eurot aastas⁴⁴⁰. Kui eeldada, et kasud jaotuvad proportsionaalselt sisemajanduse kogutoodangule ning lähtuda EL-27 ja Eesti 2011. a sisemajanduse kogutoodangu hinnangutest vastavalt 12638 ja 16 mlrd eurot⁴⁴¹, saame avaliku sektori avaandmete otsese kasu hinnanguks Eesti jaoks 50 miljonit eurot aastas; otsese ja kaudse kasu kogumaht oleks vastavalt 177 miljonit eurot aastas.

Toodud kasude hinnangutesse võib siiski suhtuda üsna ettevaatlikult. Esiteks pole päris kindel, kas hinnangutes rakendatud meetodika on laiendatav Eesti oludele. Teiseks, isegi kui avaliku sektori informatsiooni taaskasutamisega seotud kasude suurusjärk on õige, võib see vähe öelda andmete lingitavale kujule viimise kasude kohta (kuigi võib eeldada, et lingitaval kujul esitamine võimaldab andmete avanemisest saadavat efekti kõige paremini realiseerida). Kolmandaks, erinevad uuringud annavad erinevaid

⁴³⁷ <http://www.struktuurifondid.ee/infouhiskond-2/>

⁴³⁸ <https://www.ria.ee/algavad-tark-e-riik-koolitused/>

⁴³⁹ <http://www.korruptsioon.ee/orb.aw/class=file/action=preview/id=58998/Strateegia+rakendusplaan.pdf>

⁴⁴⁰ <http://ec.europa.eu/digital-agenda/en/news/economic-analysis-psi-impacts>

⁴⁴¹ http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/National_accounts_and_GDP

tulemusi. Nii väidab ajakirjas *European Journal of ePractice* 2011. a ilmunud uuring⁴⁴², et tõendid avaandmete mõju kohta on ebausaldatavad või puuduvad. EPSIPlatform poolt 2013. a läbiviidud uuring avaliku sektori avaandmete võimalikust mõjust ja selle aspektidest⁴⁴³ eristas avaandmete majanduslikku, poliitilist ja sotsiaalset mõju ning jõudis järeldusele, et avaliku sektori andmete avalikustamise ja taaskasutuse mõjude hindamine jääb raskeks ülesandeks.

Kuigi väited ava- ja linkandmete mõju kohta on erinevad, pole käesoleva töö autorite teada ükski uuring eeltoodud Graham Vickery analüüsi otseselt kahtluse alla seadnud. Isegi kui see analüüs või selle ülekanne Eesti oludele on suurusjärgu võrra ülepaistatud, näitab see siiski enamiku käesolevas uuringus toodud soovitude rahastamise otstarbekust.

Lõpuks märgime, et rahvusvaheline kogemus ei paku siinkohal ühtseid soovitusi – erinevatel riikidel on erinev lähenemine. Nii on Taani jaoks juhendamine ja koolitused tähtis avaandmete strateegia osa, samas kui näiteks UK ja USA puhul on suurem osatähtsus majanduslikel hoobadel ja õiguslikel regulatsioonidel⁴⁴⁴.

Ettepanekud ja poliitikasoovitused

- Linkandmete, semantilise veebi jt käesolevas uuringus käsitletud tehnoloogiate temaatikat on seni Eesti ülikoolides käsitletud väheste kursuste ja tudengiprojektide raames õppejõudude initsiatiivil. Ülikoolid peaksid seda initsiatiivi võimendama läbi temaatiliste õppe-, koolitus- ja teadusprojektide, mille raames kaasataks nii eksperte välismaalt kui kohalikke tudengeid. Ka on vaja ülikoolide õppekavasid uute trendide valguses regulaarselt kaasajastada.
- Koolituste esimene etapp peaks keskenduma üldistele (orienteeruva mahuga üks päev, koolitatavate suurusjärg viissada inimest ehk orienteeruvalt 20 koolituspäeva) ja süvendatud (tehniline konkreetse probleemi linkandmetega lahendamise lisakoolitus orienteeruva mahuga kaks päeva - suuremate andmekogude haldajatele, kes avaldavad oma andmeid ava- ja linkandmetena, orienteeruvalt 200 inimest ehk kokku 20 süvendatud koolituse päeva) koolitustele.
- Erinevad uuringud on üksmeelel selle kohta, et ava- ja linkandmetel on positiivsed mõjud – näiteks, võimaldades pakkuda paremaid teenuseid, luues paremat investeerimiskliimat ja vähendades korruptsiooni. Riik peaks selliseid mõjusid toetama, isegi kui täpsed kulude/tulude uuringud ei ole kättesaadavad. Kindlasti õigustavad pakutud hinnangud ülaltoodud koolituste läbiviimist – nende orienteeruv esimese etapi kogumaht jääb suurusjärku 50 000 eurot aastas.

7.3.2 Linkandmete tehnoloogia Eesti IKT sektori väljatöötatavates lahendustes

Mida saaks teha Eesti IKT sektor oma väljatöötatavates lahendustes linkandmete tehnoloogia kasutuselevõtu suurendamiseks ja oma vastavate kompetentside tõstmiseks?

Intervjuudest nähtub, et Eesti IKT sektori võimalusena nähakse eelkõige teenuste loomist. Innovatsiooni väga ei usaldata ning kõige olulisemate tegevustena tuuakse välja riigi poolset eestvedamist ning vastutuse võtmist. Erasektori kaasamiseks tuleks riigihangetele lisada ava- ja linkandmeid käsitlevad reeglid, mis on

⁴⁴² Noor Huijboom, Tijs Van den Broek, Open data: an international comparison of strategies, *European Journal of ePractice*, www.epracticejournal.eu N° 12 March/April 2011,

http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

⁴⁴³ Karolis Granickas, Understanding the Impact of Releasing and Re-using Open Government Data, *European Public Sector Information Platform*, Topic Report No. 2013/08, http://www.epsiplatform.eu/sites/default/files/2013-08-Open_Data_Impact.pdf

⁴⁴⁴ Noor Huijboom, Tijs Van den Broek, Open data: an international comparison of strategies, *European Journal of ePractice*, www.epracticejournal.eu N° 12 March/April 2011,

http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

kohustuslikud iga infosüsteemi tellimisel. Ka ankeedivastustes hinnati avaandmete rakenduste mõju (näiteks teenustes) kõrgelt.

Samal ajal näitavad globaalsed trendid kasvavat turgu (suur)andmete analüüsimise lahendustele. Globaalsed trendid viitavad märkimisväärsele ja kasvavale puudusele andmete töötlemise ja analüüsimise oskustega tööjõu osas lähima 5 aasta perspektiivis. Sellest tulenevalt on suurenenud (suur)andmeid kasutavate analüütiliste tööriistade ja SaaS lahenduste tootmine nii globaalses mastaabis. Lingitud ja puhastatud andmete kasutamine lihtsustab taoliste lahenduste loomist. Veelgi enam, kasutades globaalselt aktsepteeritud metamudeleid linkandmetes muutub lihtsamaks nende lahenduste eksportimine teistele turgudele. Eesti IKT sektor saab tõsta oma ekspordivõimekust linkandmetel põhinevat tarkvara tootes ning avalik sektor saab aidata sellele kaasa nõudes enda poolt tellitud või toetatud toodetes järjepidevalt linkandmete kasutamist.

Linkandmete kasutuselevõtu suurendamisele aitavad kaasa temaatilised teadusarendus-, tootearendus- ja innovatsioonitoetused ja Garage48-tüüpi ettevõtmised. Need võivad aidata kaasa olukorra muutumisele Eesti IKT sektoris, kus on suhteliselt vähe sektorisisest tootearendust ning riskide maandamiseks tehakse üldiselt tööd klientide tellimisel. Temaatilised rakendusuuringud ja tootearendusprogrammid võiksid tagada tootearendusele suunatud projektide jätkusuutlikkuse. Samal ajal saab avalik sektor tellijana suunata IKT sektorit linkandmetesse rohkem panustama, nõudes uutes või jätkuprojektides mõistlikkuse piires linkandmete kasutamist. Erasektor saab tellijana suunata IKT sektorit linkandmeid kasutama, kui avalikult sektorilt vajaminevad andmed on linkandmetena kättesaadavad.

Vastavalt avaliku teabe seadusele tuleb avalikustamisele kuuluv teave avalikustada ajakohasena ning viisil ja vormis, mis võimaldab selle allalaadimist masinloetaval kujul. Seepärast ei too täiendav linkandmetena publitseerimise nõue, kui see lisatakse kohe projekti algatamise faasis, kaasa väga suuri lisakulusi. Innovatsioonitoetuste ja Garage48-tüüpi meetmete maht sõltub pakutavatest toetust vajavatest projektidest ning seda saab hoida valitud piires, näiteks samas suurusjärgus koolituste ja teavitamise mahuga. Arvestades eelmises jaotises toodud linkandmete kasude hinnanguid võib ülalpakutud meetmeid seega hinnata kuluefektiivseiks.

Ettepanekud ja poliitikasoovitused

- Eesti IKT sektor saab pakkuda riigihangetele ava- ja linkandmeid käsitlevad lahendusi ja rakendusi (näiteks teenustes), käivitada innovatsiooni- ja Garage48-tüüpi ettevõtmisi, teha temaatilisi rakendusuuringuid ja tootearendust ning kasutada loodud linkandmeid.
- Lisaks saab IKT sektor luua ja töös hoida avatud, valdkondlikke linkandmete pilveplatvorme ja/või valdkondlikke/valdkonnaüleseid olemasolevale avatud linkandmete pilveplatvormile rajanevaid teenuseid (sealhulgas e-äri jaoks CRMid, turundusanalüütika ja tootearenduse planeerimise tooted, mis kasutavad linkandmeid) ning pakkuda neid nii siseriiklikult kui välismaale teenusena.
- Erasektori kaasamiseks tuleks riigihangetele lisada ava- ja linkandmeid käsitlevad reeglid, mis on kohustuslikud iga infosüsteemi tellimisel, tellida avaandmete rakendusi (näiteks teenustes) ning nõuda uutes või jätkuprojektides mõistlikkuse piires linkandmete kasutamist.

7.3.3 Eesti avaliku sektori täiendavad ülesanded infosüsteemide tellijana

Millised võiksid olla on Eesti avaliku sektori täiendavad ülesanded infosüsteemide tellijana seoses linkandmete tehnoloogiate rakendamisega?

Mitmed Eesti avaliku sektori ülesanded infosüsteemide tellijana on käsitletud eelmistes jaotistes. Nii peaks Eesti avalik sektor käivitama linkandmetega seotud pilootprojekte, tõstma linkandmete alast teadlikkust, maandama linkandmete kasutuselevõtuga kaasnevaid ohte ja riske, nõudma linkandmetega seotud avalike standardite kasutamist ja ise neid aktiivselt järgima, looma ja täiendama tehnilise infrastruktuuri lahendusi,

korraldama koosloomel põhinevate lahenduste loomist, hoolitsema vajalike semantikaressursside tekke eest, hankima linkandmete tehnoloogiate rakendamist avaandmetena publitseerimisel, kasutama linkandmeid sisutöötlusel ning lisama riigihangetele ava- ja linkandmeid käsitlevad reeglid.

Muuhulgas rõhutati intervjuudes, et avalik sektor peaks kehtestama kõikidele tellitavatele infosüsteemidele koosvõimelisuse, avaandmete ja linkandmete osa kohustuslikkuse ning semantiliste klassifikaatorite ja URI-de olemasolu nõuded. Nõuete järelevalve peaks struktuurifondidest rahastatavate projektide puhul olema seotud rahade väljamaksmisega. Muude projektide üle võiks järelevalvet korraldada RISO, kes kooskõlastaks projekti rahade väljamaksmist sõltuvalt nõuete täidetusest.

Eriti oluline on ontoloogiate loomine ja uute linkandmete tekitamine, sest linkandmete väärtus kasvab hüppeliselt võrreldes lingitud andmete mahu kasvuga. Seetõttu võiks iga uus projekt, kus andmeid kasutatakse ja/või uusi andmeid luuakse, lisada uusi linkandmeid läbi tekkinud andmete linkimise. Ontoloogiate loomisele on kaasa aidanud tellijapoolsed nõuded täitjatele luua uute infosüsteemide jaoks ka sobivad ontoloogiad. Lisaks on veel teisigi näiteid, nt ehitussektoris, kus tellija nõuete kaudu saab suunata valdkonna arengut.

Ontoloogiate loomise maksumust analüüsitakse mitmete EL projektide poolt toetatud töös „*ONTOCOM: A cost estimation model for ontology engineering*“⁴⁴⁵. Maksumus sõltub töömahust, mille määravad ontoloogia suurus ja mitmesugused muud parameetrid, nagu valdkonna keerukus, arendaja kogemus, vahendite võimsus jne. Artiklis pakutud ontoloogiate maksumuse hindamise meetod põhineb intervjuudel 36 ontoloogiate arendamise eksperdiga. Uuritavad ontoloogiad sisaldasid keskmiselt 830 komponenti (mediaan oli 330). Algusest peale ehitati 40% ontoloogiatest; ülejäänud ontoloogiates korduvkasutati keskmiselt 50% komponente. Analüüsitud ontoloogiate loomine võttis keskmiselt 5,3 kuud (mediaan 2,5). Toodud hinnangud annavad ettekujutuse ontoloogiate loomise orienteeruvast mahust ja maksumusest. Loodud ONTOCOM mudelit saab kasutada täpsemate hinnangute saamiseks sõltuvalt konkreetsest hankest.

Intervjuudes pakuti ka ettevõtluse toetamise innovatsiooniosakute toetusmeedet, mille puhul projektides eeldatakse linkandmete tehnoloogiate rakendamist, ontoloogiate loomist jms.

Ettepanekud ja poliitikasoovitused

- Eesti avalik sektor peaks kehtestama kõikidele tellitavatele infosüsteemidele koosvõimelisuse, avaandmete ja linkandmete osa kohustuslikkuse ning semantiliste klassifikaatorite ja URI-de olemasolu nõuded, millede järelevalve peaks olema seotud rahade väljamaksmisega. Eriti tuleks nõuda ontoloogiate loomist ja uute linkandmete tekitamist.
- Avalik sektor peaks suurendama analüütiliste rakenduste osakaalu, kus kasutatakse andmeid ka väljastpoolt nende endi haldusala ja kasu linkandmetest on otseselt näha. Taoliste lahenduste tekkimise soodustamiseks on mõistlik infosüsteemide lähteülesannete püstitustes eraldi välja tuua väljaspool haldusala paiknevate andmete võimalik mõju rakendusele valdkondlike andmekomplektide ja avaandmete kasutamise lõikes. Sel juhul saab infosüsteemide loomise finantseerimise otsuseid tehes arvestada nende valdkondadele mõju.

7.3.4 Soovitusi erasektorile oma infosüsteemide tellimisel ja arendamisel

Mida soovitada erasektorile oma infosüsteemide tellimisel ja arendamisel, et linkandmete kasutuselevõtt Eestis edeneks?

⁴⁴⁵ Elena Paslaru, Bontas Simperl, Christoph Tempich, York Sure. ONTOCOM: A cost estimation model for ontology engineering (2006), Proceedings of fifth ISWC, 2006, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.2432&rep=rep1&type=pdf>

Intervjuude ja ankeetide põhjal ei ole erasektori esindajad oma rolli hindamisel eriti optimistlikud. Ei usuta, et uued tehnoloogiad toovad suurt edu – innovatsiooni on tehtud, kuid see on enne pooleli jäetud, kui tulemus tekib. Kõige olulisemate tegevustena tuuakse välja riigi poolset eestvedamist ning vastutuse võtmist. Samuti nimetatakse edulugude vajadust, mis võimaldaks näidata innovatiivset probleemilahendust ning mida Euroopasse skaleerida. Erasektori rolli nähakse eelkõige teenuste loomises, taandades avalik sektor teenuseosutaja rollist. Ka ankeetides on turunduse ja müügi, erasektorile tüüpiliste avaandmete kasutusvõimaluste mõju hinnatud keskmisest madalamana.

Samas on avatud standardite kasutamine ning standardiseerimises osalemine arenenud tööstusriikides andnud olulise osa SKP kasvust. Erasektorile võiks soovitada linkandmete käsitlemist avatud standardina ning nende positiivse mõju arvestamist sarnaselt valdkondlikele metamudelitele, nagu nt SID, HR-XML jne. Tuleks olla kursis linkandmete tehnoloogia arenguga, osaleda koolitustel, jälgida linkandmete edulugusid ja kasutada avaliku sektori poolt publitseeritavaid linkandmeid. Eestis on näiteid linkandmete kasutuselevõttust, näiteks on Riikliku Ehitisregistri lingitud kujul andmed leidnud kasutust kinnisvaraga seotud portaalides.

Näiteid selle kohta, kuidas erasektor saab linkandmete tehnoloogiast kasu oma infosüsteemide tellimisel ja arendamisel:

- arendades standardiseeritud metamudelitel ja linkandmetel baseeruvat tarkvara⁴⁴⁶;
- pakkudes avaliku sektori andmete publitseerimise ja linkimise teenust ning näidislahendusi⁴⁴⁷;
- pakkudes linkandmetel põhinevaid andmeanalüüsi ja pilveteenuseid⁴⁴⁸, sh spetsialiseeritud teenuseid kindlale valdkonnale⁴⁴⁹, sealjuures arvestades teenuste ekspordipotentsiaali;
- kasutades ava- ja linkandmeid teenuste pakkumiseks^{450,451};
- rakendades linkandmete tehnoloogiat oma ettevõtte andmete (linkandmed ei pruugi olla avaandmed) haldamise lihtsustamiseks;
- rakendades linkandmete tehnoloogiat firmade vaheliseks andmete vahetamiseks (praegu kasutatakse linkandmeid pigem oma platvormi või ökosüsteemi piires);
- kasutades võimalusi, mida pakuvad innovatsioonitoetused ja Garage48-tüüpi ettevõtmised.

Ettepanekud ja poliitikasoovitused

- Kuna erasektori esindajad ei ole oma rolli hindamisel linkandmete kasutuselevõttus eriti optimistlikud, tuleks soovitada riigi poolset eestvedamist ning vastutuse võtmist ning edulugude esiletoomist. Erasektorile võiks soovitada linkandmete käsitlemist avatud standardina ning nende positiivse mõju arvestamist. Tuleks olla kursis linkandmete tehnoloogia arenguga, osaleda koolitustel, jälgida linkandmete edulugusid ja kasutada avaliku sektori poolt publitseeritavaid linkandmeid. Erasektor saab linkandmete tehnoloogiaid oma infosüsteemide tellimisel ja arendamisel kasutada, propageerides edukaid linkandmete projekte nii era- kui ka avalikus sektoris, arendades tarkvara linkandmete esitamiseks ja päringute tegemiseks, pakkudes avaliku sektori andmete publitseerimise ja linkimise teenust ning näidislahendusi, pakkudes ekspordipotentsiaaliga linkandmetel põhinevaid andmeanalüüsi ja pilveteenuseid, kasutades ava- ja linkandmeid teenuste pakkumiseks, rakendades linkandmete tehnoloogiat oma ettevõtte andmete haldamisel (linkandmed ei pruugi olla avaandmed), rakendades linkandmete tehnoloogiat firmade vaheliseks andmete

⁴⁴⁶ <http://data.gov.uk/library/geowise>

⁴⁴⁷ <http://data.gov.uk/library/swirl>

⁴⁴⁸ <http://data.gov.uk/library/flying-binary>

⁴⁴⁹ <http://data.gov.uk/library/harvey-walsh>

⁴⁵⁰ <http://data.gov.uk/library/mudlark>

⁴⁵¹ <http://umusic.co.uk/artists/>

vahetamiseks ning kasutades võimalusi, mida pakuvad innovatsioonitoetused ja Garage48-tüüpi ettevõtmised.

7.3.5 Vabasektor ja kodanikkonna kaasamine

Kuidas kaasata linkandmete tehnoloogia kasutamisse Eesti vabasektor ja kodanikkond?

Vabasektor (ühiskonna see osa, mis ei kuulu ei avalikku ega ärisektorisse, hõlmates mittetulundusühinguid, sihtasutusi, seltsinguid ja muid eraalgatusel põhinevaid ühendusi⁴⁵²) ja kodanikkond on üle maailma olnud linkandmete alal üsna aktiivne. Nii käsitletakse ülal arhitektuurianalüüsi osas DBpedia Live koosloome initsiatiivi, mille eesmärk on teha Wikipedia andmed linkandmetena kättesaadavateks ja kasutatavateks. DBpedia põhjal on loodud mitmeid rakendusi, näiteks päringuvahendeid⁴⁵³ ja veebisaite⁴⁵⁴. Paljusid linkandmetega seotud vaba tarkvara projekte on algatanud vabasektori organisatsioonid. Näiteks arendatakse Apache Software Foundation⁴⁵⁵ raames projekti Apache Jena, mis pakub Java põhist raamistikku semantilise veebi ja linkandmete rakenduste ehitamiseks⁴⁵⁶.

Eestis on koosloome läbi vastavate rakenduste leidnud kasutuse isikute poolt linkandmete loomisesse. Näiteks lisatakse veebilehtedele juba praegu märgendeid (nt *microdata* kujul), mida otsimootorid kasutavad linkimiseks, tehes samas saadud andmed otsitulemustes kättesaadavaks. Laiemale avalikkusele on Euroopa initsiatiivides linkandmete eksponeerimiseks kasutusel linkandmete brausimise keskkonnad.

Üldiselt saab vabasektori ja eraisikute huvi (link)andmete vastu tulla nende seesmistest ajenditest, näiteks soovist teha midagi paremaks, isiklikust huvist, soovist saada tuntuks vms. Linkandmete tehnoloogia on suhteliselt keeruline ja kompetentsi puudumine avaldab siinkohal negatiivset mõju. Inimesed, kes seda tehnoloogiat hästi valdavad, on suure tõenäosusega juba mingis organisatsioonis hõivatud ja tegelevad sellega oma tööülesannete raames. Lisaks on vabasektori ja kodanikkonna jaoks olemas ka palju muid huvitavaid teemasid.

LOD2 projekti küsitluste aruandest „*The LOD2 Open Government Data Stakeholder Survey 2010*“⁴⁵⁷ selgub, et peamiselt kasutatakse linkandmeid kas uuringute läbiviimisel või andmete visualiseerimisel. Vabasektor on muuhulgas huvitatud avaliku sektori avaandmete kasutamisest avaliku sektor kontrolliks ja läbipaistvuse saavutamiseks.

Eelnevast ning intervjuudest nähtub, et Eesti vabasektori ja kodanikkonna kaasamiseks oleks vaja linkandmete alast teavitamist ja koolitamist, huvipakkuvaid rakendusi, suhteliselt lihtsaid tööriistu ja kättesaadavaid linkandmeid (valik tööriistadest on toodud peatükis 7.3.7). Teavitamist, koolitust ja linkandmete loomist tuleks teha avaliku sektori tööülesannete täitmise ning eelmistes jaotistes toodud algatuste raames, näiteks eraldades koolitustele teatud protsendi kohti vabasektori ja kodanikkonna jaoks. Sellise eraldamise mõistlik maht selgub koolituste käigus ning sellest lähtudes võib arvestada vajalikke lisaressursse.

Ka huvipakkuvatele ülesannetele saab viidata teavitamise ja koolituste käigus, kuid eelkõige peaks vabasektor ja kodanikkond need ise enda jaoks leidma. Mõned näited on lastehoidmise teenused⁴⁵⁸ ja tööpakkumise ja töötamise kuulutused⁴⁵⁹.

⁴⁵² <http://portaal.eki.ee/uuedsonad/Uued-s%C3%B5nad-1/V/vabasektor-10/>

⁴⁵³ <http://wiki.dbpedia.org/OnlineAccess>

⁴⁵⁴ <http://www.aboutthisday.com/>

⁴⁵⁵ <http://www.apache.org/foundation/>

⁴⁵⁶ <http://projects.apache.org/projects/jena.html>

⁴⁵⁷

<http://wiki.lod2.eu/download/attachments/3473453/AnalysisSurvey.pdf?version=1&modificationDate=1329941495000>

⁴⁵⁸ <http://lastehoid.net>

⁴⁵⁹ <http://kirka.ee>

Vabasektorile ja kodanikule tuleks pakkuda suhteliselt lihtsaid tööriistu ja keskkondi andmete lingitavale kujule viimiseks, publitseerimiseks ning rakenduste loomiseks. Kaasamine eeldab kindlasti ka piisava hulga ava- ja linkandmete olemasolu ning teadlikkust võimalustest ja edulugudest.

Mitmes intervjuus avaldati arvamust, et era- ja vabasektori vahel pole olulist erinevust ning soovitud nende kaasamiseks on sarnased. Kodanikule on oluline pigem teenuse sisu ja kvaliteet, mitte tehnoloogia. Paljude küsimuste vastustes toodi olulisena välja edulugude olemasolu ja nende tutvustamise vajadus.

Ettepanekud ja poliitikasoovitused

- Eesti vabasektori ja kodanikkonna kaasamiseks oleks vaja linkandmete alast teavitamist ja koolitamist, teadlikkust võimalustest ja edulugudest, huvipakkuvaid rakendusi, suhteliselt lihtsaid tööriistu ja kättesaadavaid linkandmeid. Teavitamist, koolitust ja linkandmete loomist tuleks teha avaliku sektori tööülesannete täitmise ning eelmistes jaotistes toodud algatuste raames, näiteks eraldades koolitustele teatud protsendi kohti vabasektori ja kodanikkonna jaoks. Vabasektorile ja kodanikule tuleks pakkuda suhteliselt lihtsaid tööriistu ja keskkondi andmete lingitavale kujule viimiseks, publitseerimiseks ning rakenduste loomiseks. Valik vastavaid tööriistu on toodud peatükis 7.3.7.
- Disainida meetmed selliste valdkondlike ja valdkonnaüleste avatud linkandmete ökosüsteemide loomise või kohandamise võimaldamiseks, kus on toetatud nii avatud kui privaatsete/konfidentsiaalsete linkandmete kooskasutamine ilma andmete privaatsust/konfidentsiaalsust ohustamata. Vastavad ökosüsteemid koondaks erinevaid linkandmete komplekte valdkondlike või valdkonnaüleste probleemide lahendamiseks.
- Eesti vabasektori ja kodanikkonna kaasamiseks oleks vaja luua koosloome saite, mille kaudu vabakond saaks nõustada riiki andmete linkimise osas osutades näiteks linkimist vajavatele andmehulkadele, hinnata avalike andmete kvaliteeti jms,

7.3.6 Tööjaotus eri huvipoolte vahel

Milline peaks olema Eestis tööjaotus eri huvipoolte vahel linkandmete tehnoloogiate kasutamise ja arendamise korraldamisel?

Kirjanduse analüüs, nt TNO avaandmete strateegiate uuring⁴⁶⁰ ning käesolevas uuringus läbi viidud intervjuud toovad välja avaliku sektori juhtiva rolli ava- ja linkandmete kasutuselevõtus. See on ka arusaadav, sest erasektor näeb andmeid kui oma strateegilisi varasid ning on väga ettevaatlik nende avalikustamisega, vabasektor ja kodanikkond omavad aga suhteliselt vähe andmeid. Samuti on näha tugeva poliitilise juhtfiguuri positiivne mõju. Ka see on mõistetav, sest avaliku sektori töötajad ei pruugi iseenesest olla motiveeritud andmete avalikustamiseks.

Toodud üldine rollijaotus kehtib ka Eestis. Avalik sektor on enamasti initsiaator, sätestab õigusakte, soovib või kehtestab standardeid, publitseerib andmeid, arendab esmaseid teenuseid, tellib pilootprojekte, koolitab ja teavitab teisi osapooli. Erasektor, vabasektor ning kodanikkond on pigem järgijate, teenusepakujate ja andmete kasutajate rollis. Toodud üldisest rollijaotusest saab tuletada ka järgneva detailsema tööjaotuse eri huvipoolte vahel Eestis.

RISO haldab ja arendab riigi infosüsteemi semantilise koosvõimega seotud dokumentatsiooni, semantika tööühma materjale, semantilise koosvõime raamistikku ning muid riigi infosüsteemi strateegilisi dokumente⁴⁶¹. Riigi Infosüsteemi Amet koordineerib riigi infosüsteemi arendamist ja haldamist, teostab

⁴⁶⁰ http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf

⁴⁶¹ <http://www.riso.ee/et/koosvoime/semantika>

järelevalvet riigi infosüsteemi haldamist reguleerivatest õigusaktidest tulenevate nõuete täitmise üle ja peab riigi infosüsteemi haldussüsteemi⁴⁶².

RISO ja RIA koostöös koordineeritakse metaandmete, sealhulgas valdkondade vaheliste metaandmete loomist, kasutades selleks valdkondade vahelisi töörühmi ning vajadusel hankides uuringuid või pilootprojekte. MKM infoühiskonna teenuste arendamise osakond (ITAO) koordineerib linkandmete kasutamise aspekte, mis on seotud avalike teenuste ning avaliku sektori asjaajamise ja elektroonilise dokumendihalduse arendamisega.

Avaandmete rakendamise järelevalvet teostatakse RIA, Andmekaitse Inspeksiooni ja RISO koostöös. Järelevalve vahenditena saab muuhulgas rakendada struktuuritoetuste taotluste kooskõlastamist või mittekooskõlastamist, samuti Andmekaitse Inspeksiooni käsutuses olevaid meetmeid vastavalt avaliku teabe seadusele.

Avaliku sektori (täpsemalt, riigi, kohaliku omavalitsuse või muu avalik-õigusliku juriidilise isiku või avalikke ülesandeid täitva eraõigusliku isiku) andmekogude vastutavad või volitatud töötajad registreerivad andmekogud ning loovad ja registreerivad vastavad semantilised varad riigi infosüsteemi haldussüsteemis⁴⁶³.

Avaliku sektori asutused vastutavad oma infosüsteemide jaoks vajalike semantikavarade väljatöötamise ning andmete publitseerimise eest vastavalt avaliku teabe seadusele, Vabariigi Valitsuse määrusele „Riigi infosüsteemi haldussüsteem“ ja muudele õigusaktidele.

Erasektor, vabasektor ning kodanikkond hoiab ennast ava- ja linkandmete valdkonnaga kursis, leiab kasulikke, tulusaid ja/või huvitavaid rakendusi, pakub teenuseid ja kasutab andmeid.

Käesoleva küsimuse seisukohast pakuvad huvi avaliku ja erasektori koostöö (*PPP, public-private partnership*) rakendamise võimalused Eesti tingimustes. Sellist koostööd on Eestis rakendatud peamiselt ehituse valdkonnas. Riigihangete seadus sätestab: „§ 6¹. Avaliku ja erasektori koostöö. Avaliku ja erasektori koostöö käesoleva seaduse tähenduses on hankemenetluse, ehitustööde või teenuste kontsessiooni andmise teel sõlmitud hankelepingu alusel ehitustööde teostamine või teenuste osutamine, kus samaaegselt on täidetud alljärgnevad tingimused: 1) koostöö eesmärgiks on ehitustöö teostamine või teenuse osutamine avalikes huvides; 2) koostöö kestab vähemalt 20 aastat; 3) hankija ja pakkuja kasutavad ühiselt koostööks vajalikke vahendeid, näiteks asju, raha, oskusteavet, kogemust; 4) hankija ja pakkuja jagavad koostöö vastutust ning riske“. Arvestades tehnoloogia ja standardite kiiret muutumist pole Eestis vähemalt 20 aastat kestvad IKT valdkonna projektid eriti realistlikud. Seega võiks riigihangete seadusele vastavaid PPP projekte linkandmete valdkonnas lugeda Eestis vähetõenäolisteks. Suure mahu ja kestvuse tõttu peaksid nendele eelnema koolitused, pilootprojektid ning muud eelmistes jaotistes pakutud tegevused. Kui seada PPP projektidele väiksemad nõuded, võiks nende hulka tinglikult lugeda ka erasektori poolt pakutavad serverimajutuse, pilve- , mobiilse parkimise jm teenused. Analoogilisi projekte tasub algselt ka linkandmete valdkonnas, arvestades siiski seda, et riigihangete seaduse mõttes nad PPP projektide all ei kuulu.

Toodud loomulik tööjaotus taaskasutab Eesti juba olemasolevaid organisatsiooni ja protsesse ning ei too kaasa täiendavaid kulutusi lisaks juba eelpool mainitutele.

Ettepanekud ja poliitikasoovitused

- Nii maailmapraktikast kui ka käesolevas uuringus läbi viidud intervjuudest lähtuvalt tuleks soovitada avaliku sektori juhtivat rolli ava- ja linkandmete kasutuselevõtus ning tugeva poliitilise juhtfiguuri positiivset mõju. Avalik sektor on enamasti initsiaator, sätestab õigusakte, soovib või kehtestab standardeid, publitseerib andmeid, arendab esmaseid teenuseid, tellib pilootprojekte, koolitab ja teavitab teisi osapooli.
- RISO haldab ja arendab riigi infosüsteemi semantilise koosvõimega seotud dokumentatsiooni, semantika töörühma materjale, semantilise koosvõime raamistikku ning muid riigi infosüsteemi

⁴⁶² <https://www.riigiteataja.ee/akt/128042011001?leiaKehitiv>

⁴⁶³ <https://www.riigiteataja.ee/akt/13147268>

strateegilisi dokumente. Riigi Infosüsteemi Amet koordineerib riigi infosüsteemi arendamist ja haldamist, teostab järelevalvet riigi infosüsteemi haldamist reguleerivatest õigusaktidest tulenevate nõuete täitmise üle ja peab riigi infosüsteemi haldussüsteemi. RISO ja RIA koostöös koordineeritakse metaandmete, sealhulgas valdkondade vaheliste metaandmete loomist, kasutades selleks valdkondade vahelisi töörühmi ning vajadusel hankides uuringuid või pilootprojekte.

- MKM infoühiskonna teenuste arendamise osakond (ITAO) koordineerib linkandmete kasutamise aspekte, mis on seotud avalike teenuste ning avaliku sektori asjaajamise ja elektroonilise dokumendihalduse arendamisega.
- Avaliku sektori andmekogude vastutavad või volitatud töötajad registreerivad andmekogud ning loovad ja registreerivad vastavad semantilised varad riigi infosüsteemi haldussüsteemis.
- Avaliku sektori asutused vastutavad oma infosüsteemide jaoks vajalike semantikavarade väljatöötamise ning andmete publitseerimise eest vastavalt avaliku teabe seadusele, Vabariigi Valitsuse määrusele „Riigi infosüsteemi haldussüsteem“ ja muudele õigusaktidele.
- Erasektor, vabasektor ning kodanikkond on pigem järgijate, teenusepakujate ja andmete kasutajate rollis. Avaliku ja erasektori koostööd tuleb igati soodustada, kuid riigihangete seadusele vastavaid PPP projekte linkandmete valdkonnas võib lugeda Eestis vähetöenäolisteks ning suure mahu ja kestvuse tõttu peaksid nendele eelnema koolitused, pilootprojektid ja muud eelmistes jaotistes pakutud tegevused.

7.3.7 Andmetöötlusprotsessid ja andmete täieliku elutsükli toetamine

Missugune peaks olema linkandmete kasutuselevõtu puhul andmetöötlusprotsessid, sh andmete täieliku elutsükli toetamine?

Üldjuhul on linkandmete loomine mitte põhieesmärk, vaid vahend põhieesmärkide saavutamiseks. Andmeid kogutakse ja hallatakse rahuldamiseks mingit vajadust ning saamaks mingit sisulist tulemust. Vajadusel ja võimalusel publitseeritakse need andmed ava- või linkandmetena. Seega peaks kasutatavate andmetöötlusprotsesside valik eelkõige lähtuma andmekogu loomise eesmärkidest. Valitud protsesse täiendatakse linkandmete loomiseks ja kasutamiseks vajalike tegevustega.

Linkandmete andmetöötlusprotsesside eripäraks andmete elutsükli mõttes on andmete semantiline hajusus. Seetõttu on avaliku sektori linkandmete jaoks pakutud erinevaid elutsükli mudeleid⁴⁶⁴. Need pakuvad erinevat elutsükli ulatust (nt kas mudel käsitleb vaid andmeid või ka äriprotsesse), etappide valikut ning etappide organisatsiooni (nt järjestikune või iteratiivne). Ühte kõikide poolt tunnustatud linkandmete elutsükli mudelit ei ole. Olukord on analoogiline tarkvara arendusega, kus kasutatakse sõltuvalt vajadusest mitmesuguseid elutsükli mudeleid.

Laiemas äriprotsesside kontekstis on otstarbekas eeskuju võtta ava- ja linkandmete valdkonna analüüsis toodud TNO (*Netherlands Organisation for Applied Scientific Research*) avaandmete elutsükli mudelist⁴⁶⁵. Mudel koosneb viiest faasist: identifitseerimine, ettevalmistamine, publitseerimine, korduvkasutamine, hindamine.

Linkandmete valdkonna analüüsis soovitasime Eesti riigi IS-de kontekstis kasutada B. Villazon-Terrarase iteratiivset linkandmete elutsükli mudelit, mis on kooskõlas nii TNO kui W3C soovitatud mudelitega hõlmates

⁴⁶⁴ http://www.w3.org/2011/gld/wiki/GLD_Life_cycle

⁴⁶⁵ https://www.tno.nl/content.cfm?context=thema&content=prop_publicatie&laag1=897&laag2=919&laag3=123&item_id=919

järgmisi tegevusi: spetsifitseerimine, modelleerimine, genereerimine, publitseerimine ja kasutamine/rakendamine.

Järgmises tabelis (vt *Tabel 10*) esitame iga toodud linkandmete elutsükli mudeli tegevuse kohta ülesanded, kaasajal olemasolevad toetavad tehnoloogilised vahendid (kui võimalik) ja soovitused ning näited.

Tabel 10. Avalike linkandmete elutsükli tegevused ja nende toetamine

Spetsifitseeri andmed		
Ülesanded	Toetav vahend	Soovitused
Identifitseeri ja analüüsi andmeallikad. URIdes disain (disaini mallid). Defineeri/kirjelda litsents.		Tuleb analüüsida juba olemasolevaid avalikke avaandmeid (ka teistes riikides) nende rakendustes taaskasutamise eesmärgil ja neid avalikke andmeid, mis pole veel avatud ja mida soovitakse avada. Ontoloogia (mudeli) URI-de mallid peavad erinema ontoloogia objektide/eksemplaride URIdes mallidest. Näiteks: URI-de baas: http://linkeddata.ee/ http://geo.linkeddata.ee/ Ontoloogia mõistete URI-d: http://geo.linkeddata.ee/ontology/{concept property} http://geo.linkeddata.ee/ontology/Linn Objektide URI-d: http://geo.linkeddata.ee/resource/{resourceType}/{resourceName} http://geo.linkeddata.ee/resource/Linn/Tartu URI-de mallide loomine ja haldamine riigis peaks olema mingi keskse organi pädevuses (nt RIA) seda nii andmeallikate kui ontoloogiatega URIdes mallide osas. Loodud URIsid peaks taaskasutama niipalju kui võimalik.
Modelleeri		
Ülesanded	Toetav vahend	Soovitused
Otsi sobivaid ontoloogiaid/sõnastikke, mis modelleerivad andmeallikaid. Loo andmeallikate mudel taaskasutades olemasolevaid ontoloogiaid/sõnastikke või loo uusi vajalikke ontoloogiaid ja sõnastikke.	Ontoloogiatega loomiseks: Protege, OntoEdit, Google Refine (SKOS-i jaoks)	Ontoloogiad on kirjeldatud OWL keeles või RDF(S) keeles, mis mõlemad baseeruvad RDF-il. Dublin Core, Friend of a Friend (FOAF), Statistical Core Vocabulary (Scovo), RDF Data Cube Vocabulary, isikute ja organisatsioonide ontoloogiaid ning muud ontoloogiaid ja sõnastikud sobivad taaskasutuseks.
Genereeri (RDF)		
Ülesanded	Toetav vahend	Soovitused
Transformeeri andmed RDF-i vastavalt ontoloogiatele. (üks võimalus on ka RDF ekstraheerimine tekstist, XML ja SQL-st).	RDF-i transformeerimiseks: CSV: Google Refine, XLWrap, RDF123, NOR2O RDB: D2R Server,	Andmete spetsifitseerimise etapil valitud andmehulkade transformeerimine RDF kujule vastavalt modelleerimise etapis loodud ontoloogiatele ja sõnastikele. Tihti kirjutatakse vastav skript, mis teisendab andmed vastavalt ontoloogiatele CSV-st RDF-i

Puhasta andmed.	ODEMapster, W3C	(näiteks UK valitsuse linkandmete projekt)
Identifitseeri andmehulga, mis sobiks linkimiseks.	RDB2RDF WG – R2RML	NB! Tihti on kasulik või lausa vajalik RDF genereerimise juurest pöörduda tagasi kas modelleerimise või andmete spetsifitseerimise juurde.
Leia seosed töödeldavate andmehulkade andmeühikute ja eelnevalt identifitseeritud andmehulkade andmeühikute vahel.	XML : GRDDL, ReDeFer	UK kogemus : on tähtis aru saada (analüüsida), kus on lingitavate andmehulkade parimad ühised punktid (andmeelemendid), sest need on eriti kasulikud linkimiseks.
Valideeri neid leitud seoseid.	Ekstraheerimiseks: PoolParty Extractor, OpenLink Virtuoso Sponger, CSVImport jt.	Linkimine:
Lingi andmed.	Puhastamiseks: Google Refine	http://dbpedia.org/resource/Tartu
	Identifitseerimiseks: ckan.net	OWL: sameAS
	Seoste leidmiseks: Limes, Silk framework	http://geo.linkeddata.ee/resource/Linn/Tartu
	Valideerimiseks: samesAS Validator	

Publitseeri

Ülesanded	Toetav vahend	Soovitused
Publitseeri andmed avatud litsentsiga	RDF andmete hoidmise ja haldusvahendid:	Riik peaks tagama järgmist:
Publitseeri vastavad metaandmed	Virtuoso Universal Server, Jena, Sesame, 4Store, YARS, OWLIM	Andmed peavad olema avatud litsentsiga, muidu ei saa neid nimetada avaandmeteks.
Taga andmete juurdepääs ja efektiivne leitavus/avastamine.	SPARQL lüüsid juurdepääsuks.	Otsene juurdepääs toorandmetele on vajalik.
	Publitseerimisvahendid : Pubby, Talis Platform, Fuseki	Metaandmete standardite (pigem üldiste kui valdkondlike) kasutamine on vajalik.
	Metaandmete halduseks: VoID	Linkandmete versioneerimine ja toorandmete pidev uuendamine on tähtis.
	Litsentside publitseerimiseks: Open Prov	
	Linkandmetele juurdepääsu tagamiseks: registreeri linkandmehulk CKAN registris, genereeri sisukaardi failid kasutades sitemap4rdf ja esita sisukaardi asukoht otsingumootoritele Google ja Sindice.	

Kasuta ja rakenda

Ülesanded	Toetav vahend	Soovitused
Loo rakendusi.	Geo-andmete jaoks map4rdf, mis võimaldab Google kaardivaadet RDF andmetele	Loo lihtsad rakendusliidesed avalike linkandmete kasutamiseks tavakodanikele, ka loomulikus keeles.
Kasuta rakendustest saadud tagasisidet uute avalike linkandmete spetsifitseerimiseks.		Tähtis on andmete visualiseerimine ja uuendamine.
		Üks võimalus linkandmete rakenduste

ehitamiseks on alustada geoandmetest (UK).

Kuigi riigiasutused ei pruugi luua rakendusi ise on rakenduste kohta tagasiside kogumine oluline selleks, et saaks spetsifitseerida uusi avalikke linkandmeid publitseerimiseks ja (taas)kasutamiseks.

Kodanikud saavad linkandmete arengule kaasa aidata kasutades nendega tekkivaid uusi võimalusi. Kodanikele saab luua ka koosloome liideseid avaandmete ja linkandmete kvaliteedi parandamiseks (näiteks bussiplaanides vigade korrigeerimiseks jms)

Kodanikele võib luua linkandmete infrastruktuuri, et nad saaksid ise luua oma linkandmete rakendusi.

Vajalikud lisakulutused on analüüsitud eelnevates ettepanekutes.

Ettepanekud ja poliitikasoovitused

- Kasutatavate andmetöötlusprotsesside valik peaks eelkõige lähtuma andmekogu loomise eesmärkidest. Valitud protsesse tuleks täiendada linkandmete loomiseks ja kasutamiseks vajalike tegevustega, võttes aluseks ülaltoodud (vt *Joonis 1* ja *Tabel 10*) avalike linkandmete elutsükli mudeli.
- URI-de mallide loomine ja haldamine riigis peaks olema mingi keskse organi pädevuses (nt RIA) seda nii andmeallikate kui ontoloogiade URI-de mallide osas. Riigiasutused peaksid loodud URIsid taaskasutama niipalju kui võimalik.
- Kuna RIHA depositooriumis juba hallatakse andmekogude ontoloogiaid, siis võiks selles hallata ka teisi riigi IS andmete linkimiseks vajalikke originaalselt loodud ontoloogiaid (nt valdkonnüleseid taaskasutatavaid ontoloogiaid, rakenduste poolt lingitavate andmete mudelitele vastavaid ontoloogiaid jm)
- Riik peaks tagama avalike linkandmete publitseerimisel järgmist:
 1. Avatud litsentsiga andmed.
 2. Otsene juurdepääs toorandmetele (vähemalt CSV formaadis, soovitatavalt RDF formaadis (vt USA õppetunnid)).
 3. Metaandmete standardite (pigem üldiste kui valdkondlike) kasutamine.
 4. Linkandmete versioneerimine ja toorandmete pidev uuendamine.
 5. Riigiasutused peaksid koguma linkandmete rakenduste kohta tagasisidet kodanikelt (näiteks koosloome saitide kaudu). See on oluline selleks, et riigiasutused saaksid spetsifitseerida uusi avalikke linkandmeid publitseerimiseks ja (taas)kasutamiseks.
 6. Kodanike aktiivseks kaasamiseks avalike linkandmete rakenduste loomisele võiks luua neile sobiva linkandmete infrastruktuuri (nt Openlink Virtuoso baasil).

8 Kasutatud kirjandus

1. 5 Stars of Linked Open Data <http://5stardata.info>
2. AKSW <http://aksw.org/Projects/TripleCheckMate.html>
3. Andmekaitse Inspeksioon. Andmekogude juhend
https://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/Andmekogude%20juhend%202013_o.rtf
4. Andmekaitse ja teadustöö. Sirp, 25.01.2013
http://www.sirp.ee/index.php?option=com_content&view=article&id=17020:andmekaitse-ja-teadustoeoe&catid=9:sotsiaalia&Itemid=13&issue=3426
5. Asset Description Metadata Schema, ADMS, <https://joinup.ec.europa.eu/asset/adms/release/all>
6. Autoriõiguse seadus, <https://www.riigiteataja.ee/akt/114062013005>
7. Avaandmete roheline raamat, AARR, Avaliku teabe avamine taaskasutamiseks, versioon 0.4 2013-11-21, Eesti Vabariigi valitsus, <http://www.riso.ee/et/avaandmete-roheline-raamat>
8. Avaliku teabe seadus: <https://www.riigiteataja.ee/akt/119122012005?leiaKehtiv>
9. BBC <http://www.bbc.co.uk/2012/>
10. BBC http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html
11. BBC <http://www.bbc.co.uk/nature/wildlife>
12. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American, 284(5), 28{37 (2001)
13. Berni kirjandus- ja kunstiteoste kaitse konventsiooniga ühinemise seadus,
<https://www.riigiteataja.ee/akt/24662>. Berni kirjandus- ja kunstiteoste kaitse konventsioon ,
<https://www.riigiteataja.ee/akt/13101723>
14. BIG – Big Data Public Private Forum <http://big-project.eu/>
15. BIG – Big Data Public Private Forum http://big-project.eu/sites/default/files/D2.4.1_FINAL_vo_8.pdf
16. Bill Goodwin, US and Canada follow UK initiative on Open Data, 28.10.2013, Computer Weekly,
<http://www.computerweekly.com/news/2240207982/US-and-Canada-follow-UK-initiative-on-Open-Data>
17. Bizer, Heath & Berners-Lee, Linked Data—The Story So Far“. International Journal on Semantic Web and Information Systems 5 (3): 1–22. doi:10.4018/jswis.2009081901.
18. Brabham Daren, “Crowdsourcing as a Model for Problem Solving: An Introduction and Cases“
http://www.clickadvisor.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf Convergence: The International Journal of Research into New Media Technologies 14 (1): 75–90, 2008
19. Brabham Daren, “Crowdsourcing as a Model for Problem Solving: An Introduction and Cases“, Convergence: The International Journal of Research into New Media Technologies 14 (1): 75–90, 2008
20. Cabinet Office. Government ICT Strategy, 2011.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/85968/uk-government-government-ict-strategy_o.pdf
21. Cabinet Office. Policy Paper. G8 Open Data Charter and Technical Annex
<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>
22. Capgemini Group, www.campegmini.com
23. CensusHub 2013. https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Census_Hub
24. Central Intelligence Agency <https://www.cia.gov/news-information/speeches-testimony/2012-speeches-testimony/in-q-tel-summit-remarks.html>
25. Chief Technology Officer Council. Designing URI Sets for the UK Public Sector
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-uri-sets-uk-public-sector.pdf
26. Data.dcs: Converting Legacy Data into Linked Data
http://events.linkedata.org/ldow2010/papers/ldow2010_paper01.pdf
27. DATA.GOV.UK <http://data.gov.uk/blog/linked-data-pilot-implementations-update>
28. DATA.GOV.UK <http://data.gov.uk/library/flying-binary>
29. DATA.GOV.UK <http://data.gov.uk/library/geowise>
30. DATA.GOV.UK <http://data.gov.uk/library/harvey-walsh>
31. DATA.GOV.UK <http://data.gov.uk/library/mudlark>
32. DATA.GOV.UK <http://data.gov.uk/library/swirrl>
33. DATA.GOV.UK <http://data.gov.uk/linked-data/UKGovLD>
34. DATA.GOV.UK <http://data.gov.uk/resources/uris>

35. Datacatalogs.org <http://datacatalogs.org/>
36. Datahub <http://datahub.io/>
37. Datahub <http://datahub.io/dataset/2000-us-census-rdf>
38. DataMarket <http://datamarket.com/>
39. DBpedia <http://dbpedia.org/>
40. DBpedia <http://dbpedia.org/Architecture>
41. DBpedia <http://wiki.dbpedia.org/DBpediaLive>
42. DBpedia <http://wiki.dbpedia.org/OnlineAccess>
43. DBpedia Live <http://wiki.dbpedia.org/About>
44. DeciderMail <http://deciderlab.com/>
45. DG Health & Consumers is a DG of the European Commission,
http://ec.europa.eu/dgs/health_consumer/about_us/who_we_are_en.htm
46. Digital agenda for Europe <https://ec.europa.eu/digital-agenda>
47. Digital agenda for Europe <https://ec.europa.eu/digital-agenda/en/european-egovernment-action-plan-2011-2015>
48. Digital agenda for Europe <https://ec.europa.eu/digital-agenda/en/open-data-portals>
49. Digital agenda for Europe <https://ec.europa.eu/digital-agenda/en/public-sector-information-raw-data-new-services-and-products>
50. Digital agenda for Europe <https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/H2020%20LEIT-ICT%20WP%202014-15%20-%2031%2010%202013.pdf>
51. Dimitris Kontokostas, Sören Auer and Jens Lehmann, Crowdsourcing Linked Data quality assessment, http://jens-lehmann.org/files/2013/iswc_crowd_qa.pdf
52. Ding, L., Peristeras V., Hausenblas M., „Linked Open Government Data,“ IEEE Intelligent Systems, vol. 27, no. 3, pp. 11-15, May-June 2012
53. Directorate-General – DG, <http://ec.europa.eu>
54. Drupal <https://drupal.org/project/neologism>
55. Eesti 2020 tegevuskava 2011-2015, <http://valitsus.ee/et/riigikantslei/eesi2020>
56. Eesti geoportaal <http://inspire.maaamet.ee/>
57. Eesti infoühiskonna arengukava 2013, p.4.3.1, Majandus- ja Kommunikatsiooniministeerium, Tallinn 2009
58. Eesti infoühiskonna arengukava 2020 (heaks kiidetud Vabariigi Valitsuse poolt 14.11.2013). Majandus- ja Kommunikatsiooniministeerium, <http://263654.edicypages.com/eesi-infohiskonna-arengukava-2020/infohikonna-arengukava-2020-loppversioon>
59. Ehitisregistri avaandmed - <http://opendata.riik.ee:8080/dataset/ehitisregistri-avaandmed>
60. EL andmekaitse direktiiv. 1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=31995L0046&m odel=guichett (22.03.2013). EL uus andmekaitse direktiiv. 2012. European new data protection law. http://www.mlawgroup.de/news/publications/detail.php?we_objectID=227 (20.03.2013)
61. EL struktuuritoetus <http://www.struktuurifondid.ee/infohiskond-2/>
62. Electronic Simple European Networked Services <http://www.esens.eu/home.html>
63. Elena Paslaru, Bontas Simperl, Christoph Tempich, York Sure. ONTOCOM: A cost estimation model for ontology engineering (2006), Proceedings of fifth ISWC, 2006.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.2432&rep=rep1&type=pdf>
64. ePractice <http://epractice.eu>
65. ePractice <http://www.epractice.eu/en/factsheets/>
66. ePractice http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf
67. ePSIplatform <http://epsiplatform.eu/>
68. ePSIplatform <http://epsiplatform.eu/content/european-psi-scoreboard>
69. ePSIplatform <http://epsiplatform.eu/topicreports>
70. ePSIplatform <http://www.epsiplatform.eu/sites/default/files/Final%20TR%20Europes%20Data%20Catalogs.pdf>
71. ePSIplatform <http://www.epsiplatform.eu/sites/default/files/Final%20TR%20Linked%20Data.pdf>
72. EUROOPA PARLAMENDI JA NÕUKOGU DIREKTIIV 2013/37/EL, 26. juuni 2013, millega muudetakse direktiivi 2003/98/EÜ avaliku sektori valduses oleva teabe taaskasutamise kohta
73. Euroopa tulevikuinterneti portal <http://www.future-internet.eu/>
74. European Commission <http://ec.europa.eu/dgs/connect/>

75. European Commission <http://ec.europa.eu/digital-agenda/en/news/economic-analysis-psi-impacts>
76. European Commission http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-15action_en.htm
77. European Commission http://ec.europa.eu/isa/actions/index_en.htm
78. European Commission http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf
79. European Commission http://ec.europa.eu/isa/news/2013/the-benefits-of-geo-location-services-for-egovernment_en.htm
80. European Commission http://ec.europa.eu/research/horizon2020/index_en.cfm
81. European Commission http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=h2020-documents
82. Eurostat <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>
83. Eurostat http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database
84. Eurostat http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/National_accounts_and_GDP
85. Eurostat <http://eurostat.linked-statistics.org/>
86. Executive Order – Making Open and Machine Readable the New Default for Government Information, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-defaultgovernment>, <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>
87. FP7 Information and Communication Technologies Work Programme, <http://cordis.europa.eu/fp7/ict/>
88. Fujitsu White Paper, March 2012. <http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-%28v1.0%29.pdf>
89. G8. Policy paper. G8 Open Data Charter and Technical Annex. Published 18 June 2013. <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>
90. Gartner http://www.gartner.com/it/content/1258400/1258425/january_6_techrends_rpaquet.pdf
91. Gartner <http://www.gartner.com/newsroom/id/2575515>
92. Gartner <http://www.gartner.com/newsroom/id/2593815>
93. Gartner <http://www.gartner.com/newsroom/id/2603215>
94. Gartner Top Predictions 2014: Plan for a Disruptive, but Constructive Future, Gartner 2013
95. Gartner, Gartner Says Potential Size and Diversity of the Internet of Things Mask Immediate Opportunities, <http://www.gartner.com/newsroom/id/2564916>, Gartner 2013
96. Gartner, Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data. <http://www.gartner.com/it/page.jsp?id=1731916>, Gartner 2012
97. Gartner, Hype Cycle for Big Data, 2012, Published: 31 July 2012
98. GeoLinked Data <http://geo.linkeddata.es/>
99. Getting Results from Crowds <http://www.resultsfromcrowds.com>
100. Getting Results from Crowds <http://www.resultsfromcrowds.com/features/crowd-business-models/>
101. Gómez-Pérez, Asunción, Mariano Fernández-López and Oscar Corcho. 2004. Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer Heidelberg.
102. GOV.UK ICT strategy resources <https://www.gov.uk/government/collections/ict-strategy-resources>
103. Haav, H.-M. 2010. Nõuded RIHA ontoloogiatele. (Requirements for ontologies in RIHA) http://www.ria.ee/public/RIHA/nouded_riha_ontoloogiatele_r26.PDF (accessed February 03, 2013).
104. Haav, H.-M. 2010. Ontoloogiate loomise meetoodika. (A methodology for ontology development) http://www.ria.ee/public/RIHA/ontoloogiate_loomise_meetoodika_v4.PDF (accessed February 03, 2012).
105. Haav, H.-M. 2011. A practical methodology for development of a network of e-government domain ontologies. In Building the e-World Ecosystem: 11th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2011, Revised Selected Papers, ed. T. Skersys, R. Butleris, L. Nemuraite, and R. Suomi, 113, Springer Heidelberg.
106. Hadapt <https://hadapt.com/>
107. Hanga, T. (Kas) riik on IPv6 kasutuselevõtuks valmis?, 05.06.2012 – IPv6 Forum, IEEE Eesti sektsioon
108. Helmut Nagy, Thomas Thurner, Martin Kaltenböck. Analyses of the LOD2 – OGD Survey. <http://wiki.lod2.eu/download/attachments/3473453/AnalysisSurvey.pdf?version=1&modificationDate=1329941495000>.
109. HM Government. A Concept Model for the UK Public Sector https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266279/A-Concept-Model-for-the-UK-Public-Sector.pdf

110. HM Government. Information Principles https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266284/Information_Principles_UK_Public_Sector_final.pdf
111. HM Government. UK Government Reference Architecture (UKRA) https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266257/UK-Reference-Architecture-V1-o-HMG-Branded.pdf
112. HP <http://www8.hp.com/us/en/hp-information/environment/cense.html#.Uue8oBD8LRY>
113. HTML Living Standard <http://www.whatwg.org/specs/web-apps/current-work/multipage/microdata.html>
114. <http://ambtenaar20.ning.com/page/welkom-op-ambtenaar-20>
115. <http://answers.semanticweb.com/questions/14029/are-semantic-technologies-ready-for-commercial-use>
116. <http://apps.dri.ie/locationLODer/>
117. <http://arstechnica.com/business/2012/05/open-government-reboot-focuses-on-apis-instead-of-data/>
118. <http://blog.schema.org/2012/11/good-relations-and-schemaorg.html>
119. <http://blog.semantic-web.at/2013/06/04/theres-money-in-linked-data/>
120. http://ceur-ws.org/Vol-628/ldow2010_paper14.pdf
121. http://ceur-ws.org/Vol-782/SimperlEtAl_COLD2011.pdf
122. <http://creativecommons.org>
123. <http://dailycrowdsourcing.com/training/crowdsourcing/what-is-crowdsourcing>
124. <http://data.europeana.eu>
125. <http://ecb.publicdata.eu>
126. <http://euclid-project.eu/>
127. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF>
128. <http://eur-lex.publicdata.eu/>
129. <http://gsablogs.gsa.gov/dsic/about/>
130. <http://hadoop.apache.org>
131. <http://hive.apache.org>
132. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>
133. http://ifgi.uni-muenster.de/~j_ortm02/publications/OrtmannetalTC2011.pdf
134. http://jens-lehmann.org/files/2013/iswc_crowd_qa.pdf
135. <http://kirka.ee> http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf
136. <http://lastehoid.net>
137. <http://latc-project.eu/datasets>
138. <http://linkeddata.future-internet.eu/images/3/3f/DeckerIntro.pdf>
139. http://linkeddata.future-internet.eu/index.php/Main_Page
140. <http://linkedgeodata.org/>
141. <http://linkedopendata.gr/>
142. <http://lists.w3.org/Archives/Public/public-lod/2009Apr/0162.html>
143. <http://logd.tw.rpi.edu/>
144. <http://microformats.org/about>
145. <http://money.cnn.com/2013/04/08/technology/security/shodan/>
146. <http://nebula-fia.org/>
147. <http://opencorporates.com/>
148. <http://open-data.europa.eu/>
149. <http://opendata.riik.ee/juhendid/ettepanekud-edasisteks-arendusteks>
150. <http://opendefinition.org/>
151. <http://planet-data.eu/>
152. <http://portaal.eki.ee/uuedsonad/Uued-s%C3%B5nad-1/V/vabasektor-10/>
153. <http://projects.apache.org/projects/jena.html>
154. <http://publications.europa.eu/>
155. <http://research.google.com/archive/mapreduce.html>
156. <http://schema.org/docs/gs.html>
157. http://semanticweb.com/sindicetech-helps-enterprises-build-private-linked-data-clouds_b30454
158. <http://sindice.com/>
- Library of Congress Linked Data Service <http://id.loc.gov/>
159. <http://thedataweb.rm.census.gov/TDW.html>
160. <http://theodi.org/case-studies/opencorporates-case-study>
161. <http://theodi.org/news/odi-startup-demand-logic-saves-king-s-college-london-390000-year-energy-costs>

162. <http://thewebindex.org/>
163. <http://umusic.co.uk/artists/>
164. <http://venturebeat.com/2013/10/09/ibm-launches-research-lab-for-big-data-applications/>
165. <http://whitedb.org>
166. <http://virtuoso.openlinksw.com/rdf-quad-store/>
167. <http://vocab.data.gov/VoID>
168. <http://www.abouthisday.com/>
169. <http://www.aifb.kit.edu/web/Inproceedings3284>
170. <http://www.apache.org/foundation/>
171. http://www.apnic.net/community/ipv6-program/ipv6-for-decision-makers/ipv6_for_decision_makers.pdf
172. <http://www.beamtoothbrush.com/toothbrush/>
173. <http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential>
174. <http://www.cam.ac.uk/research/news/future-internet-aims-to-sever-links-with-servers>
175. http://www.cio.gov/documents/Shared_Services_Strategy.pdf
176. http://www.clickadvisor.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf
177. <http://www.cloudera.com>
178. <http://www.cmswire.com/cms/customer-experience/semantic-technology-the-future-of-search-for-digital-marketers-022209.php>
179. http://www.computerworld.com/s/article/9230801/Walmart_rolls_out_semantic_search_engine_sees_business_boost
180. <http://www.constitutionproject.org/pdf/DataMiningPublication.pdf>
181. <http://www.crowdsourcing.org/editorial/eyekas-francois-petavy-five-predictions-for-crowdsourcing-in-2014/30116>
182. <http://www.dailymail.co.uk/sciencetech/article-2115871/The-CIA-wants-spy-TV-Agency-director-says-net-connected-gadgets-transform-surveillance.html>
183. <http://www.data-forum.eu/>
184. <http://www.devttyso.com/2013/10/reverse-engineering-a-d-link-backdoor/>
185. <http://www.dri.ie/projects>
186. <http://www.earthobservatory.eu/>
187. <http://www.ebaltics.com/00904523>
188. <http://www.envision-project.eu/>
189. <http://www.evs.ee/Standardimine/Eestistandarditearv/tabid/96/Default.aspx>
190. <http://www.fp7-pursuit.eu/>
191. <http://www.franz.com/>
192. <http://www.garlik.com/>
193. http://www.ibm.com/smarterplanet/us/en/business_analytics/article/cognitive_computing.html
194. <http://www.iks-project.eu/projects/apache-stanol>
195. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020
196. <http://www.iso.org/iso/fr/home/standards/benefitsofstandards/benefits-detail.htm?emid=7>
197. http://www.itworld.com/slideshow/129283/how-cops-and-robbers-are-using-google-earth-383397?source=ITWNLE_nlt_tonight_2013-11-19
198. <http://www.legislation.gov.uk>
199. <http://www.letsdoitworld.org>
200. <http://www.marklogic.com>
201. <http://www.marklogic.com/solutions/government/>
202. <http://www.networkworld.com/slideshow/114134/15-most-powerful-big-data-companies.html#slide1>
203. <http://www.ntia.doc.gov/legacy/ntiahome/ntiageneral/ipv6/final/IPv6final2.htm>
204. <http://www.nuigalway.ie/about-us/news-and-events/news-archive/2012/april2012/us-government-joins-the-dots-with-irish-linked-data-technologies-1.html>
205. <http://www.ontotext.com>
206. <http://www.ontotext.com/news/gartner-predicts-exciting-2013-for-semantic-tech>
207. <http://www.opendataresearch.org/barometer>
208. <http://www.opengovpartnership.org/>
209. <http://www.openlinksw.com>
210. <http://www.optique-project.eu/>
211. <http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/rdfsemantic-graph-1902016.html>

212. <http://www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html>
213. <http://www.platfora.com>
214. <http://www.productontology.org/>
215. <http://www.rdfabout.com/demo/census/sparql.xpd>
216. <http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf>
217. <http://www.sdmx.org>
218. <http://www.seco.tkk.fi/events/2014/2014-01-24-ldf/>
219. <http://www.seco.tkk.fi/events/2014/2014-01-24-ldf/ap/frosterus-koko.pdf>
220. <http://www.sindicetech.com/>
221. http://www.sis.se/pdf/Economic_impact_of_standardization_France.pdf
222. <http://www.sisense.com/>
223. <http://www.slideshare.net/3groundstones/progress-update-on-government-linked-data-world>
224. http://www.slideshare.net/fabien_gandon/w3c-tutorial-on-semantic-web-and-linked-data-at-www-2013
225. <http://www.slideshare.net/onlyjiny/linkeddata>
226. <https://www.zone.ee/et/2013/06/06/zone-ee-kaivitas-ipv6-toe/>
227. IBM <http://www.research.ibm.com/client-programs/accelerated-discovery-lab/index.shtml>
228. IBM <http://www.research.ibm.com/cognitive-computing/index.shtml#fbid=-pHjUq5w8zr>
229. IBM <http://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml#fbid=-pHjUq5w8zr>
230. IBM <http://www.research.ibm.com/cognitive-computing/why-cognitive-systems.shtml#fbid=-pHjUq5w8zr>
231. IBM <http://www-03.ibm.com/innovation/us/watson/>
232. IBM. SyNAPSE: a cognitive computing project from IBM research. 2011; Available from:
233. Intellektuaalomandi õiguste kaubandusaspektide leping, <https://www.riigiteataja.ee/akt/79299>
234. Interactive Konowledge Stack, <http://www.iks-project.eu/>
235. IoT6 European research project, Deliverable D1.5: „IoT6 Reference Model”, White Paper(2012), <http://www.iot6.eu>
236. IQT <https://www.iqt.org/>
237. Isikuandmete kaitse seadus
238. Iteraction <http://iteration.ee/>
239. Joinup <https://joinup.ec.europa.eu>
240. Joinup <https://joinup.ec.europa.eu>
241. Joinup <https://joinup.ec.europa.eu/community/cesar/description>
242. Joinup <https://joinup.ec.europa.eu/community/osor/description>
243. Joinup <https://joinup.ec.europa.eu/community/semic/description>
244. Joinup <https://joinup.ec.europa.eu/community/semic/document/case-study-how-linked-data-transforming-e-government>
245. Joinup <https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>
246. Joinup <https://joinup.ec.europa.eu/community/semic/highlights/asset/all>
247. Joinup <https://joinup.ec.europa.eu/news/isa-member-state-representatives-endorse-key-specifications-e-government-interoperability>
248. Joinup. How Linked Data is transforming eGovernment https://joinup.ec.europa.eu/sites/default/files/D4.3.2_Case_Study_Data_Integration_vo.15.pdf
249. Justiitsministeerium <http://www.korruptsioon.ee/orb.aw/class=file/action=preview/id=58998/Strateegia+rakendusplaan.pdf>
250. K. Bereta, P. Smeros and M. Koubarakis. Representing and Querying the Valid Time of Triples for Linked Geospatial Data. In the 10th Extended Semantic Web Conference (ESWC 2013). Montpellier, France. May 26-30, 2013.
251. Karolis Granickas, Understanding the Impact of Releasing and Re-using Open Government Data, European Public Sector Information Platform, Topic Report No. 2013 / 08, http://www.epsiplatform.eu/sites/default/files/2013-08-Open_Data_Impact.pdf
252. Keskkonnainfo <http://www.keskkonnainfo.ee/>
253. Kevin Ashton: That 'Internet of Things' Thing. In: RFID Journal, 22 July 2009. <http://www.rfidjournal.com/articles/view?4986>
254. Kinnistusraamatuseadus, <https://www.riigiteataja.ee/akt/102052013002>
255. Kohalike omavalitsuste raamatupidamise avaldamise seadus <http://www.riigipilv.ee/>
256. Kotkas, Vahur; Haav, Hele-Mai; Tepandi, Jaak; Õunapuu, Enn; Grauberg, Jaanus (2013). Uurimisprojekti „Andmeaitade (teiseste andmekogude) loomise põhimõtete väljatöötamine“

- lõpparuanne. Tallinn: TTÜ Küberneetika Instituut
<http://valitsus.ee/UserFiles/valitsus/et/riigikantselei/strateegia/poliitika-analuusid-ja-uuringud/tarkade-otsuste-fondi-uuringute-kokkuvotted/L%C3%B5pparuanne.pdf>
257. Koubarakis, M., et al.: Real-time wildfire monitoring using scientific database and linked data technologies. In: In the 16th International Conference on Extending Database Technology (EDBT 2013). Genoa, Italy (March 18-22 2013)
 258. Küngas, P. 2010. Semantilise kirjeldamise juhised vo.4. (Guidelines for semantic enrichment of data and services) http://ftp.ria.ee/pub/riha/Semantilise_kirjeldamise_juhis_vo4.pdf (accessed February 01, 2013).
 259. Küngas, P. 2010. Semantilise kirjeldamise juhised vo.4. (Guidelines for semantic enrichment of data and services) http://ftp.ria.ee/pub/riha/Semantilise_kirjeldamise_juhis_vo4.pdf (accessed February 01, 2013).
 260. L. Ding et al. TWC LOGD: A Portal for Linked Open Government Data Ecosystems. Journal of Web Semantics 9(3) (2011).
 261. Lebo et al. Producing and Using Linked Open Government Data. In D. Wood (ed.) Linking Government Data, Springer, 2011
 262. Legal barriers to linked open data <http://linkedup-project.eu/2013/05/23/legal-barriers-to-linked-open-data/>
 263. Li Ding, et al (2011). TWC LOGD: A Portal for Linked Open Government Data Ecosystems. Journal of Web Semantics: Science, Services and Agents on the World Wide Web (in press).
<http://dx.doi.org/10.1016/j.websem.2011.06.002>
 264. Library of Congress Linked Data Service <http://id.loc.gov/>
 265. Linked Data - Connect Distributed Data across the Web <http://linkeddata.org/>
 266. Linked Data - Connect Distributed Data across the Web <http://linkeddata.org/guides-and-tutorials>
 267. Linked Data API, <http://code.google.com/p/linked-data-api>
 268. Linked Data service of the DNB, <http://www.dnb.de/EN/lds>
 269. Linked data, <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#linked-data>
 270. Linked Open Data Around-The-Clock, <http://latc-project.eu/>
 271. Linked Open Government Data. Li Ding Qualcomm, Vassilios Peristeras and Michael Hausenblas. IEEE Intelligent Systems, May/June 2012 (Vol. 27, No. 3) pp. 11-15 1541-1672/12/© 2012 IEEE <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>
 272. Linked Open Government Data. Li Ding Qualcomm, Vassilios Peristeras and Michael Hausenblas. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>
 273. Linking data about applications and decisions for authorisation of plant protection products, <http://health.testproject.eu/PPP/>
 274. Linking UK Government Data http://events.linkeddata.org/ldow2010/papers/ldow2010_paper14.pdf
 275. Listpoint 2012. Listpoint open datasurvey, <http://www.publictechnology.net/news/open-datas-potential-remains-closed-public-sector-staff/37572>
 276. LOD2 <http://data.lod2.eu/2010/ogd-survey/>
 277. LOD2 <http://lod2.eu/>
 278. LOD2
<http://wiki.lod2.eu/download/attachments/3473453/AnalysisSurvey.pdf?version=1&modificationDate=1329941495000>
 279. LOD2 <http://www.slideshare.net/lod2project/the-semantic-data-web-sren-auer-university-of-leipzig>
 280. LOD2 Stack <http://stack.lod2.eu/>
 281. Läbipaistev Eesti Omavalitsus
<http://www.riigipilv.ee/QvAJAXZfc/opensdoc.htm?document=LEO.qvw&host=local&anonymous=true>
 282. M. G. Zuiderwijk, et al., „The necessity of metadata for open linked data and its contribution to policy analyses,” presented at the Conference on E-Democracy and Open Government (CeDEM12), Krems, Austria, 2012
 283. Maa-amet <http://geoportaal.maaamet.ee>
 284. Maakatastriseadus, <https://www.riigiteataja.ee/akt/13335388>
 285. Majandus- ja Kommunikatsiooniministeerium. Uuring „E-äri ja e-kaubanduse kasutamine Eestis ja laiendamise võimalused“ http://www.mkm.ee/public/Lopparuaranne_-_E-ari_ja_e-kaubandus_1_6_avalik_2013.pdf
 286. Manyika, J., M. Chui, B. Brown, B. et al. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

287. MICUS Management Consulting report on –Assessment of the Re-use of Public Sector Information (PSI) in the Geographical Information, Meteorological Information and Legal Information Sectorsl, published March 2009, www.micus.de/pdf/MICUS-Studie_PSI_EU_March_2009.pdf
288. MIT Technology Review <http://www.technologyreview.com/news/514066/what-happened-when-one-man-pinged-the-whole-internet/>
289. Mitchell, I., Wilson, M., 2012. Linked Data. Connecting and exploiting big data. Fujitsu White Paper, March 2012. <http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-%28v1.0%29.pdf>
290. Mitchell, I., Wilson, M., 2012. Linked Data. Connecting and exploiting big data.
291. Mäehans, H. Registriandmete avalikustamisest Ehitisregistri näitel. Bakalaureusetöö, Informaatika instituut, Tallinna Ülikool, 2012
292. Nikolaos Loutas, Christophe Colas, Stijn Goedertier, Linking data about applications and decisions for authorisation of plant protection products, PwC EU Services, 2013, https://joinup.ec.europa.eu/asset/core_business/document/linking-data-about-applications-and-decisions-authorisation-ppp
293. Nikolaou, K. Dogani, K. Kyzirakos and M. Koubarakis. Sextant: Browsing and Mapping the Ocean of Linked Geospatial Data. In the 10th Extended Semantic Web Conference (ESWC 2013). Montpellier, France. May 26-30, 2013. Demo paper.
294. Noor Huijboom, Tijs Van den Broek, Open data: an international comparison of strategies, European Journal of ePractice, www.epracticejournal.eu N° 12 March/April 2011, http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf
295. NY Times http://www.nytimes.com/2013/11/23/us/politics/tension-and-woes-before-health-website-crash.html?_r=0
296. Obama 2011. B. Obama, Technology, <http://www.whitehouse.gov/issues/technology>
297. Obama, Technology, <http://www.whitehouse.gov/issues/technology>, 2011
298. OECD <http://www.oecd.org>
299. OECD <http://www.oecd.org/eco/surveys/>
300. OECD <http://www.oecd.org/estonia/>
301. OECD <http://www.oecd.org/statistics/>
302. OECD http://www.oecd-ilibrary.org/governance/government-at-a-glance-2013_gov_glance-2013-en
303. Open Data Challenge Website <http://opendatachallenge.org/>
304. Open Data on The Web, 2013, Report, <http://www.w3.org/2013/04/odw/report#closed>
305. Open Data Research Network, <http://www.opendataresearch.org/>
306. Open Data White Paper – Unleashing the Potential, HM UK Government, June 2012, http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf
307. Open Geospatial Consortium: OGC GeoSPARQL - A geographic query language for RDF data. OGCR Implementation Standard (2012) GeoSPARQL <http://www.geosparql.org/>
308. Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) Systematic Mapping Studies in Software Engineering. In Proceedings of 12th International Conference on Evaluation and Assessment in Software Engineering.
309. Phil Archer, Makx Dekkers, Stijn Goedertier, Nikolaos Loutas, Study on Business Models for Linked Open Government Data - BM4LOGD, PwC EU Services, <https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>
310. Postimees <http://e24.postimees.ee/1225740/erki-kerdi-firma-hakkab-sotsiaalmeedia-pohjal-laenuvoimet-hindama>
311. Postimees <http://e24.postimees.ee/2054986/uuring-kolmveerand-balti-ettevotetest-leiab-et-suurandmed-aitavad-paremaid-otsuseid-teha>
312. Protokoll Eesti ühinemise kohta Maailma Kaubandusorganisatsiooni Asutamislepinguga (Marrakeši lepinguga), <https://www.riigiteataja.ee/akt/79272>
313. PSI-directive 2003. DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information, http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf
314. PwC, Harnessing the Power of Crowdsourcing, PwC Technologyforecast, 2011
315. PwC, Internet of Things: Evolving transactions into relationships, PwC Technologyforecast. Issue 1, 2013
316. PwC, Spinning a data Web. Making Semantic Web connections, PwC Technologyforecast, Spring 2009

317. PwC. Big Data Analysis A CFO's Toolkit
318. PwC. Big Data for Financial Services. BCM Learning Insights (2011)
319. PwC. Decoding innovation's DNA
320. PwC. Technology Forecast: Making sense of Big Data (2010, Issue 3)
321. Rahvastikuregistri seadus, <https://www.riigiteataja.ee/akt/114032011005>
322. Rahvastikuregistrist andmete väljastamise hindade kinnitamine. Regionalministri määrus, RT I, 10.11.2010, 15, <https://www.riigiteataja.ee/akt/110112010015>
323. RIA <https://www.ria.ee/algavad-tark-e-riik-koolitused/>
324. RIA https://www.ria.ee/ipv6-koolituspraktikum/?op=training_detailview
325. Rich Quinnell, Predicting 2014's IoT Trends, http://www.iotworld.com/author.asp?section_id=3150&doc_id=562274
326. Riigi Infosüsteemi Ameti põhimäärus <https://www.riigiteataja.ee/akt/128042011001?leiaKehtiv>
327. Riigi infosüsteemi haldussüsteem <https://www.riigiteataja.ee/akt/13147268>
328. Riigi maakatastri pidamise korra kinnitamine, <https://www.riigiteataja.ee/akt/24854>
329. Riigi Teataja <https://www.riigiteataja.ee/lyhendid.html?sorteeri=pealkiri&kasvav=true>
330. Riiklik Ehitisregister <http://ehitisregister.ee/sparql>
331. Riikliku statistika seadus, <https://www.riigiteataja.ee/akt/13338093>
332. RISO . Infoühiskonna arengukava 2020 http://www.riso.ee/sites/default/files/elfinder/article_files/infoyhiskonna_arengukava_2020_f.pdf
333. RISO <http://www.riso.ee/et/koosvoime/raamistik>
334. RISO <http://www.riso.ee/et/koosvoime/semantika>
335. RISO <http://www.riso.ee/et/koosvoime/veebide-raamistik.odt>
336. RISO http://www.riso.ee/sites/default/files/elfinder/article_files/RISsemantikaVo7-loplik.pdf
337. Ruumiandmete seadus, <https://www.riigiteataja.ee/akt/128022011002>
338. S. Auer, L. Bühmann, C. Dirschl, O. Erling, M. Hausenblas, R. Isele, J. Lehmann, M. Martin, P. N. Mendes, B. Van Nuffelen, C. Stadler, S. Tramp, H. Williams. Managing the Life-Cycle of Linked Data with the LOD2 Stack. In Proc. International Semantic Web Conference 2, (7650):1-16, Springer, 2012. http://svn.aks.org/iod2/Paper/ISWC2012-InUse_LOD2-Stack/public.pdf
339. S. Speicher, J. Arwe, A. Malhotra. Linked Data Platform 1.0, W3C Last Call Working Draft 30 July 2013. <http://www.w3.org/TR/ldp/>.
340. Sébastien Ziegler, et al, IoT6 – Moving to an IPv6-Based Future IoT. Galis and A. Gavras (Eds.): FIA 2013, LNCS 7858, pp. 161–172, 2013.
341. Shadbolt, Nigel, O'Hara, Kieron, Berners-Lee, Tim, Gibbins, Nicholas, Glaser, Hugh, Hall, Wendy and schraefel, m.c. (2012) Linked open government data: lessons from Data.gov.uk. IEEE Intelligent Systems, 27, (3), Spring Issue, 16-24. (doi:10.1109/MIS.2012.23).
342. Space-Time Insight <http://www.spacetimeinsight.com>
343. Statistikaameti andmebaas <http://pub.stat.ee/px-web.2001/dialog/statfile2.asp>
344. stSPARQL Reference <http://www.strabon.di.uoa.gr/stSPARQL>
345. Zettaset <http://www.zettaset.com/>
346. T. Anderson et al, The NEBULA Future Internet Architecture , Galis and A. Gavras (Eds.): FIA 2013, LNCS 7858, pp. 16–26, 2013
347. T. Heath, C. Bizer. Linked Data: Evolving the Web into a Global Data Space, Morgan & Claypool Publishers, 2011. <http://www.uni-koblenz-landau.de/campus-koblenz/fb4/web/teaching/ws1213/seminar-web-science/linked-data.pdf>
348. Tassilo Pellegrin, Economics of Big Data: A Value Perspective on State of the Art and Future Trends, In: Big Data Computing, R. Akerkar (Ed), pp 344-367, 2014 by Taylor & Francis Group, LLC
349. Tech Trends 2014, Deloitte, 2014, <http://dupress.com/articles/2014-tech-trends-crowdsourcing/?coll=6210>
350. Teeme ära! <http://www.teemeara.ee/et>
351. Telecoms Tech <http://www.telecomstechnews.com/news/2013/dec/19/what-are-deutsche-telekoms-m2m-predictions-2014/>
352. Tervise infosüsteemi põhimäärus. Vabariigi Valitsuse määrus. Vastu võetud 14.08.2008 nr 131, <https://www.riigiteataja.ee/akt/13251011>
353. The Digital Agenda of the EU, <http://ec.europa.eu/digital-agenda/>
354. The Directive on the re-use of public sector information (Directive 2003/98/EC tuntud kui 'PSI Directive' <https://ec.europa.eu/digital-agenda/en/legislative-measures>
355. The Government Linked Data (GLD) Working Group, <http://www.w3.org/2011/gld>
356. The Human Brain Project <https://www.humanbrainproject.eu/>

357. The Human Brain Project
<https://www.humanbrainproject.eu/documents/10180/17646/Vision+Document/8bb75845-8b1d-41e0-bcb9-d4de69eb6603>
358. The Interoperability Solutions for European Public Administrations, ISA, ISA teabevärv
<http://ec.europa.eu/isa/>
359. The Open Data Institute (ODI) <http://www.theodi.org/>
360. The White House <http://www.whitehouse.gov/omb/e-gov>
361. The White House <http://www.whitehouse.gov/open>
362. The White House <http://www.whitehouse.gov/open/documents/open-government-directive>
363. The White House <http://www.whitehouse.gov/share/brain-initiative>
364. The White House http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/s200.pdf
365. The White House http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/federal-cloud-computing-strategy.pdf
366. The White House http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/25-point-implementation-plan-to-reform-federal-it.pdf
367. The White House
http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf
368. The White House <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>
369. The White House <http://www.whitehouse.gov/sites/default/files/omb/infoereg/for-agencies/informing-consumers-through-smart-disclosure.pdf>
370. The White House <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>
371. The White House http://www.whitehouse.gov/sites/default/files/uploads/2012digital_mem_rel.pdf
372. The White House <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>
373. Tim Davies, Open data barometer, 2013 Global Report, 31.october, 2013, ODI and Web Index,
<http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>
374. TNO
https://www.tno.nl/content.cfm?context=thema&content=prop_publicatie&laag1=897&laag2=919&laag3=123&item_id=919
375. Tomasz Janowski, Jeanne Holm, Elsa Estevez, Open Government Data W3C Discussion Summary,
http://www.w3.org/egov/wiki/images/5/5b/Open_government_data_5.pdf
376. Tsentraalseid portaale mitteomavad riigid puuduvad joonisel. Allikas:
<http://dx.doi.org/10.1787/888932942868>
377. TWC LOGD <http://logd.tw.rpi.edu/>
378. Tööstusomandi õiguskorralduse aluste seadus, <https://www.riigiteataja.ee/akt/128122011046>. Vt ka Tööstusomandi kaitse Pariisi konventsioon, <https://www.riigiteataja.ee/akt/13088459>.
379. Ubaldi, B. (2013), „Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives”, OECD Working Papers on Public Governance, No. 22, OECD Publishing.
<http://dx.doi.org/10.1787/5k46bj4f03s7-en>
380. UNdata <http://data.un.org>
381. W3C <http://dir.w3.org>
382. W3C <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
383. W3C <http://www.w3.org/>
384. W3C <http://www.w3.org/2005/Incubator/lld/wiki/Scope>
385. W3C <http://www.w3.org/2011/gld>
386. W3C <http://www.w3.org/2011/gld/charter>
387. W3C http://www.w3.org/2011/gld/wiki/228_Best_Practices_Pragmatic_Provenance
388. W3C http://www.w3.org/2011/gld/wiki/Best_Practices_Discussion_Summary
389. W3C http://www.w3.org/2011/gld/wiki/GLD_Life_cycle
390. W3C http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook
391. W3C http://www.w3.org/2012/06/pmod/pmod2012_submission_7.pdf
392. W3C <http://www.w3.org/Consortium/membership-faq#project>
393. W3C <http://www.w3.org/DesignIssues/LinkedData.html>
394. W3C <http://www.w3.org/egov>
395. W3C <http://www.w3.org/RDF/>
396. W3C <http://www.w3.org/standards/>

397. W3C <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>
398. W3C <http://www.w3.org/TR/2013/WD-ldp-ucr-20130131/#scope-and-motivation>
399. W3C <http://www.w3.org/TR/cooluris/>
400. W3C <http://www.w3.org/TR/ldp-ucr/>
401. W3C <http://www.w3.org/TR/owl-features/>
402. W3C <http://www.w3.org/TR/rdf-schema/>
403. W3C <http://www.w3.org/TR/rdf-sparql-query/>
404. W3C <http://www.w3.org/TR/uri-clarification>
405. W3C <http://www.w3.org/TR/xhtml-rdfa-primer/>
406. W3C http://www.w3.org/wiki/HCLSIG_BioRDF_Subgroup
407. W3C <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
408. W3C <http://www.w3c.org/RDF>
409. W3C <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>
410. W3C <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html>
411. W3C <https://dvcs.w3.org/hg/ldpwg/raw-file/default/ldp-bp/ldp-bp.html#bib-LDP-PRIMER>
412. Vabariigi Valitsus http://valitsus.ee/UserFiles/valitsus/et/riigikantselei/strateegia/politika-analuusid-ja-uuringud/tarkade-otsuste-fondi-uuringute-kokkuvotted/Andmeaidad_1%C3%B5pparuanne.pdf
413. Whitehouse 2012, Big Data Press Release, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf
414. Wikipedia http://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users
415. Wikipedia <http://www.wikipedia.org/>
416. Vinod Baya and Bo Parker , The Thing Stack: Technologies that guide customers to their goals, in Fred Cripe , Internet of Things: Evolving transactions into relationships, PwC Technology Forecast, 2013, No 1
417. World Bank <http://data.worldbank.org>
418. www.idc.com
419. Välisministeeriumi teadaanne, <https://www.riigiteataja.ee/akt/13267527>. WIPO (Maaailma Intellektuaalse Omandi Organisatsiooni) autoriõiguse leping, [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=DD:11:33:22000A0411\(01\):ET:PDF](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=DD:11:33:22000A0411(01):ET:PDF)
420. Äriregistri, mittetulundusühingute ja sihtasutuste registri ning kommertspandiregistri elektrooniliste andmete väljastamise eest võetava tasu suuruse ja nende isikute loetelu, kellele väljastatakse äriregistri, mittetulundusühingute ja sihtasutuste registri ning kommertspandiregistri elektroonilisi andmeid tasuta, kinnitamine. Justiitsministri määrus, RT I, 08.11.2013, 12, <https://www.riigiteataja.ee/akt/108112013012>
421. Äriseadustik, <https://www.riigiteataja.ee/akt/102072013063>
422. X-tee andmeteenuste analüüs (Haav, Hele-Mai; Küngas, Peep (2013). Semantic data interoperability: the key problem of big data. In: Big Data Computing: (Toim.) Akerkar, Rajendra. Chapman and Hall / CRC Press, 2013. [ilmumas]
423. YarcData <http://www.yarcdata.com>

9 Lisad

9.1 Kvantitatiivküsitluse ankeet

Q1 Uuring „Lingitud Eesti“

Riigikantselei on Tarkade otsuste fondi ja Euroopa Sotsiaalfondi toel tellinud koostöös Majandus- ja Kommunikatsiooniministeeriumi ning Eesti Infotehnoloogia ja Telekommunikatsiooni Liiduga uuringu „Lingitud Eesti“.

„Lingitud Eesti“ all mõistame IKT-lahenduste toel loodud kogu Eesti inforuumi hõlmavat hajutatud isikute, organisatsioonide, tarkade seadmete, teadmiste, infosüsteemide ning andmete koosvõimelist võrgustikku. Linkandmete tehnoloogia võimaldab seni hajutatult eksisteerivad andmebaasid, sotsiaalmeedia, dokumendid, faktid, seadmed, teadmised jms integreerida ühtseks tervikuks ning avab seeläbi riikidele, era- ja vabasektorile seni võimatud olnud ehk põhimõtteliselt uued võimalused uut tüüpi teenuste loomiseks. Selle võrgustiku loomine eeldab läbimurret mitmetes suundades nagu ülikiire andmesidevõrk ja targad seadmed, koosvõimelised infosüsteemid ja teadmusvõrgustikud, semantilised võrgustikud – kõige olulisemana aga linkandmete kasutuselevõtt.

Käesoleva ankeetküsitluse eesmärgiks on uurida Eesti riigi-, era- ja vabasektori valmisolekut linkandmete loomiseks ja kasutuselevõtuks ning sellega seonduvate tehnoloogiate kasutuselevõtu võimekust Eesti riigis. Teie arvamus selle IT-ala probleemidest ja arenguvõimalustest on meie jaoks oluline. Sestap loodame, et leiate võimaluse 15-20 minutit oma väärtuslikku aega uuringusse panustada.

Uuringu läbiviijateks on SOA Trader OÜ, PricewaterhouseCoopers Advisors AS ja Tepinfo OÜ. Küsitlus on anonüümne ning kogutud andmeid kasutatakse ainult üldistatud kujul.

Teid ette tänades,

Peep Küngas
Hele-Mai Haav
Jaak Tepandi
Lauri Tepandi
Mihkel Lauk
Triin Tars

Q2 Et tagada ühtne mõistetest arusaamine, on alljärgnevalt lahti seletatud nimetatud tehnoloogiad:

Linkandmed. Linkandmed (Linked Data) on struktureeritud andmed, mis on kirjeldatud ja avaldatud viisil, mis lubab neid automaatselt seostada. Linkandmed on esitatud ühtses formaadis, näiteks RDF. Linkandmed võivad olla avaandmed ja võivad ka mitte olla. Vt ka <http://www.w3.org/DesignIssues/LinkedData>

Avaandmed. Avaandmed (Open Data) on andmed, mis on kõigile antud vabalt kasutada, taaskasutada ja jagada, kitsendatult maksimaalselt viitamise või samadel tingimustel taasjagamise nõudega. Avaandmed võivad olla esitatud linkandmetena (Linked Open Data), aga ei pruugi. Tim Bernes Lee poolt pakutud vormingute ja kodeeringute süsteemis avaandmete „kasutajasõbralikkuse“ hindamiseks on linkandmed kõrgeimal tasemel.

Semantilised tehnoloogiad. Semantilised tehnoloogiad (Semantic Technologies) võimaldavad lisada andmetele tähendust või seda andmetest ekstraheerida ning esitada andmeid koos nende tähendustega masinloetaval kujul. Põhilised kasutusküpsed standardid on URI, RDF, OWL, SPARQL, RDFS, RDFa, jt.

Asjade Internet. Asjade Internet (Internet of Things) tähendab füüsiliste objektide (näiteks esemete, seadmete) varustamist sensoritega, mis on lingitud traadiga või traadita võrgu abil, kasutades selleks tihti Interneti protokoll (IP). Teisalt võivad nn asjadeks olla ka abstraktsed mõisted ja andmed.

Suurandmete töötlemise ja hõive tehnoloogiad. Suurandmeid (Big Data) iseloomustavad järgmised omadused: mahukus, muutuvus, mitmekesisus ja lisaks ka keerukus. Üks suurandmete lahendamata probleeme on semantiline koosvõime, sest andmete mitmekesisus on väga suur ja andmed pärinevad väga erinevatest allikatest.

Koosloome tehnoloogiad. Koosloome (Crowdsourcing) on praktika, kus teenuste, ideede ja andmete saamiseks kasutatakse suurt hulka tavaliselt online-režiimis ühendatud inimesi. Koosloome protsess eeldab, et teenused, andmed, seadmed jne. kasutaks ühist mõistete süsteemi, ontoloogiat, semantikat ja sisaldaks linke (s.t viitaks) viimastele.

Q3 1. Palun hinnake oma teadmisi loetletud tehnoloogiate osas:

	Ei tea sellest tehnoloogiast midagi (1)	Üldisel tasemel teadmised valdkonnast (2)	Keskised teadmised valdkonnast (3)	Mul on põhjalikud teadmised valdkonnast (4)	Olen valdkonna ekspert (5)
Avaandmed (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmed (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Semantilised tehnoloogiad (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asjade Internet (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suurandmete töötlemise ja hõive tehnoloogiad (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Koosloome tehnoloogiad (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 2. Palun hinnake oma praktilisi kogemusi loetletud tehnoloogiate osas:

	Kogemused puuduvad (1)	Kogemused 1-2 projektis (2)	Kogemused 3 või enamas projektis (3)
Avaandmed (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmed (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Semantilised tehnoloogiad (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asjade Internet (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suurandmete töötlemise ja hõive tehnoloogiad (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Koosloome tehnoloogiad (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5 3. Palun hinnake järgnevate tehnoloogiate olulisust Eestile linkandmete kasutuselevõtu kontekstis:

	Ebaoluline (1)	Väheoluline (2)	Keskmiselt oluline (3)	Pigem oluline (4)	Väga oluline (5)	Ei oska öelda (6)
Koosloome tehnoloogiad (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asjade Internet (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Seantilised tehnoloogiad (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avaandmed (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suurandmete töötlemise ja hõive tehnoloogiad (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q6 4. Kuivõrd olete teadlik valitsuse plaanidest linkandmete vallas?

- ☐ Pole teadlik (1)
- ☐ Olen kuulnud terminit „linkandmed“, aga ei tea enam (2)
- ☐ Olen teadlik valitsuse plaanidest (3)
- ☐ Olen ekspert, tean hästi valitsuse plaane (4)

Q7 5. Milline on Teie arvates Eesti tänane tehnoloogiline võimekus/kompetents nimetatud tehnoloogiatrendide raames?

	Väga halb (1)	Halb (2)	Keskmine (3)	Hea (4)	Väga hea (5)	Ei oska öelda (6)
Avaandmed (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmed (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Semantilised tehnoloogiad (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asjade Internet (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suurandmete töötlemise ja hõive tehnoloogiad (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Koosloome tehnoloogiad (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q8 6. Milline on Teie arvates Eesti arenguvajadus järgmiste tehnoloogiatrendide raames?

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Avaandmed (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmed (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Semantilised tehnoloogiad (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asjade Internet (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suurandmete töötlemise ja hõive tehnoloogiad (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Koosloome tehnoloogiad (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q9 7. Palun hinnake järgmisi potentsiaalseid kasusid linkandmete kasutamisest riigile, erasektorile ja elanikkonnale tervikuna:

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Riigi läbipaistvuse ja efektiivsuse suurenemine (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Analüüsivõimekuse kasv, paremate otsuste tegemise võimalus (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eesti konkurentsivõime paranemine (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Paremad ja odavamad teenused (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uutlaadi teenused, mida seni ei ole saanud või pole olnud otstarbekas luua (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Riigi kulude vähenemine, kuna erasektor/vabakond saab ise teenuseid luua (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Majanduskasv (eelnevatest tulenevalt) (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muud (palun kirjeldage järgneval lehel vaba tekstina) (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
IKT sektori kasv ja areng (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q10 Muud kasud? (kirjeldage vaba tekstina)

Q11 8. Palun hinnake linkandmete loomist ja kasutamist pärssivaid tegureid Eestis (st kuivõrd loetletud takistus on segav Eesti kontekstis):

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Selge riikliku visiooni, strateegia ja tegevusplaani puudumine (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vastutaja ja eestvedaja puudumine (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Puudulik rahastamine (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Konkreetsed vajaduse puudumine (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Poliitilise koosvõime ja tahte probleemid (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Puudulik analüüs ja info, mille põhjal otsuseid teha (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Organisatoorsed takistused (nt ametkondade vahelise koostöö probleemid) (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Õiguslikud takistused (nt puudulik seadusandlus, vastuolud seadustes) (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tehnoloogilise kompetentsi puudumine riigisektoris (Tellija) (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Riigisektori infosüsteemide üldine hetkeseis, sh aegunud arhitektuur (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Semantilise ühtlustamise probleemid (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muud (palun kirjeldage järgneval lehel vaba tekstina)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(12) Tehnoloogilise kompetentsi puudumine erasektoris (Pakkuja) (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vähene linkandmete alane teadlikkus (14)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avaliku sektori avaandmete eest tasu küsimine (15)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q12 Muud takistused?(kirjeldage vaba tekstina)

Q13 9. Palun hinnake järgmiste ohtude suurust seoses potentsiaalse linkandmete laialdase kasutuselevõutuga Eestis:

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Privaatsuse vähenemine, sh isiku kaudse tuvastamise oht (mosaiikandmed) (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Madal andmete kvaliteet ja usaldusväärsus (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andmete vale interpreteerimine (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Süsteemi loomise ja ülalpidamise kulukus suurem kui reaalne kasu (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Keerukate sõltuvuste tekkimine, probleemid andmete ja protsesside muutumisel (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muud (palun kirjeldage järgneval lehel vaba tekstina) (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q14 Muud ohud? (kirjeldage vaba tekstina)

Q15 10. Palun hinnake järgmiste valdkondade põhiandmehulkade linkandmetena avamise potentsiaalset kasu Eestile tervikuna:

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Globaalne areng (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Haridus (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tervishoid (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teadus ja uurimistöö (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Äri (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Transport ja infrastruktuur (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Finants ja lepingud (nt eelarve ja kulutused, lepingud, pakkumused) (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Energeetika ja keskkond (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Georuumiline (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Maa seire (nt meteoroloogia, metsandus, kalandus, jahindus, põllumajandus) (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sotsiaalne mobiilsus ja heaolu (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kuritegevus ja õigus (12)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Statistika (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Valitsuse aruandlus ja demokraatia (14)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muud (palun kirjeldage järgneval lehel vaba tekstina) (15)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q16 Muud valdkonnad? (kirjeldage vaba tekstina)

Q17 11. Palun hinnake, milline mõju on järgmist laadi projektide/tegevuste käivitamisel Eestile (kui vastusevariandis on kasutatud terminit avaandmed, siis peetakse silmas avaandmeid, mis pole esitatud linkandmetena):

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Avaandmete loomine/tootmine, haldus ja publitseerimine (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete jaoks metaandmete (sh ontoloogiatega) loomine (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete loomise metoodika ja juhendite koostamine ning nende publitseerimine (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete turvalisus, privaatsus ja kvaliteet (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete ja teiste seonduvate tehnoloogiate alaste koolituste pakkumine/laiendamine (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete alase teadlikkuse suurendamine, sh kasutuselevõtu propageerimine (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete loomine/tootmine, haldus (koos metaandmetega/ontoloogiatega) ja publitseerimine (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Avaandmete rakendused (näiteks teenustes) (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete rakendused (sh valdkonnaspetsiifilised liidesed) (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Koosloome kasutamine linkandmete mahu suurendamisel, nende semantilise tähenduse tõstmisel või kvaliteedi määramisel (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linkandmete temaatilised rakendusuringud ja tarkvaraarendusprogrammid (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Konkreetsed selge praktilise efektiga linkandmetel põhineva(te) pilootrakendus(t)e väljatöötamine (12)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muud (palun kirjeldage järgneval lehel vaba tekstina) (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Osalemine rahvusvahelistes avaandmete projektides/algatustes (14)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Osalemine rahvusvahelistes linkandmete projektides/algatustes (15)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
--	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Q18 Muud projektid/tegevused/algatused? (kirjeldage vaba tekstina)

Q19 12. Linkandmete kontekstis peetakse üha enam oluliseks esitada avaandmeid linkandmetena. Palun hinnake järgnevate avaandmete kasutusvõimaluste atraktiivsust enda jaoks (kuivõrd kasutaksite avaandmeid loetletud viisidel):

	Väga väike (1)	Väike (2)	Keskmine (3)	Suur (4)	Väga suur (5)	Ei oska öelda (6)
Andmete kasutamine töötlemata kujul (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uute seoste avastamine (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uute teenuste loomine (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andmete parandamine (nt vananenud andmed) (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andmete publitseerimine (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Turundus, müük (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Integreerimine või linkimine teiste andmetega (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uuringud ja analüüs (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Muud (palun kirjeldage järgneval lehel vaba tekstina) (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q20 Muud kasutusviisid? (kirjeldage vaba tekstina)

Q21 13. Millises sektoris (või sektorites) töötate või olete peamiselt aktiivne?

- ☐ Teadus (1)
- ☐ Avalik haldus (riigi- ja valitsusasutused) (2)
- ☐ Erasektor (äri) (3)
- ☐ Vabasektor (eraalgatuslik, vabakond) (4)

Q22 14. Milline on Teie positsioon organisatsioonis, kus Te hetkel töötate?

- ☐ Juhataja/direktor (1)
- ☐ Osakonna/meeskonna juht (2)
- ☐ Juhtiv- või peaspetsialist (3)
- ☐ Spetsialist (4)
- ☐ Projektijuht (5)
- ☐ Töötaja (6)
- ☐ Muu (7)

Q23 15. Millises valdkonnas (valdkondades) tegutseb Teie ettevõtte/asutus/organisatsioon?

- ☐ Põllumajandus, metsamajandus ja kalapüük (1)
- ☐ Mäetööstus (2)
- ☐ Töötlev tööstus (3)
- ☐ Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine (4)
- ☐ Veevarustus, kanalisatsioon, jäätme- ja saastekäitlus (5)
- ☐ Ehitus (6)
- ☐ Hulgi- ja jaekaubandus, mootorsõidukite ja mootorrataste remont (7)
- ☐ Veondus ja laondus (8)
- ☐ Majutus ja toitlustus (9)
- ☐ Info ja side (10)
- ☐ Finants- ja kindlustustegevus (11)
- ☐ Kinnisvaraala tegevus (12)
- ☐ Kutse-, teadus- ja tehnikaala tegevus (13)
- ☐ Haldus- ja abitegevused (14)
- ☐ Haridus (15)
- ☐ Tervishoid ja sotsiaaltoetused (16)
- ☐ Kunst, meelelahutus ja vaba aeg (17)
- ☐ Muu (18)

Q24 16. Palun märkige oma kõrgeim omandatud haridustase:

- ☐ Kõrgharidus (doktor) (1)
- ☐ Kõrgharidus (magister, diplom) (2)
- ☐ Kõrgharidus (bakalaureus) (3)
- ☐ Kutseharidus (4)
- ☐ Keskharidus või madalam (5)
- ☐ Muu (6)

Q25 17. Kas eelnevad küsimused olid Teie jaoks arusaadavalt esitatud?

- ☐ Jah (1)
- ☐ Ei (2)
- ☐ Valdavalt arusaadavalt (3)
- ☐ Valdavalt mitteamusaadavalt (4)

Q26 18. Kas Teil on käesoleva uurimuse temaatika osas muid kommentaare, mida sooviksite jagada?