

Data management, kvalita

Datová architektura a správa dat

Otevřená data

INS_2020_5. přednáška

Architektura

• **Architektura** (dle ISO/IEC 42010:2007, revidována normou ISO/IEC/IEEE 42010:2011) - fundamentální uspořádání systému, které tvoří komponenty a vztahy mezi nimi, včetně vztahu k prostředí, a principy, které řídí jeho návrh a rozvoj.

ISO / IEC / IEEE 42010: 2011 se zabývá tvorbou, analýzou a udržováním architektur systémů pomocí popisů architektury.

- Je vytvořen koncepční model popisu architektury.
- Je zadán požadovaný obsah popisu architektury.
- Jsou zavedena hlediska architektury, architekturní rámce a jazyky popisu architektury pro kodifikaci konvencí a běžných postupů popisu architektury.
- Je zadán požadovaný obsah hledisek architektury, architekturních rámců a jazyků popisu architektury.

Architektura

- **Architektura v podnikové informatice (Enterprise Architecture, EA)**

je přístup, koncept, prostředek a nástroj vyjadřující fundamentální uspořádání vztahu mezi byznysem a jeho informačním systémem, které vede k naplnění cíle organizace a zároveň respektuje okolní prostředí a konzistentně dodržuje formulované principy návrhu a rozvoje systému (Voříšek a kol., 2008) – zavedeno jako pojem „Podniková architektura“

- **Datová architektura**

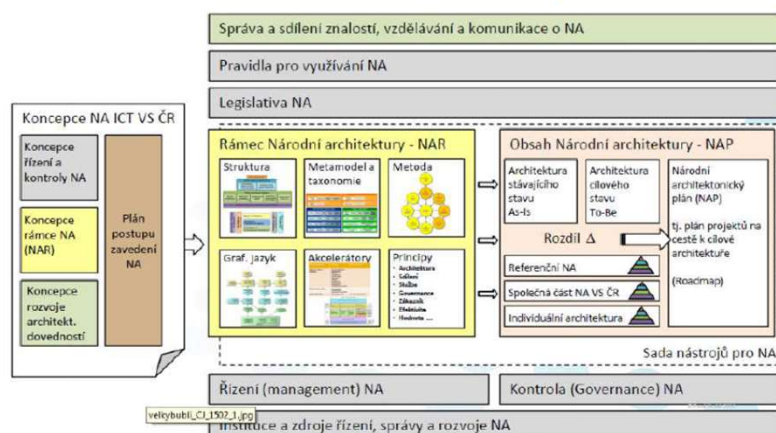
popisuje strukturu datové základny organizace. Datová architektura popisuje jednotlivé datové sady, datové zdroje, vazby mezi nimi a vazby na další relevantní prvky podnikové architektury (Enterprise Architecture)

Národní architektura ICT VS ČR - 1



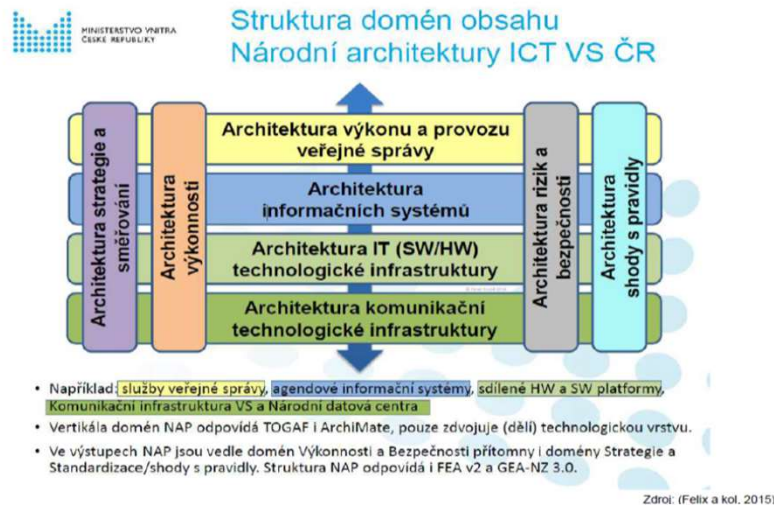
MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

Koncepce zavedení Národní architektury ICT VS ČR



Zdroj: (Felix a kol. 2015)

Národní architektura ICT VS ČR - 2



Cíle analýzy a návrhu datové architektury

- Analyzovat potřeby v oblasti dat a jejich správy
- Identifikovat a popsat požadavky na datovou architekturu
- Navrhnout datovou architekturu, která bude v souladu s definovanými požadavky a která naplní definovanou strategii v oblasti řízení a využívání dat

Obsah datové architektury

Konkrétní obsah se může lišit v závislosti na použitém rámci podnikové architektury a na potřebách organizace.

Typické součásti datové architektury dle **Data Management Association International (DAMA) Releases DAMA Dictionary of Data Management** - slovník obsahující více než 800 výrazů definujících společný slovník správy dat pro IT profesionály, správce dat a vedoucí pracovníky:

- Datový model organizace – na konceptuální úrovni
- Zachycení vazby prvků datového modelu na další prvky podnikové architektury jako procesy, organizační jednotky, role, aplikace, cíle, strategie, projekty, technologie.
- Specifické architektury
- Architektura souvisejících technologií – architektura popisující jaké technologie a nástroje jsou v rámci organizace používány pro zpracování a ukládání dat
- Integrovaná architektura
- Architektura datových skladů a Business Intelligence
- Architektura metadat
- Datová architektura zachycuje také referenční a master data

Data Management - Správa dat

- Jedna z domén řízení informatiky
- Anglicky **Data Management**, používají se i pojmy (Enterprise) **Information Management**, Enterprise Data Management, Information Asset/Resource Management apod.
- **Data Management** - představuje plánování a provádění politik, praktik a projektů pořízení, kontroly, ochrany, dodávky a zvyšování hodnoty dat a informačních aktiv (Mosley a kol., 2010)
- **Information management** - představuje prostředky pomocí kterých organizace zajišťuje plánování, identifikaci, vytvoření, získání, sběr, organizování, řízení (governance), zabezpečení, použití, sledování, distribuci, výměnu, údržbu, uchování a vyřazení informací jakož i prostředky, kterými zajišťuje, že hodnota informací je rozpoznána a maximálně využita (QGCI, 2009)

Cíle

Hlavním cílem je zajistit, že data, resp. informace jsou k dispozici

- tomu, kdo je potřebuje,
- v podobě či formátu, v jakém je potřebuje a
- v okamžiku, kdy je potřebuje.

Upraveno dle: (QGCI0. 2009)

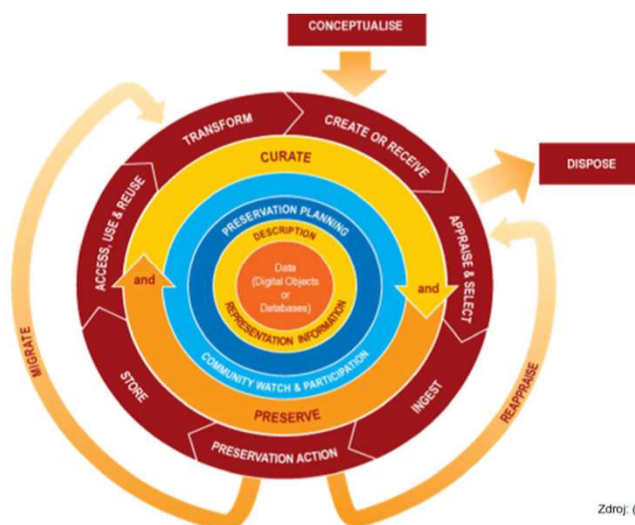
Hlavní oblasti správy dat (Data Management)

1. Strategické řízení
2. Analýza a návrh datové architektury
3. Pořizování, provoz a údržba dat, poskytování dat, skartace dat
4. Řízení nákladů a přínosů dat
5. Řízení bezpečnosti dat
6. Řízení kvality dat
7. Řízení metadat
8. Řízení referenčních dat a master dat

Životní cyklus informačních aktiv dle QGEA



The DCC Curation Lifecycle Model



Základní stadia životního cyklu dat

1. Plánování
2. Specifikace
3. Pořízení
4. Uložení a klasifikace
5. Použití a údržba
6. Archivace a obnova
7. Skartace

Open Data – Otevřená data

Data zveřejněná na Internetu, která jsou:

- úplná,
- snadno dostupná,
- strojově čitelná,
- používající standardy s volně dostupnou specifikací,
- zpřístupněna za jasně definovaných podmínek užití dat s minimem omezení,
- dostupná uživatelům při vynaložení minima možných nákladů.

Požadavky na podmínky užití otevřených dat

- Neomezují jejich uživatele ve způsobu použití dat.
- Opravňují uživatele k jejich dalšímu šíření.
- Musí být uveden autor dat (i při dalším šíření).
- Při dalším šíření musí i ostatní uživatelé mít stejná oprávnění s daty nakládat - během šíření dat nesmí dojít např. k omezení jejich využití pouze pro nekomerční účely.

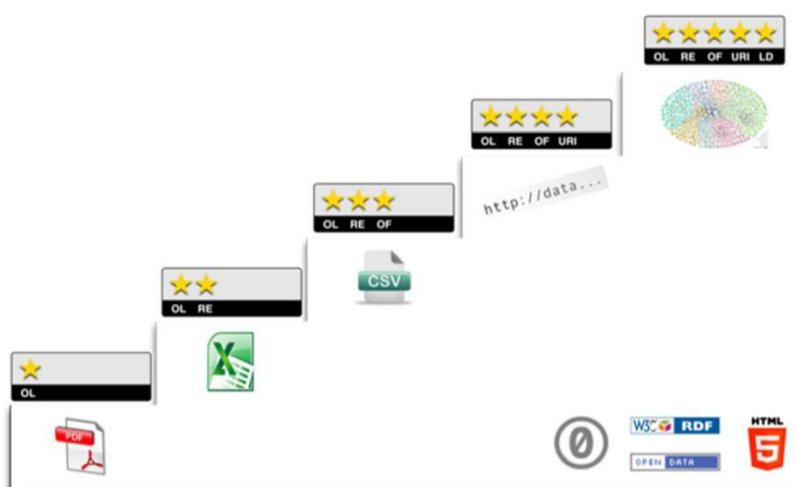
Přínosy otevřených dat

- | | |
|---|---|
| <ul style="list-style-type: none"> • (P1) posílení transparentnosti, • (P2) podpora ekonomického růstu, • (P3) zlepšení služeb veřejné správy, • (P4) podpora opětovného použití dat, • (P5) zlepšení vnímání veřejné správy veřejností, • (P6) zlepšení procesů a dat veřejné správy, • (P7) zamezení chyb vzniklých při práci s daty | <ul style="list-style-type: none"> • (P8) snížení počtu dotazů dle zák. č. 106/1999 Sb. • (P9) pořádek ve vlastních datech • (P10) možnost transparentně informovat veřejnost o subjektech, které neplní své zákonné povinnosti • (P11) zvýšení hodnoty dat • (P12) budování dobrého jména |
|---|---|

Rizika otevřených dat

- (R1) Zveřejnění dat v rozporu se zákonem
- (R2) Riziko porušení ochrany obchodního tajemství
- (R3) Riziko porušení ochrany osobních údajů
- (R4) Zveřejnění nevhodných dat či informací
- (R5) Dezinterpretace dat
- (R6) Absence konzumentů dat
- (R7) Překrývání dat

Otevřená data – stupně otevřenosti



Zdroj: (Hausenblas, 2012)

- Převzato z <https://5stardata.info/cs>
- stupeň 1 - datová je dostupná v síti WWW s vhodnými podmínkami užití otevřených dat (viz [Jak stanovit podmínky užití datových sad?](#)),
- stupeň 2 - datová sada je poskytována ve strojově čitelném formátu, který umožňuje automatizované strojové zpracování,
- stupeň 3 - datová sada je poskytována v otevřeném formátu, tj. ve formátu s volně dostupnou specifikací,
- stupeň 4 - na identifikaci entit v datové sadě se používají IRI,
- stupeň 5 - datová sada splňuje standard propojených dat.

Otevřená data – stupeň 5

- Otevřená data na stupni 5 jsou také nazývána **Propojená otevřená data** (angl. **Linked Open Data**, zkratka LOD)

Dostupnost dat na webu

Regionální statistiky ČSÚ



<http://www.czso.cz>

RÚIAN



<http://www.cuzk.cz>

ARES



<http://www.mfer.cz>

Sbírka zákonů



<http://portal.gov.cz>

RPP



<http://www.szcr.cz>

Kontroly SZPI



<http://www.potravinynapranyri.cz/>

Kontroly ČOI



<http://www.coi.cz>

Kontroly ČIZP



<http://www.cizp.cz>

Kontroly NKÚ



<http://data.nku.cz>

Web nabízí řadu dat o kontrolách. Můžeme s nimi efektivně (strojově) pracovat?

- Data o kontrolách a související data jsou publikována v různých dokumentech mnoha organizacemi a na mnoha místech.
- Dokumenty jsou až na výjimky určeny pro lidi, ne pro strojové zpracování.
- Souvislosti nejsou dostatečně popsány a zaznamenány.

Zdroj: Upraveno dle (Nečaský a kol., 2014)


Dostupnost dat na webu

Regionální statistiky ČSÚ




<http://www.czso.cz>

RÚIAN



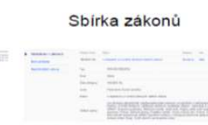
<http://www.cuzk.cz>

ARES




<http://www.mfer.cz>

Sbírka zákonů




<http://portal.gov.cz>

RPP



<http://www.szcr.cz>

Kontroly SZPI



<http://www.potravinynapranyri.cz/>

Kontroly ČOI




<http://www.coi.cz>

Kontroly ČIZP



<http://www.cizp.cz>

Kontroly NKÚ



<http://data.nku.cz>

Jak nalézt odpovědi na následující otázky?

- Kontroly provedené v okrese Semily
 - + u subjektů provozujících obor činnosti „Pronájem a půjčování věcí movitých“
- Kontroly provedené v rámci agendy „Odpadové hospodářství“
- Počty kontrol a objemy pokut v okresech přepočtené na 1000 obyvatel

Zdroj: Upraveno dle (Nečaský a kol., 2014)

Principy propojených dat

1. Identifikujte datové entity pomocí http:// URL.
2. Používejte URL tak, aby je ostatní mohli vyhledat v síti WWW, přistupovat k němu a odkazovat se na něj.
3. Pokud někdo přistoupí k URL entity, poskytněte o ní údaje ve formátu RDF.
 - Případně nabídněte RDF datové API (SPARQL).
4. Mezi údaji poskytněte odkazy na URL souvisejících entit k doplnění kontextu a umožnění procházení entit v rámci sítě WWW.