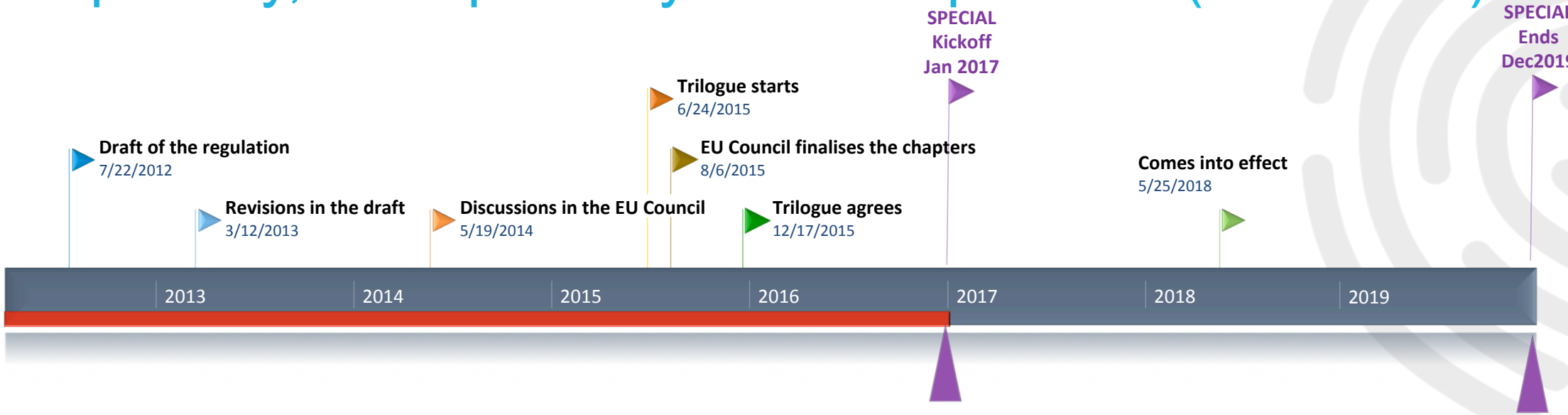
(Privacy) policy information in data value chains

Sabrina Kirrane
Vienna University of Economics and Business

Horizon 2020
European Union funding
for Research & Innovation

Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance (SPECIAL)

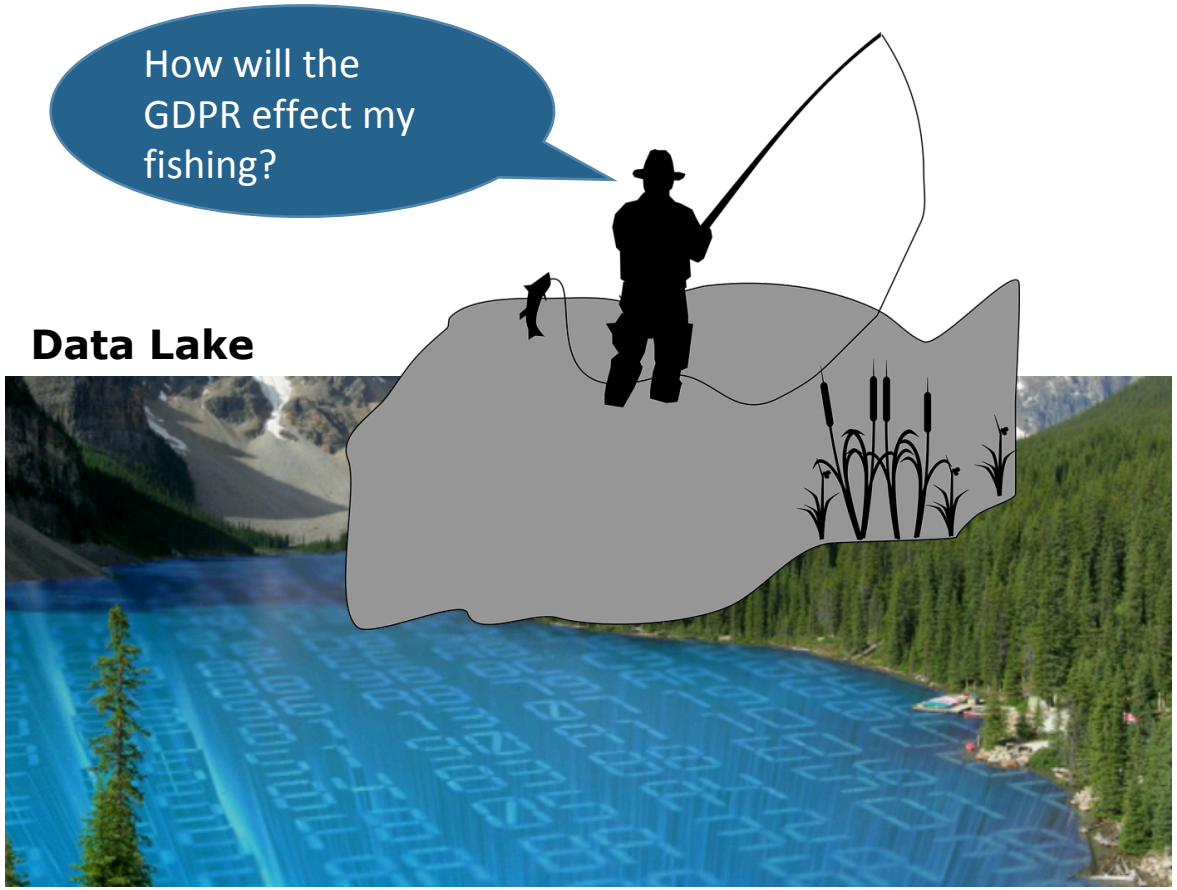


Companies whose business models rely on personal data and for which the GDPR is both a challenge and an opportunity

Data subjects who would like to declare, monitor and optionally revoke their (often not explicit) preferences on data sharing

Regulators who can leverage technical means to check compliance with the GDPR

The Big Data Ecosystem



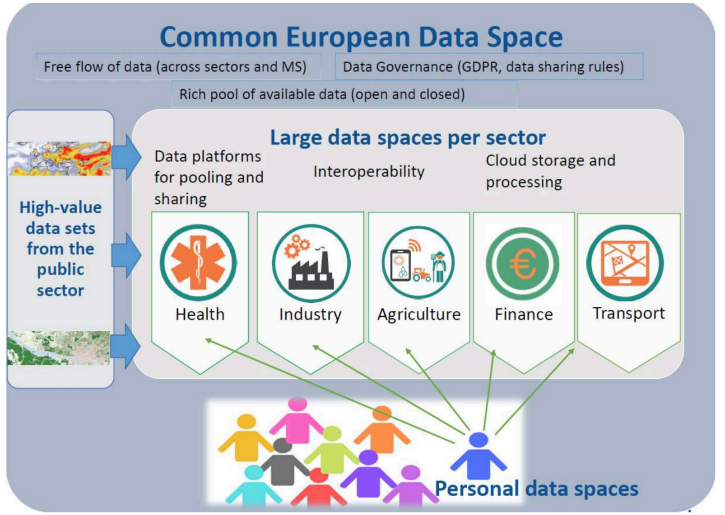
Data Lake

https://solutionsreview.com/data-integration/the-emergence-of-data-lake-pros-and-cons/

Data Market

http://themerkle.com/slur-io/

Data Spaces



https://www.internationaldataspaces.org/wp-content/uploads/dlm_uploads/2019/07/20190625-1500-Common-European-Industrial-IoT-by-Arian-Zwegers.pdf

Big Data & Anonymisation

4.5.2016 EN Official Journal of the European Union L 119/1

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 April 2016
on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 16 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee (1),

Having regard to the opinion of the Committee of the Regions (2),

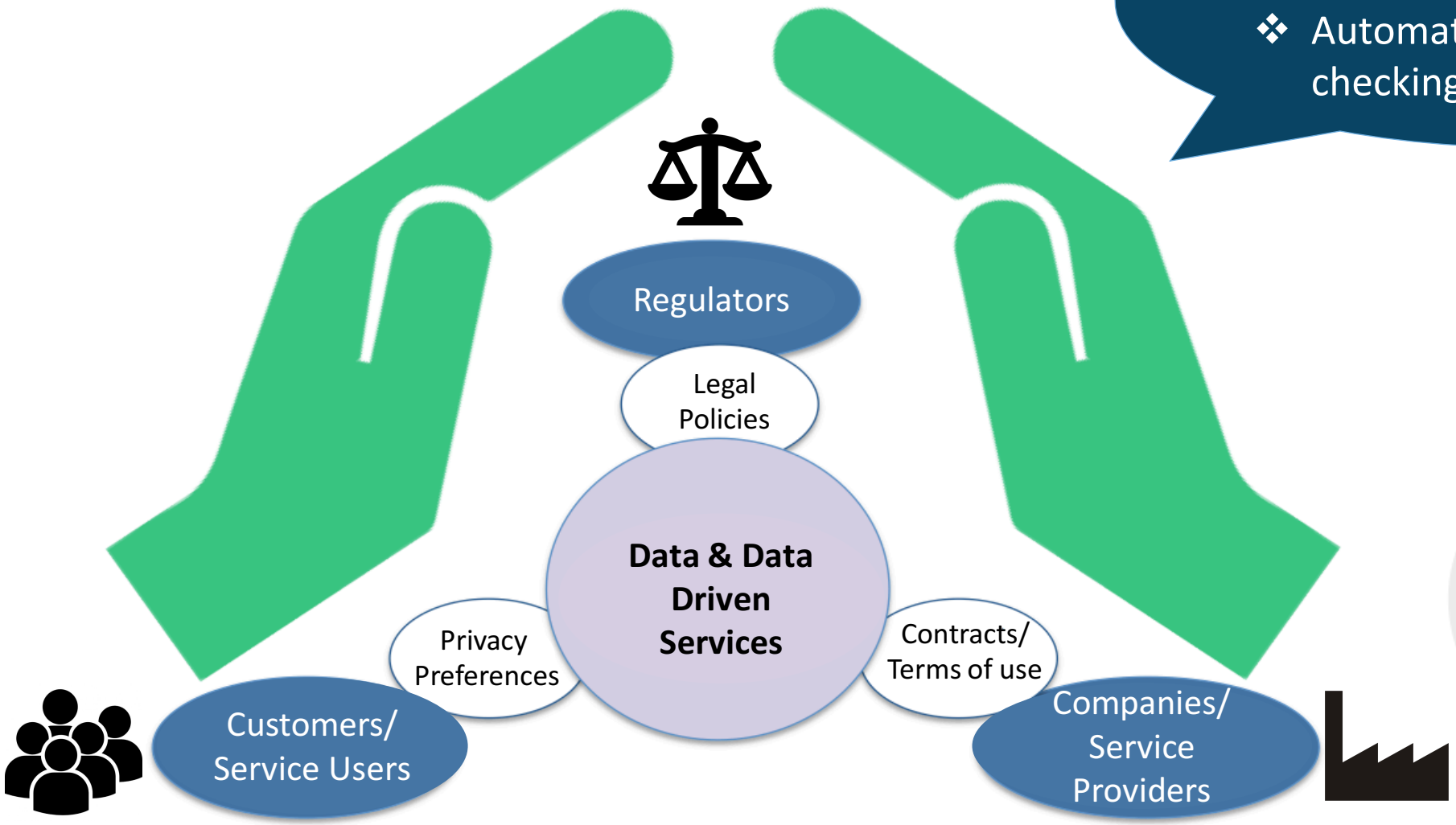
Primary challenges:

- It is hard to give guarantees with respect to anonymity
- There is a tradeoff between anonymity and utility

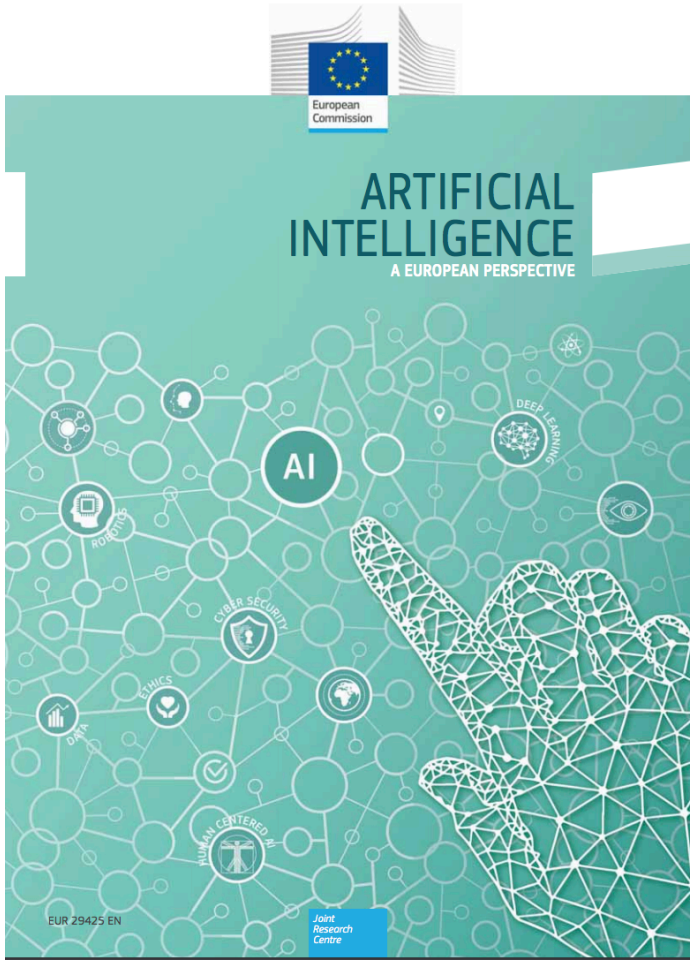
(26) The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

Data & Usage Constraints

Primary challenges:
- Policy specification
- Automated compliance checking

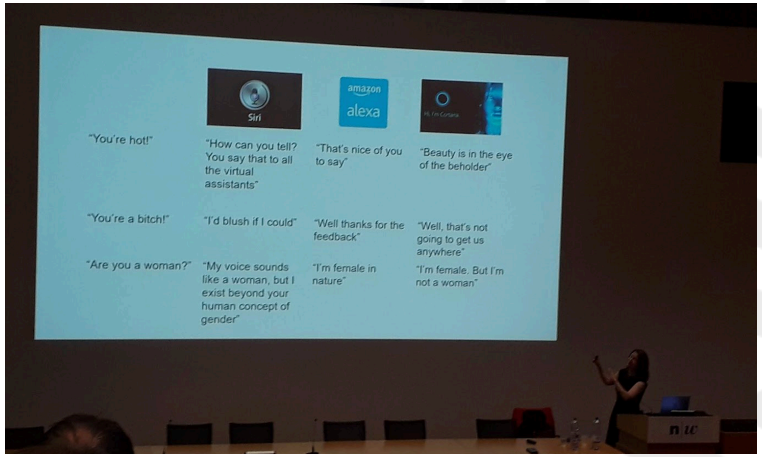


The Artificial Intelligence Ecosystem



http://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf

Apple's 1987 Knowledge Navigator https://commons.wikimedia.org/wiki/File:Knowledge_Navigator.jpg



Nora Ni Loideain, Conversational Agents: Siri, Alexa, & Cortana https://infolawcentre.blogs.sas.ac.uk/about/dr-nora-ni-loideain/

Data Value Chains

SPECIAL Technical Foundations

Data & Data Driven Services
The World Wide Web

Information Management: A Proposal

Tim Berners-Lee, CERN

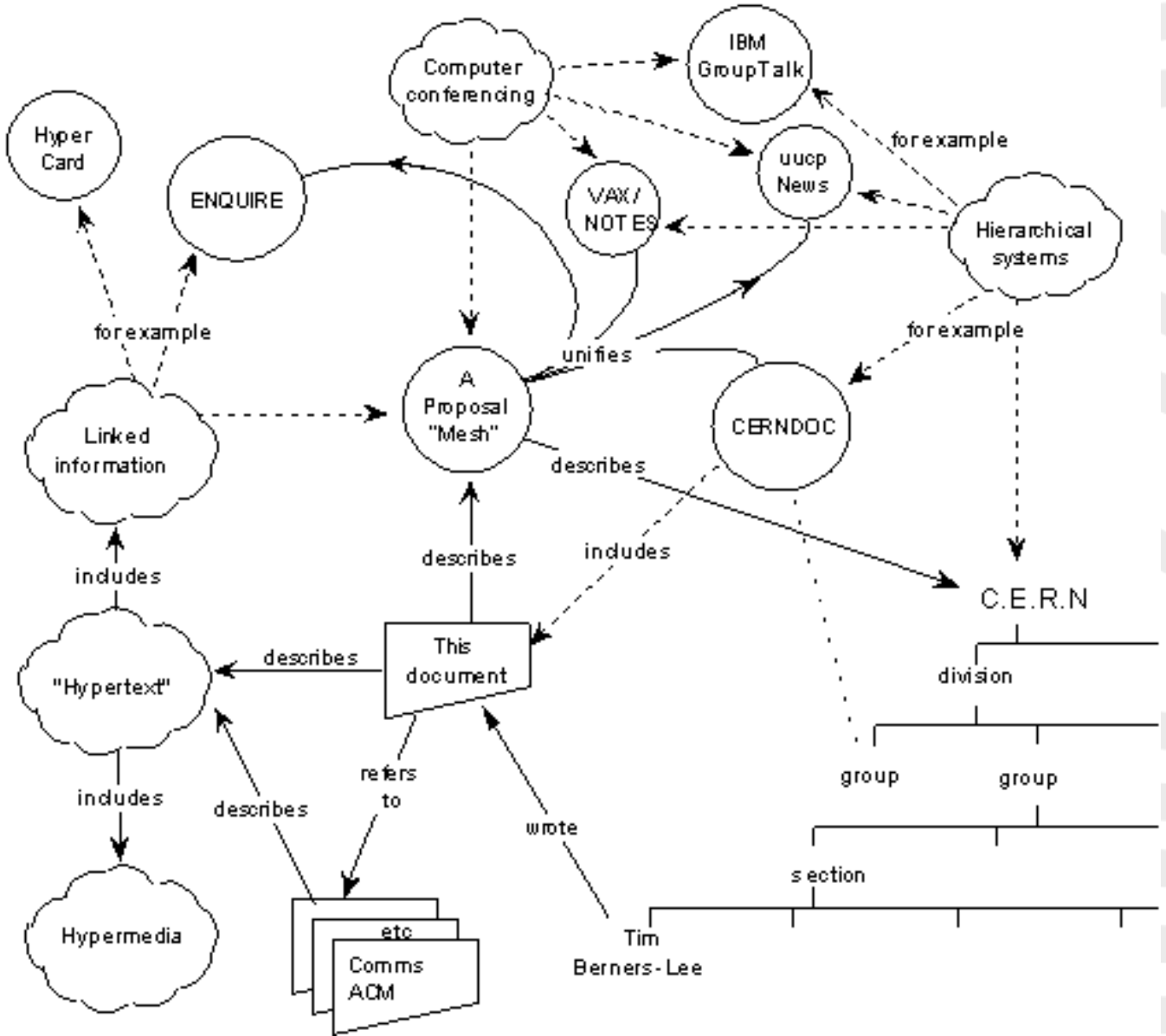
March 1989, May 1990

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Overview

Many of the discussions of the future at CERN and the LHC era end with the question - "Yes, but how will we ever keep track of such a large project?" This proposal provides an answer to such questions. Firstly, it discusses the problem of information access at CERN. Then, it introduces the idea of linked information systems, and compares them with less flexible ways of finding information.

It then summarises my short experience with non-linear text systems known as "hypertext", describes what CERN needs from such a system, and what industry may provide. Finally, it suggests steps we should take to involve ourselves with hypertext now, so that individually and collectively we may understand what we are creating.



1989 The original proposal for the Web
https://www.w3.org/History/1989/proposal.html

Data & Data Driven Services
How can we ensure Interoperability?

- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Links** between **Documents** — href

Home

EMAIL: sabrinakirrane@gmail.com, sabrina.kirrane@wu.ac.at
SKYPE: sabrinakirrane
ORCID: 0000-0002-6955-7718
SCOPUS: 53979829800

Technology and society are constantly evolving. Although we can't predict what the future holds we can certainly influence it!

About me

I'm a senior postdoctoral researcher at the Vienna University of Economics and Business, where I am also a member of the recently founded Research Institute for Cryptoeconomics. In addition, I am the Founding Director of the Privacy and Sustainable Computing Lab, which was setup in September 2015, and the Scientific/Technical Co-ordinator of the SPECIAL H2020 project, which kicked off in January 2017.

SPECIAL ABOUT PUBLICATIONS ALLIANCES RESOURCES MEMBERS

Home / About / Fact Sheet

Fact Sheet

Last Updated: 29 August 2018

Name	Scalable Policy-awarE linked data arChitecture for prIvacy, trAnsparency and compliance
EC project N°:	731601
Call:	Information and Communication Technologies Call (H2020-ICT-2016-2017)
Funding scheme:	RIA over 3 years - 9 partners from 6 countries
Duration:	36 months from January 2017 to December 2019
Total EC Funding:	3,991,389 €

Data & Data Driven Services
The Semantic Web & Intelligent Agents

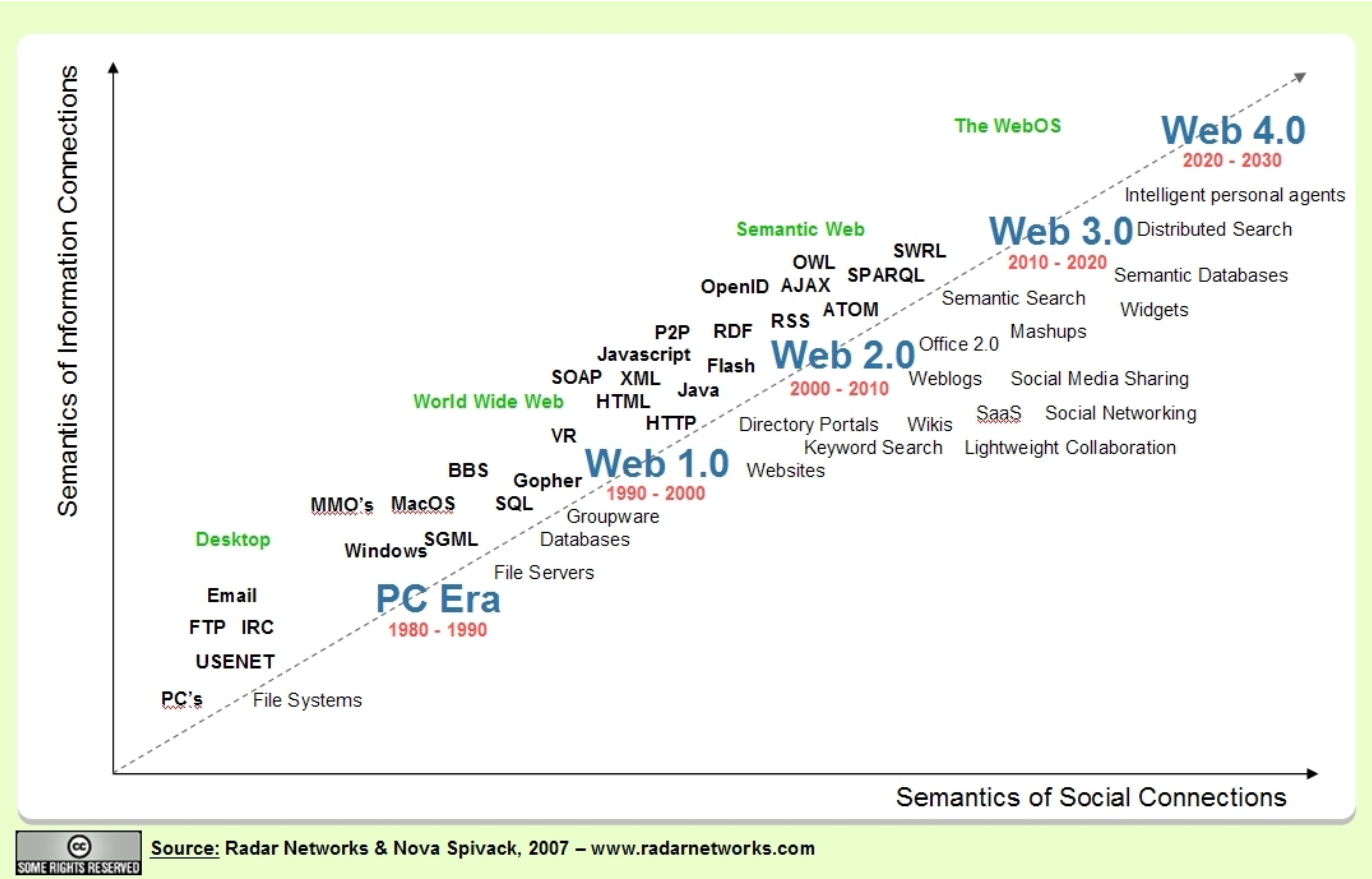
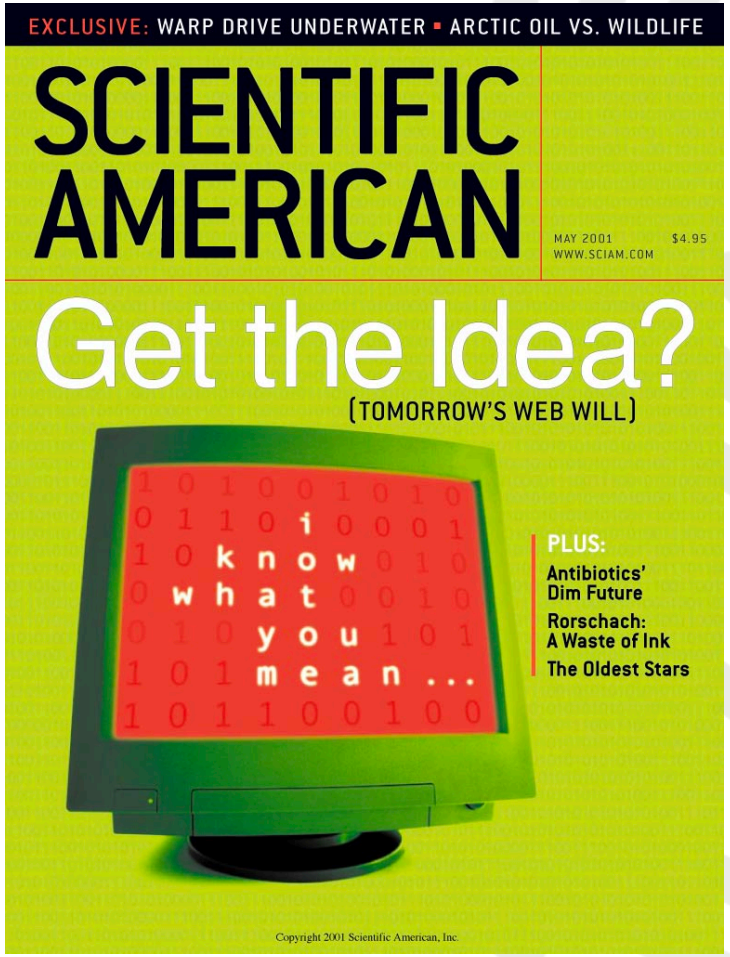


Image by Radar Networks; Nova Spivak
http://memebox.com/futureblogger/show/824



2001 The Semantic Web
https://www.scientificamerican.com

Data & Data Driven Services
How can we ensure Interoperability?

- Globally Unique identifiers
- A common protocol
- **Links** between **Documents**

URIs
HTTP
href

- Globally Unique identifiers
- A common protocol
- **Typed Links** between **Entities**

URIs
HTTP
RDF

www.sabrinakirrane.com#me

foaf: workplaceHomepage

https://www.wu.ac.at/en/Infobiz/

Home

EMAIL: sabrinakirrane@gmail.com, sabrina.kirrane@wu.ac.at
SKYPE: sabrinakirrane
ORCID: 0000-0002-6955-7718
SCOPUS: 53979829800

Technology and society are constantly evolving. Although we can't predict what the future holds we can certainly influence it!

About me

I'm a senior postdoctoral researcher at the Vienna University of Economics and Business, where I am also a member of the recently founded Research Institute for Cryptoeconomics. In addition, I am the Founding Director of the Privacy and Sustainable Computing Lab, which was setup in September 2015, and the Scientific/Technical Co-ordinator of the SPECIAL H2020 project, which kicked off in January 2017.

FOAF Vocabulary Specification 0.99

Namespace Document 14 January 2014 - *Paddington Edition*

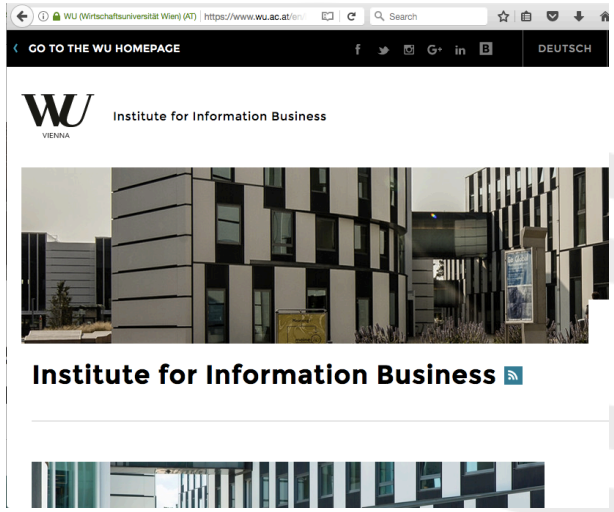
This version:
http://xmlns.com/foaf/spec/20140114.html (rdf)
Latest version:
http://xmlns.com/foaf/spec/ (rdf)
Previous version:
http://xmlns.com/foaf/spec/20100809.html (rdf)
Authors:
Dan Brickley, Libby Miller
Contributors:
Members of the FOAF mailing list (foaf-dev@lists.foaf-project.org) and the wider RDF and Semantic Web developer community. See acknowledgements.

Copyright © 2000-2014 Dan Brickley and Libby Miller

This work is licensed under a Creative Commons Attribution License. This copyright applies to the *FOAF Vocabulary Specification* and accompanying documentation in RDF. Regarding underlying technology, FOAF uses W3C's RDF technology, an open Web standard that can be freely used by anyone.

Abstract

This specification describes the FOAF language, defined as a dictionary of named properties and classes using W3C's RDF technology.



GO TO THE WU HOMEPAGE DEUTSCH

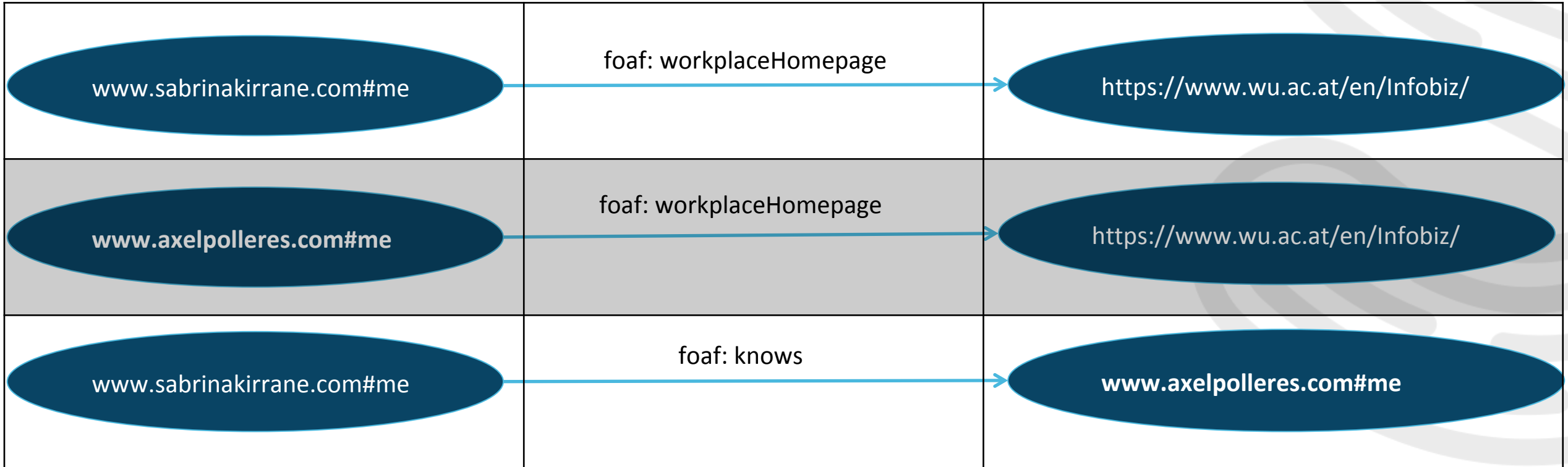
Institute for Information Business

Institute for Information Business

Data & Data Driven Services
How can we ensure Interoperability?

- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Links** between **Documents** — href

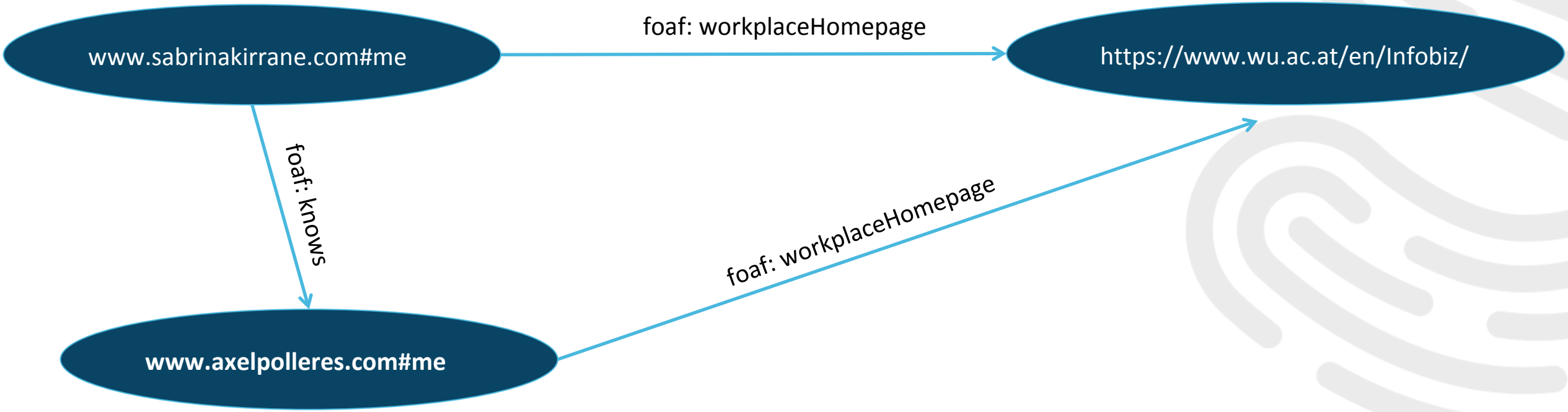
- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Typed Links** between **Entities** — RDF



Data & Data Driven Services
How can we ensure Interoperability?

- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Links** between **Documents** — href

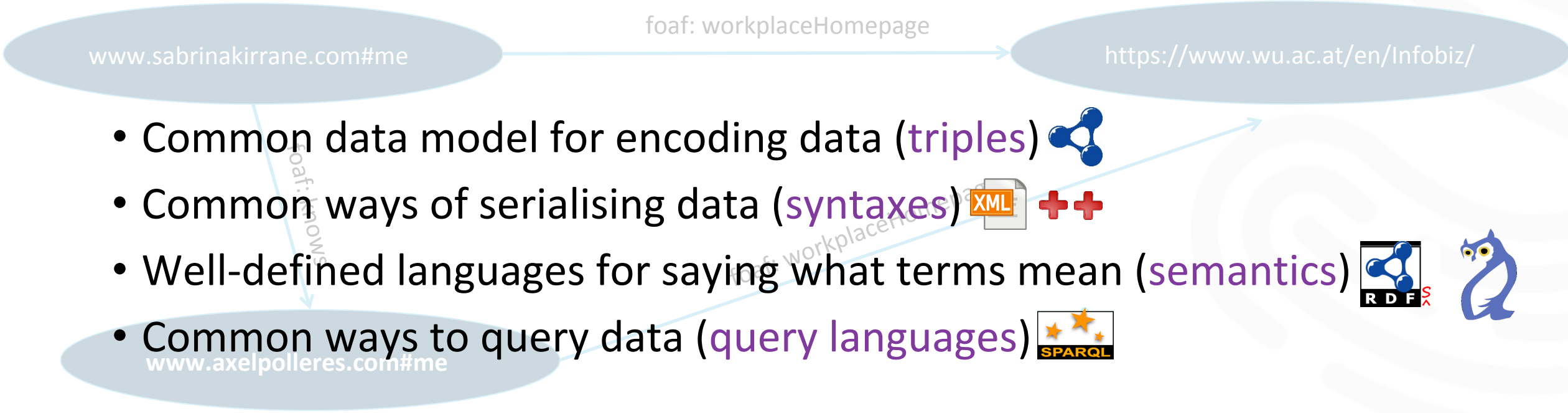
- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Typed Links** between **Entities** — RDF



Data & Data Driven Services
How can we ensure Interoperability?

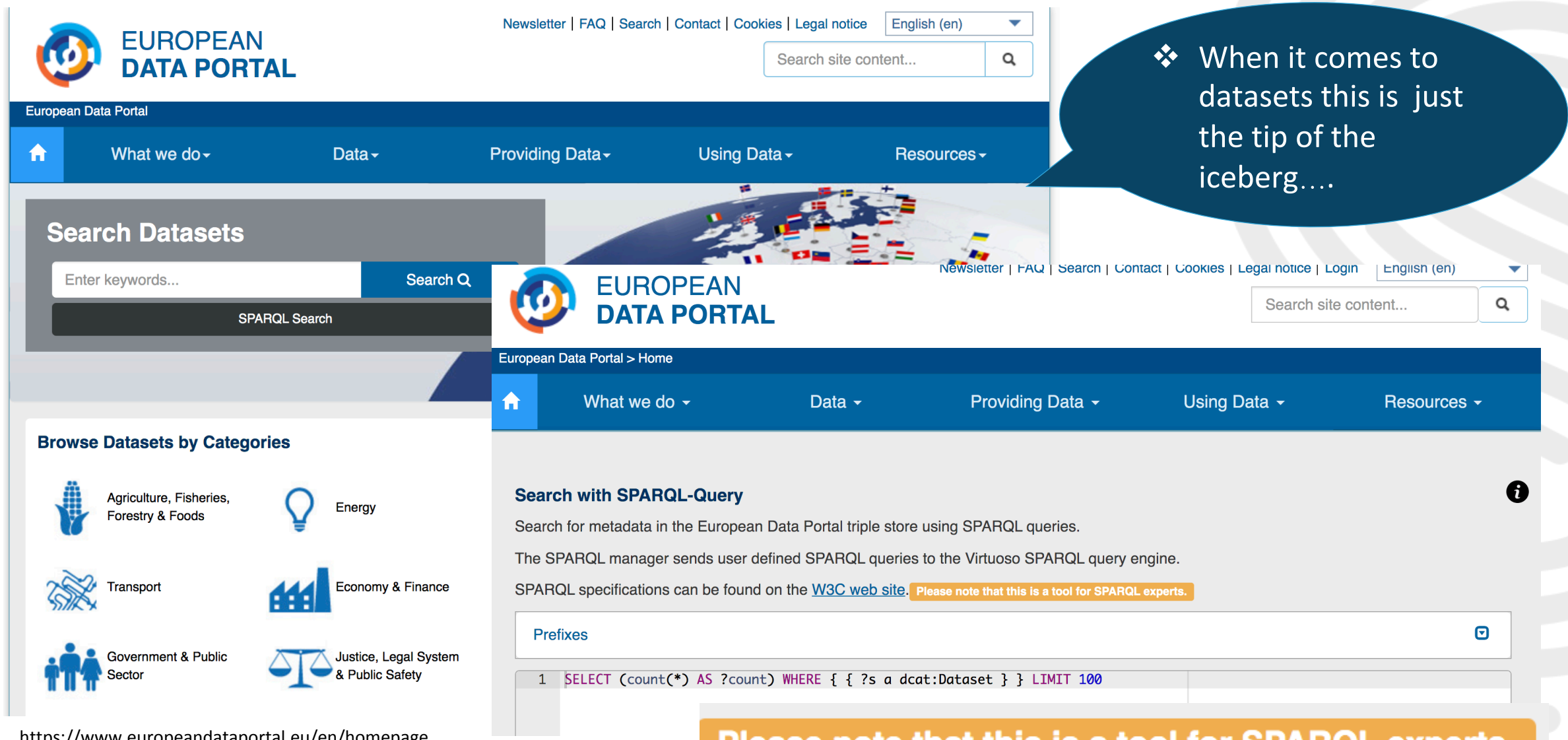
- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Links** between **Documents** — href

- Globally Unique identifiers — URIs
- A common protocol — HTTP
- **Typed Links** between **Entities** — RDF



- Common data model for encoding data (triples)
- Common ways of serialising data (syntaxes)
- Well-defined languages for saying what terms mean (semantics)
- Common ways to query data (query languages)

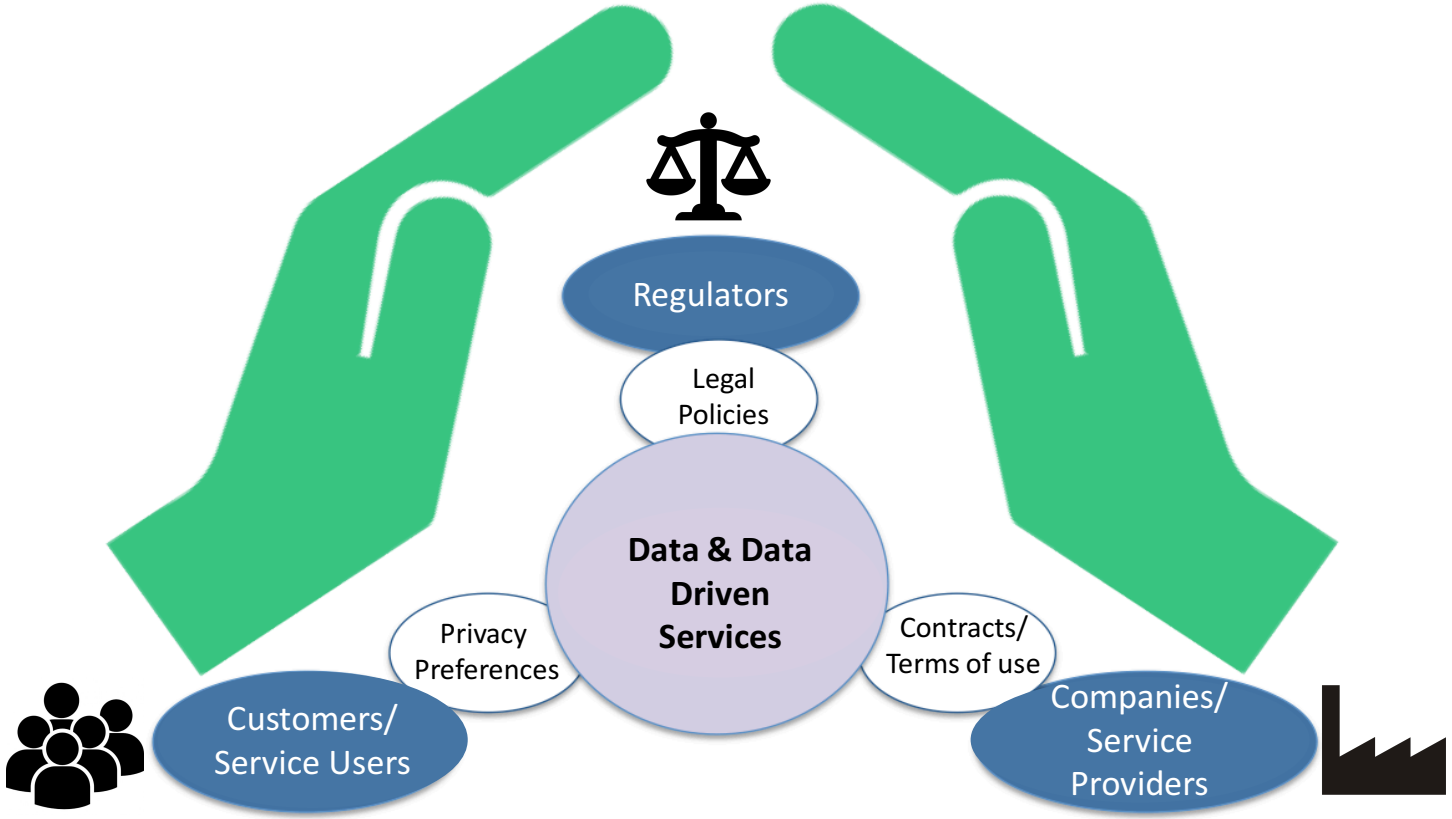
Data & Data Driven Services
Distributed Data Sources



- When it comes to datasets this is just the tip of the iceberg….

https://www.europeandataportal.eu/en/homepage

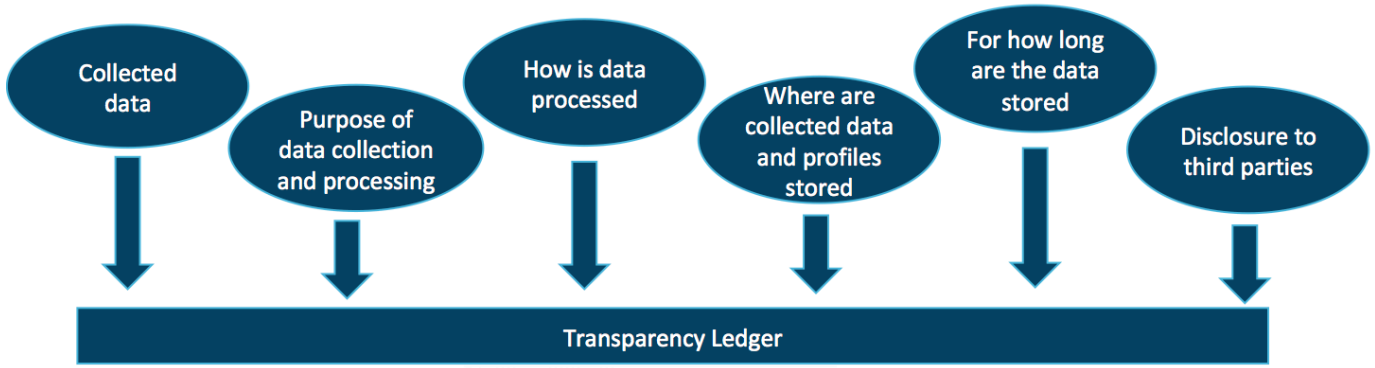
Please note that this is a tool for SPARQL experts.

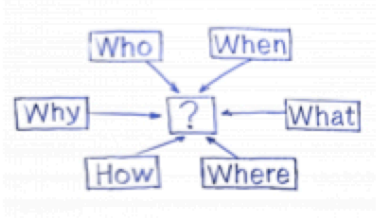


Data & Data Driven Services

Privacy Preferences

Privacy Preferences
How can we ensure Interoperability?





- Available for download via the SPECIAL website: https://www.specialprivacy.eu/publications/public-deliverables
- An unofficial draft specification has been published online https://www.specialprivacy.eu/platform/ontologies-and-vocabularies

The SPECIAL Usage Policy Language
version 0.1

Unofficial Draft 06 April 2018

Editor:
Javier D. Fernández (Vienna University of Economics and Business)

Authors:
Piero Bonatti (Università di Napoli Federico II)
Sabrina Kirrane (Vienna University of Economics and Business)
Iliana Mineva Petrova (Università di Napoli Federico II)
Luigi Sauro (Università di Napoli Federico II)
Eva Schlehahn (Unabhängies Landeszentrum für Datenschutz (ULD))

This document is licensed under a Creative Commons Attribution 3.0 License.

Abstract

This document specifies usage policy language of SPECIAL. The usage policy language is meant to express both the data subjects' consent and the data usage policies of data controllers in formal terms, understandable by a computer, so as to automatically verify that the usage of personal data complies with data subjects' consent.

The ontology defined in this document is publicly available at http://www.specialprivacy.eu/langs/usage-policy.

The SPECIAL Policy Log Vocabulary
A vocabulary for privacy-aware logs, transparency and compliance - version 0.3

Unofficial Draft 06 April 2018

Editor:
Javier D. Fernández (Vienna University of Economics and Business)

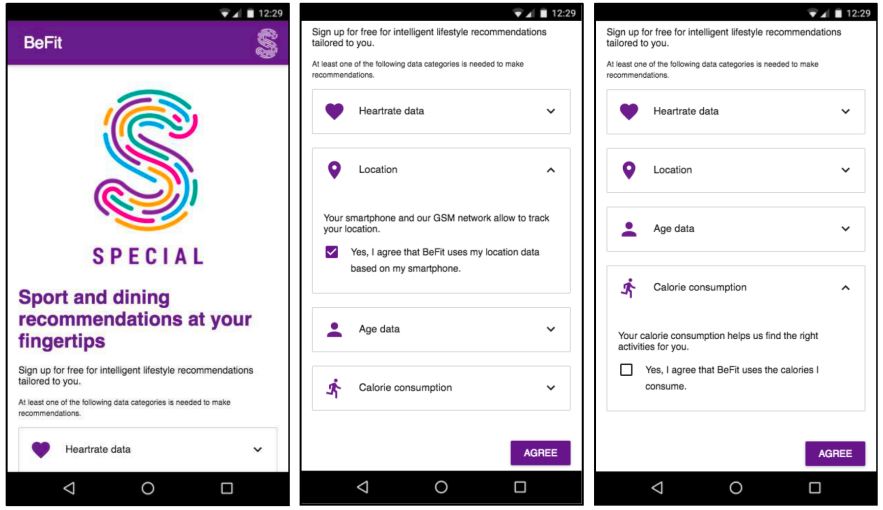
Authors:
Piero Bonatti (Università di Napoli Federico II)
Wouter Dullaert (Tenforce)
Javier D. Fernández (Vienna University of Economics and Business)
Sabrina Kirrane (Vienna University of Economics and Business)
Uros Milosevic (Tenforce)
Axel Polleres (Vienna University of Economics and Business)

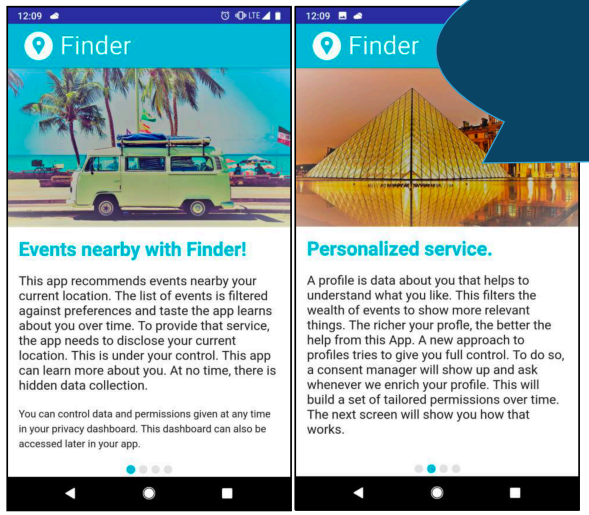
This document is licensed under a Creative Commons Attribution 3.0 License.

Abstract

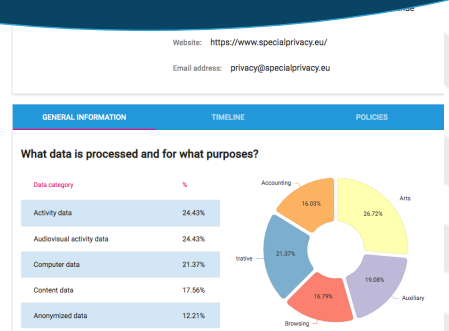
This documents specifies *splog*, a vocabulary to log data processing and sharing events that should comply with a given consent provided by a data subject. We also model the consent actions related to consent giving and revocation

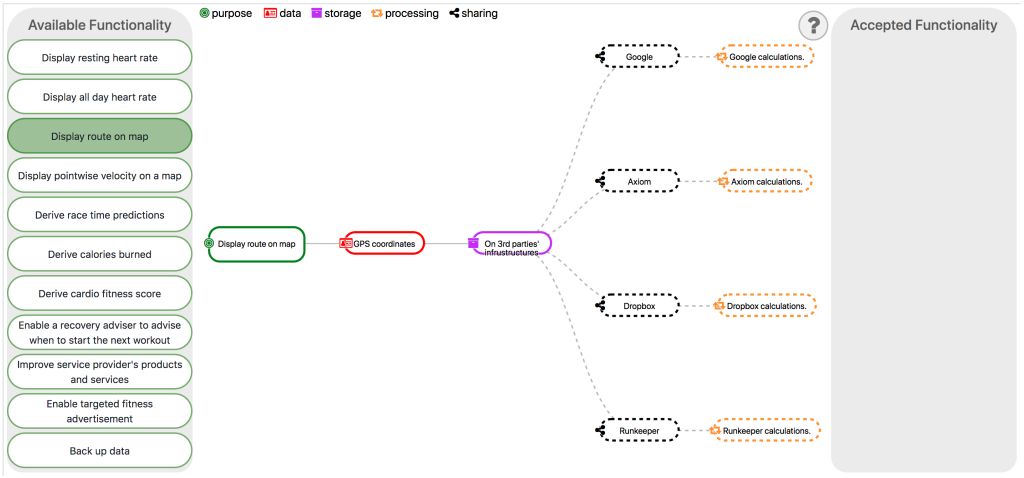
Privacy Preferences
Challenge: Human Computer Interaction

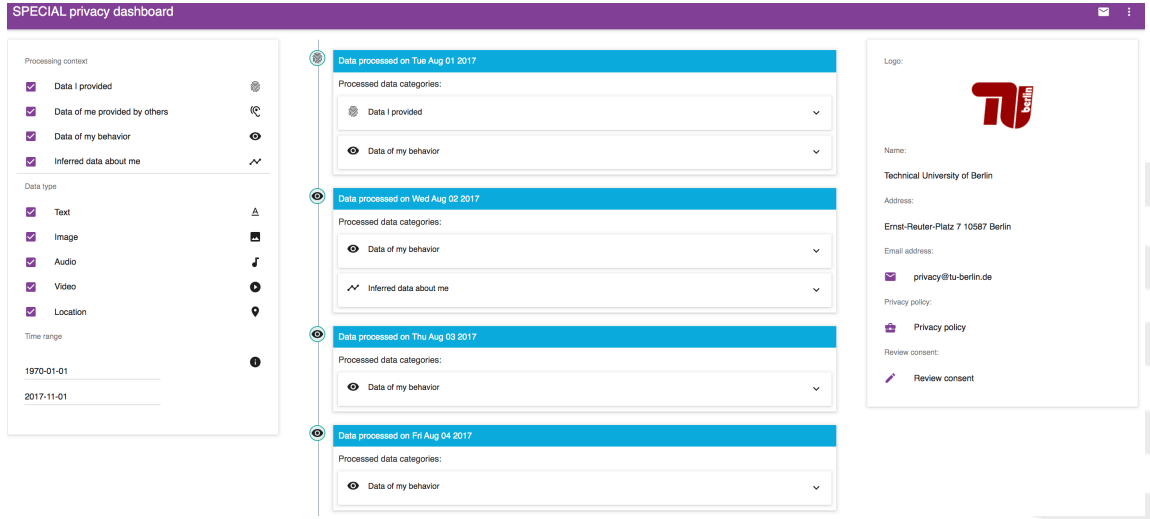


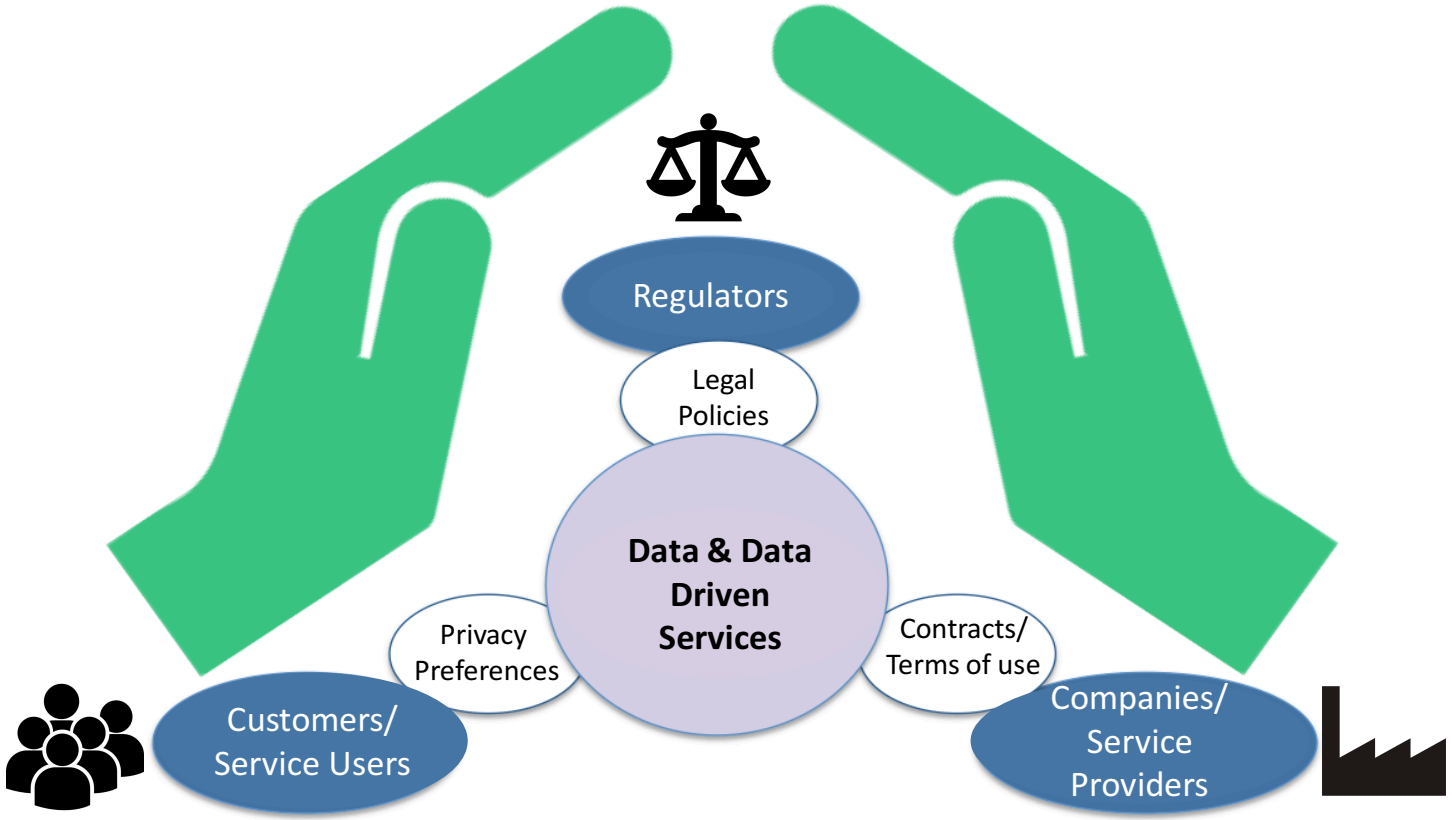


❖ We need to make it easy for individuals to manage their personal data





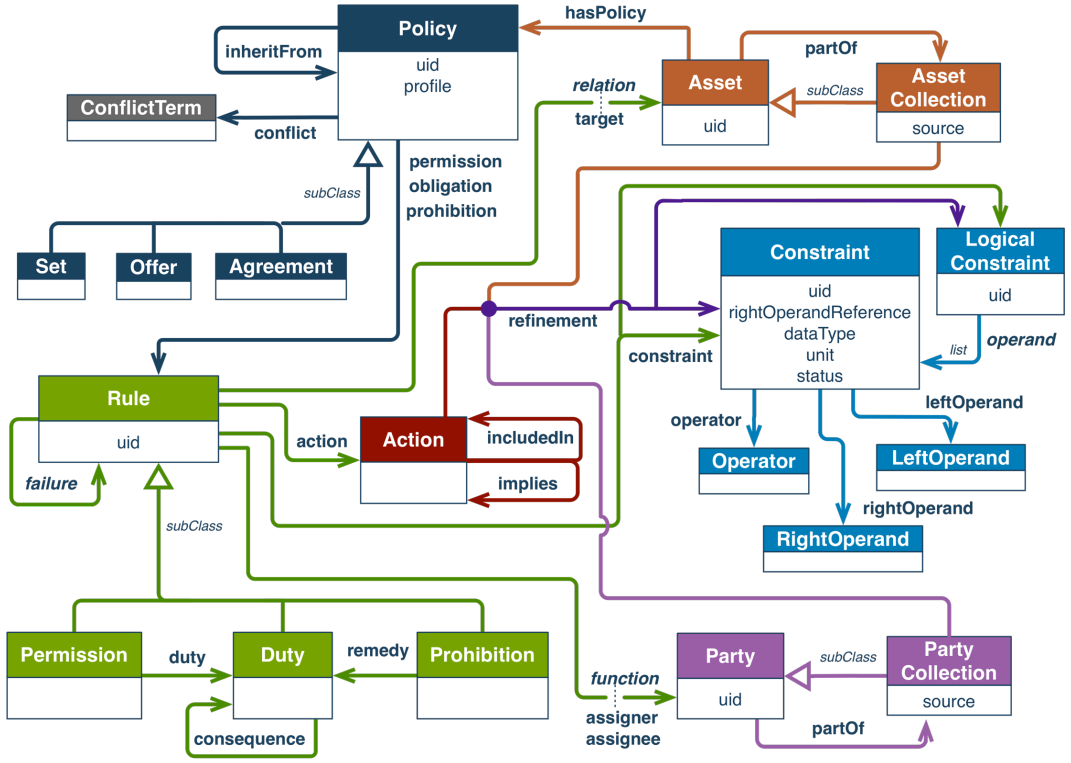




Data & Data Driven Services

Terms of Use

Contracts & Terms of Use
How can we ensure Interoperability?



- Modeling licenses using the Open Digital Rights Language
- Dependency modeling
- Conflict detection & Resolution

W3C Recommendation

W3C Recommendation

ODRL Information Model 2.2

W3C Recommendation 15 February 2018

This version:
https://www.w3.org/TR/2018/REC-odrl-model-20180215/

Latest published version:
https://www.w3.org/TR/odrl-model/

Latest editor's draft:
https://w3c.github.io/poe/model/

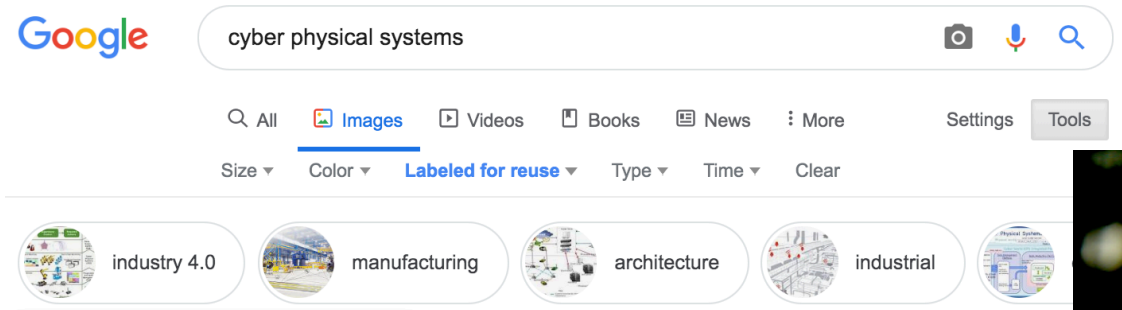
Implementation report:
https://w3c.github.io/poe/test/implementors

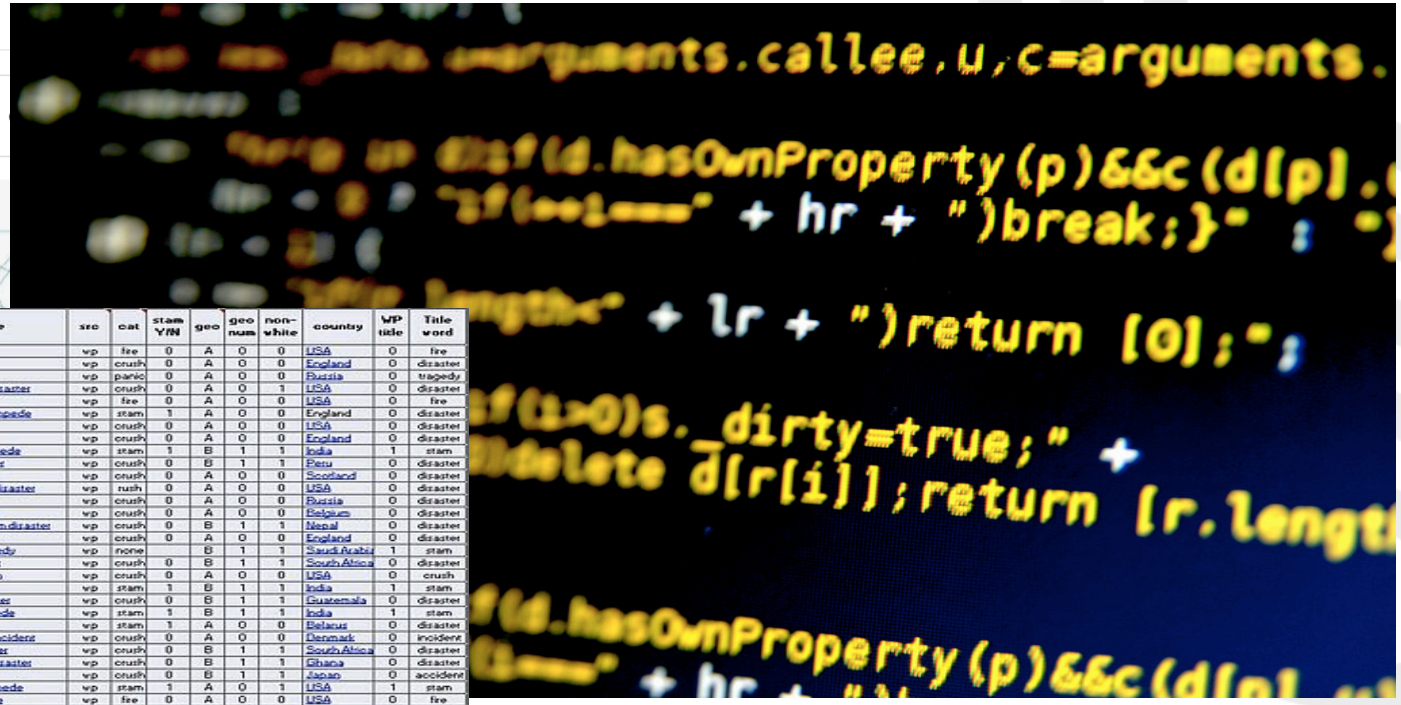
Previous version:
https://www.w3.org/TR/2018/PR-odrl-model-20180104/

Automated Rights Clearance Using Semantic Web Technologies: The DALICC Framework, Tassilo Pellegrini, Victor Mireles, Simon Steyskal, Oleksandra Panasiuk, Anna Fensel, and Sabrina Kirrane, Semantic Applications: Methodology, Technology, Corporate Use, ***Springer Semantic Applications, Methodology, Technology, Corporate Use****, 2018*

Contracts & Terms of Use Challenge: Compliance

- There are many resources without any terms of us
- We need compliance tools

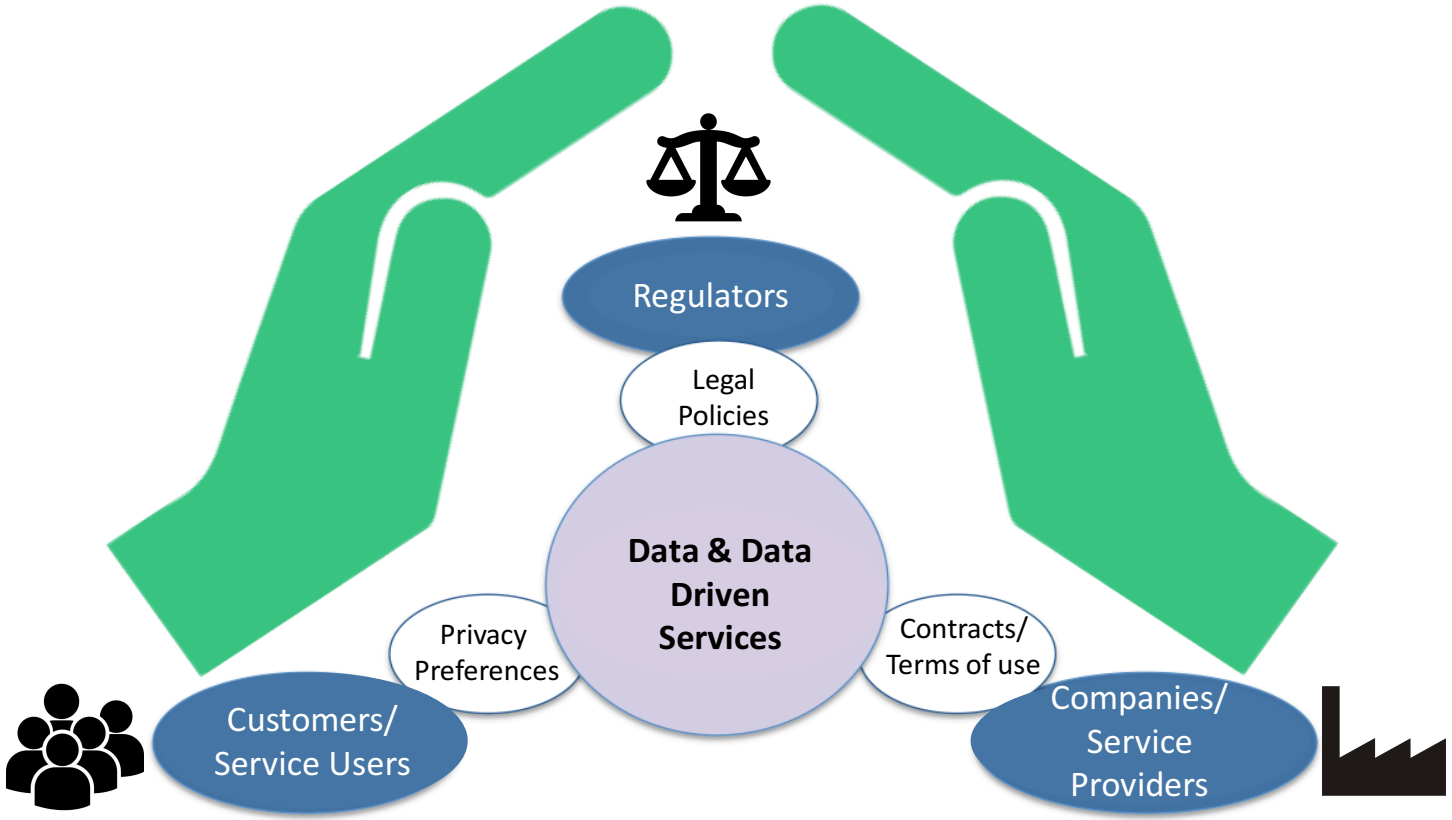




event	year	dead	injured	crowd size	event type	event type num	mod type num	relig.	indoor	counterflow	bridge	event population	population killed	specific trigger	crowd mgmt failure	escape panic	Event name	src	cat	stam Y/N	geo	geo num	non-white	country	WP title	Title word
5	1876	278		1,000	F	1	1	0	Y	N	N	A	A	Y		Y	Brooklyn Theatre fire	wp	fire	0	A	0	0	USA	0	fire
7	1883	183		1,100	T	8	8	0	Y	N	N	C	C	Y		N	Victoria Hall disaster	wp	crush	0	A	0	0	England	0	disaster
8	1896	1389	1300	500,000	P	3	3	0	N	N	N	G	G	Y		N	Khodynka Tragedy	wp	panic	0	A	0	0	Russia	0	tragedy
9	1902	115		6,000	R	5	5	1	Y	N	N	A	A	Y		Y	Shiloh Baptist Church disaster	wp	crush	0	A	0	1	USA	0	disaster
10	1903	602		2,150	F	1	1	0	Y	N	N	G	G	Y		Y	Iroquois Theatre Fire	wp	fire	0	A	0	0	USA	0	fire
11	1908	16	40		T	8	8	0	Y	N	N	C	C	Y		N	Barnsley Public Hall stampede	wp	stam	1	A	0	0	England	0	disaster
12	1913	73		400	T	8	8	0	Y	N	N	G	C	Y		Y	Italian Hall Disaster	wp	crush	0	A	0	0	USA	0	disaster
14	1946	33	400	85,000	S	7	7	0	N	N	N	G	G	N	Y	N	Burnden Park disaster	wp	crush	0	A	0	0	England	0	disaster
15	1954	800	2000	5,000,000	R	5	5	1	N	N	N	G	G	N	Y	N	1954 Kumbh Mela stampede	wp	stam	1	B	1	1	India	1	stam
17	1964	328		53,000	S	7	7	0	N	N	N	G	G	Y	Y	Y	Estadio Nacional disaster	wp	crush	0	B	1	1	Peru	0	disaster
18	1971	66	200	80,000	S	7	7	0	N	N	N	G	G	N	Y	N	1971 Ibrox disaster	wp	crush	0	A	0	0	Scotland	0	disaster
19	1979	11	26	14,770	M	2	2	0	N	N	N	G	G	Y		N	1979 The Who concert disaster	wp	rush	0	A	0	0	USA	0	disaster
20	1982	66	61	16,643	S	7	7	0	N	N	N	G	C	N	Y	N	Luzhniki disaster	wp	crush	0	A	0	0	Russia	0	disaster
21	1985	39	600	59,000	S	7	7	0	N	N	N	G	G	Y	Y	Y	Heysel Stadium disaster	wp	crush	0	A	0	0	Belgium	0	disaster
22	1988	93	100		S	7	7	0	N	Y	N	G	G	Y	Y	Y	1988 Kathmandu stadium disaster	wp	crush	0	B	1	1	Nepal	0	disaster
24	1989	96	766	55,000	S	7	7	0	N	Y	N	G	G	N	Y	N	Hillsborough disaster	wp	crush	0	A	0	0	England	0	disaster
25	1990	1426			R	5	5	1	N	N	N	A	A	N	Y	N	1990 Mecca tunnel tragedy	wp	none		B	1	1	Saudi Arabia	1	stam
26	1991	40		30,000	S	7	7	0	N	N	N			Y		Y	Orkney Stadium Disaster	wp	crush	0	B	1	1	South Africa	0	disaster
32	1993	0	73	77,745	S	7	7	0	N	N	N	G		Y	Y	N	The Camp Randall Crush	wp	crush	0	A	0	0	USA	0	crush
34	1994	113	500	50,000	PO	4	4	0	N	N	N	G	WC	Y	Y	Y	1994 Govari stampede	wp	stam	1	B	1	1	India	1	stam
35	1996	83	147	50,000	S	7	7	0	N	N	N	G	G	N	Y	N	The October 16 disaster	wp	crush	0	B	1	1	Guatemala	0	disaster
37	1999	53			R	5	5	1	N					Y			1999 Sabarimala stampede	wp	stam	1	B	1	1	India	1	stam
38	1999	53	100		M	2	2	0	N					Y		Y	Nyamiha stampede	wp	stam	1	A	0	0	Belarus	0	disaster
39	2000	9			M	2	2	0	N					N		N	Roskilde Festival 2000 incident	wp	crush	0	A	0	0	Denmark	0	incident
41	2001	43		120,000	S	7	7	0	N			N		Y	Y	Y	Ellis Park Stadium disaster	wp	crush	0	B	1	1	South Africa	0	disaster
42	2001	127			S	7	7	0	N	N	N			Y	Y	Y	Accra Sports Stadium disaster	wp	crush	0	B	1	1	Ghana	0	disaster
43	2001	11	247		P	3	3	0	N	N	Y	G	C	N	Y	N	Akashi crowd crush	wp	crush	0	B	1	1	Japan	0	accident
46	2003	21	50	1,500	M	2	2	0	Y	Y	N	A	A	Y	Y	Y	2003 E2 nightclub stampede	wp	stam	1	A	0	1	USA	1	stam
47	2003	100	230	462	F	1	1	0	Y	N	N	A	A	Y	Y	Y	The Station nightclub fire	wp	fire	0	A	0	0	USA	0	fire
52	2005	291		300,000	F	1	1	1	N	Y	N	G	WC	N			Mandhar Devi temple stampede	wp	stam	1	B	1	1	India	1	stam
53	2005	953		1,000,000	R	5	5	1	N	N	Y	G	G	Y	Y	Y	2005 Al-Aaimmah bridge stampede	wp	stam	1	B	1	1	Iraq	1	stam
55	2006	345	289	2,130,594	R	5	5	1	N	N	N	A	A	N		N	Hajj 2006	wp	alt		B	1	1	Saudi Arabia	1	stam
56	2006	73	400	30,000	S	7	7	0	N	N	N			Y	Y	N	PhilSports Stadium stampede	wp	stam	1	B	1	1	Philippines	1	stam
64	2008	162	47	3,000	R	5	5	1	N		N	G	G	Y	Y	Y	Naina Devi stampede	wp	stam	1	B	1	1	India	1	stam
66	2008	224	425	25,000	R	5	5	1	N		N	M	A	Y	Y	Y	2008 Jodhpur stampede	wp	stam	1	B	1	1	India	1	stam
69	2009	19	135	36,000	S	7	7	0	N	N	N			Y	Y	Y	Houphouet-Boigny Arena stampede	[121]	crush	0	B	1	1	Côte d'Ivoire	1	stam
73	2010	21	500	250,000	M	2	2	0	N	Y	N	G	G	N	Y	N	Love Parade disaster	wp	crush	0	A	0	0	Germany	0	disaster
74	2010	347	755	4,000,000	P	3	3	1	N		Y	G	G	N	Y	N	Phnom Penh stampede	wp	stam	1	B	1	1	Cambodia	1	stam
75	2011	102	100		R	5	5	1	N	N	N				Y	N	2011 Sabarimala stampede	wp	stam	1	B	1	1	India	1	stam
79	2013	60	200	50,000	P	3	3	0	N	Y	N	G	WC				2013 Houphouet-Boigny stampede	wp	stam	1	B	1	1	Côte d'Ivoire	1	stam
81	2013	242	168	1,500	F	1	1	0	Y	N	N	A	A	Y	Y	Y	Kiss nightclub fire	wp	fire	0	B	1	1	Brazil	0	fire
82	2013	36	39		R	5	5	1	Y	N	Y	G	G		Y		2013 Kumbh Mela stampede	wp[25]	stam	1	B	1	1	India	1	stam
83	2013	115	110	500,000	R	5	5	1	N	N	Y	G	WC		Y		2013 Madhya Pradesh stampede	wp[26]	stam	1	B	1	1	India	1	stam
84	2014	32	26		R	5	5	1	N	N	N	G	WC		N		2014 Patna stampede	wp[27]	stam	1	B	1	1	India	1	stam
85	2014	7	40		PO	4	4	0	N	N	N		A	N	Y	N	2014 Multan stampede	wp	stam	1	B	1	1	Pakistan	1	stam
86	2014	11	40	30,000	R	5	5	1	N	N	N	G	G	N	Y	N	Korekove stadium stampede	wp	stam	1	B	1	1	Zimbabwe	1	stam
87	2014	36	49	300,000	P	3	3	0	N	Y	N	G	G	N	Y	N	2014 Shanghai stampede	wp[28]	crush	0	B	1	1	China	1	stam
88	2015	28		5,000	S	7	7	0	N	N	N	G	G	Y	Y	N	30 June Stadium Stampede	wp[29]	stam	1	B	1	1	Egypt	1	stam
89	2015	18	78		P	3	3	0	N	N	N			Y		Y	2015 Haiti Carnival stampede	wp	stam	1	B	1	1	Haiti	1	stam
92	2015	1554	934	2,000,000	R	5	5	1	N	Y	N	A	A	N	Y	N	2015 Mina stampede	en	stam	1	B	1	1	Saudi Arabia	1	stam

https://en.wikipedia.org/wiki/File:Data_for_2015_stampede_analysis.jpg

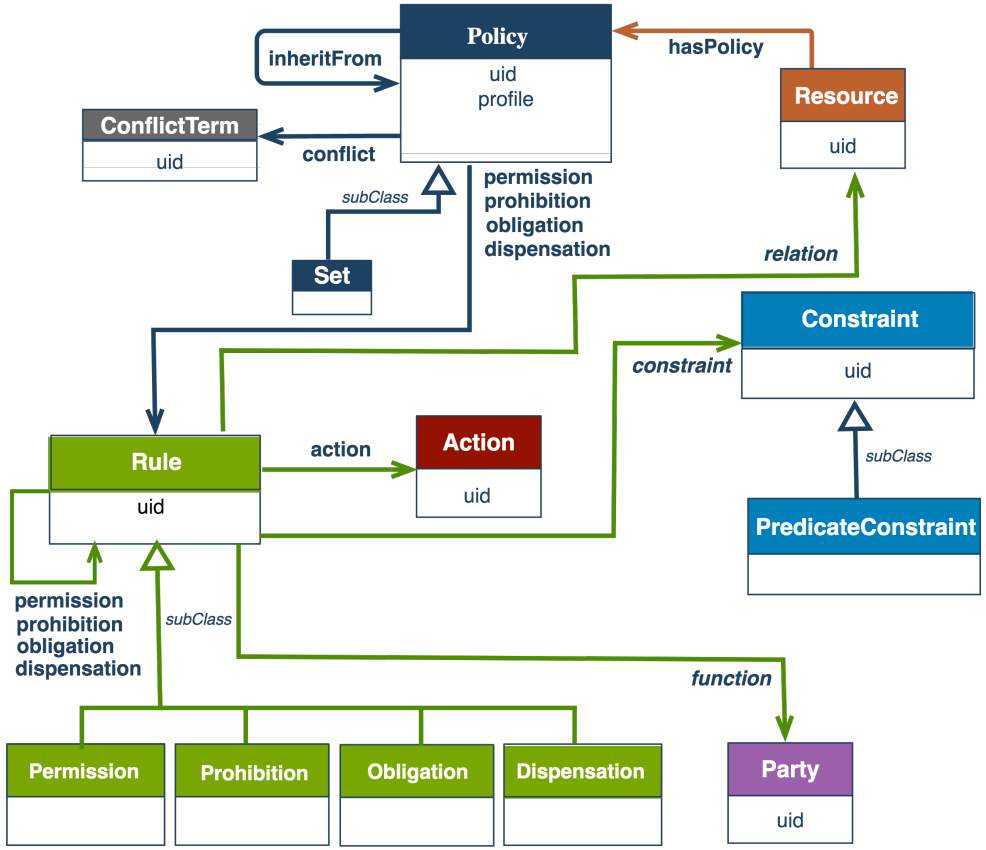
60095534



Data & Data Driven Services

Legal Policies

Legal Policies
How can we ensure Interoperability?



- Modeling regulatory obligations using an adaption of the Open Digital Rights Language
- Automated compliance checking for business policies

Draft Specification

ODRL Regulatory Compliance Profile
version 0.1

Unofficial Draft 29 May 2019

Editor:
Sabrina Kirrane (Vienna University of Economics and Business)

Authors:
Sabrina Kirrane (Vienna University of Economics and Business)
Marina De Vos (University of Bath)
Julian Padget (University of Bath)

This document is licensed under a Creative Commons Attribution 4.0 License.

Abstract

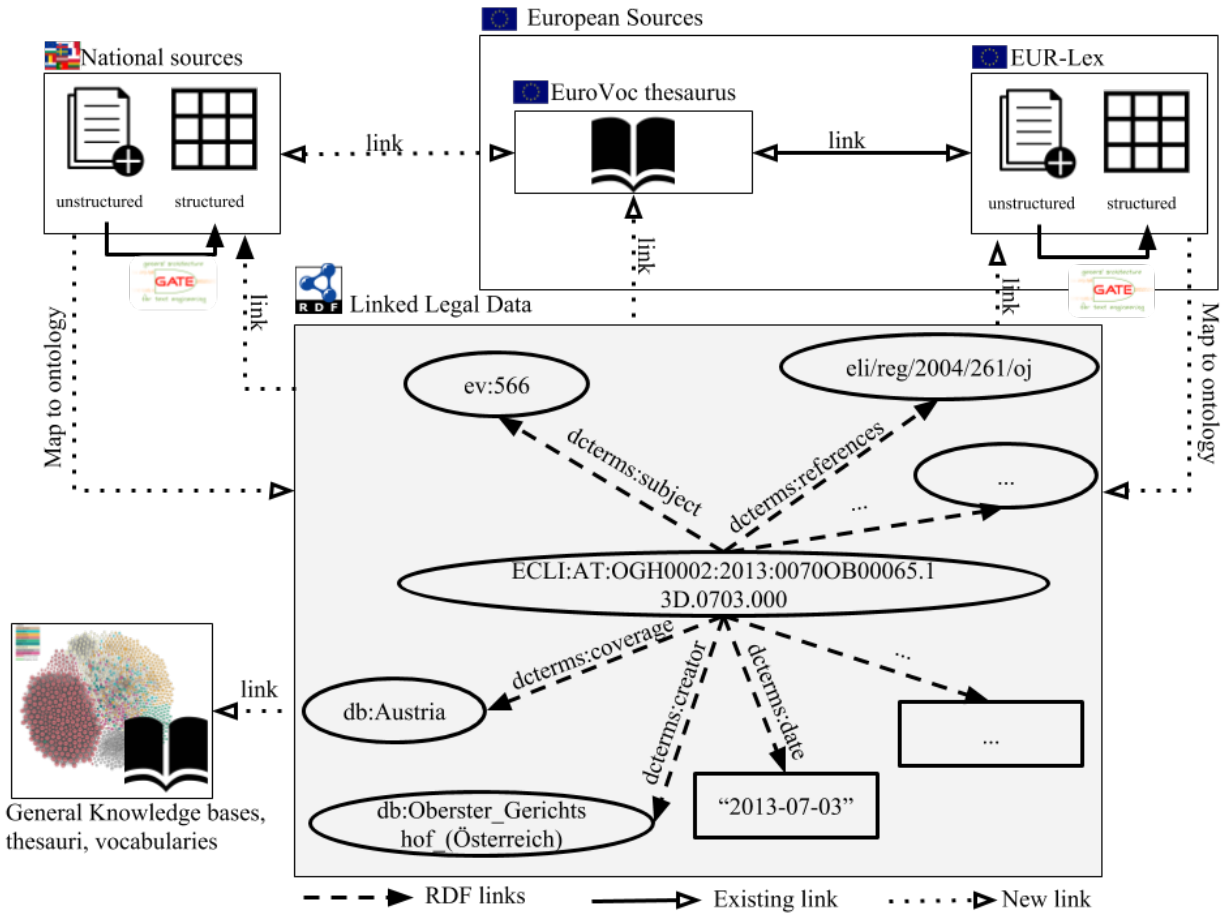
The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and encoding mechanisms for representing statements about the usage of content and services.

This document constitutes an ODRL profile that adapts the ODRL Core Model and Vocabulary with concepts and terms to support regulatory compliance checking of business policies.

In essence, ODRL Regulatory Compliance Profile policies are used to represent regulatory permissions, prohibitions, obligations, and dispensations, which may be limited by constraints (e.g., temporal, spatial).

*ODRL policy modelling and compliance checking, Marina De Vos, Sabrina Kirrane, Julian Padget and Ken Satoh, Proceedings of the **3rd International Joint Conference on Rules and Reasoning** (RuleML+RR 2019)*

Legal Policies
Challenge: Cross Jurisdiction



- Multilingual Cross Jurisdiction
- Compliance Tools

- Extracting temporal and event data from legal text
- Modeling metadata relating to legislation and cases in a legal knowledge graph

Legal Knowledge Graph
Version 0.1

Interlinking Legal Data, Erwin Filtz, Sabrina Kirrane and Axel Polleres, Proceedings of the Posters and Demos Track of the 14th International ***Conference on Semantic Systems*** *(SEMANTiCS 2018)*

Data & Data Driven Services

Challenges & Opportunities

Data Value Chains
Challenges & Opportunities

- **Standardisation** of vocabularies (e.g., privacy, legal, licensing) is difficult
- Privacy is only the tip of the iceberg, from a usage control perspective we also need to consider other **regulations, licenses, social norms, cultural differences**
- There are **cognitive limitations** in terms of understanding how data is /will be used
- Ensuring such systems are **comply with usage constraints** is a crucial to success (i.e., all usage policies are adhered to and the system as a whole works as expected)
- We need to embrace **distributed and decentralised systems**, which complicates things further

Thank you / contact details

Technical/Scientific contact

Sabrina Kirrane

Vienna University of Economics and Business

sabrina.kirrane@wu.ac.at

Administrative contact

Jessica Michel Assoumou

ERCIM W3C

jessica.michel@ercim.eu

Horizon 2020
European Union funding
for Research & Innovation

The project SPECIAL (Scalable Policy-awarE linked data arChitecture for prIvacy, trAnsparency and compLiance) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731601 as part of the ICT-18-2016 topic Big data PPP: privacy-preserving big data technologies.