

Data sharing with knowledge graphs for sustainable and healthy diets

PLDN, Hilversum

27.09.2022, Anna Fensel



Outline

- On Semantic Technology
 - Semantic Web
 - Knowledge Graphs
 - FAIR Data
 - Data Access: Licenses, Consent, Contracts
- Linking Healthy Food to Linked Data
 - Food in Knowledge Graphs and FAIR Data
 - Need for Behavioural Change
- Outlook

Outline

- On Semantic Technology
 - Semantic Web
 - Knowledge Graphs
 - FAIR Data
 - Data Access: Licenses, Consent, Contracts
- Linking Healthy Food to FAIR Data
 - Food in Knowledge Graphs and FAIR Data
 - Need for Behavioural Change
- Summary of Research Activities

Global Challenges

■ Digital transformation challenges e.g.:

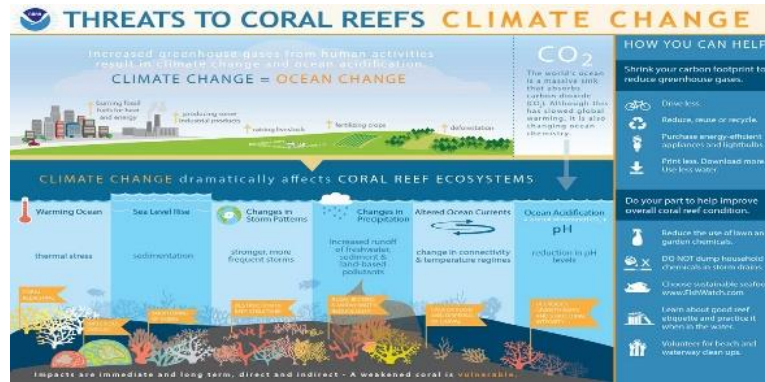
- Publishing content and information, its dissemination
- Marketing, communication and sales online

(e.g. tourism, entertainment, media...)

■ Behavioral change challenges such as like energy efficiency, sustainability, climate change, food security

■ Sustainable and healthy diets

-> facilitated by linked data



What is the Semantic Web?

» *"The Semantic Web is an extension of the current web in which information is given **well-defined meaning**, better enabling computers and people to work in cooperation."*

T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", Scientific American, May 2001



Ontology example

Concept

conceptual entity of the domain

Property

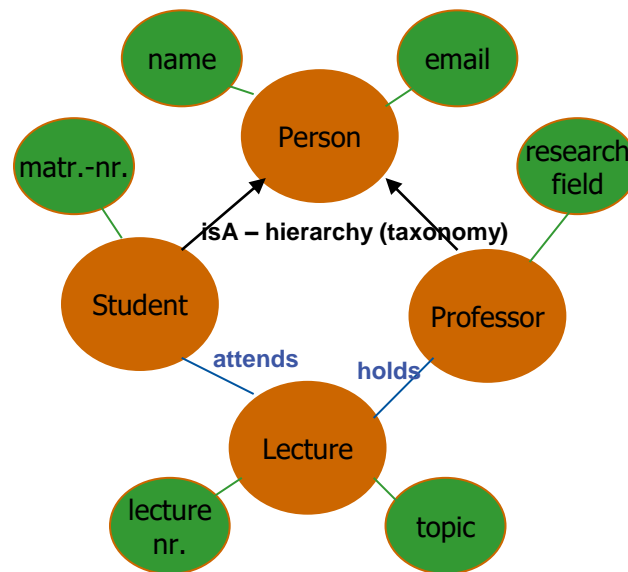
attribute describing a concept

Relation

relationship between concepts or properties

Axiom

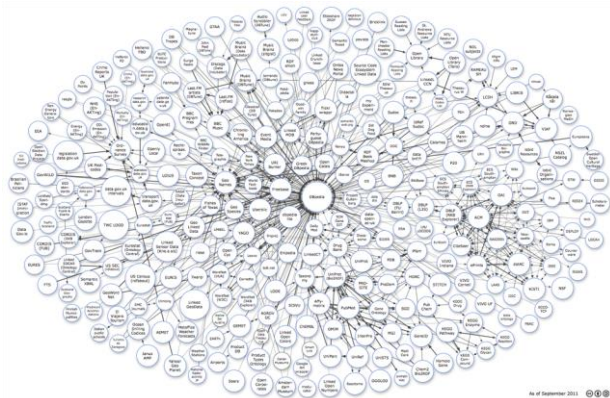
coherency description between Concepts / Properties / Relations via logical expressions



$\text{holds}(\text{Professor}, \text{Lecture}) \Rightarrow$
 $\text{Lecture.topic} = \text{Professor.researchField}$

Practical cases:

Linked Data, Open government data



A screenshot of the HM Government data.gov.uk website. The header features the HM Government logo and the URL "data.gov.uk". Below the header is a navigation bar with links: Home, Blog, Data, SPARQL, Apps, Ideas, Forum, Wiki, Resources, and About. The main content area has a green banner with the text "Unlocking innovation" and "Working with UK Public Sector information and data". To the right of the banner is a blue molecular model. Below the banner, there is a paragraph of text about government data reuse and a "Search Data" section with a search bar and a "Search" button. To the right of the search bar is a "Browse for Data" section with links to "List all datasets" and "Common tags". On the far right, there is a sidebar with a "Subscribe by RSS" button, a "User login" section with fields for "Username" and "Password", and a "Log in" button. At the bottom of the sidebar is a "What is the Semantic Web?" section with a logo and the text "Combining different data".

Data Quality: 5-star Linked OPEN Data



★ Available on the web (whatever format) but with an open license, to be Open Data

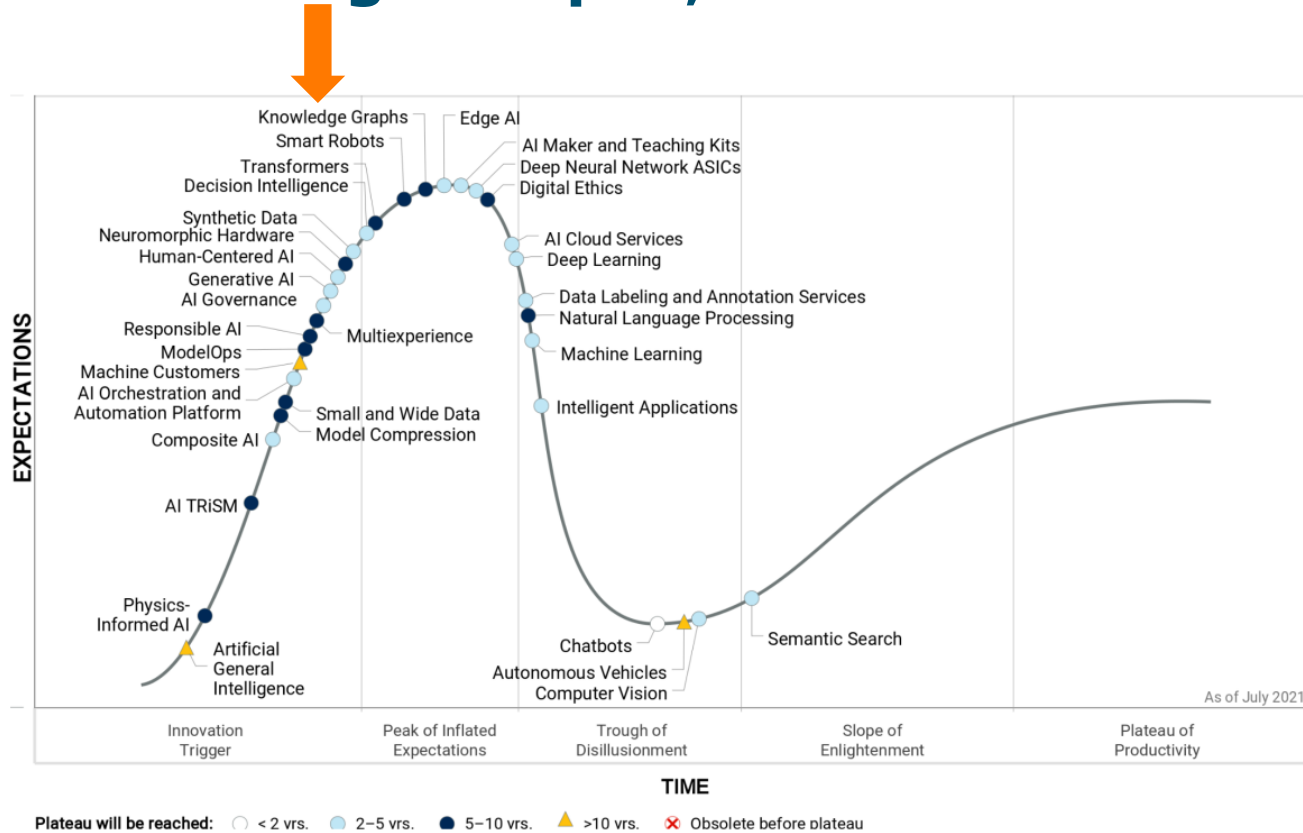
★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (URIs, RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

Knowledge Graphs, in 2021



The Semantic Web as an integrated graph of knowledge graphs

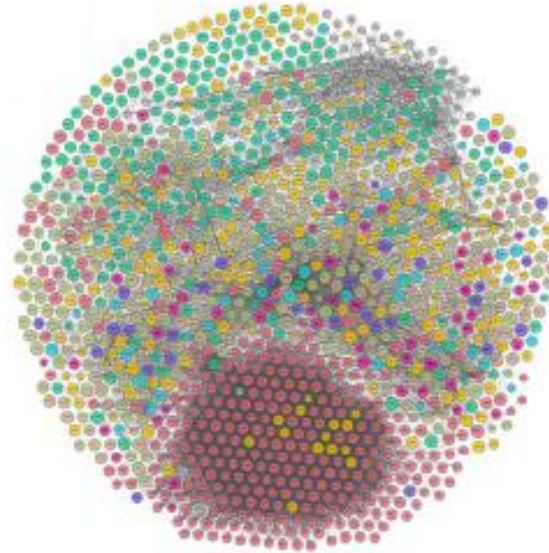
Knowledge Graphs

Name	Entities	Relations	Types	Facts
Freebase	40M	35K	26.5K	637M
DBpedia (en)	4.6M	1.4K	735	580M
YAGO3	17M	77	488K	150M
Wikidata	15.6M	1.7K	23.2K	66M
NELL	2M	425	285	433K
Google KG	570M	35K	1.5K	18B
Knowledge Vault	45M	4.5K	1.1K	271M
Yahoo! KG	3.4M	800	250	1.39B





- **Manual Construction** - curated, collaborative
- **Automated Construction** - semi-structured, unstructured

Right: **Linked Open Data cloud** - over 1200 interlinked KGs encoding more than 200M facts about more than 50M entities.

Spans a variety of domains - Geography, Government, Life Sciences, Linguistics, Media, Publications, Cross-domain..



FAIR Data

 F indable	Data and materials enriched with metadata assigned with a unique identifier
 A ccessible	Data and metadata stored in a trusted repository with an open and free protocol. Accessible by machines and humans
 I nteroperable	Using vocabularies and public domain ontologies the metadata can be referenced and linked
 R eusable	Additional documentation and protocols describing the acquisition of the data, licensed with a detailed provenance

Essentially,
FAIR data is about semantics
and knowledge graphs

Key paper: Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.

Data! Lots of it!

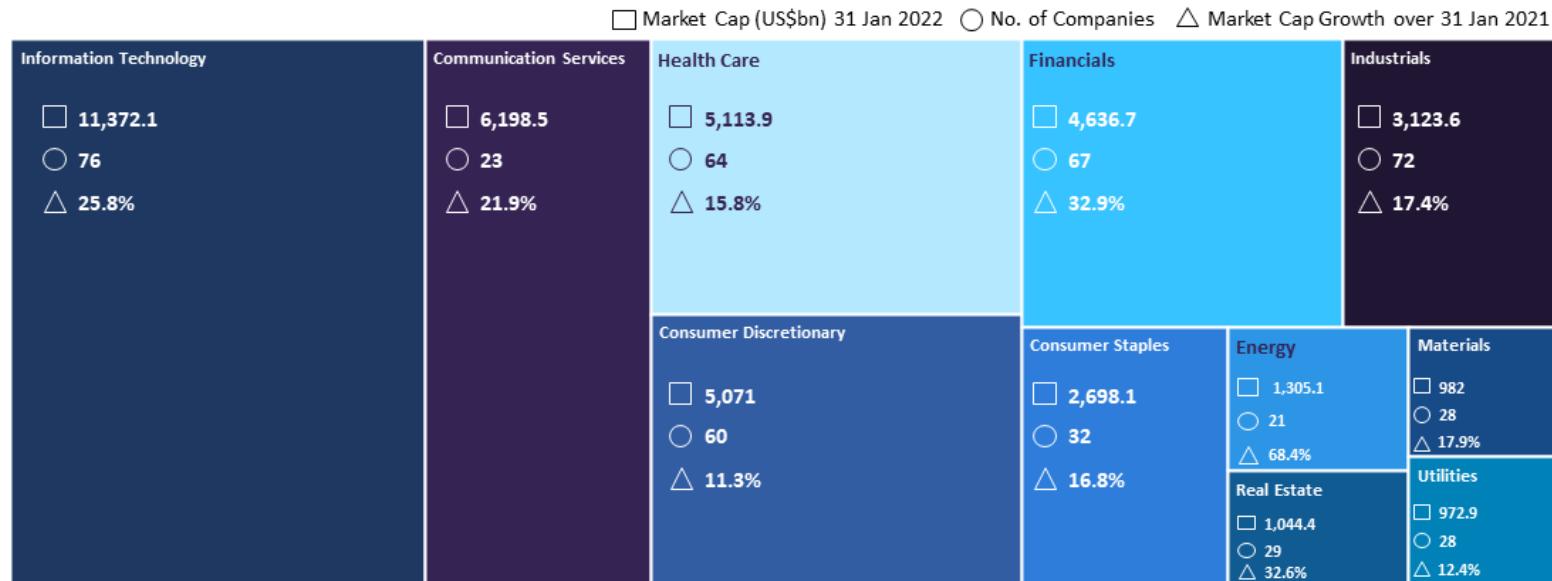


starting 1999

Data (technology & communications) - is the new gold?



Sector Composition of S&P 500 as on 31 Jan 2022



2022 - Europe is in trouble



Not all data and knowledge are to be open



Towards de-centralization

■ In communication



Edward Snowden



Edward Joseph „Ed“ Snowden ist ein US-amerikanischer Whistleblower und ehemaliger CIA-Mitarbeiter. Seine Enthüllungen gaben Einblicke in das Ausmaß der weltweiten Überwachungs- und Spionagepraktiken von Geheimdiensten – überwiegend jenen der Vereinigten Staaten und Großbritanniens. [Wikipedia](#)

Geboren: 21. Juni 1983 (Alter 35 Jahre), Elizabeth City, North Carolina, Vereinigte Staaten

Lebenspartnerin: Lindsay Mills

Bücher: *Everything You Know about the Constitution is Wrong*, *Superners: Gespräche mit Helden*

Ausbildung: Anne Arundel Community College (2004–2005), MEHR

Doku: Citizenfour



Pawel Walerjewitsch Durow

Unternehmer

Pawel Walerjewitsch Durow ist ein russischer Unternehmer und Gründer des in Russland populärsten sozialen Netzwerks vk.com und des Instant-Messengers Telegram. Forbes schätzt sein Vermögen auf umgerechnet 1,7 Milliarden US-Dollar. Bisweilen wurde er in Medien als „russischer Mark Zuckerberg“ genannt. [Wikipedia](#)

Geboren: 10. Oktober 1984 (Alter 33 Jahre), Sankt Petersburg, Russland

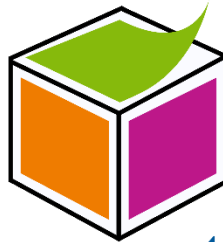
Größe: 1,76 m

Beruf: Entrepreneur

Vermögen: 1,7 Milliarden USD (2018)

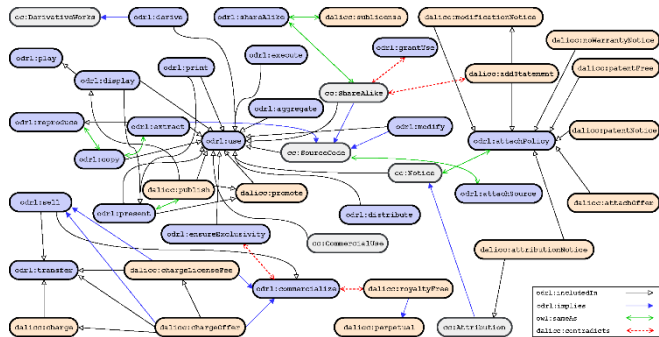
Ausbildung: Staatliche Universität Sankt Petersburg (2002–2006)

■ In payments



Semantic
blockchain

Licenses

**LICENSE SEARCH**

PROVENANCE INFORMATION

What kind of asset do you want to license?

☒ Creative work (i.e. text, picture, sound, movie)

CLEAR

Dataset

☒ Software

PERMISSIONS

  Do you allow to copy the work?

☒ ☐ Do you allow others to modify the work?

DUTIES

☒ Others must state changes

☒ ☐ ☒ Do you allow others to distribute the work?

☒ ☐ ☒ Do you allow to create derivatives from your work?

☒ ☐ ☐ Do you allow commercial use?

☒ ☐ ☐ Do you want to apply a share-alike clause?

FULL MATCHING LICENSES

- » Attribution 4.0 International
- » The MIT License
- » GNU Lesser General Public License

LICENSE LIBRARY

[Search for a license](#)

GNU GENERAL PUBLIC LICENSE, VERSION 2

INFO

- » uri: https://dalicc.net/license-library/GNU_GPL_v2
- » jurisdiction: Worldwide » title: GNU General Public License , version 2
- » hasVersion: 2.0 » alternative: GNU GPLv2 » validityType: Perpetual
- » source: <https://www.gnu.org/licenses/old-licenses/gpl-2.0.html>
- » publisher: Free Software Foundation » target: Software
- » legalcode: <https://www.gnu.org/licenses/old-licenses/gpl-2.0.html>

✓ PERMISSIONS

» Reproduce	» Display	❶ DUTIES	» ChargeLicenseFee
» Distribute	» Present	» Notice	» Grant Use
» Modify	» Derive	» Source Code	
» Derivative Works		» Share Alike	
» Commercial Use			

* PROHIBITIONS

- ChargeLicenseFee
- Grant Use

! DUTIES

- » Notice
- » Source Code
- » Share Alike

LICENSE COMPOSER

PROVENANCE INFORMATION

PERMISSIONS

- ☒ ☒ Do you allow to copy the work?
- ☒ ☒ Do you allow others to modify the work?
- ☒ ☒ Do you allow others to distribute the work?
- ☒ ☐ ☒ Do you allow to create derivatives from your work?
- ☒ ☐ ☒ Do you allow commercial use?
- ☒ ☐ ☒ Do you want to apply a share-alike clause?

ACCEPTING WARRANTY OR ADDITIONAL LIABILITY

LIMITATION OF LIABILITY

DISCLAIMER

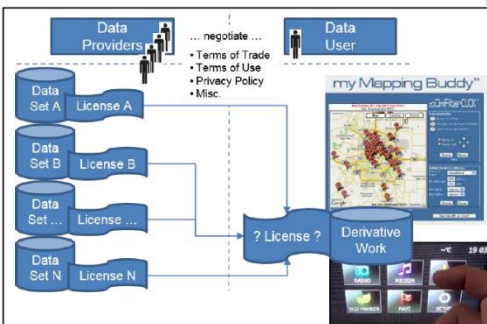
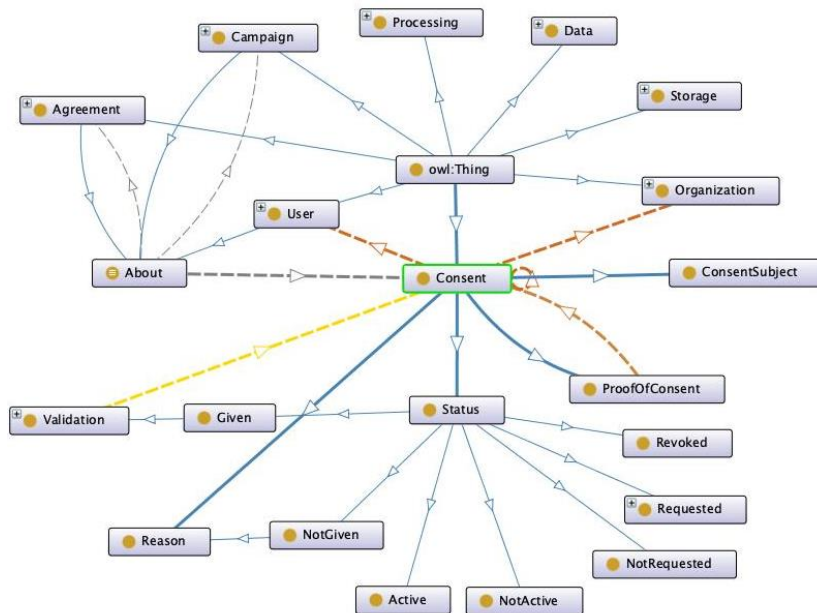
[DOWNLOAD PDF](#)

Figure 1: Problem of License Clearance

Consent

- Based on the CampaNeo project ontology, which models **consent**, **campaigns**, **contracts**, **data**, **data processing**
- Reuses the GConsent¹, FIBO² and DPV³ ontologies
- Provides common understanding between all entities in the CampaNeo project
- Supports the interoperability, traceability and transparency of data processing



Metrics

Axiom	696
Logical axiom count	290
Declaration axioms count	224
Class count	166
Object property count	42
Data property count	8
Individual count	0
Annotation Property count	12

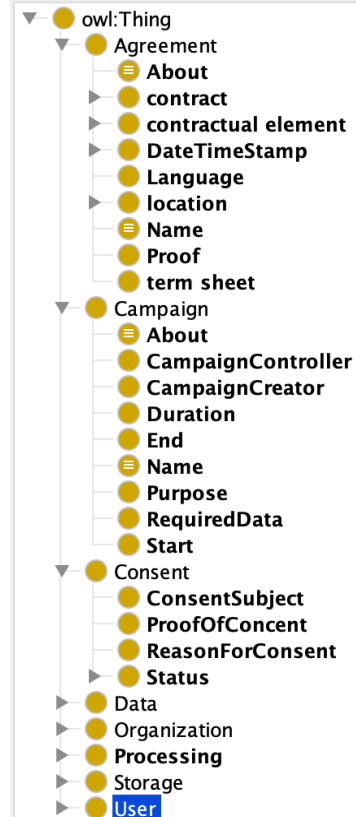
Class axioms

SubClassOf	171
EquivalentClasses	1
DisjointClasses	3
GCI count	0
Hidden GCI Count	2

Object property axioms

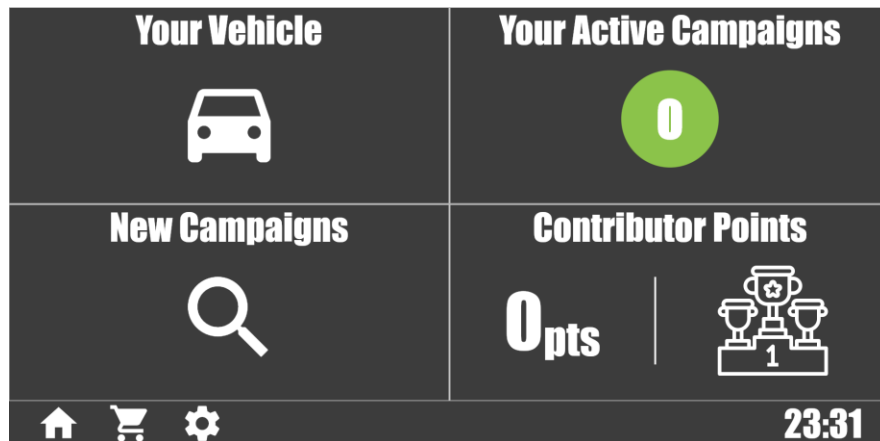
SubObjectPropertyOf	24
EquivalentObjectProperties	0
InverseObjectProperties	0
DisjointObjectProperties	0
FunctionalObjectProperty	0
InverseFunctionalObjectProperty	0

The CampaNeo Ontology and Metrics

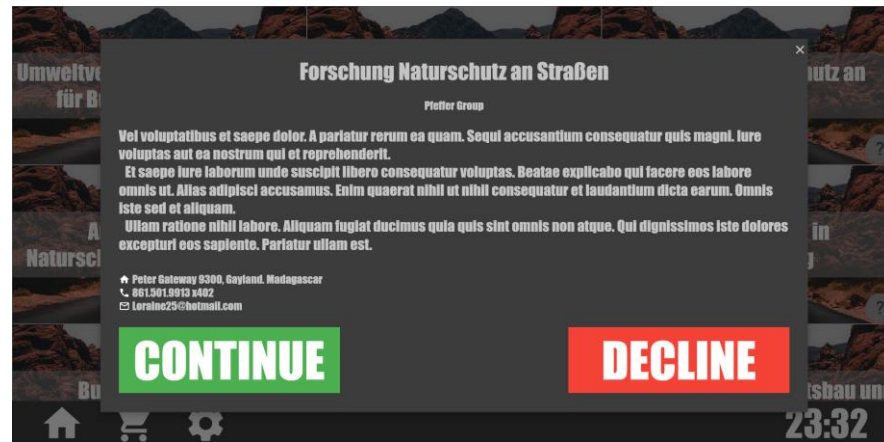


The CampaNeo User Interface (UI): Consent Request

- A cross-platform tool/web application for informed consent management (e.g. allows requesting, giving and revoking consent)
- Improves user experience (ease the comprehension of consent)
- Breaks down formalities (e.g. complex privacy policies written in legalese)
- Increases transparency regarding data processing
- Uses gamification to raise awareness about data sharing



CampaNeo UI's Main Screen

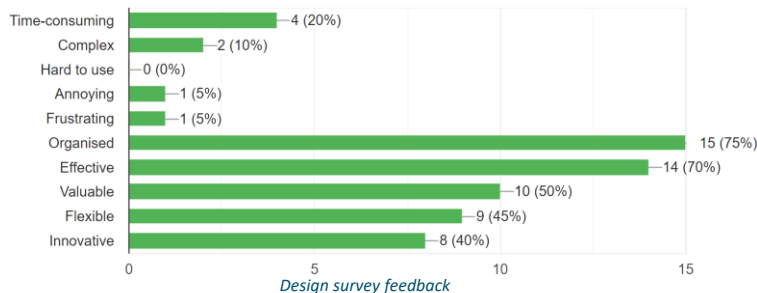


Campaign Description Dialog

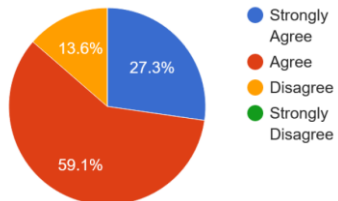
The CampaNeo User Interface (UI): Consent Request – Evaluation Results

Research Questions:

- "Does a graphical visualisation of an informed consent request via UI help raise one's awareness?"
 - Yes! "According to GDPR, I have the right to..." - 82.55% average correctness
- "Do incentives affect one's willingness to consent?"
 - Yes! 81.9% are motivated to engage in campaigns
 - Yes! 63.7% feel inclined to recommend CampaNeo to others

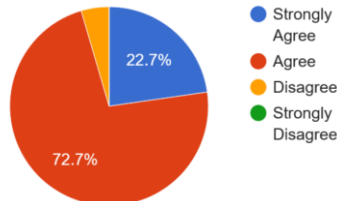


I feel more confident in my knowledge of GDPR



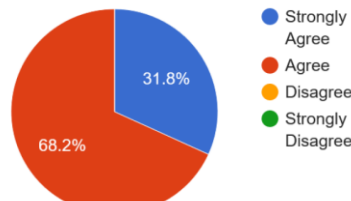
Increased confidence in the users' knowledge of GDPR

I understand what giving consent means.



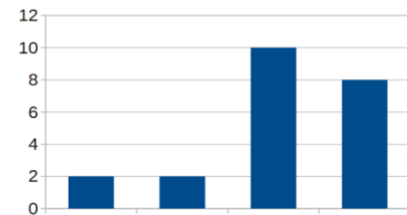
Evaluation results of the consent comprehension

I find the incentives attractive and tempting.



Overall satisfaction with incentives

I feel motivated to participate in a campaign.



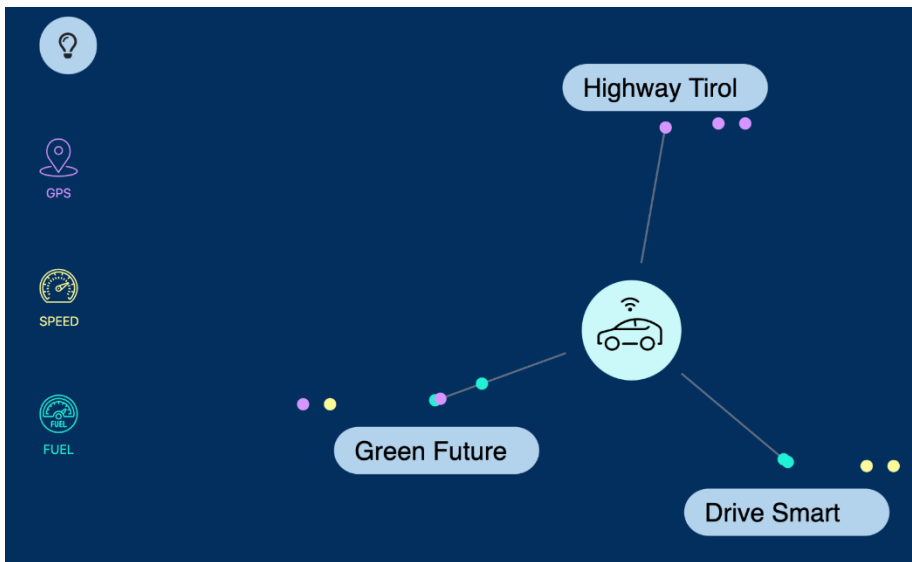
Motivation to participate in a campaign. 1-the most disagreeable, 4-the most agreeable

The CampaNeo User Interface (UI): Post-Consent Visualisation of Data Flows

- Visualisation of what happens after consent is given

[Live Demo](#)

- Provides transparency of data sharing



A screenshot of the CampaNeo UI showing a list of data sharing events. The title 'Green Future' is at the top right. The list has five rows, each with an icon, a timestamp, and a source. The icons are: location pin, speedometer, fuel gauge, location pin, and fuel gauge. The timestamps are: 'GPS (retrieved on 2021-01-11T08:48:00)', 'Speed (retrieved on 2021-01-11T09:01:00)', 'Fuel (retrieved on 2021-01-12T09:02:00)', 'GPS (retrieved on 2021-01-12T08:35:00)', and 'Fuel (retrieved on 2021-01-12T09:05:00)'. The sources are: 'University of Innsbruck', 'Volkswagen', 'Volkswagen', 'University of Innsbruck', and 'Volkswagen'.

Icon	Timestamp	Source
Location pin	GPS (retrieved on 2021-01-11T08:48:00)	University of Innsbruck
Speedometer	Speed (retrieved on 2021-01-11T09:01:00)	Volkswagen
Fuel gauge	Fuel (retrieved on 2021-01-12T09:02:00)	Volkswagen
Location pin	GPS (retrieved on 2021-01-12T08:35:00)	University of Innsbruck
Fuel gauge	Fuel (retrieved on 2021-01-12T09:05:00)	Volkswagen

For more details:

Bless, C., et al. "Raising Awareness of Data Sharing Consent Through Knowledge Graph Visualisation", Studies on the Semantic Web, Volume 53: Further with Knowledge Graphs. IOS Press, 2021. 44-57.

The CampaNeo User Interface (UI): Post-Consent Visualisation of Data Flows

Research Questions:

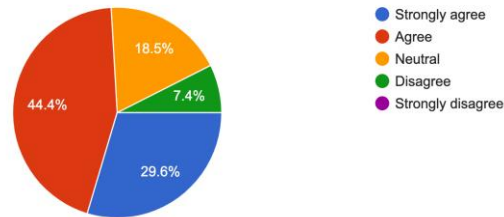
- "Can we raise legal awareness of the data sharing process through transparent visualisation?"
- "Does the tool make the data processor more trustworthy to users?"

Before use test

1. Would you agree to share vehicle data like GPS location, speedometer data or fuel gauge readings with certain companies?

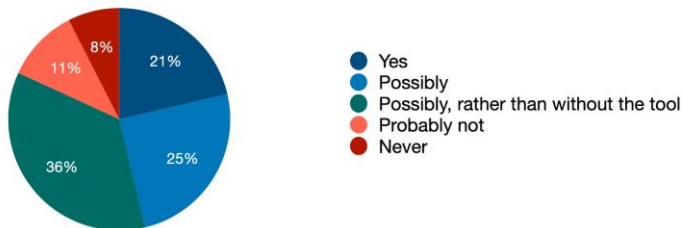


The graph helped me understand what happens to my data.
27 responses

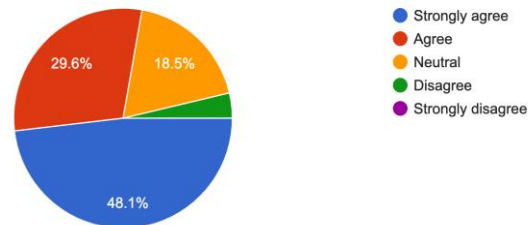


After user test

2. Would you agree to share vehicle data like GPS location, speedometer data or fuel gauge readings, if you had the tool available to control the sharing activities?



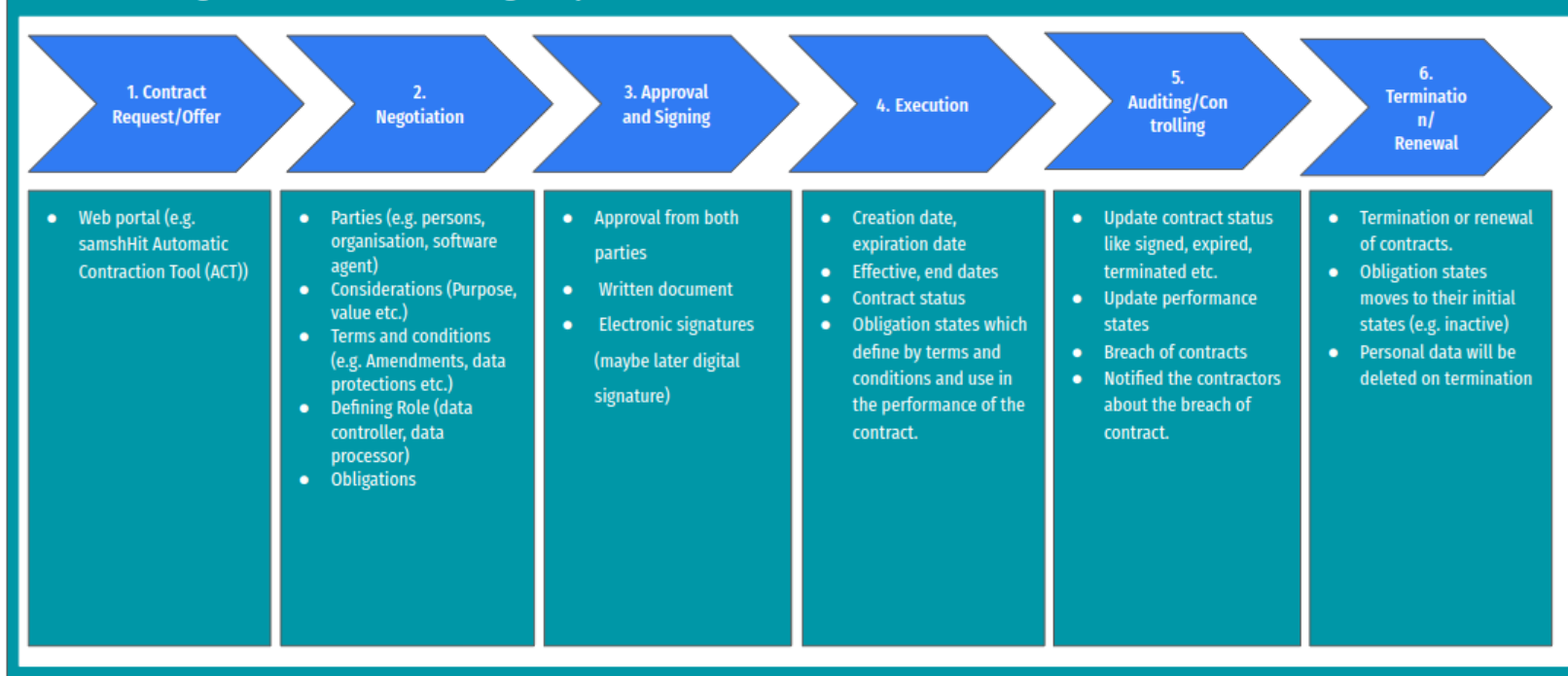
I can easily see what happened to my data at a specific time.
27 responses



Data Sharing Contracts – Automatic Contracting Tool

(<https://smashhit.eu>, Twitter: smashHitP)

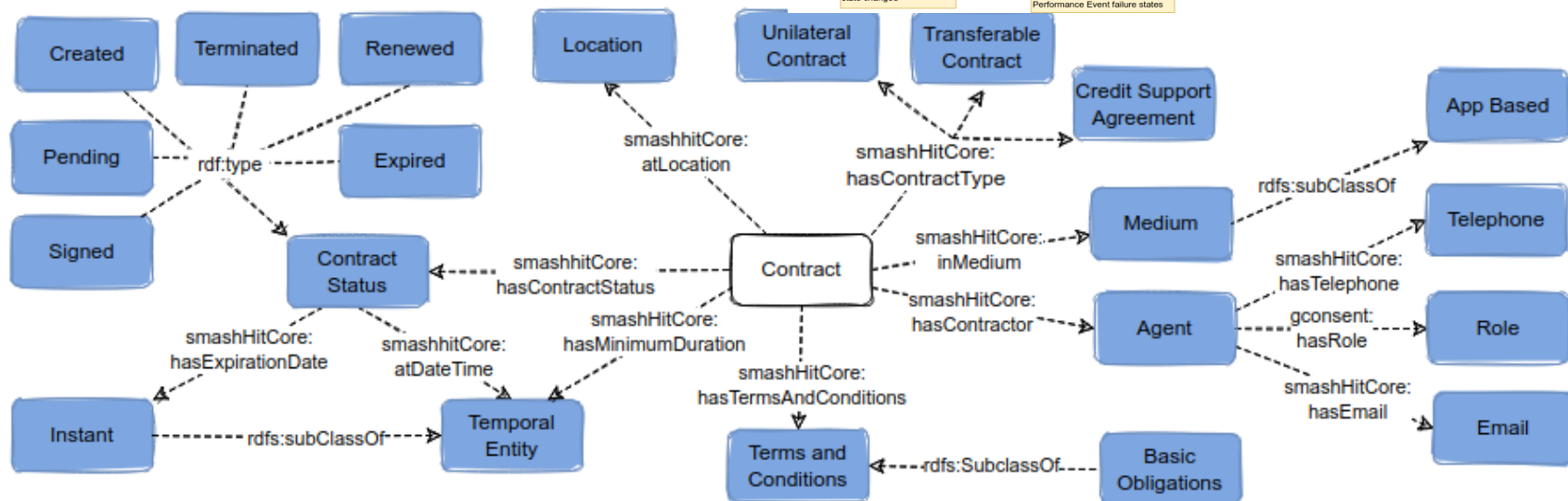
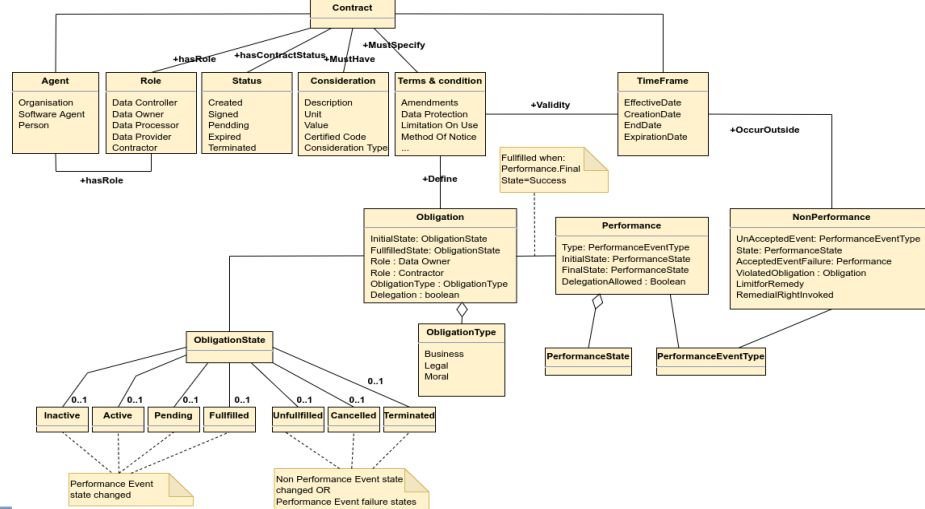
Contract Management (Based on Knowledge Graph)



Contract Semantic Model

(<https://smashhit.eu>;

Twitter: smashHitP)



From our smashHitCore ontology

Contracts Compliance Reasoning

(<https://smashhit.eu>; Twitter: smashHitP)



- Automatic detection of contract conflict/breach (based on obligations):

- Validity of contracts

- Fulfilment of obligations
- Contract expiration, termination and renewal

- Automated rule-based reasoning

- (e.g. IF {(contract status="signed")} THEN {(contract cannot be modified)}.)
- Automatic updates of contractors through notifications

- Compliant with License

- Licensing digital assets (e.g. Digital contracts)
- Support the automated clearance of rights using DALICC¹

New Contract Obligation

Contract	1
Term	Payment Terms
Description	Obligation to pay
Start Date	2022-02-22
End Date	2022-02-25
Party	Tim

Contract ID	Description	Contractor	Term	Status	Action
1	Obligation to pay	Brade	Payment Terms	Pending	
1	Obligation to deliver	Scott	Delivery Terms	Pending	
1	Selling data to third party	Scott	Payment Terms	Pending	
1	Selling data to third party	Brade	Payment Terms	Fulfilled	
2	Obligation to pay	Tim	Payment Terms	Violated	

Outline

- On Semantic Technology
 - Semantic Web
 - Knowledge Graphs
 - FAIR Data
 - Data Access: Licenses, Consent, Contracts
- Linking Healthy Food and Linked Data
 - Food in Knowledge Graphs and FAIR Data
 - Need for Behavioural Change
- Outlook

How healthy are bananas?

(Note linked open data & knowledge graphs!)

bananas calories

Alle Bilder Shopping News Maps Mehr Einstellungen Suchfilter

Ungefähr 26 600 000 Ergebnisse (0,64 Sekunden)

Banane / Energie

89 Kalorien

Typ: Bananen Menge: 100 Gramm

Quellen: USDA Feedback geben



[www.healthline.com > nutrition > Diese Seite übersetzen](#)
How Many Calories and Carbs Are in a Banana? - Healthline
25.12.2016 — A medium-sized **banana** contains 105 **calories**, on average. However, different sizes of **bananas** contain varying amounts of **calories**. Below are ...

[www.healthline.com > nutrition > Diese Seite übersetzen](#)
Are Bananas Fattening or Weight Loss Friendly? - Healthline
10.05.2016 — **Bananas** Are High in Fiber, But Low in **Calories**. **Calorie** for **calorie**, **bananas** contain a lot of fiber. One medium **banana** provides around 12% of ...

[www.calorieking.com > foods > ca... > Diese Seite übersetzen](#)
Calories in Bananas, raw | CalorieKing
There are 89 **calories** in 1 small [5.5 oz with skin, 6" long] (3.5 oz) of **Bananas**, raw. You'd need to walk 25 minutes to burn 89 **calories**. Visit CalorieKing to see ...
Carbs: 22.6 g Protein: 1.1 g
Fat: 0.3 g Fiber: 2.6 g

[www.nutritionix.com > food > ban... > Diese Seite übersetzen](#)
Calories in Banana - Nutritionix
Banana. Nutrition Facts. **Banana**. Serving Size: medium (7" to 7-7/8" long) (118g grams). Amount Per Serving. **Calories** from Fat 3.5. **Calories** 105. % Daily Value ...

[www.verveelfit.com > are-banana... > Diese Seite übersetzen](#)



Banane
Obst

Aus dem Englischen übersetzt - Eine Banane ist eine längliche, essbare Frucht - botanisch eine Beere -, die von verschiedenen Arten großer krautiger Blütenpflanzen der Gattung Musa produziert wird. In einigen Ländern können zum Kochen verwendete Bananen als "Kochbananen" bezeichnet werden, was sie von Dessertbananen unterscheidet. [Wikipedia \(Englisch\)](#)

[Ursprüngliche Beschreibung aufrufen >](#)

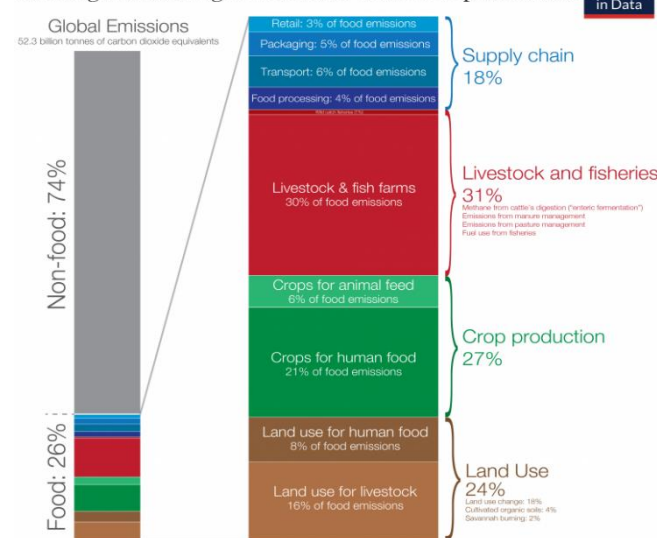
Nährwertangaben
Bananen
Quellen: USDA

Menge pro 100 Gramm
Kalorien 89
Fettgehalt 0,3 g
Gesättigte Fettsäuren 0,1 g
Mehrfach ungesättigte Fettsäuren 0,1 g
Einfach ungesättigte Fettsäuren 0 g
Cholesterin 0 mg
Natrium 1 mg
Kalium 358 mg
Kohlenhydrate 23 g

Making discoveries by following links, or how to create near-zero CO2 emission diets

- Food production accounts for ca. 26 % of greenhouse gas emissions, food waste for ca. 9% (2021 Food Waste Index report, UN)
- We want to make **diets more sustainable and adapting to climate change**
- Food data is mostly not transparent and not actionable
- We will bring in explicit semantics to the data, interlink different diverse disperse data sources and reason basing on them, enabling making of better decisions
- Contributions to climate change, food data infrastructures, artificial intelligence and data science
- Collaborative project in WUR climate change investment theme: WU (WDCC, CHL,..), WR (WBFR,...)
- We look for: relevant data providers, food, sustainability, production, logistics expertise,...

Global greenhouse gas emissions from food production 



Data source: Joseph Poore & Thomas Nemecek (2018). Reducing food's environmental impacts through producers and consumers. Published in Science, OurWorldinData.org - Research and data to make progress against the world's largest problems. Licensed under CC-BY by the author Hannah Ritchie



An apple

0 CO₂e plucked from the garden
32g CO₂e local and seasonal
80g CO₂e shipped, seasonal
290g CO₂e shipped, out-of-season, frozen

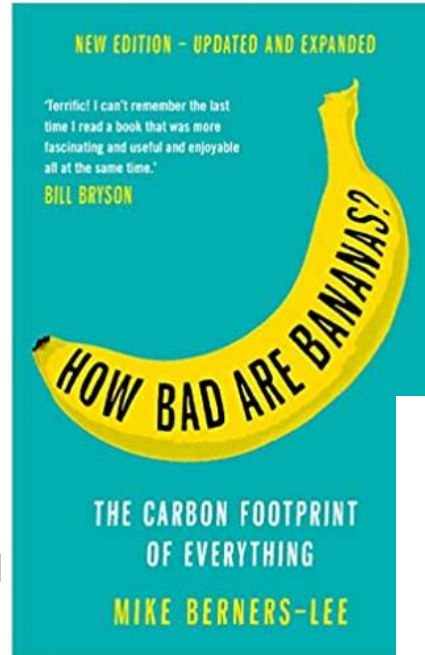
Apples are low-carbon food wherever they come from but local and seasonal is best²

The statement above seems obvious and incontrovertible. But to give you an idea of the complexity and uncertainty of calculating carbon footprints, consider this. A New Zealand university study claimed to have found that their

Application example: sustainable and healthy food... or “what to eat”



Mike Berners-Lee
- professor and fellow
of the Institute for Social
Futures at Lancaster
University, UK
- brother of Tim-Berners
Lee (inventor of the Web)



*Food production accounts for ca. 26 %
of greenhouse gas emissions, food
waste for ca. 9% (2021 Food Waste Index report,
UN)*



An apple

0 CO₂e plucked from the garden
32g CO₂e local and seasonal
80g CO₂e shipped, seasonal
290g CO₂e shipped, out-of-season, frozen

**Apples are low-carbon food wherever they
come from but local and seasonal is best²**

The statement above seems obvious and incontrovertible. But to give you an idea of the complexity and uncertainty of calculating carbon footprints, consider this. A New Zealand university study claimed to have found that their country's apples exported to the UK market had a footprint of just 185g CO₂e per kilo – significantly lower than UK apples for local consumption, which came in at 271g per kilo.³ The argument was that UK production entailed greater use of fossil fuels on the farm and required more cold storage. The study also pointed to the

Developments with relevant, data, semantics and research data infrastructures are many



recipe ingredients oliebolle

Alte Bilder Shopping News Videos Mehr Einstellungen Suchfilter

Ungefähr 236 000 Ergebnisse (0,69 Sekunden)

www.thespruceeats.com · Dutch Food · Dutch Desserts

Old-Fashioned Dutch Doughnuts (Oliebollen) Recipe



18.09.2020 — Ingredients · Steps to Make It · Prepare the Yeast Dough · Fry and Serve the Oliebollen.

★★★★★ Bewertung: 4 · 119 Ergebnisse · 50 Min. · Kalorien: 133

Rezepte



Echte holländische oliebolle Teigspezialität...
Kochbar.de
5,0 ★★★★★ (1f)



Oliebollen (Holländisches...)
Pinterest
Kleine Rezensionen



Old-Fashioned Dutch Doughnuts (Oliebollen)
The Spruce Eats
4,0 ★★★★★ (119)
50 Min.

Mehr anzeigen

www.allrecipes.com · ... · Bread · Yeast Bread Recipes

Oliebollen (Dutch Doughnuts) Recipe | Allrecipes



Ingredients · 1 (0.6 ounce) cake compressed fresh yeast · 1 cup lukewarm milk · 2 ¼ cups all-purpose flour · 2 teaspoons salt · 1 egg · ¾ cup dried currants · ¾ cup ...

★★★★★ Bewertung: 4,3 · 50 Ergebnisse · 2 Std. 8 Min. · Kalorien: 269,8

honestcooking.com · oliebolle-tr... · Diese Seite übersetzen

Oliebollen -Dutch Doughnuts Recipe - Honest Cooking

https://www.health.gov.au/initiatives-and-programs/research-data-infrastructure-initia



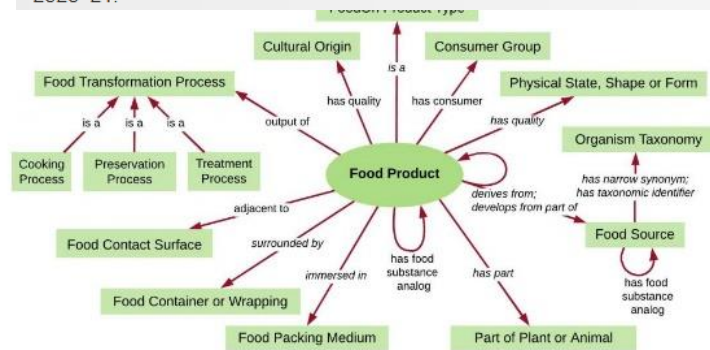
Australian Government
Department of Health

Home Health topics Initiatives and programs Resources

Home > Initiatives and programs

Research Data Infrastructure initiative

The Research Data Infrastructure initiative will fund the creation or extension of national research data infrastructure with a focus on data registries, biobanks and data linkage platforms to support Australian medical research. It will provide \$80 million over 8 years, starting in 2020–21.



- FoodOn: <https://foodon.org> – integrated with life science ontologies, such as the Gene Ontology
- Food Taxonomy (WUR): <http://www.foodvoc.org>, AgroVoc, and many other ontologies
- Research data management initiatives in the EU and world-wide (CERN, Open Data Pilot)



WAGENINGEN
UNIVERSITY & RESEARCH


Research Data Infrastructure - Example


- Comfocus project: <https://www.youtube.com/watch?v=tLOy0zE1E4Y>

https://twitter.com/ComfocusP?ref_src=twsrc^tfw|twcamp^embeddedtimeline|twe

← **COMFOCUS project** 7 Tweets **Follow**



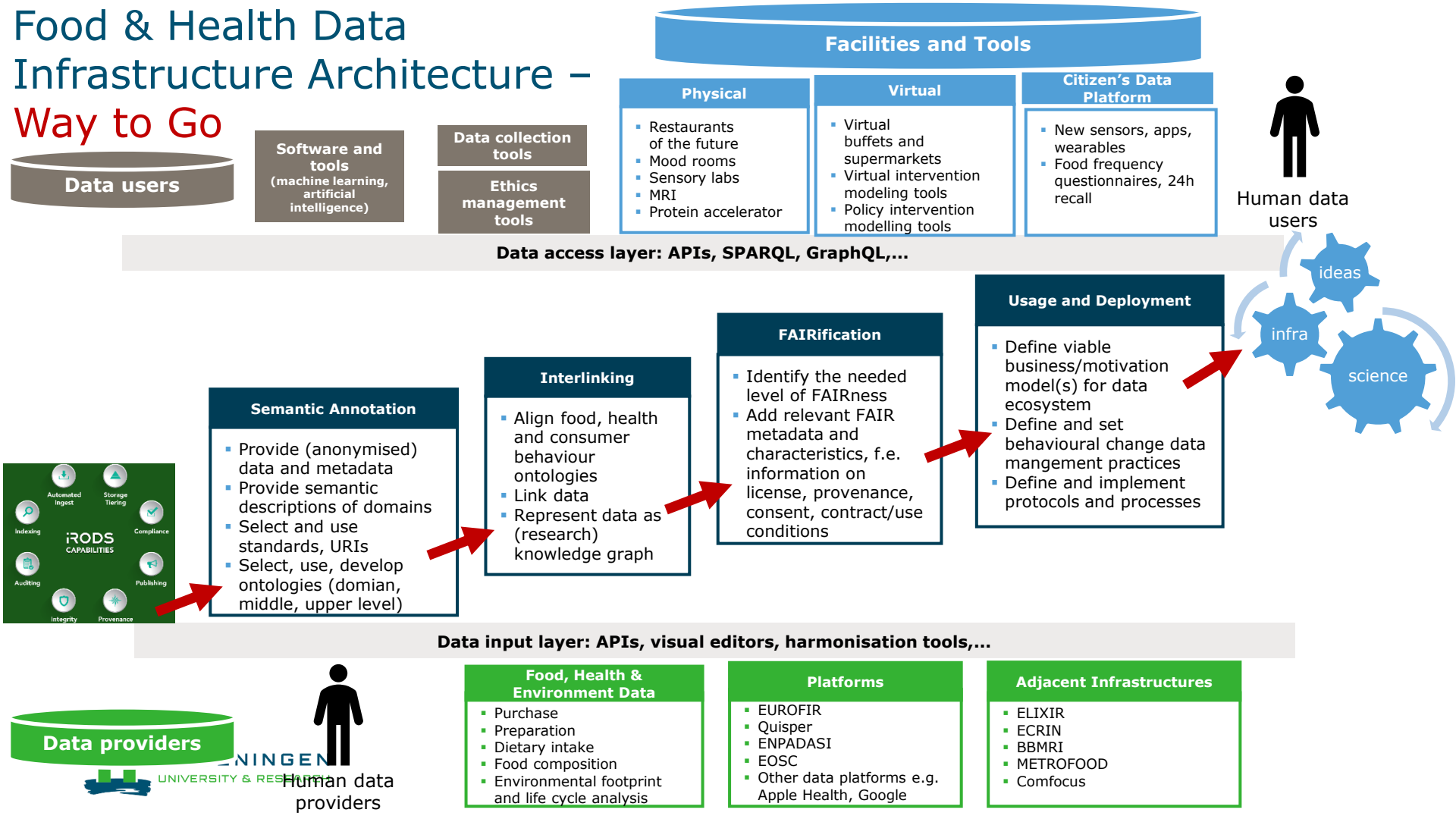
COMFOCUS project @ComfocusP · Feb 24
presentation video -  COMFOCUS community on food consumer science youtu.be/tLOy0zE1E4Y via @YouTube

 youtube.com
COMFOCUS - Community on Food Consumer Scie...
The COMFOCUS project, funded from the European Union's Horizon 2020 research and innovation ...

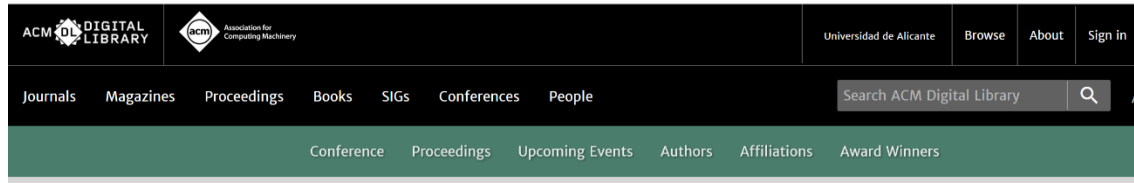
How do we derive knowledge out of data?



Food & Health Data Infrastructure Architecture – Way to Go



Example: Easier Access to the Data Translating from machine to machine



Home > Conferences > SAC > Proceedings > SAC '22 > *GraphSPARQL: a GraphQL interface for linked data*

RESEARCH-ARTICLE



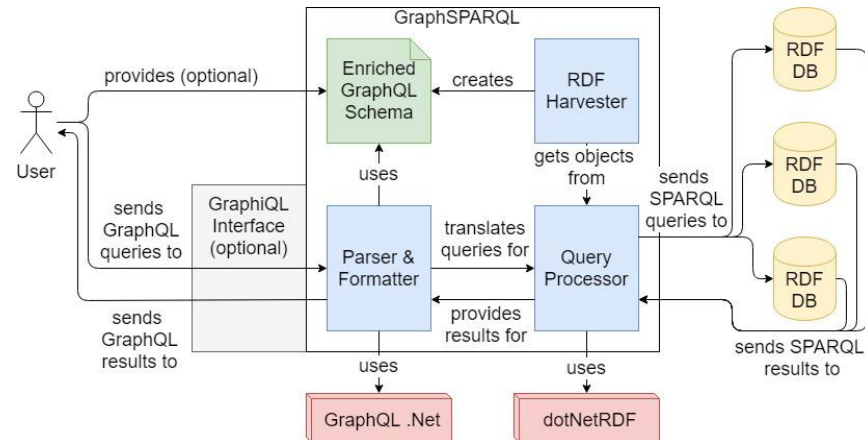
GraphSPARQL: a GraphQL interface for linked data

Authors: Kevin Angele, Manuel Meitinger, Marc Bußjäger, Stephan Föhl, Anna Fensel [Authors Info & Claims](#)

SAC '22: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing • April 2022 • Pages 778–785 • <https://doi.org/10.1145/3477314.3507655>

Online: 06 May 2022 [Publication History](#)

0 10



Societal dimension – Roadmapping activities

Construction of Big Data research roadmaps at EU and national level
(projects OntoCommons, BYTE, BIG, KnowledgeWeb)

Technology in Society 54 (2018) 74–86



www.ontocommons.eu



Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc



-> adoption of ontology technologies

The societal impact of big data: A research roadmap for Europe

Martí Cuquet*, Anna Fensel

Semantic Technology Institute, Department of Informatics, University of Innsbruck, Technikerstraße 21a, 6020 Innsbruck, Austria¹



ARTICLE INFO

Keywords:

Big data
Research roadmap
Societal externalities
Skills development
Standardisation

ABSTRACT

With its rapid growth and increasing adoption, big data is producing a substantial impact in society. Its usage is opening both opportunities such as new business models and economic gains and risks such as privacy violations and discrimination. Europe is in need of a comprehensive strategy to optimise the use of data for a societal benefit and increase the innovation and competitiveness of its productive activities. In this paper, we contribute to the definition of this strategy with a research roadmap to capture the economic, social and ethical, legal and political benefits associated with the use of big data in Europe. The present roadmap considers the positive and

Outline

- On Semantic Technology
 - Semantic Web
 - Knowledge Graphs
 - FAIR Data
 - Data Access: Licenses, Consent, Contracts
- Linking Healthy Food and Linked Data
 - Food in Knowledge Graphs and FAIR Data
 - Need for Behavioural Change
- Outlook

Ongoing AI and Data Science research

- 1. *Data Sharing: Intelligent Data Value Chain Production and Consumption Ecosystems.*** Enabling automated exchange of the information across systems (handling of semantic interoperability, consent, contracts/monetization, privacy). For:
 1. research data, platforms, infrastructures,
 2. industrial data for the applications creation, such as in the domains of food, health, consumer understanding, behavior intervention...
- 2. *Combining Symbolic and Sub-Symbolic AI: Hybrid Methods based on Knowledge Graphs, Data Analytics and Machine Learning.*** Making the sensor, image and stream data understandable and usable - f.e. processing and understanding of image data of food, or of large numbers of events e.g. movements at a supermarket, or events on social media, survey results, also applying meta machine learning tools like Autokeras, h2o.ai, decision making systems combining symbolic and sub-symbolic reasoning.
- 3. *Responsible AI: Visualization of Data, Behavioral Change and Explainable AI.*** Data visualization and communication techniques based on semantic technology, specifically in applications allowing end users and the researchers to understand the data and system/algorithm complexity, make data more FAIR, make better decisions - e.g. which food to consume to be more sustainable and healthy, also taking into account the data of the researcher or end user, such as the data about previous nutrition, medical conditions.

Collaborations and broader picture

■ **Data sharing**

- Domains such food, health, biotechnology, production, sustainability, any research infrastructures in need of FAIR data
- Open science, research knowledge graphs e.g. WUR open research knowledge graph (WORKG)

■ **Combining Symbolic and Sub-Symbolic AI:** computer vision - robotics, farms, environment...

■ **Responsible AI:** Behavioural change, interventions, consumption, marketing research, ethics...

Digital Twin (of a Living System)

- **Human Sensing - Decision Making Processes and Communication** Basing f.e. on health image analysis and measurements, Brain Computer Interfaces (BCIs), sensors for physical characteristics.
- An infrastructure for semantic transfer of thoughts, body sensory input, health measurements, thinking processes and communication directly to the Internet would resolve these challenges.

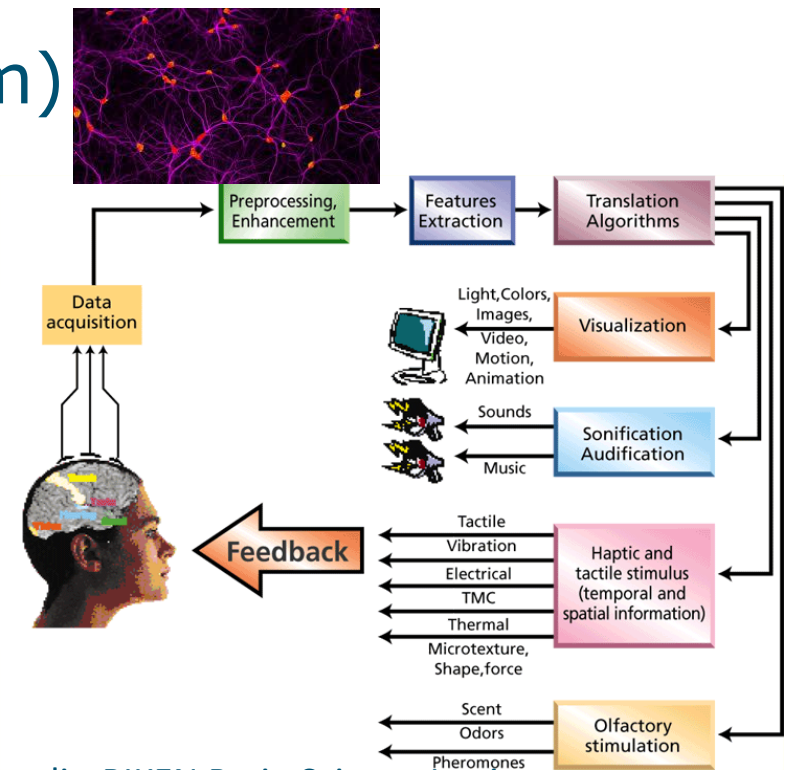


Image credit: RIKEN Brain Science Institute

How do we derive knowledge out of data?



Making data meaningful with semantics

<https://de.wikipedia.org/wiki/Sushi>



Thank you for attention!

