Data sharing with knowledge graphs for sustainable and healthy diets

PLDN, Hilversum

27.09.2022, Anna Fensel

WAGENINGEN
UNIVERSITY & RESEARCH

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

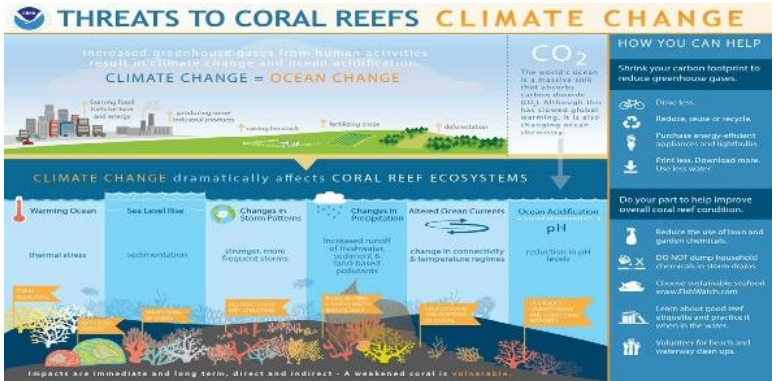
Global Challenges

- 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**



WAGENINGEN
UNIVERSITY & RESEARCH

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

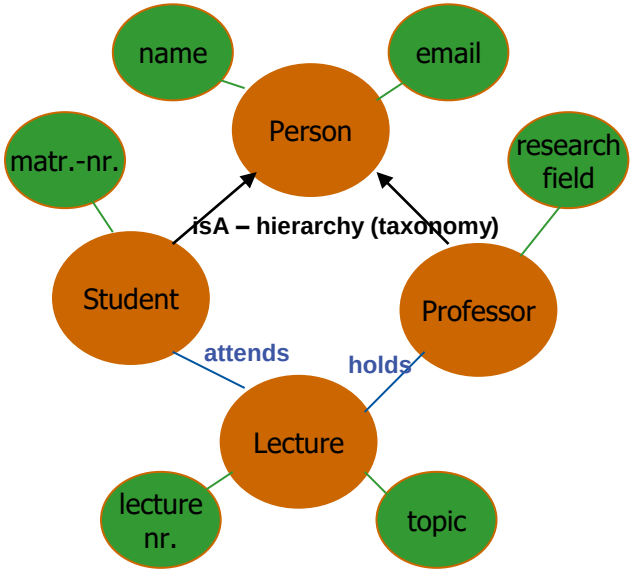
attribute describing a concept

Relation

relationship between concepts or properties

Axiom

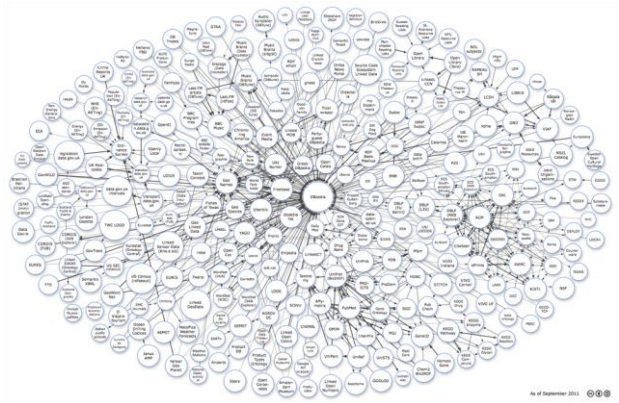
coherency description between Concepts / Properties / Relations via logical expressions

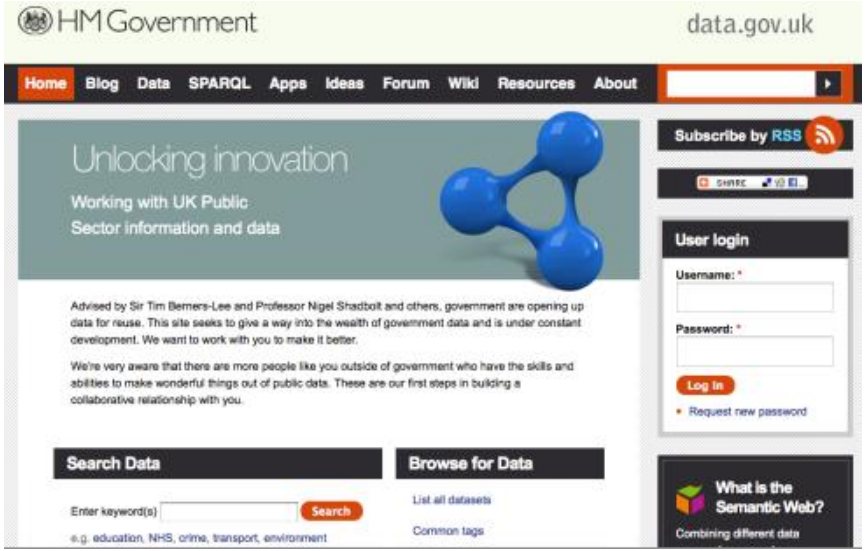


holds(Professor, Lecture) =>
Lecture.topic = Professor.researchField

Practical cases:
Linked Data, Open government data

DBpedia





WAGENINGEN
UNIVERSITY & RESEARCH

Data Quality: 5-star Linked OPEN Data

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

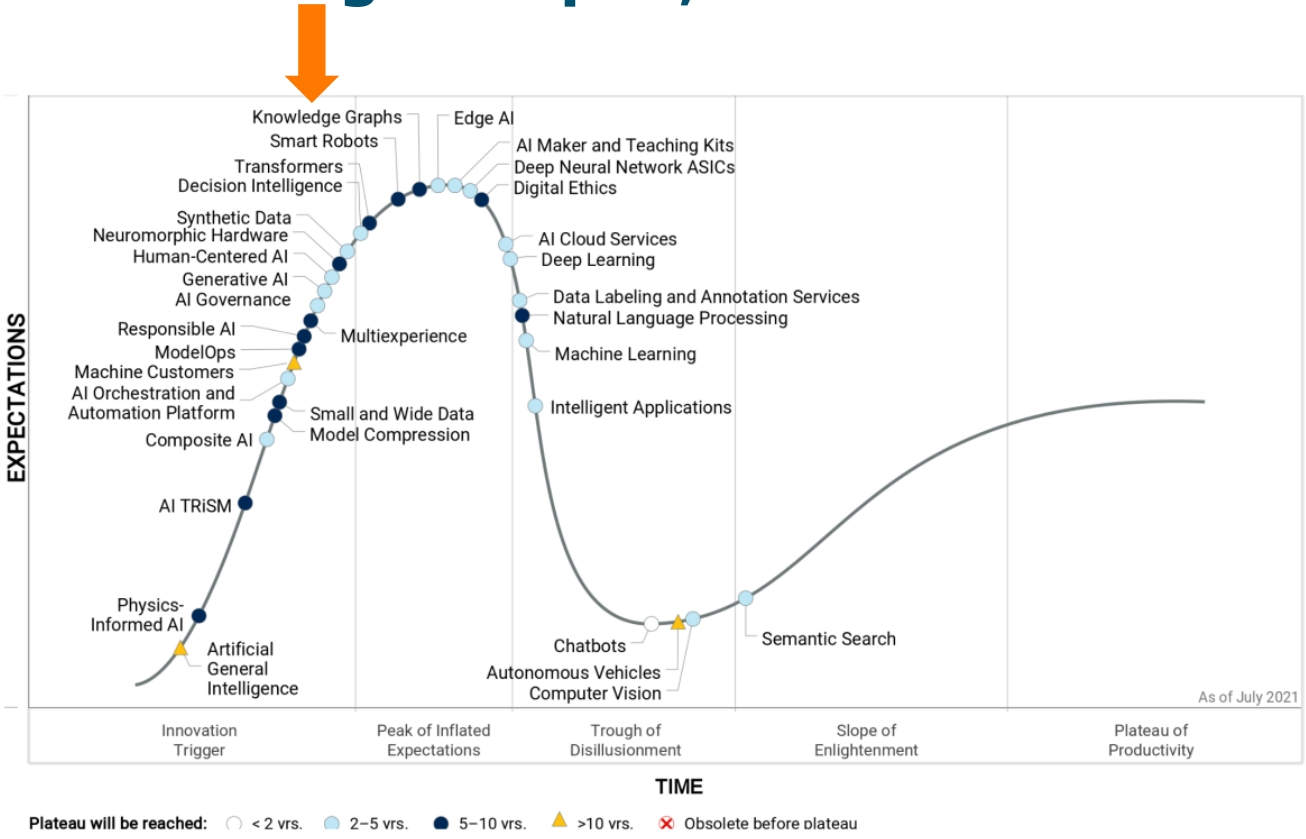
★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ 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



WAGENINGEN
UNIVERSITY & RESEARCH

Gartner

The Semantic Web as an integrated graph of knowledge graphs

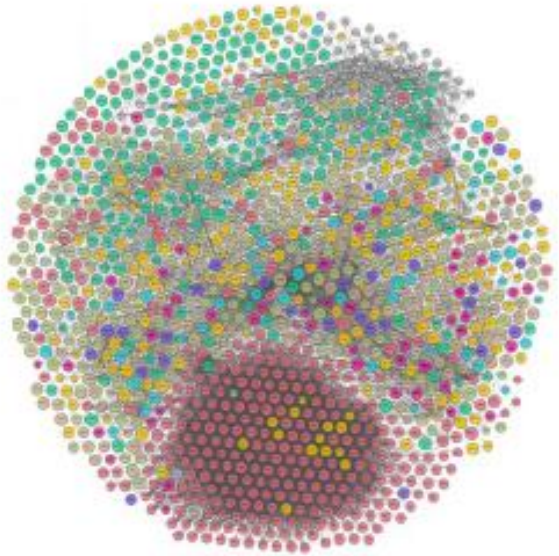
Knowledge Graphs

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

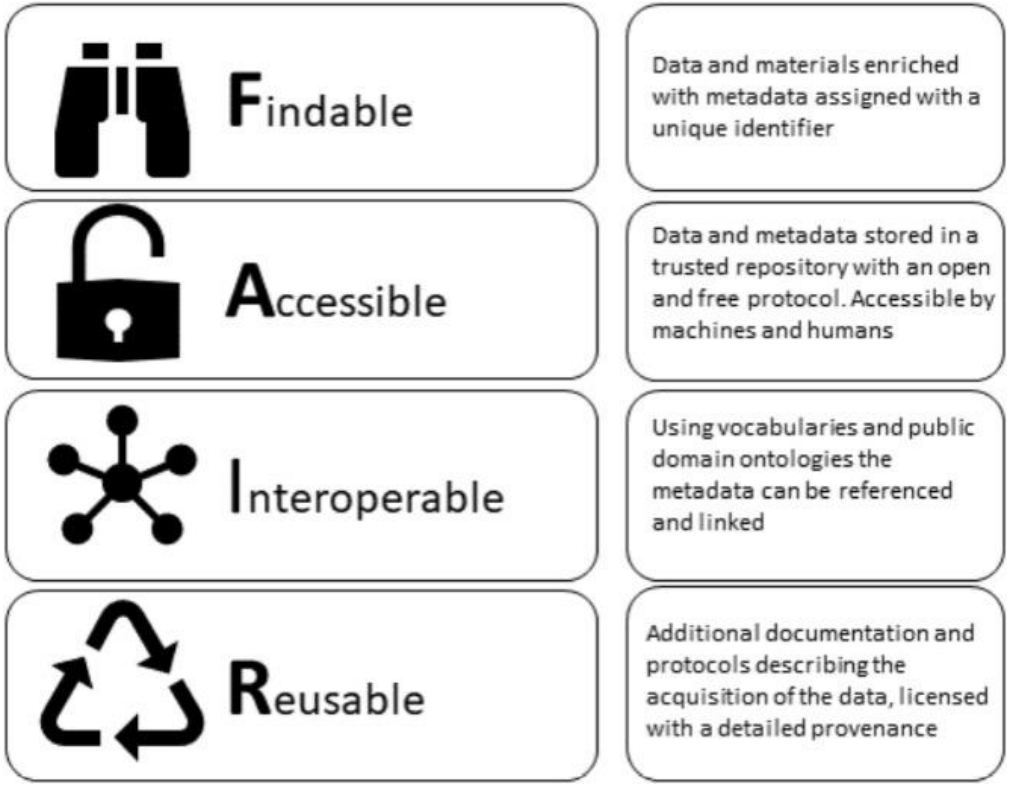
- **Manual Construction** - curated, collaborative
- **Automated Construction** - semi-structured, unstructured

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



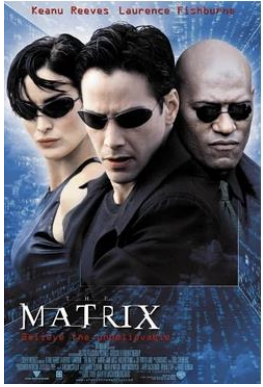
Essentially,
FAIR data is about semantics and knowledge graphs

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.

WAGENINGEN
UNIVERSITY & RESEARCH

See also: https://www.go-fair.org/fair-principles/

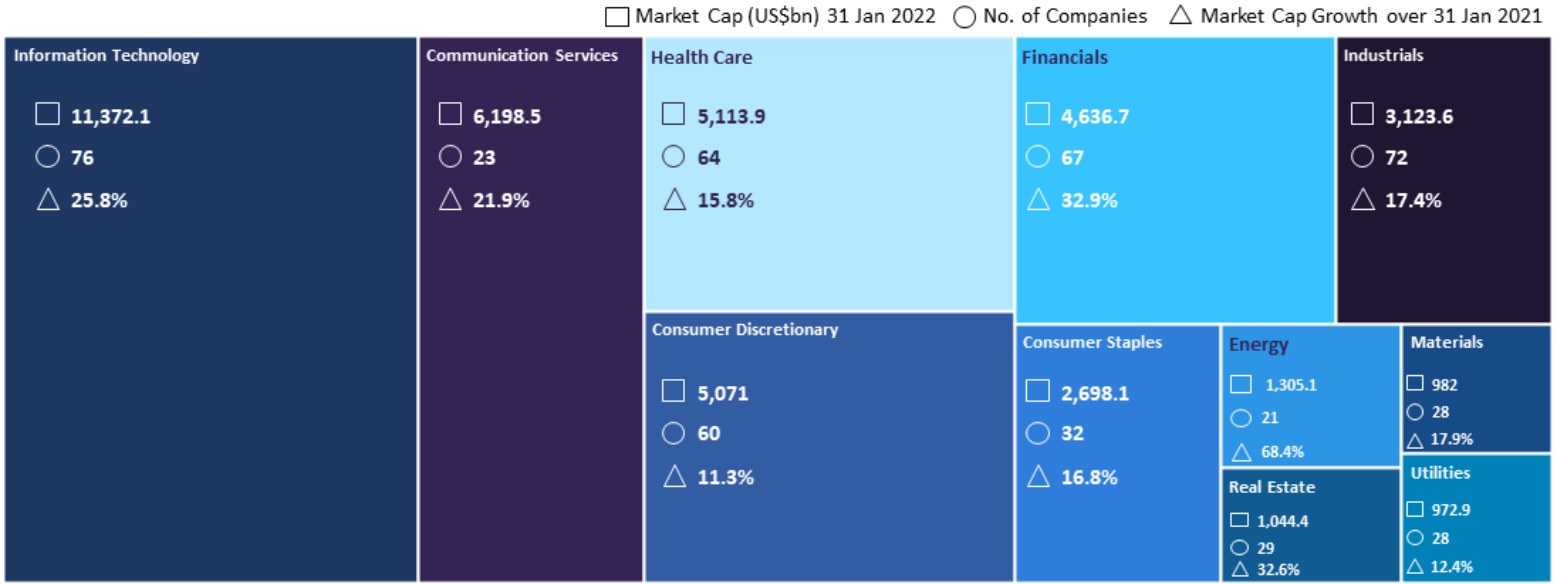
Data! Lots of it!



starting 1999

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

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



Source: GlobalData Intelligence Center and Stock Exchanges

GlobalData.

2022 - Europe is in trouble

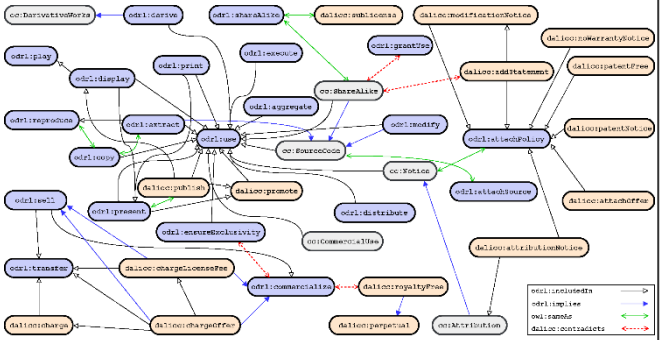
Not all data and knowledge are to be open

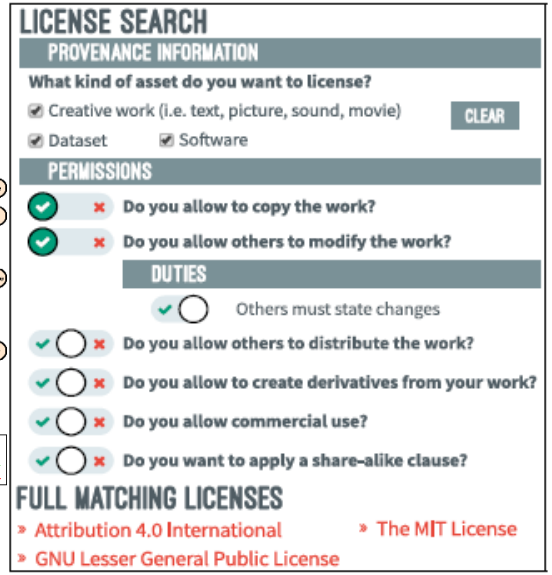
Towards de-centralization

- In communication

- In payments

Licenses





Pellegrini, T., Mireles, V., Steyskal, S., Panasiuk, O., Fensel, A., & Kirrane, S. (2018). Automated rights clearance using semantic web technologies: The DALICC framework. In *Semantic Applications* (pp. 203-218). Springer Vieweg, Berlin, Heidelberg.

https://www.dalicc.net

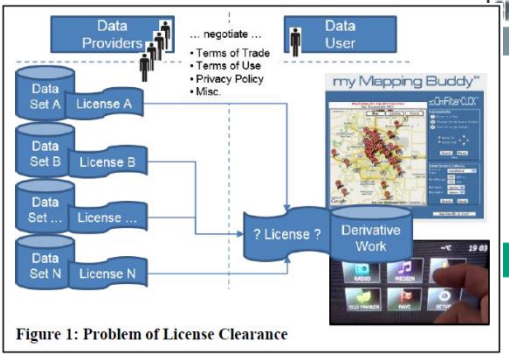
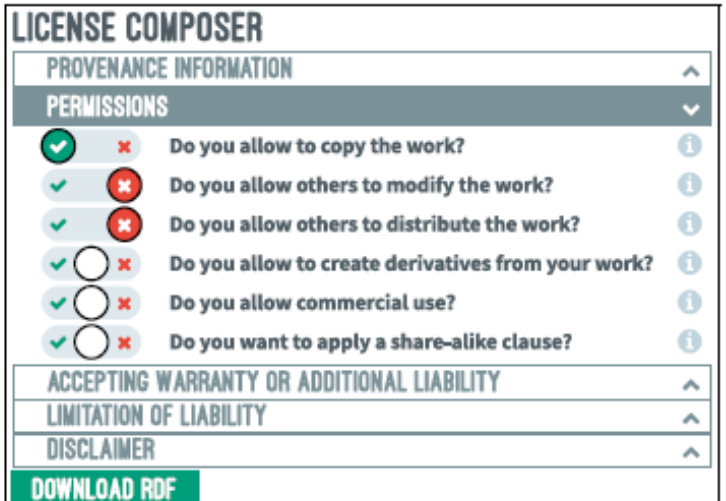


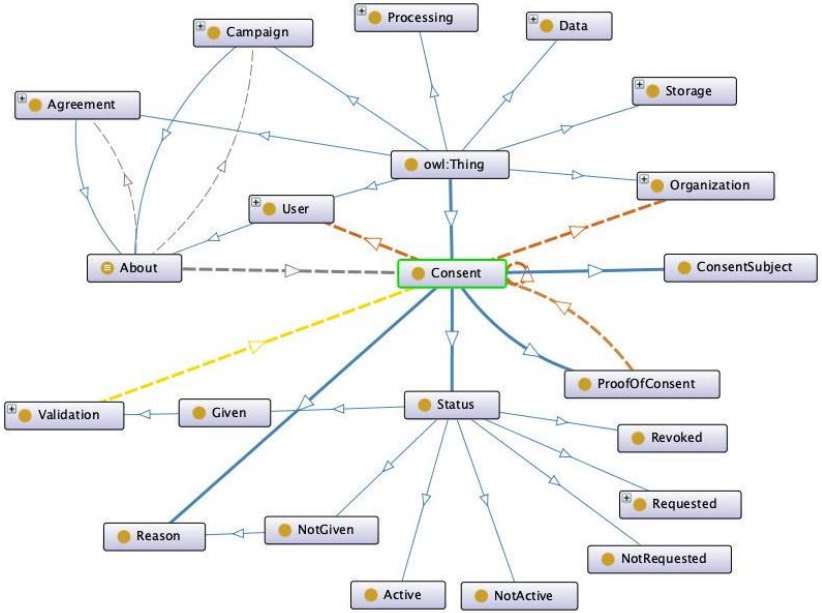
Figure 1: Problem of License Clearance



UNIVERSITY & RESEARCH

Consent

- Based on the CampaNeo project ontology, which models **consent, campaigns, contracts, data, data processing**
- Reuses the GConsent[1], FIBO[2] and DPV[3] ontologies
- Provides common understanding between all entities in the CampaNeo project
- Supports the interoperability, traceability and transparency of data processing



CAMPA NEO
consent. telematics. campaigns!

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

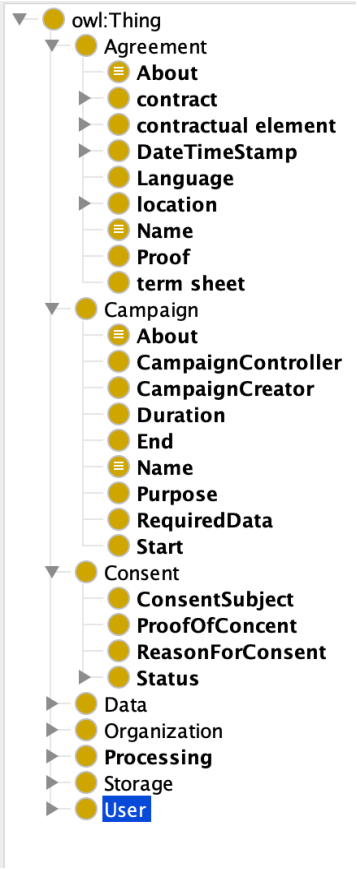
Class axioms

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



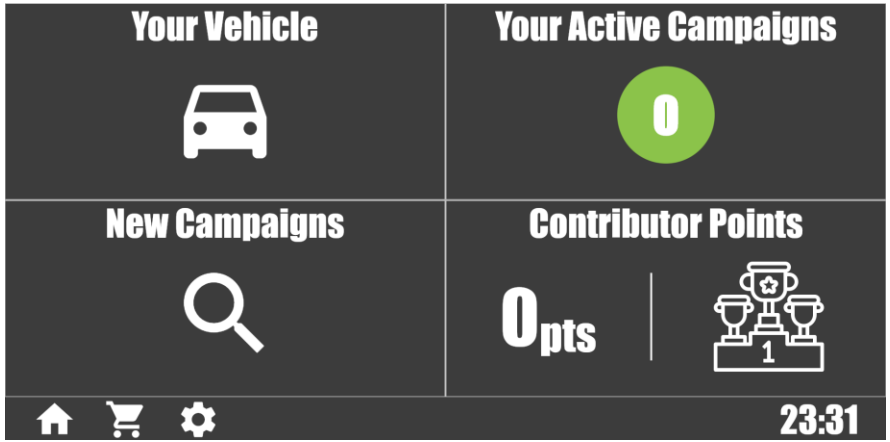
1 http://openscience.adaptcentre.ie/ontologies/GConsent/docs/ontology

2 https://spec.edmcouncil.org/fibo/ontology

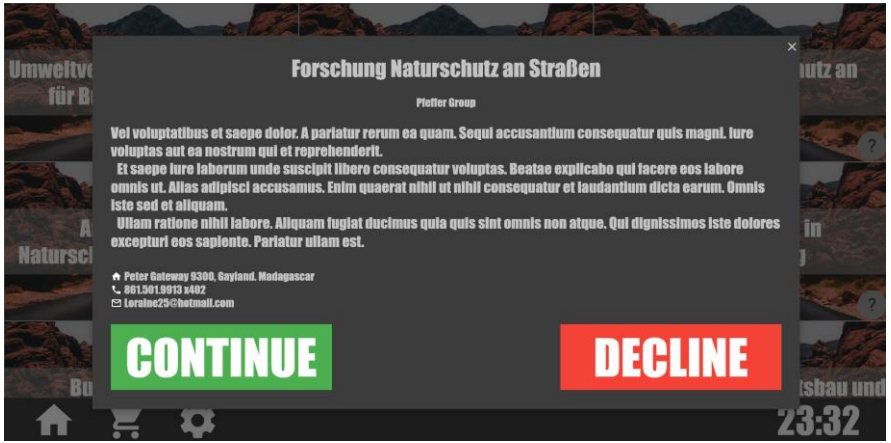
3 https://w3c.github.io/dpv/

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

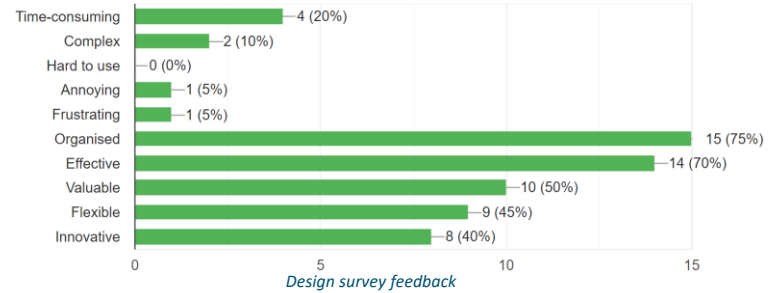
Live Demo

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

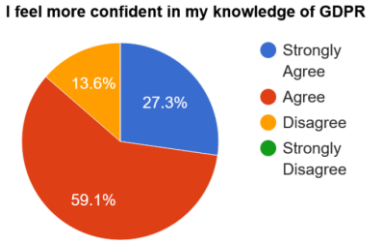
VOLKSWAGEN AKTIENGESELLSCHAFT

Research Questions:

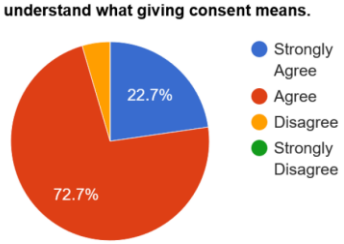
- "*Does a graphical visualisation of an informed consent request via UI help raise one's awareness?*"
  - Yes! "*According to GDPR, I have the right to...*" - 82.55% average correctness
- "*Do incentives affect one's willingness to consent?*"
  - Yes! 81.9% are motivated to engage in campaigns
  - Yes!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



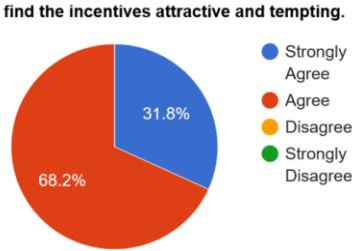
Design survey feedback



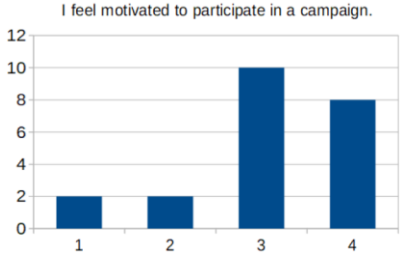
Increased confidence in the users' knowledge of GDPR



Evaluation results of the consent comprehension



Overall satisfaction with incentives



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

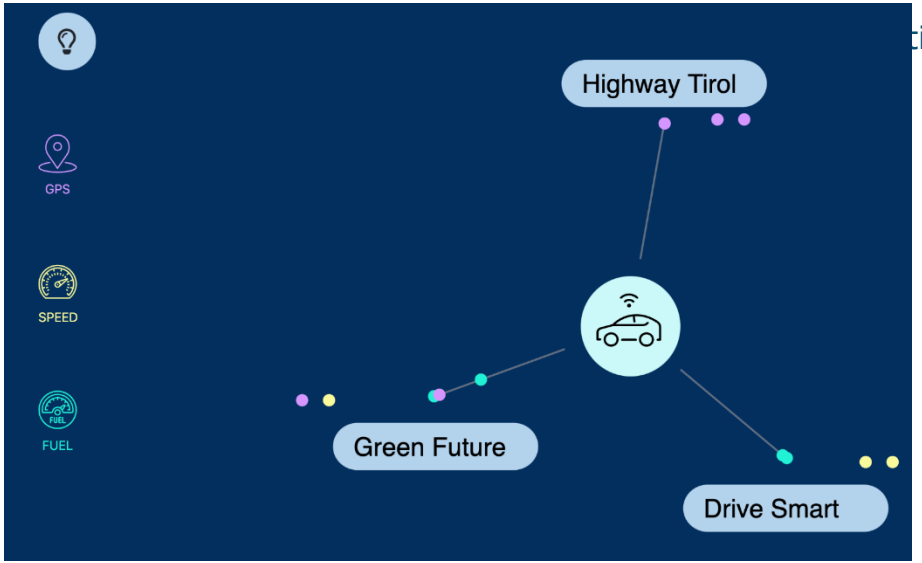
WAGENINGEN UNIVERSITY & RESEARCH

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



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:

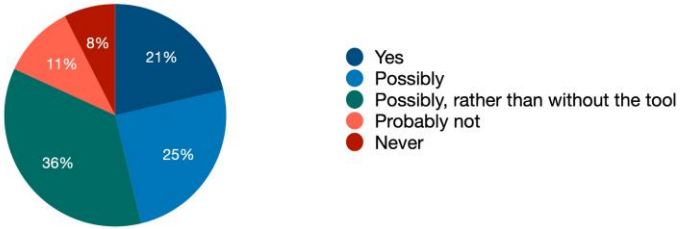
- "*Can we raise legal awareness of the data sharing process through transparent visualisation?*"
- "*Does the tool make the data processor more trustworthy to users?*"

Before use test

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

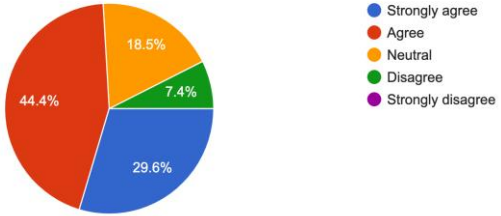
After user test

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?



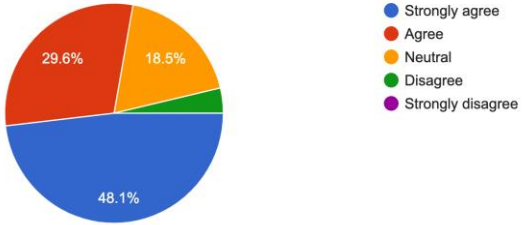
The graph helped me understand what happens to my data.

27 responses



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

27 responses

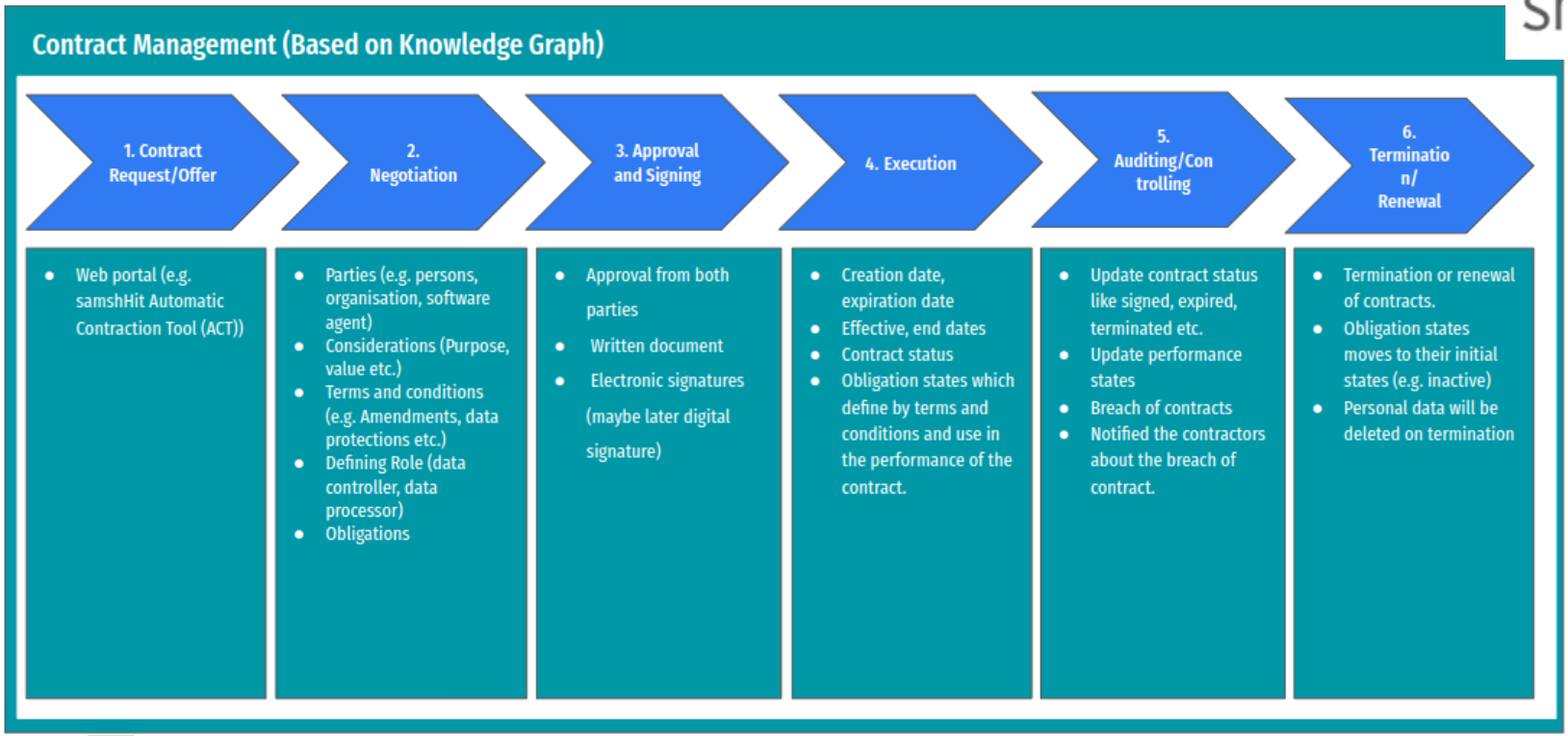


UNIVERSITY & RESEARCH

Data Sharing Contracts – Automatic Contracting Tool

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

smashHit



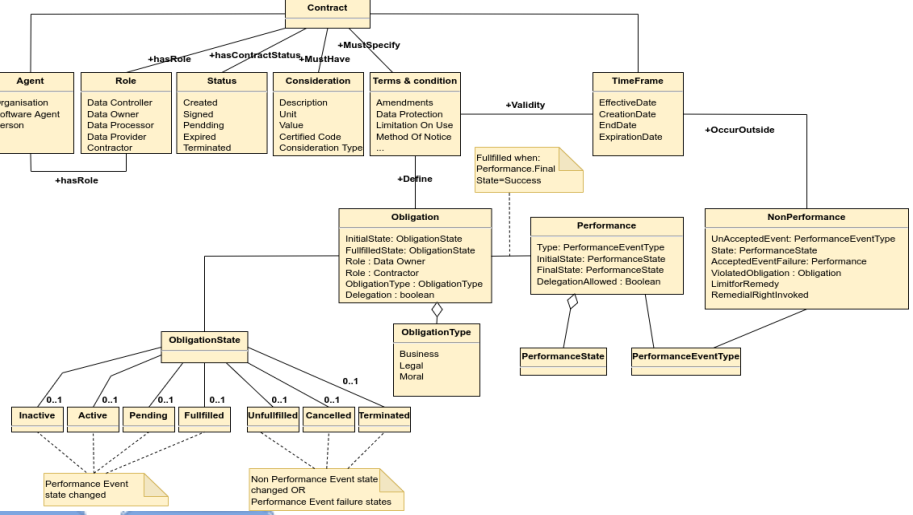
WAGENINGEN UNIVERSITY & RESEARCH

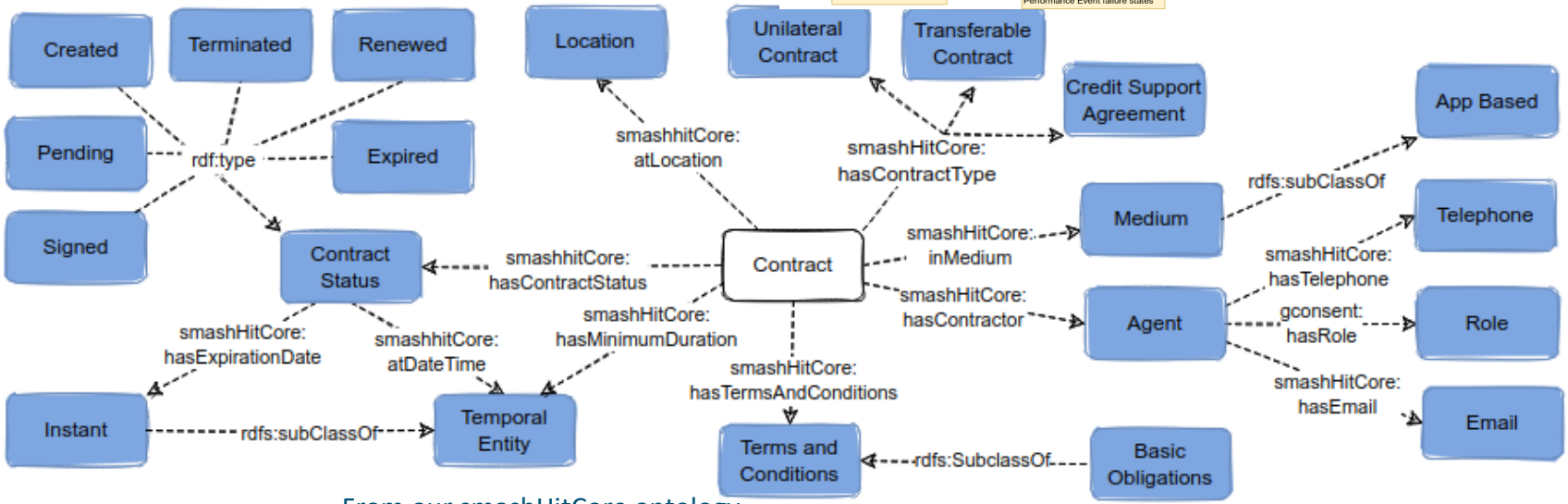
Tauqeer, A.; Kurteva, A.; Chhetri, T.R.; Ahmeti, A.; Fensel, A. Automated GDPR Contract Compliance Verification Using Knowledge Graphs. *Information* **2022**, *13*, 447. https://doi.org/10.3390/info13100447

Contract Semantic Model

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

smashHit





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 DALICC[1]

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	

WAGENINGEN UNIVERSITY & RESEARCH

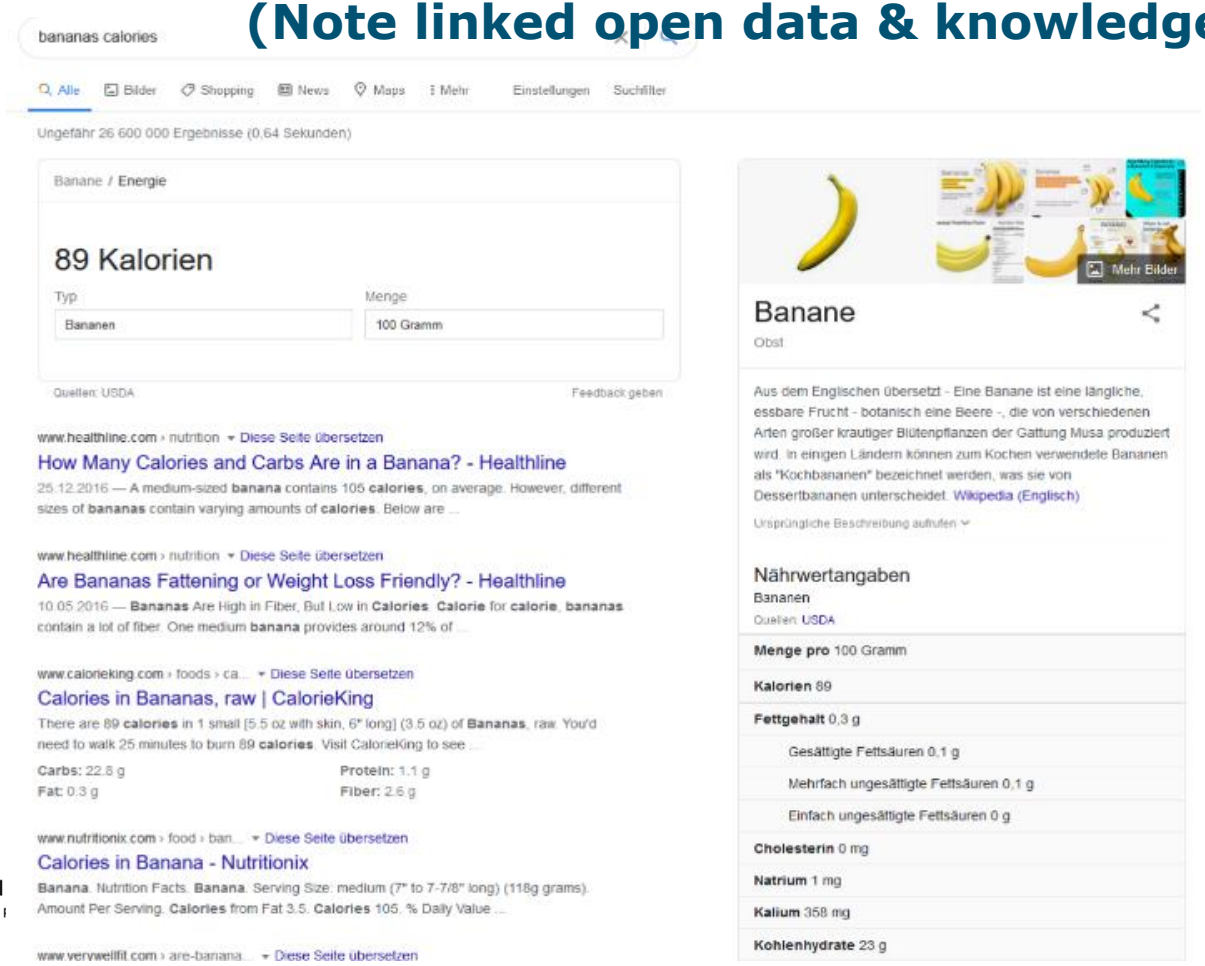
[1] https://www.dalicc.net/

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?

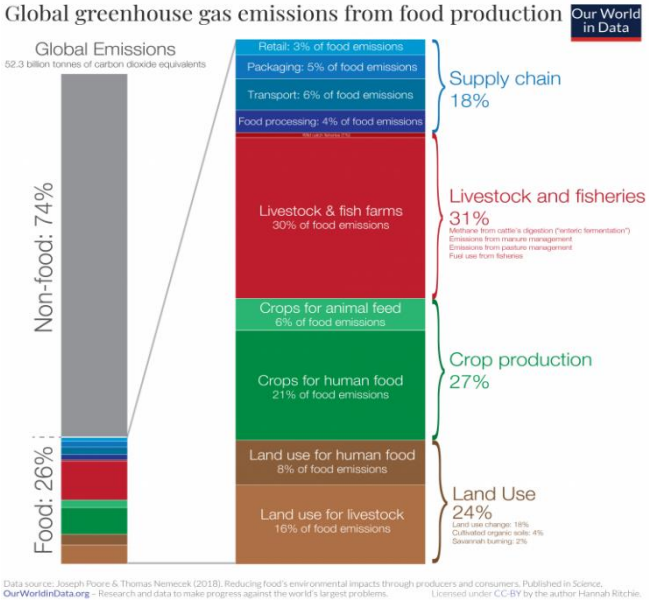
(Note linked open data & knowledge graphs!)



WAGENI
UNIVERSITY & I

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,..



An apple

0 CO_2e plucked from the garden
32g CO_2e local and seasonal
80g CO_2e shipped, seasonal
290g CO_2e shipped, out-of-season, frozen

Apples are low-carbon food wherever they come from but local and seasonal is best[2]

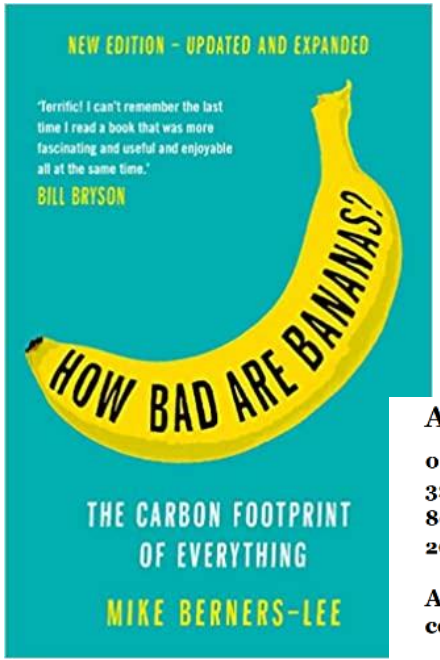
The statement above seems obvious and incontrovertible. But to give you an idea of the complexity and uncertainty of calculating carbon footprints, consider this. A New

WAGENINGEN
UNIVERSITY & RESEARCH

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)*

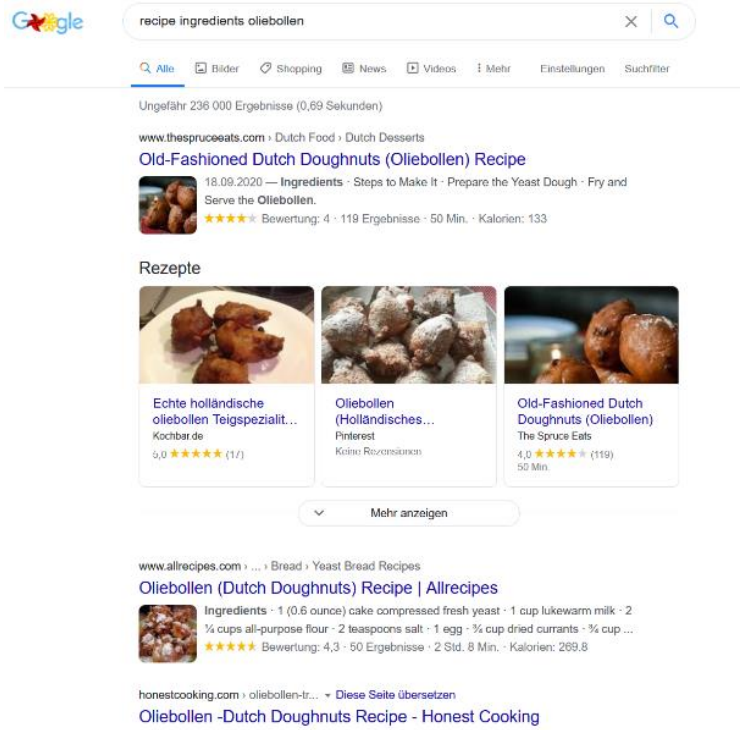
An apple

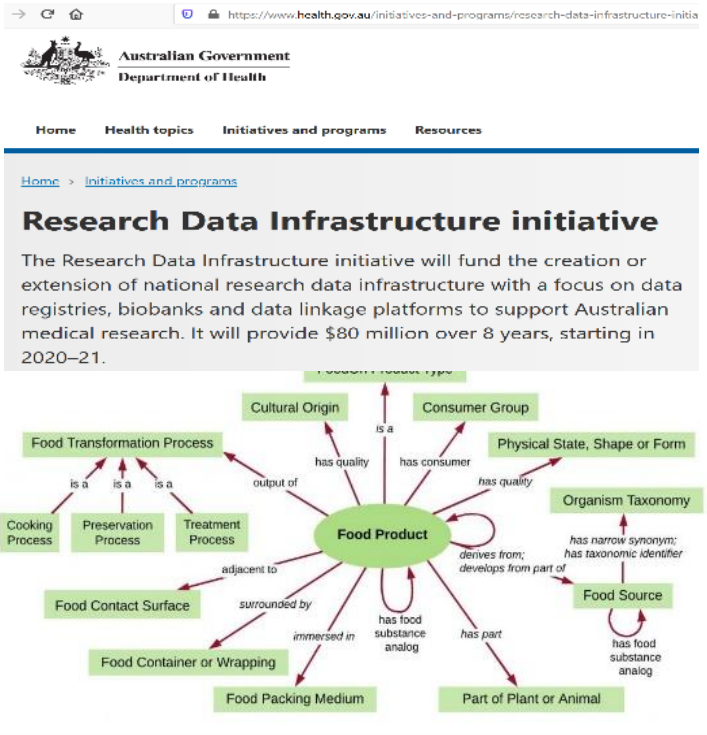
0 CO_2e plucked from the garden
32g CO_2e local and seasonal
80g CO_2e shipped, seasonal
290g CO_2e shipped, out-of-season, frozen

Apples are low-carbon food wherever they come from but local and seasonal is best[2]

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 CO_2e 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





- 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

WAGENINGEN
UNIVERSITY & RESEARCH

How do we derive knowledge out of data?

Food & Health Data Infrastructure Architecture – Way to Go

Facilities and Tools

Physical
- Restaurants of the future
- Mood rooms
- Sensory labs
- MRI
- Protein accelerator

Virtual
- Virtual buffets and supermarkets
- Virtual intervention modeling tools
- Policy intervention modelling tools

Citizen's Data Platform
- New sensors, apps, wearables
- Food frequency questionnaires, 24h recall

Human data users

Data users

Software and tools (machine learning, artificial intelligence)

Data collection tools

Ethics management tools

Data access layer: APIs, SPARQL, GraphQL,...

ideas

infra

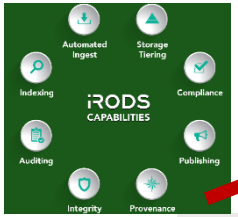
science

Usage and Deployment
- Define viable business/motivation model(s) for data ecosystem
- Define and set behavioural change data mangement practices
- Define and implement protocols and processes

FAIRification
- Identify the needed level of FAIRness
- Add relevant FAIR metadata and characteristics, f.e. information on license, provenance, consent, contract/use conditions

Interlinking
- Align food, health and consumer behaviour ontologies
- Link data
- Represent data as (research) knowledge graph

Semantic Annotation
- Provide (anonymised) data and metadata
- Provide semantic descriptions of domains
- Select and use standards, URIs
- Select, use, develop ontologies (domian, middle, upper level)



Data input layer: APIs, visual editors, harmonisation tools,...

Data providers

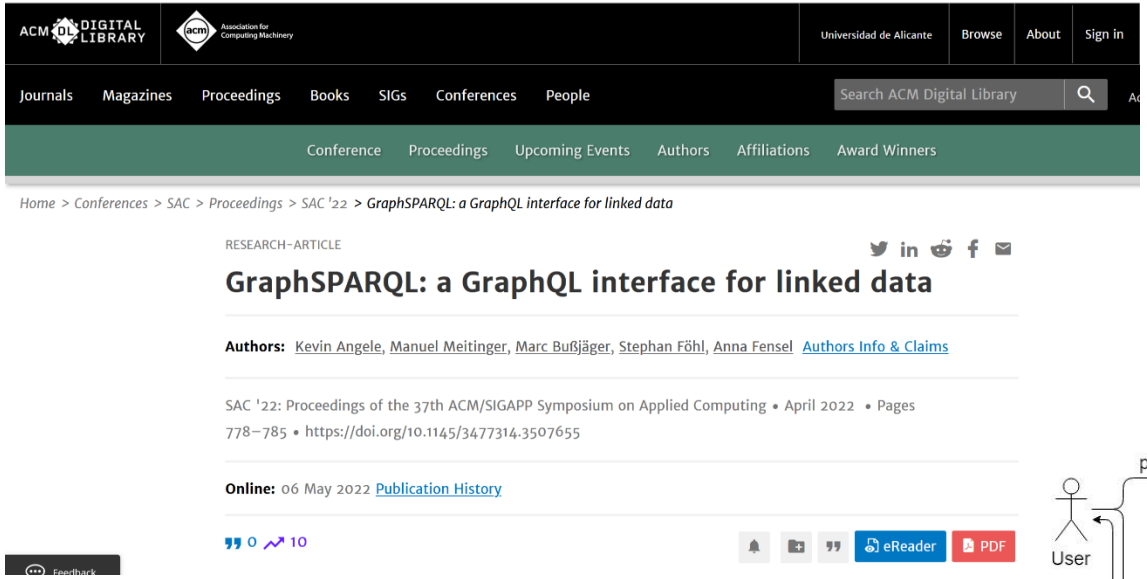
Human data providers

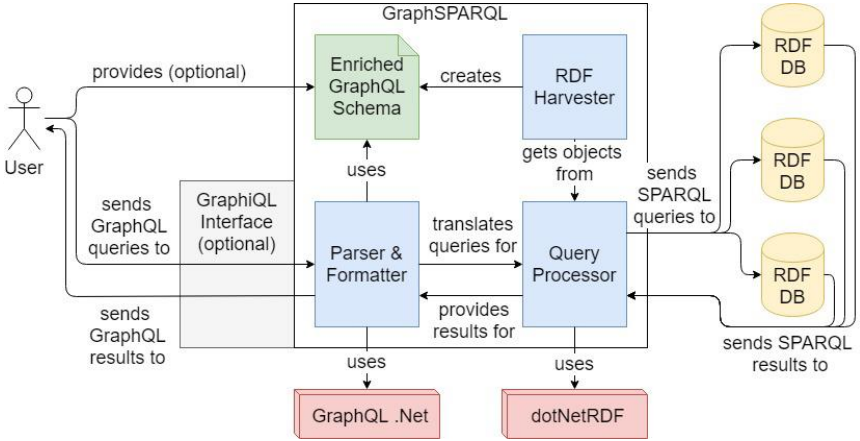
Food, Health & Environment Data
- Purchase
- Preparation
- Dietary intake
- Food composition
- Environmental footprint and life cycle analysis

Platforms
- EUROFIR
- Quisper
- ENPADASI
- EOSC
- Other data platforms e.g. Apple Health, Google

Adjacent Infrastructures
- ELIXIR
- ECRIN
- BBMRI
- METROFOOD
- Comfocus

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





WAGENINGEN
UNIVERSITY & RESEARCH

Societal dimension – Roadmapping activities

Construction of Big Data research roadmaps at EU and national level

(projects OntoCommons, BYTE, BIG, KnowledgeWeb)

ONTO COMMONS — ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

www.ontocommons.eu

-> adoption of ontology technologies

Technology in Society 54 (2018) 74–86

Contents lists available at ScienceDirect

Technology in Society

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

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, Austria[1]

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

WAGENINGEN UNIVERSITY & RESEARCH

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

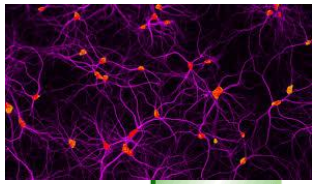
Ongoing AI and Data Science research

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.

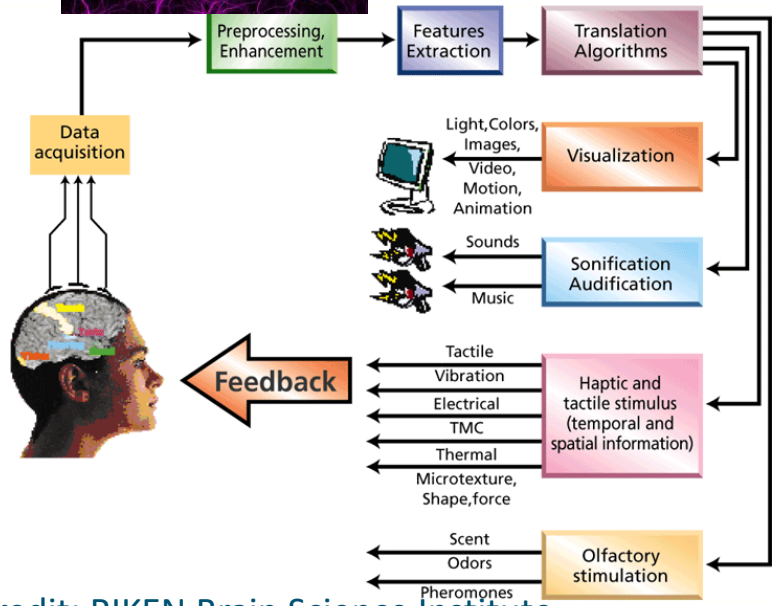
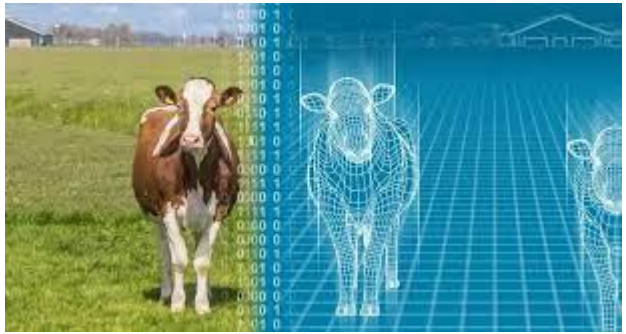


Image credit: RIKEN Brain Science Institute



Fensel, A. "**Towards an Automated Semantic Data-driven Decision Making Employing Human Brain**", In Proceedings of the 2nd International Conference on Advanced Research Methods and Analytics (CARMA 2018), pp. 167-175, Valencia, Spain (2018).

How do we derive knowledge out of data?

WAGENINGEN
UNIVERSITY & RESEARCH

Making data meaningful with semantics

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

Thank you for attention!

WAGENINGEN
UNIVERSITY & RESEARCH