Allheilmittel oder Basis-Technologie
Was ist dran an Linked Data?

Sören Auer

Did we succeed yet?

From the LOD2 - Creating Knowledge out of Interlinked Data Proposal 2009:

"*The Linked Data paradigm has evolved from a practical research idea into a very promising candidate for addressing one of the biggest challenges in the area of intelligent information management: the exploitation of the Web as a platform for data and information integration in addition to document search. To translate this initial success into a world-scale disruptive reality, encompassing the Web 2.0 world and enterprise data alike, the following research challenges need to be addressed: …*"

Unfortunately the answer is No ☹

Why did we not succeed yet

External

- In the beginning: **disadvantage of the late birth** – XML was first
- Now: **Big Data** is dominating everything

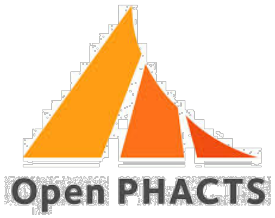
Internal

- The Linked Data community did not manage to engage and support a **critical mass of application domains** yet
- the advantage of Linked Data and semantic representations only outweighs the additional overhead/complexity if data is truly heterogeneous and under diverse control – **data value chains**

Sucessful application domains
Linked Data & Semantic Integration

- Search Engine Optimization & Web-Commerce

- Pharma, Lifesciences



- Digital Libraries

… but what about

Manufacturing, Logistics and Industrial Applications

Banking, Finance, Insurance, Compliance, Risk Assessment

Enterprise Information Integration & Architecture

Retail, Product Data

Automotive, Transportation

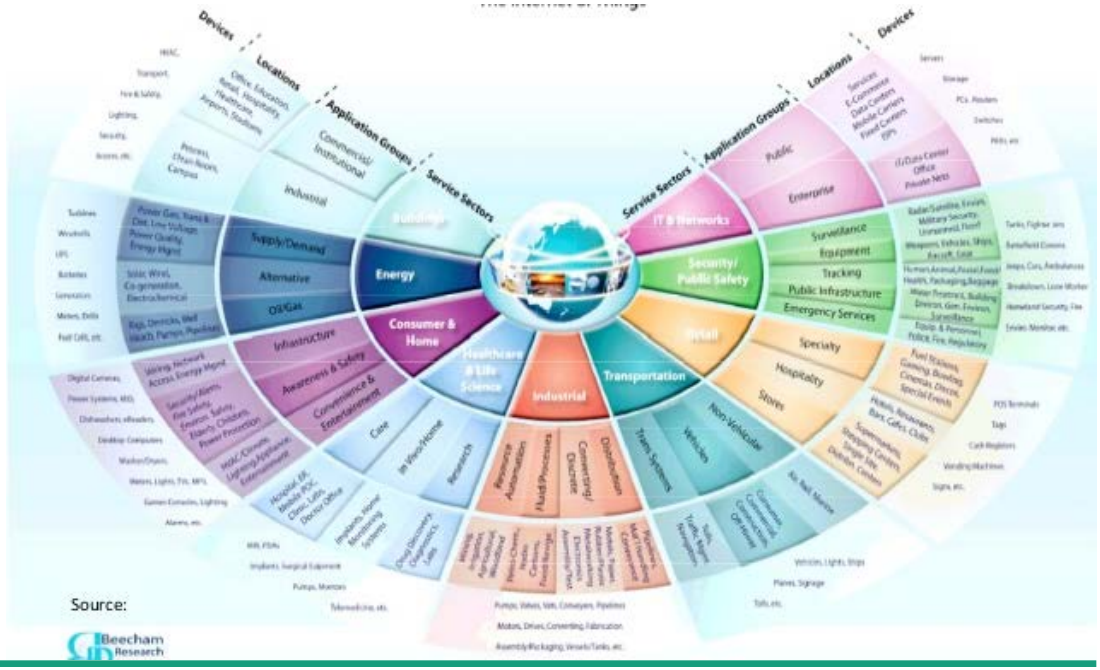
Travel, Mobility

Energy

Agriculture, Food

Research data

…



Fraunhofer IAIS

How can we bring Linked Data to more domains?

DATA VALUE CHAINS

Fraunhofer
IAIS

The rolling Smartphone

New Business Models for the Automotive Industry with Data Value Chains

Quelle: GTÜ

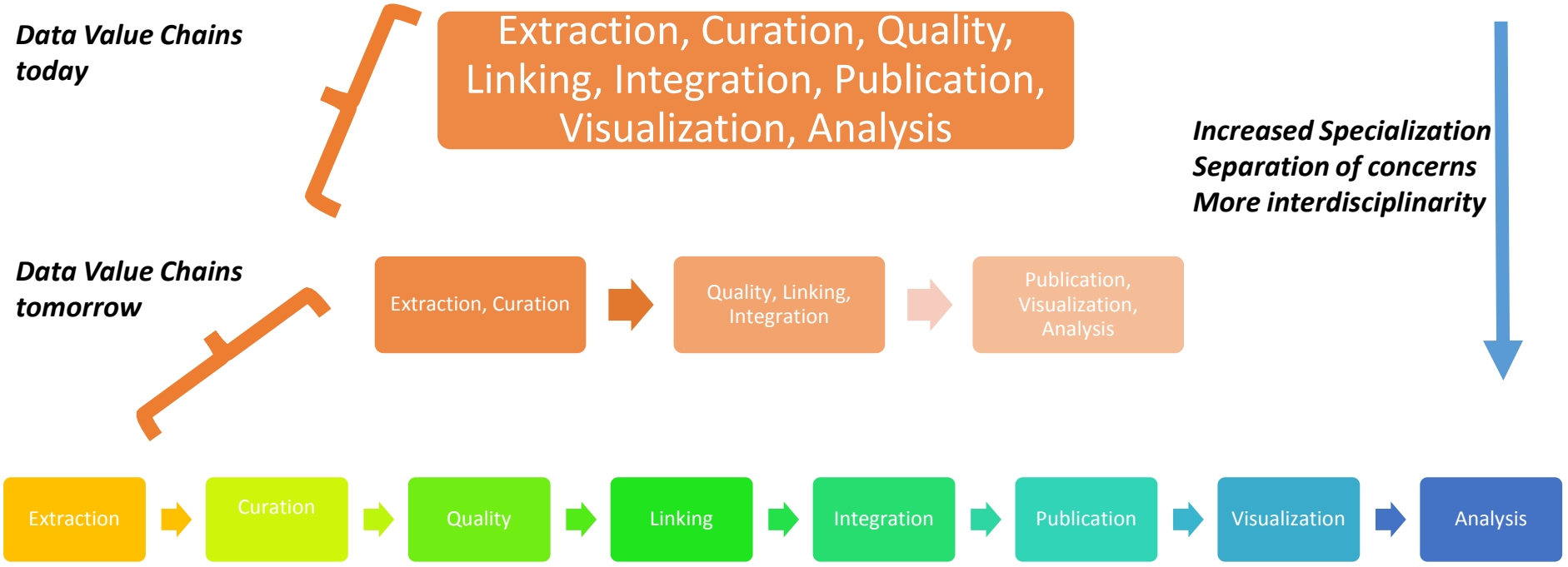
Quelle: www.farming-simulator.com

Windshield wiper as rain sensors for micro wether prognosis

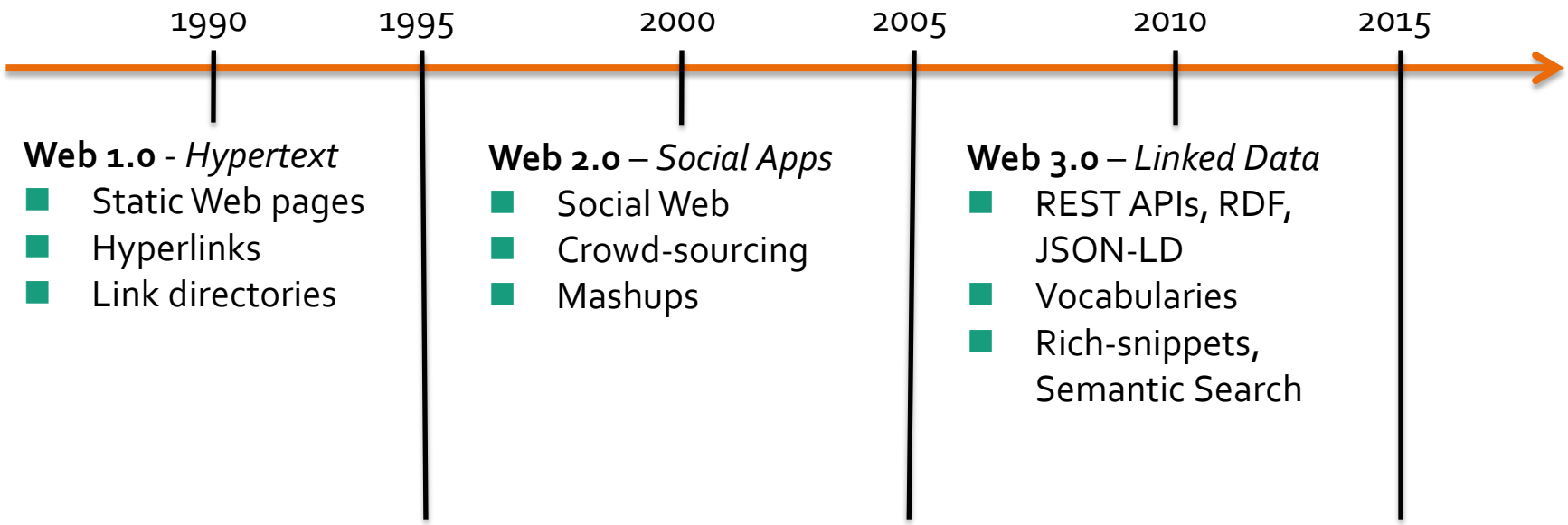
Automotive industry can become data provider for other industries

Fraunhofer IAIS

Towards Data Value Chains



The Evolution of the Web & Intranets



Web 1.0 - *Hypertext*
- Static Web pages
- Hyperlinks
- Link directories

Web 2.0 – *Social Apps*
- Social Web
- Crowd-sourcing
- Mashups

Web 3.0 – *Linked Data*
- REST APIs, RDF, JSON-LD
- Vocabularies
- Rich-snippets, Semantic Search

Intranet 1.0 - *Hypertext*
- Static Intranet pages
- Keyword search
- Hyperlinks

Intranet 2.0 – *Social Enterprise Apps*
- Salesforce
- Crowd-sourcing
- Mashups

Intranet 3.0 – *Enterprise Data Intranet*
- URI Scheme
- Enterprise taxonomies / knowledge bases
- RDB2RDF Mapping

Fraunhofer IAIS

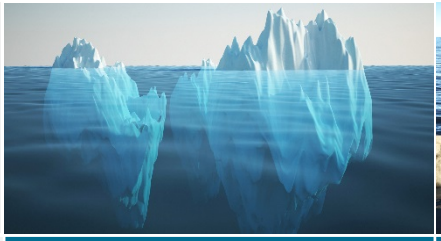
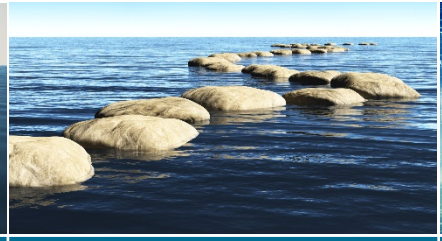
The Industrial Data Space Initiative

Community of >30 large German and European Companies

Pre-competitive, publicly funded innovation project involving 11 Fraunhofer institutes for developing IDS reference architecture

Current signatories of the MoU to support the **Industrial Data Space Association**

Semantic Data Linking for Enterprise Data Value Chains

	Data Lake	Industrial Data Space	Pure Internet
			
	centralized, monopolistic	federated, secure, „trusted", standard-based	completely dezentral, open, unsecure
Data management	Central Repository	Decentral	Decentral
Data Ownership	Central	Decentral	Decentral
Data Linking	Single provider	Federated, on demand	Missing
Data Security	Bilateral	Certified system	Bilateral
Market structure	Central Provider	Role system	Unstructured
Transport infrastructure	Internet	Internet	Internet

Bilder: ©Fotolia
Francesco De Paoli, Nmedia, hakandogu

Fraunhofer

Basic principles of the Industrial Data Space

On Demand Interlinking

Linked Light Semantics

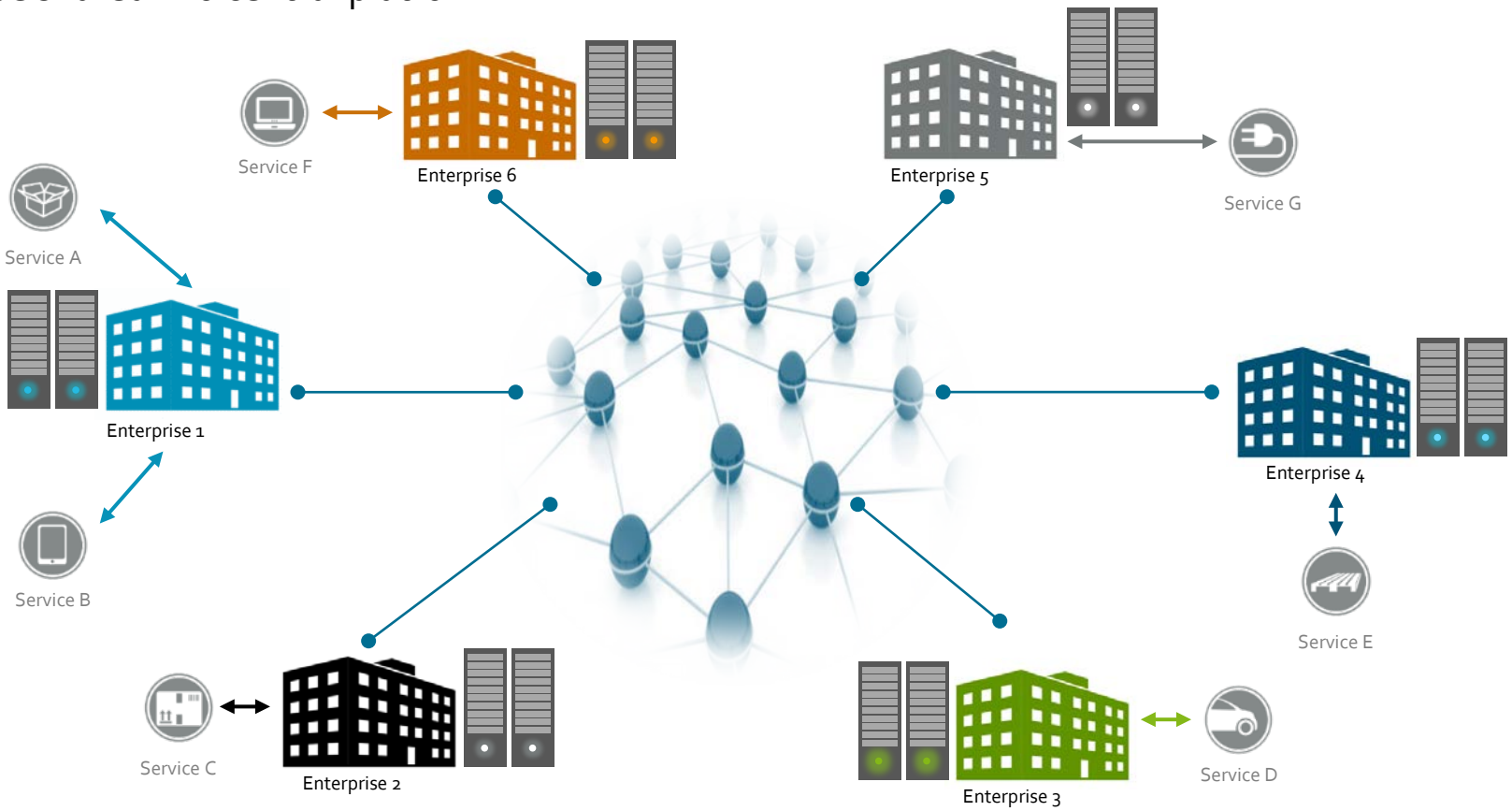
Security with Industrial Data Container

Certified Roles

Fraunhofer

Industrial Data Space: On Demand Interlinking

All Data **stays with its Ownern** and are controlled and secured. Only on request for a service data will be shared. No central platform.



Fraunhofer

Linked Light Semantics

A lighweight approach for Data Interlinking

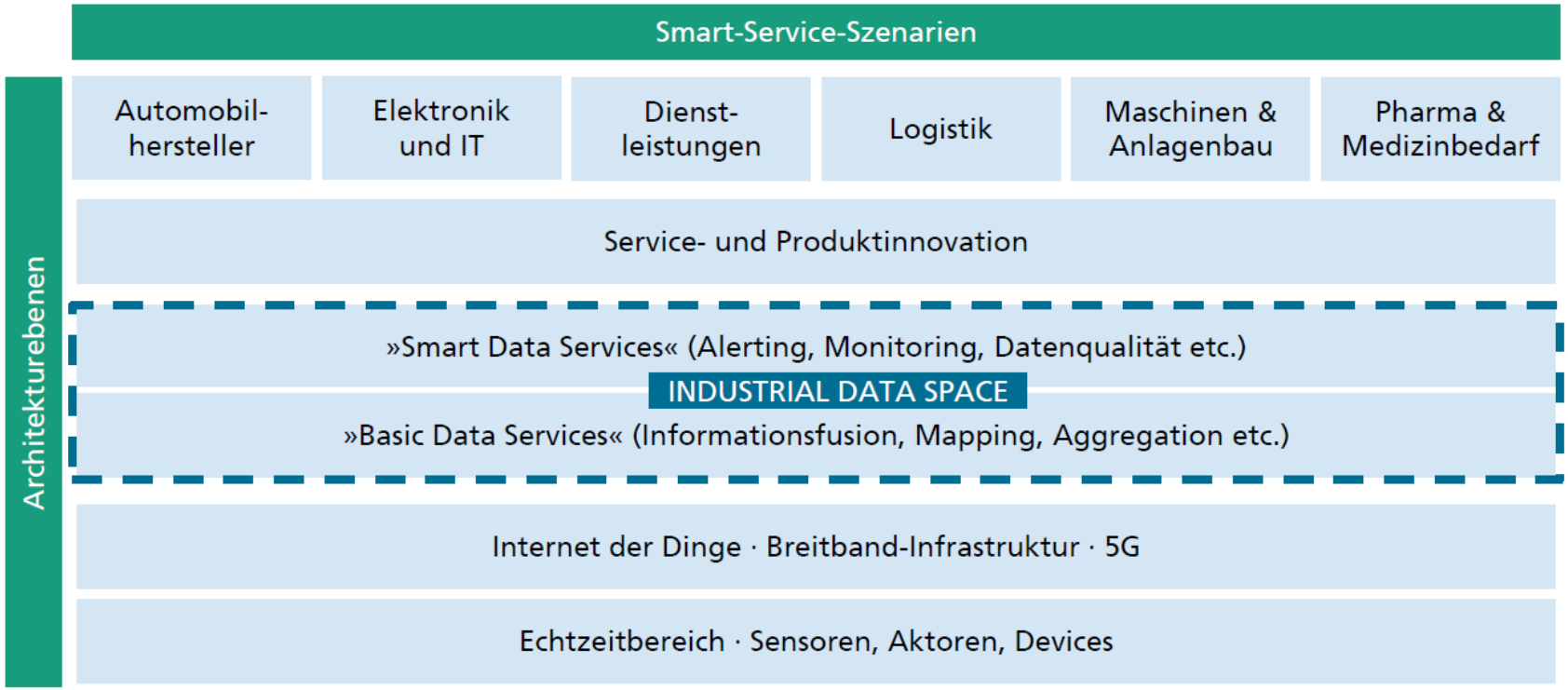
Classical Enterprise systems	Linked Light Semantics	Internet / WWW
Fixed Data schema	Reference vocabularies	Web pages
Globale Enforcement	Bridge between local Representations	Only Links
Closed	Intelligent and structured interlinked	Completely open
Manuel Transformation	Automatic translation/mapping	Lack of standardization
High cost	Leight-weight	No structure

Q: istockphoto.com

IDS leverages Linked Data for semantic interoperability

Problem	Linked Data Approach	Example
1. Unique Identifikation of (data) objects	URIs (analog Web addresses) for Identification of arbitrary objects	`https://data.vw.de/car/Golf7`
2. Adressability and Data access	Web Protokols HTTP/HTTPS for De-Referencing and access of data	
3. Semantic Data Representation	Triple & Graph-based RDF Data Model	
4. Wide Interlinking of Data	URIs serve as "Data Links" between distributed Databases	
5. Domain-specific Data structures	Creation of interlinked, modular, reuseable vokabularies	
6. Security-by-Design	Certificates, Encryption, Authentification as Internet Banking	

IDS focusses on Data and Data-driven Service Architecture



IDS Reference Architecture

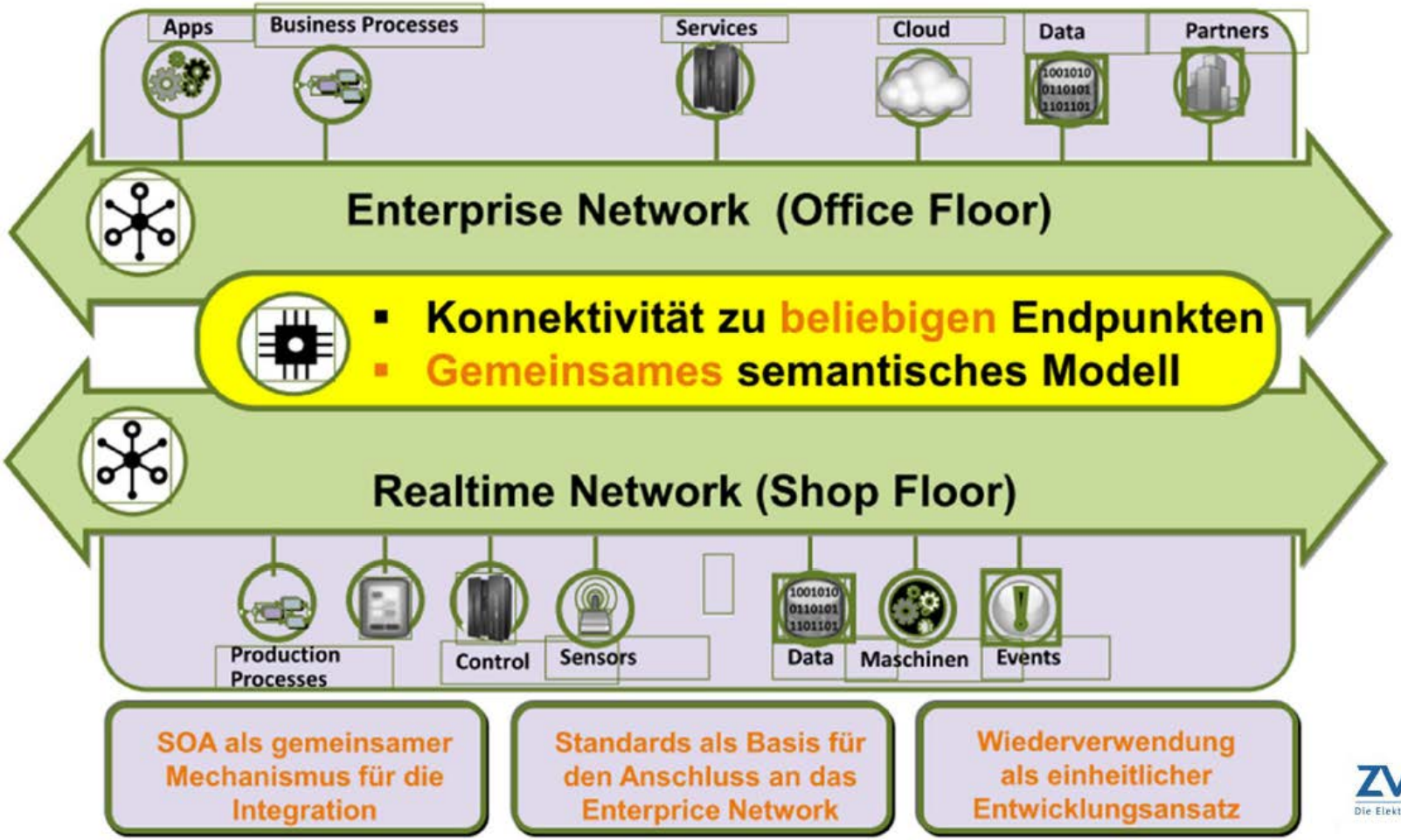
Defines …

- Governance
- Data services architecture
- Functional domain architecture
- Security architecture

Envisions six roles

- **Data Owner**
- **Data User**
- **Broker** - registry for data sources
- **IDS-AppStore Provider**
- **Clearing House** – usage tracking, accounting/invoicing
- **Certification**

IDS for Industry 4.0:
Semantic Models bridge between Shop & Office Floor



Quelle: ZVEI Führungskreis Industrie 4.0; 14.01.2014

* SOA: Service - Oriented - Architecture

Currently Prioritized IDS Use Cases

Pharma

- »Real-Life Evidence«
- More effective and efficient treatment
- Personalized medicine

Automobil

- Traffic management 2.0
- Dynamic route planning
- »Connected Drive Services«

Retail

- Autonomous transparecy in the supply chain
- Consumer-oriented Supply Chain

Produktion

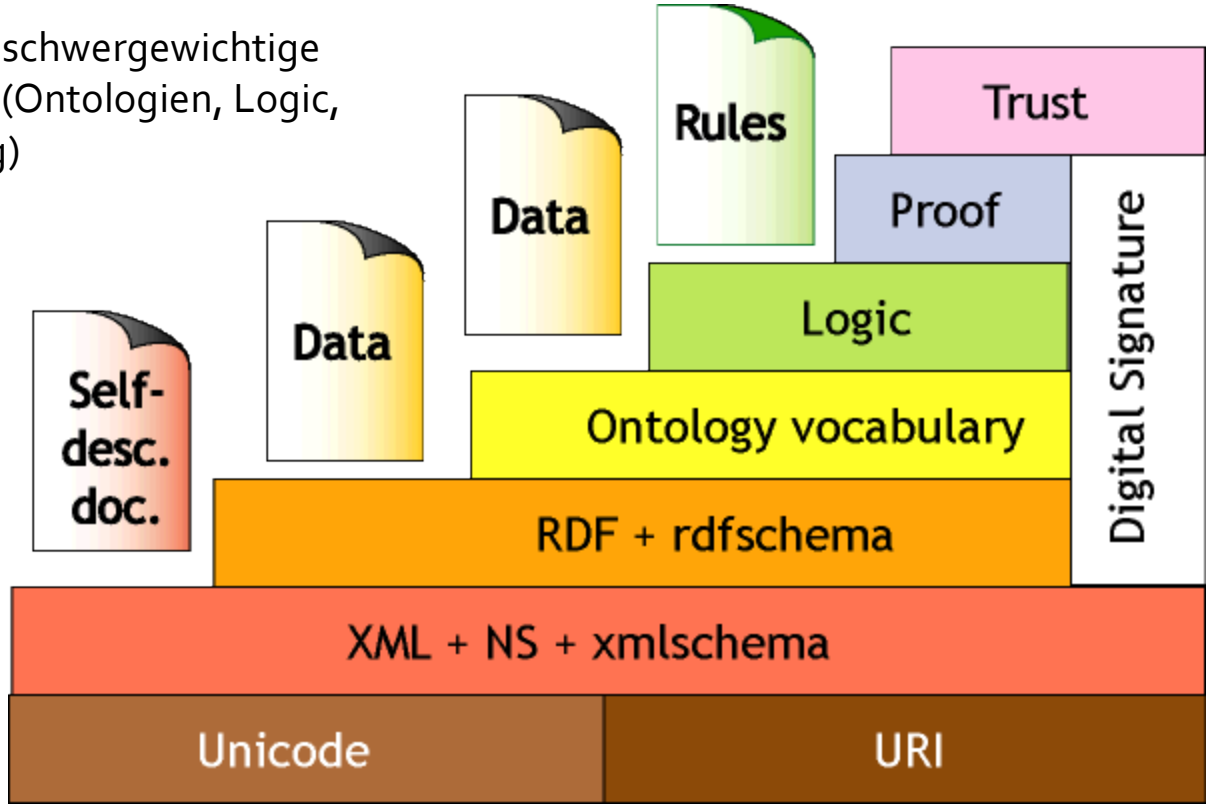
- Intelligent production scenarios for small series
- Self-control of production

RDF MUST BECOME THE LINGUA FRANCA OF DATA INTEGRATION

Fraunhofer
IAIS

Die Semantic Web Layer Cake 2001

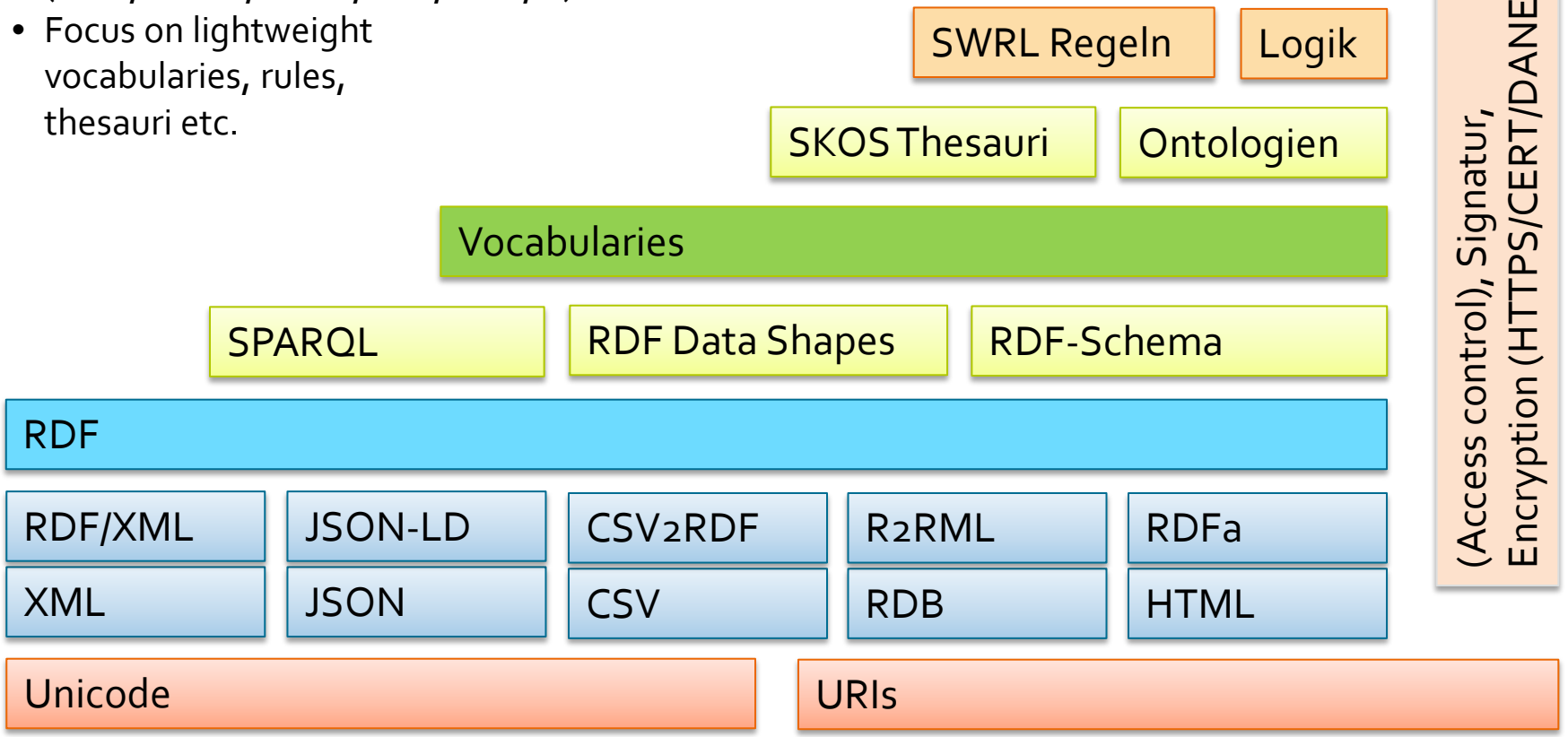
- Monolithisch basierend auf XML
- Fokus auf schwergewichtige Semantik (Ontologien, Logic, Reasoning)



http://www.w3.org/2001/10/03-sww-1/slide7-0.html

The Semantic Web Layer Cake 2015 – "A Little Semantics Goes a Long Way"

- Lingua Franca of Data integration with many technology interfaces (XML, HTML, JSON, CSV, RDB,...)
- Focus on lightweight vocabularies, rules, thesauri etc.



Comprehensive, Evolving, Community-curated, Executable

WE NEED BETTER VOCABULARIES

Existing vocabularies are not suited for cross-domain applications

For the 457 vocabularies listed in LOV:

- average number of classes: 42
  average number of properties: 59
- Omitting the four vocabularies with the highest number of classes and properties, these figures decrease to 31 classes and 37 properties

A large number of crucial domains is not or only superficially covered by existing vocabularies:

Table 1: Identified gaps in the coverage of existing vocabularies

Domain	Information	Vocabulary support missing for ...
Spatial data	Maps	map data including road networks, traffic signs
Education	School Grades	levels of achievement, and their scales, depending on the country
Calendar	Holidays	fixed or floating holidays in each country, duration, types (e.g. non-working days, religious holidays, state holidays)
Labour market	Professions	qualifications, degrees, levels of experience; in general, and specific to countries
Nutrition	Cooking recipes	ingredients and their properties (e.g. allergenes), how to process them, possible substitutes
Do it yourself	Building/ repairing guides	required tools and materials, steps, dangerous warnings
Supply chain	Logistics, Suppliers, Products	contact information, product descriptions, production/demand forecasts

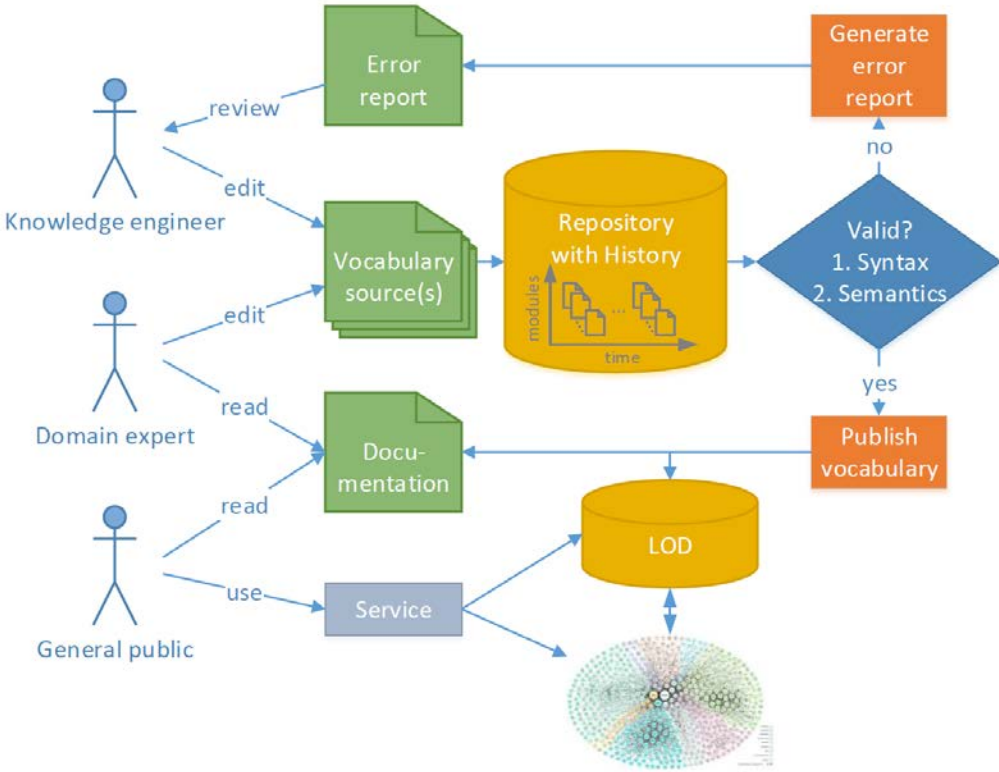
VoCol: Git basieres Kollaboratives Vokabulary-Engineering

Methodologie & Infrastruktur

Inspiriert durch agile Software-Entwicklu
und modern Versionskontrolle

**Idea: commit vocabulary (file) changes t
Git, VoCol will take care of everythir
else (dereferencability,
documentation, visualization,
versioning, SPARQL endpoint with
example data …)**

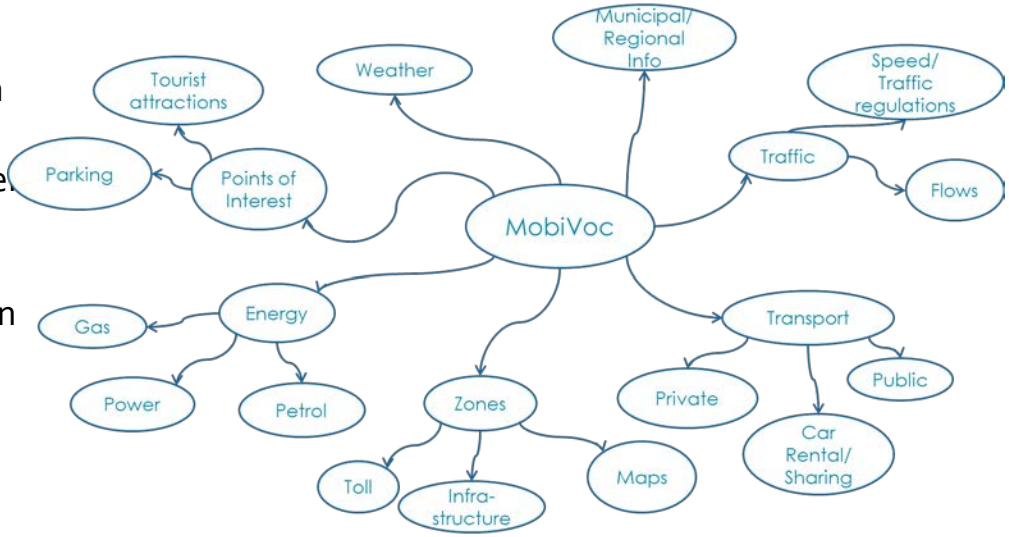
http://eis.iai.uni-bonn.de/Projects/VoCol.html

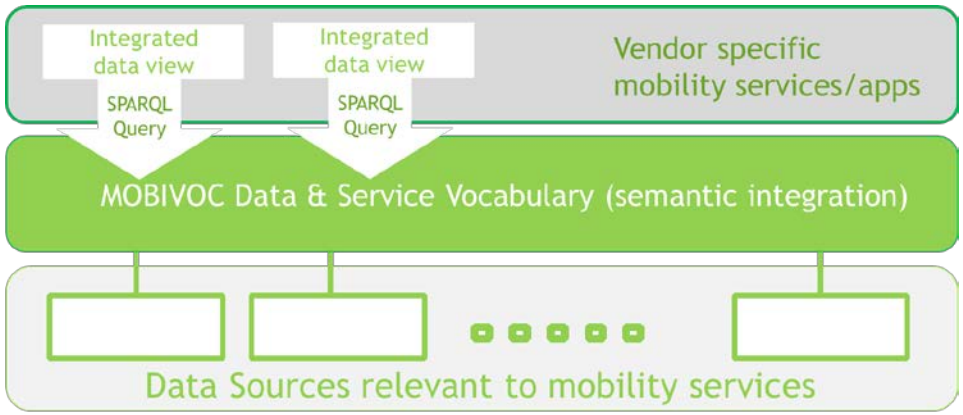


Fraunhofer IAIS

Beispiel Mobility Vokabular MobiVoc – Supporting the mobility of humans by the mobility of data

- Vernetzung und Integration von Informationen aus einer Vielzahl verschiedener Quellen (Kartendaten, Fahrzeugdaten, Wetterdaten, Mobilitätsangebote, Ereignisdaten,...) Diese Daten werden von einer Vielzahl verschiedener Akteure und Organisationen vorgehalten und haben oft proprietäre Datenstrukturen.
- Ziel der MobiVoc-Initiative ist es, mit Vokabularen und unter Nutzung von Semantic-Web-Technologien und Ontologien die Mobilität von Daten zwischen den zahlreichen beteiligten Akteuren entscheidend zu verbessern
- Initiative des ITA Automotive Service Partner e.V. unter Beteiligung von BMW, Microsoft, Accenture, Fraunhofer
- Kollaborative, agile Vokabularentwicklung auf GitHub





Fraunhofer IAIS

VoCol Configuration Page

Search

Home

General Info

Vocabulary Name: Enter vocabulary name

Domain name: Enter domain name

Client Side Hooks ☑

Web Hook ☑

Turtle Editor ☑

Repository info

Repository Name: Enter repository address

Branch Name: Enter branch name to be monitore

User: Enter repository user

Password: Enter repository password

Syntax Validation

Rapper ◉

Jena Riot ○

Documentation Generation

SchemaOrg ◉ ?

Widoco ○ ?

Additional Services

Visualization ☑

Sparql EndPoint ☑

Syntax Validation Report ☑

Schema Evolution Report ☑

Homepage description

File Edit Insert View Format Table Tools

Formats B I

article » div

Save Configuration

Supporting human mobility by data mobility

New mobility concepts and better data networking are both crucial factors for global economic development. To invent innovative and sustainable mobility concepts, new data-based value-added services are required.

For example:

Route Planning

Route planning taking into account places of interest, energy saving aspects (e.g. for electric vehicles) or complex travel arrangements

Multimodal Mobility

Linking various multimodal mobility services with regard to complex environment conditions and current or forecast events like weather or traffic volume

Alternative Scenarios

Dynamic and interactive planning of alternative mobility services in the event of unforeseen circumstances (e.g. breakdown in public transport, congestions)

To implement such innovative data-based services for mobility support it is necessary to link and integrate data coming from a vast number of different sources. Some examples are map data, vehicle data, weather data, mobility service descriptions or events information. These data sets come from various stakeholders and organisations and they often have proprietary data structures.

Our goal is to significantly improve the data mobility between all stakeholders by providing a standardized vocabulary using Semantic Web technologies and ontologies. For the open vocabulary covering various mobility aspects we use RDF (Resource Description Framework) – a recommended specification of the World Wide Web Consortium (W3C) and the so-called lingua franca for the integration of data and web. We invite everyone who is interested to join our MobiVoc initiative and to participate in the development of the Open Mobility Vocabulary.

Organization of Schemas

The schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy.

Classes

- openingHours
- Engine
- Plug
- ParkingFacilityLocation
- ChargingPointFees
- PlugType
- MeansOfTransport
- ChargingSpeed
- ChargingPointOwner
- VehicleType
- ParkingFacilityConnection
- AccessType
- EmissionZone
- EmissionZone
- BookingType
- Charger
- ConvexHull
- ParkingFacilityFeature
- ChargingPointLocation
- polygon
- ParkingFacilityConfiguration
- AccessInformation
- ParkingFacilityStatus
- ParkingFacility

openingHours

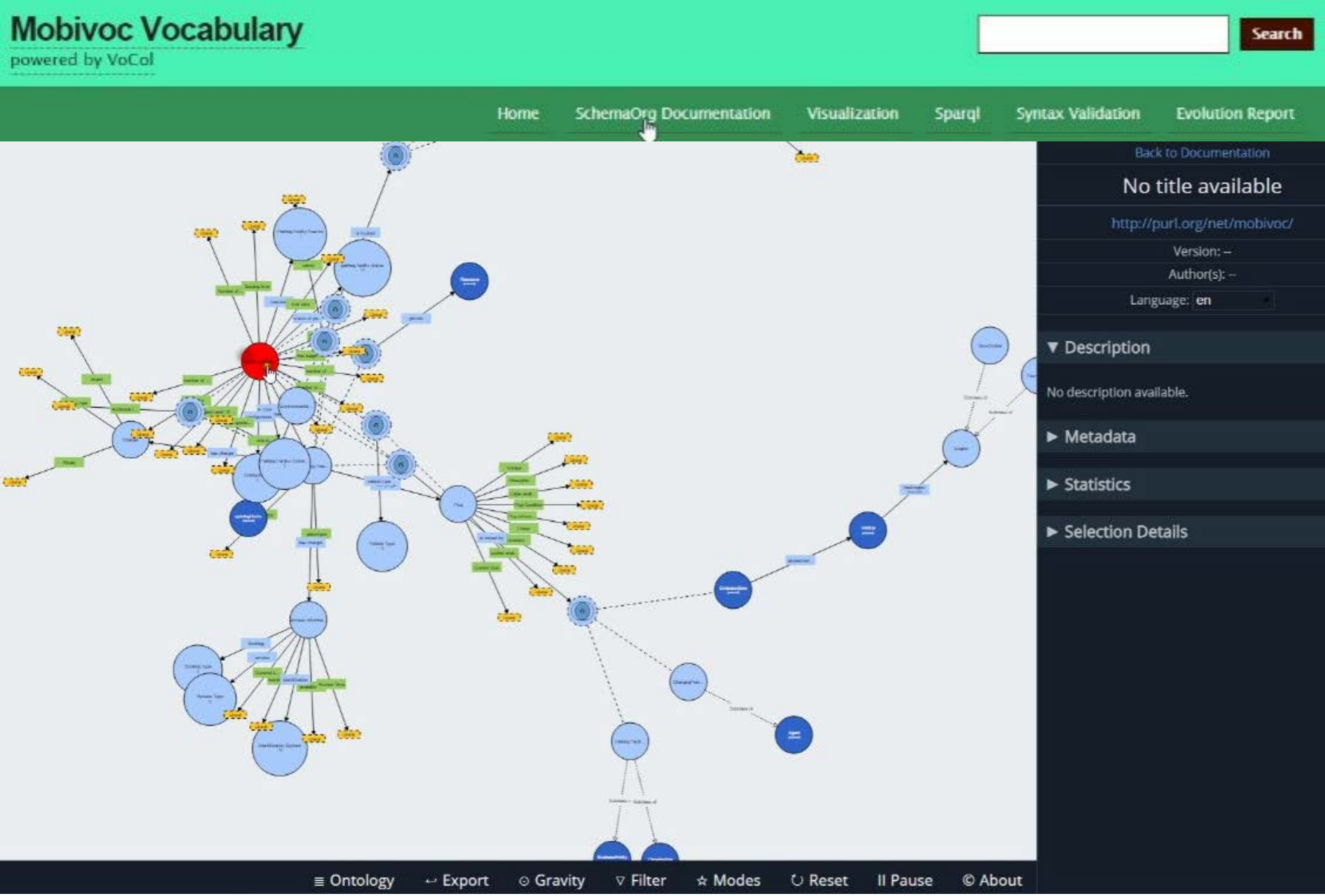
Thing > Property > openingHours

The opening hours for a business. Opening hours can be specified as a weekly time range, starting with days, then times per day. Multiple days can be listed with commas ',' separating each day. Day or time ranges are specified using a hyphen '-'.
- Days are specified using the following two-letter combinations: `Mo, Tu, We, Th, Fr, Sa, Su`.
- Times are specified using 24:00 time. For example, 3pm is specified as `15:00`.
- Here is an example: `<time itemprop="openingHours" datetime="Tu,Th 16:00-20:00">Tuesdays and Thursdays 4-8pm</time>`.
- If a business is open 7 days a week, then it can be specified as `<time itemprop="openingHours" datetime="Mo-Su">Monday through Sunday, all day</time>`.

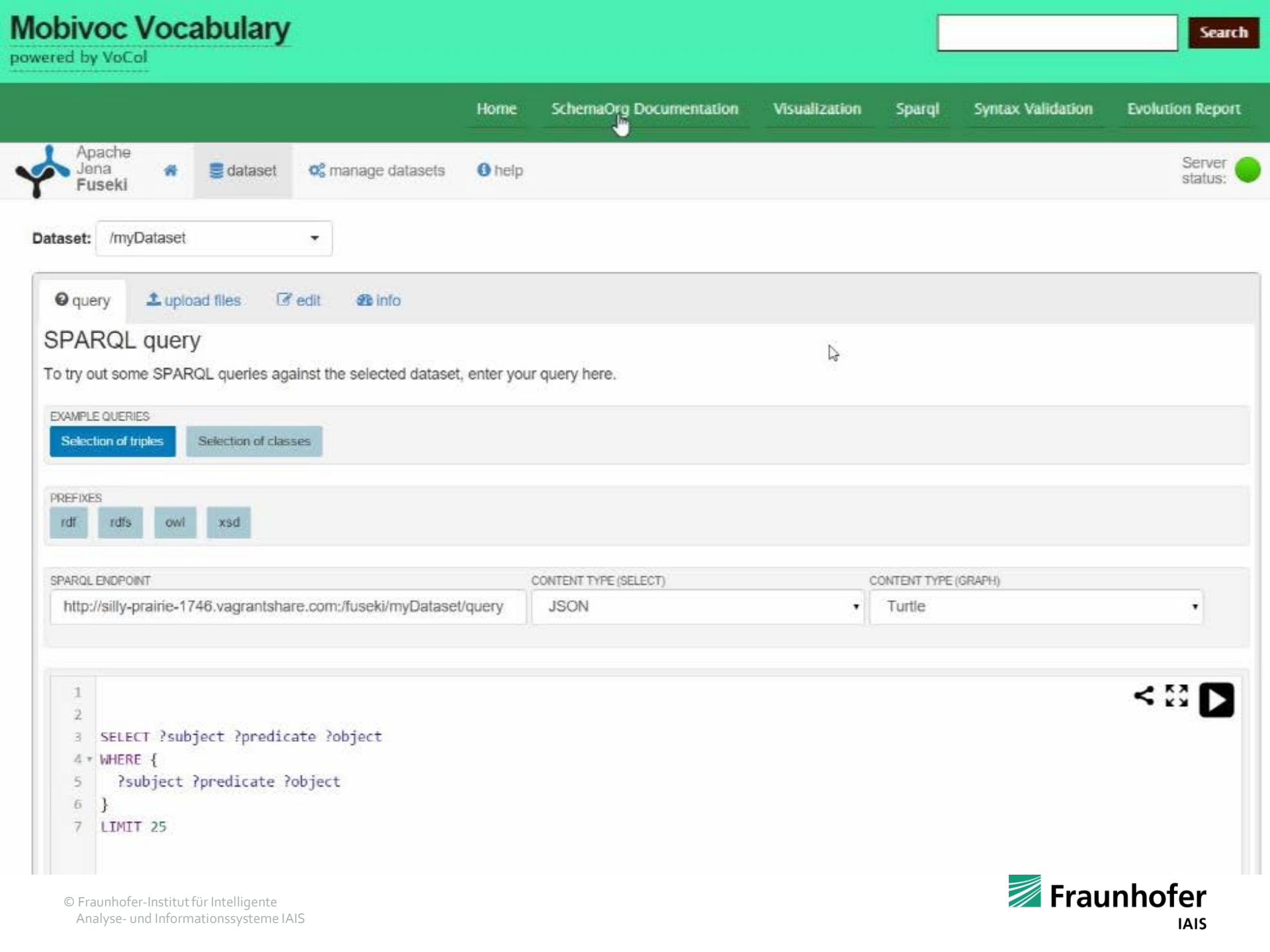
Usage: Between 100,000 and 250,000 domains

[more...]

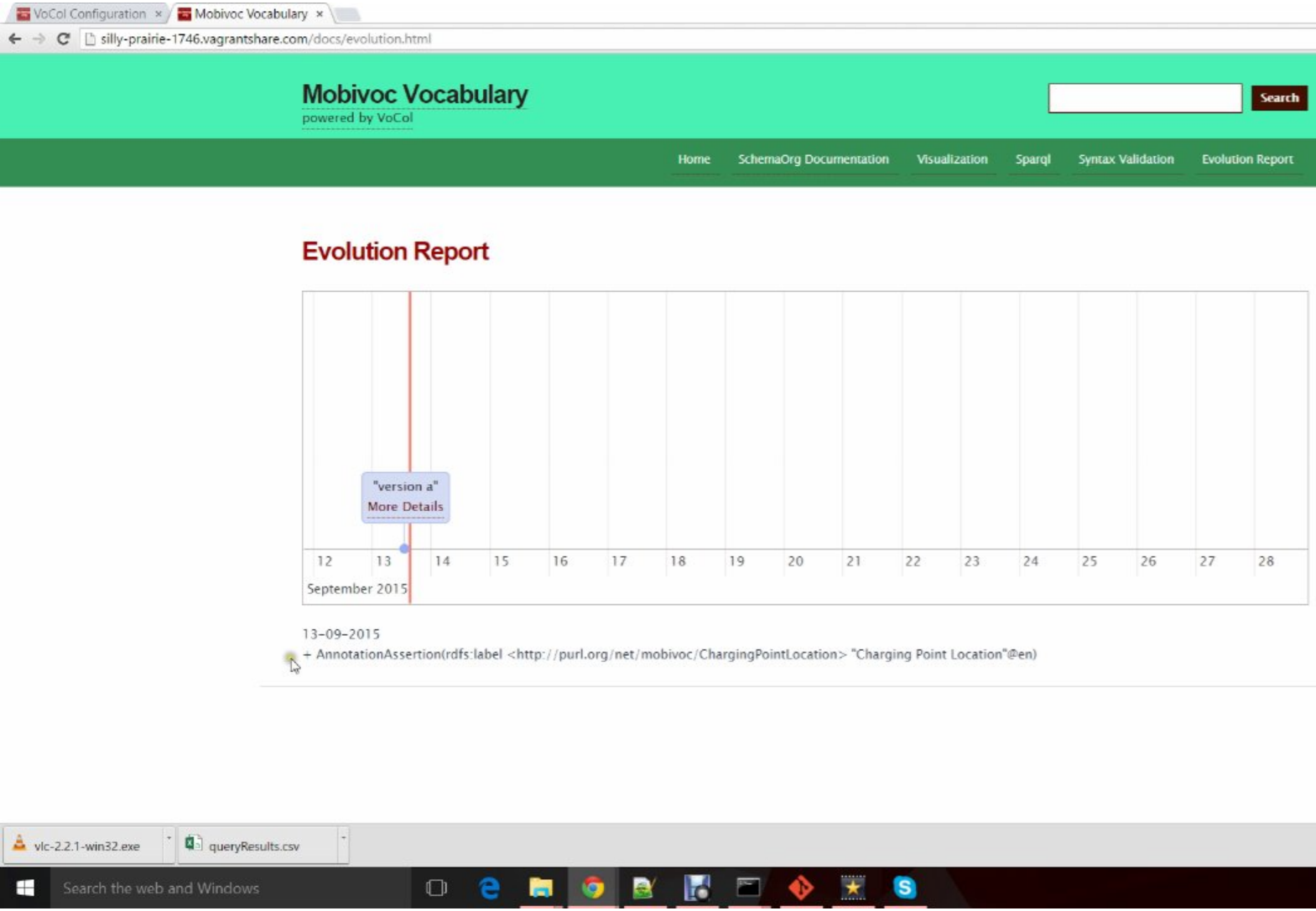
Property	Expected Type	Description
Properties from Thing		
additionalType	URL	An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax - the 'typeof' attribute - for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.
alternateName	Text	An alias for the item.
description	description or Text	Description of charging point.
image	ImageObject or URL	An image of the item. This can be a URL or a fully described ImageObject.
name	Text	The name of the item.
	URL	URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page,



Fraunhofer
IAIS



Fraunhofer IAIS



Fraunhofer
IAIS

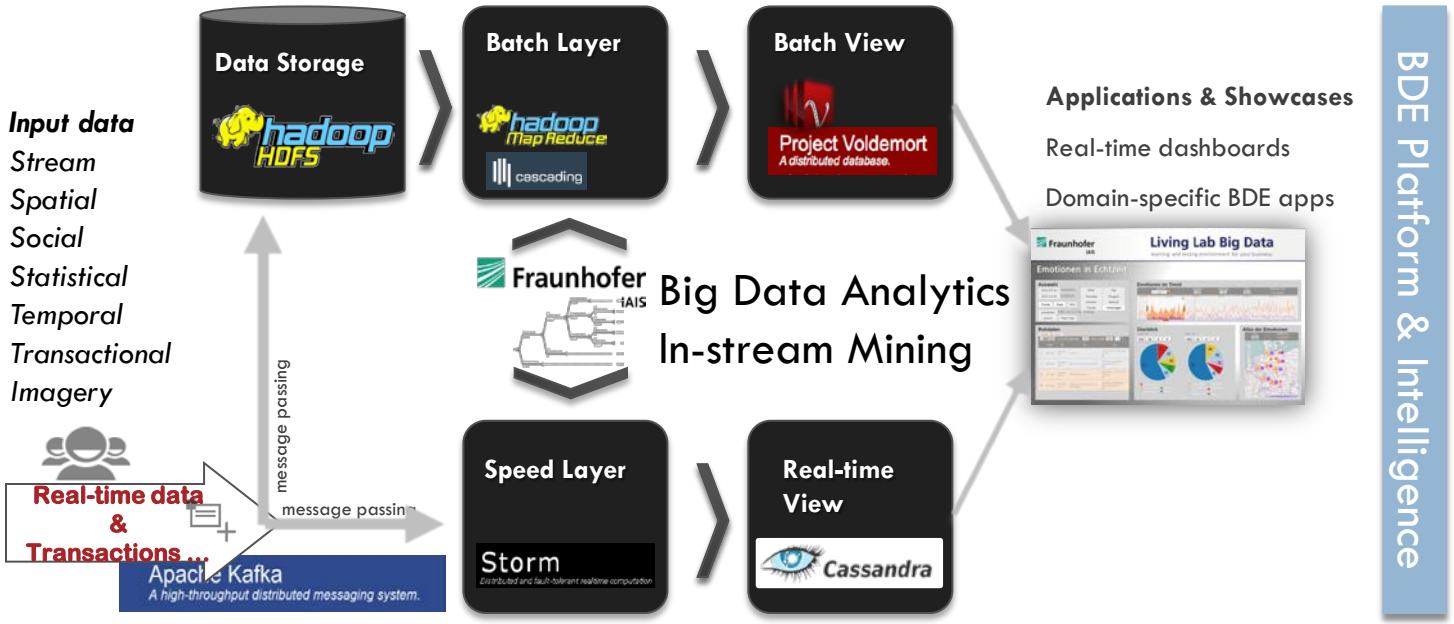
Challenge

HOW CAN WE COMBINE BIG DATA & LINKED DATA

Fraunhofer IAIS

Blueprint of the Data Aggregator Platform

- Follows typical Lambda Architecture



- Possibly integrated on top of existing Big Data distribution

Hortonworks

Data Aggregator Platform Challenges

- Ingest semantic (RDF) and non-semantic (CSV, JSON, XML, …) data
  - Integrate various mapping techniques (R2RML, CSV on the Web, JSON-LD)
- preserve semantics, provenance and metadata in Big Data processing chains
  - Preserve URI/IRIs
  - Preserve triples
- Exploit semantics for aggregations

Challenge

HOW CAN WE EXPLOIT LINKED DATA (FOR QUESTION ANSWERING)

Fraunhofer
IAIS

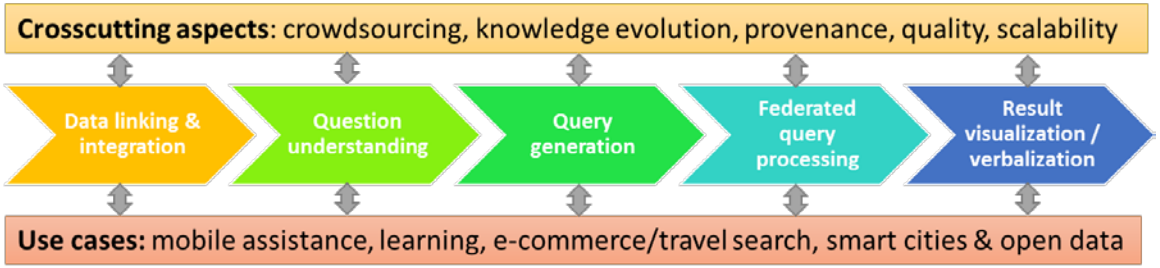
Deep AI for Question Answering

Grand Challenge for Computer Science

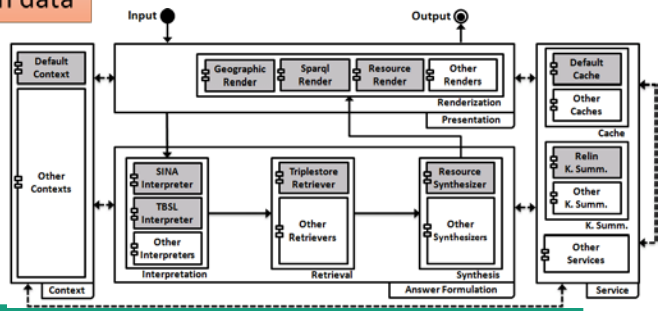
Question Answering is currently more smart information retrieval or based on inflexible query templates

We work on a next generation of QA, which uses structured data (databases, big data, open data) as background knowledge

Many hard AI problems in NLP, data integration/quality, query processing



OpenQA - Design and develop an open architecture for question answering



Fraunhofer IAIS

Take home messages

We are not done yet ☹

Linked Data gains slowly traction – could be/must be faster

More effort has to be spend on

- Developing **cross-domain (Linked) Data value chains**
- Bringing Linked Data to **more domains**
  (finance, transport, research data, …)

Challenges include

- Big Data vs. Linked Data
- Comprehensive, Evolving, Community-curated, Executable vocabularies
- Question answering

„DO MORE WITH [BIG|LINKED|OPEN] DATA!“

Prof. Dr. Sören Auer, soerean.auer@iais.fraunhofer.de

Fraunhofer-Allianz Big Data | Fraunhofer IAIS

Schloss Birlinghoven

53757 Sankt Augustin

www.iais.fraunhofer.de

www.bigdata.fraunhofer.de

© vege | Fotolia

european data forum

Luxembourg, 16-17 Nov 2015

http://2015.data-forum.eu/