Datasets of the Federal Environment Agency for
Linked Environment Data

Maria Rüther
Federal Environment Agency (UBA), Germany

Thomas Bandholtz, Till Schulte-Coerne
innoQ Deutschland GmbH

Open Environment Data in the 90s

- The Aarhus Convention 1998
  - "Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters"
  - United Nations Economic Commission for Europe (UNECE).

>> European Environment Agency (EEA)

Catalogue of Data Sources (CDS, discontinued)

GEneral Multilingual Environment Thesaurus (EEA)

>> Environment Agencies in Germany

Umwelt-Thesaurus UMTHES

Umweltdatenkatalog (UDK)

German Environmental Information Network (GEIN)

Open Environment Data > 2003

- Directive 2003/4/EC on public access to environmental information
    - Headed by "**Freedom of access to information**".
    - Requests for information may be refused for several reasons
        - the confidentiality of the proceedings of public authorities or of commercial or industrial information,
        - public security or national defense,
        - the course of justice,
        - intellectual property rights,
        - the confidentiality of personal data,
        - the interests of the person who supplied the information on a voluntary basis,
        - or the protection of the environment.
- National law in Germany since 2005

Current „official" Approaches

- From GEIN to PortalU in Germany

  - Extended information scope and contributants

- INSPIRE

  - Directive 2007/2/EC Infrastructure for Spatial Information in the European Community

  - Based on OGC Standards (CSW, Feature, Sensor Web, …)

- EEA: Shared Environmental Information System (SEIS)

  - Commission Communication 2008

- Many public sectoral portals (water, soil)

- Web pages, PDF, CSV, XML over Web Services
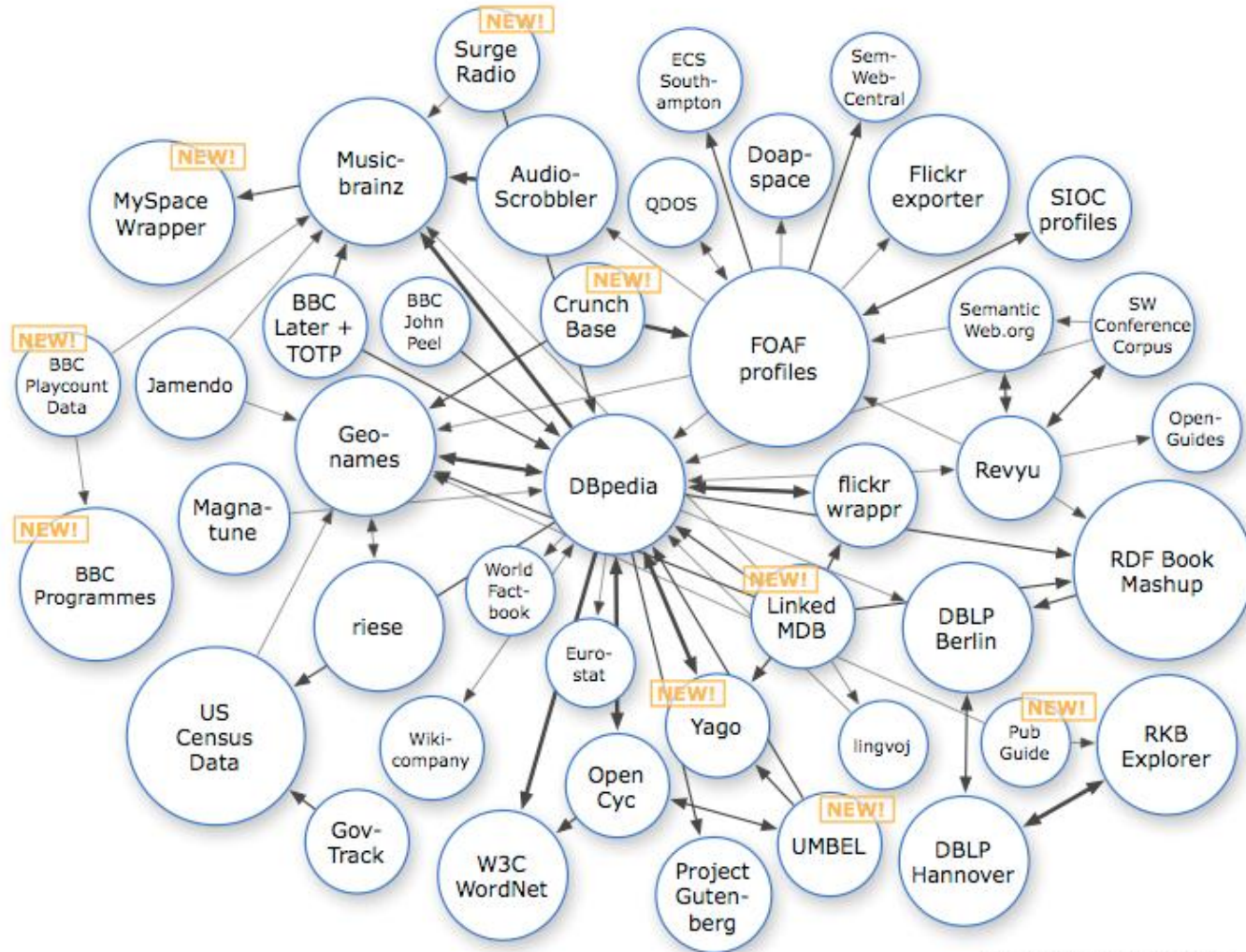
- Exhausting harmonization processes!!!

Suddenly …

2007-10-08
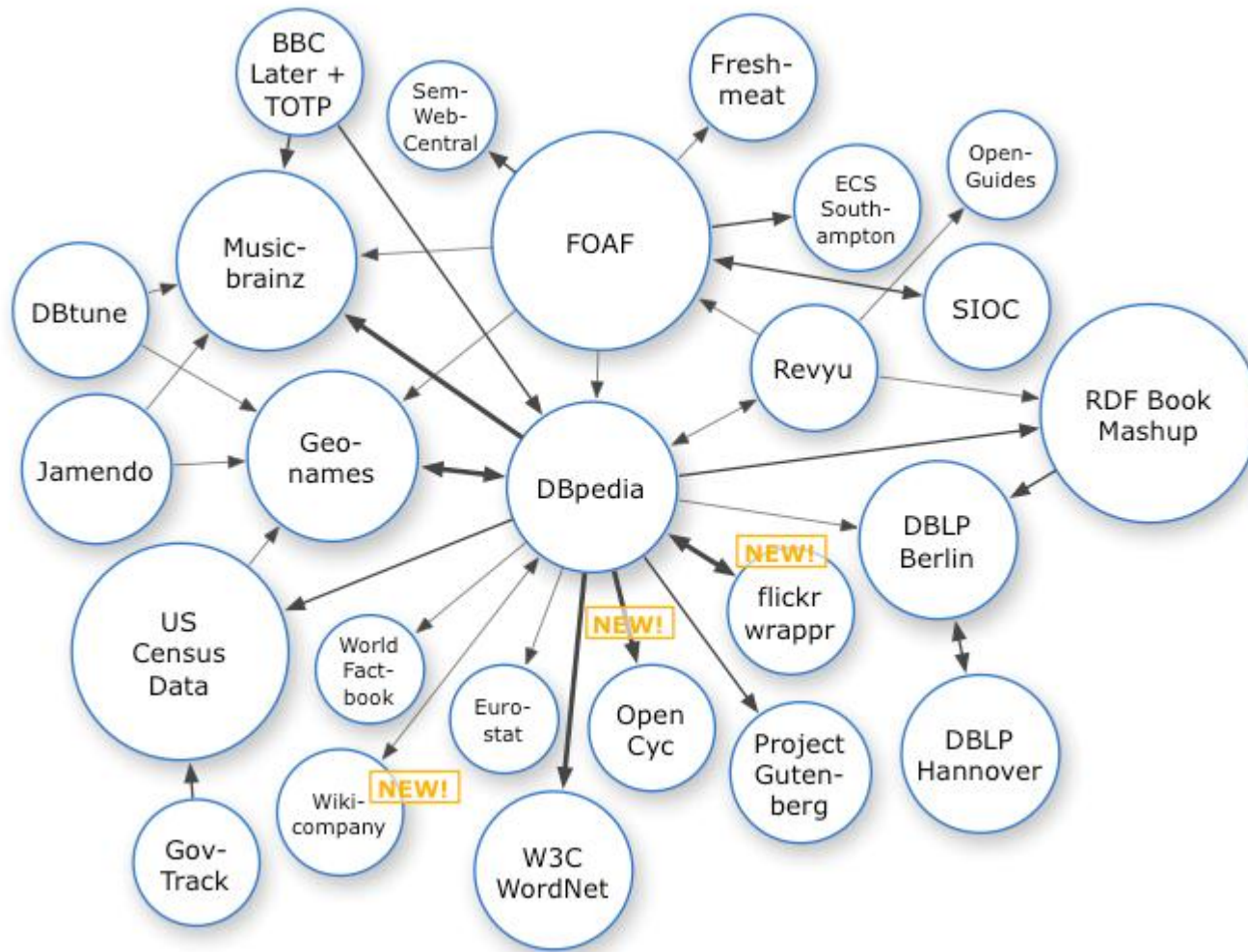
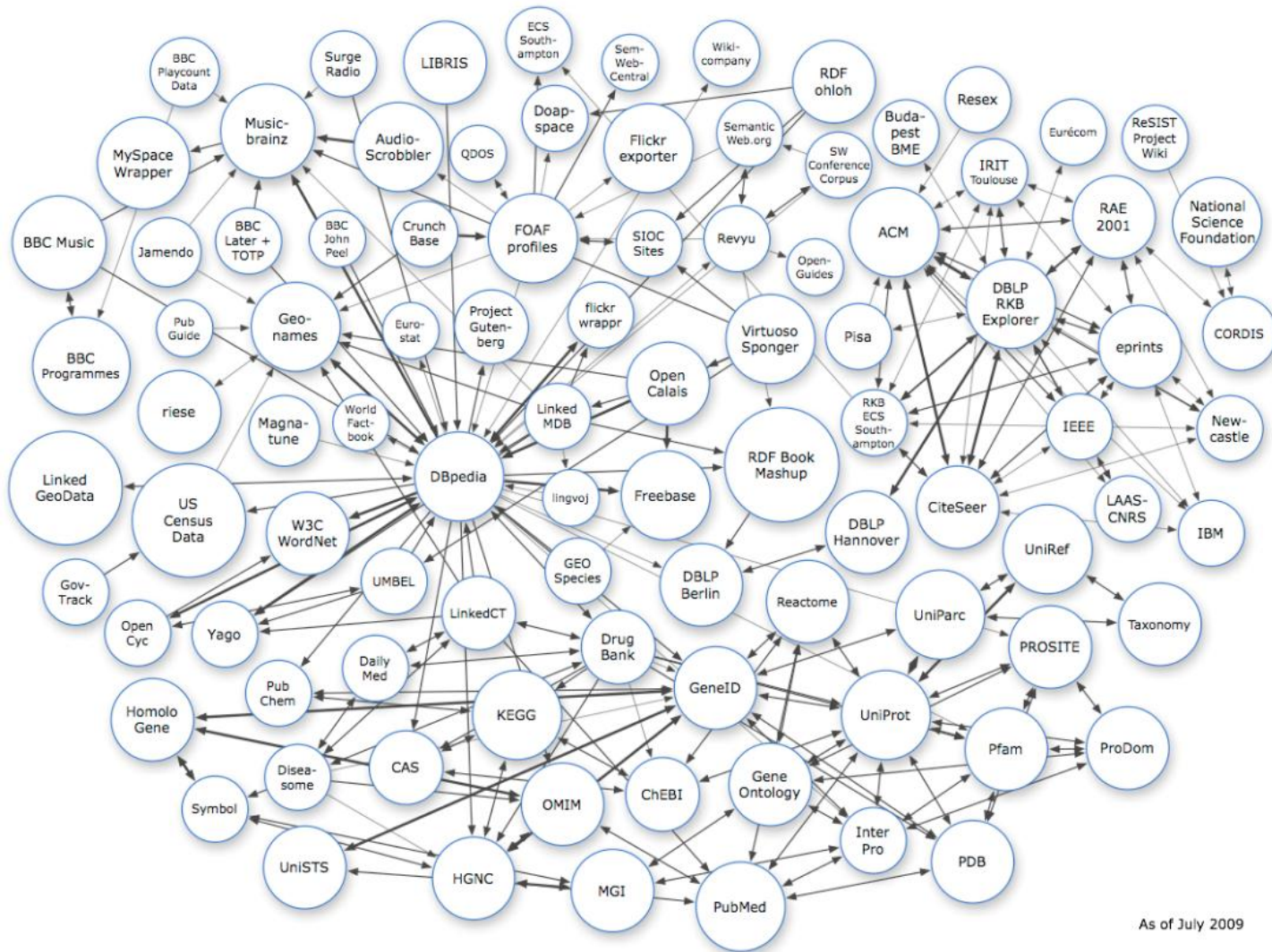Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

2008-09-18

As of September 2008

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

Umwelt Bundes Amt
Für Mensch und Umwelt

innoQ

2009-07-14

As of July 2009

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

Umwelt Bundes Amt
Für Mensch und Umwelt

innoQ

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

Sub-Clouds: Linked Open Drug Data

http://esw.w3.org/HCLSIG/LODD
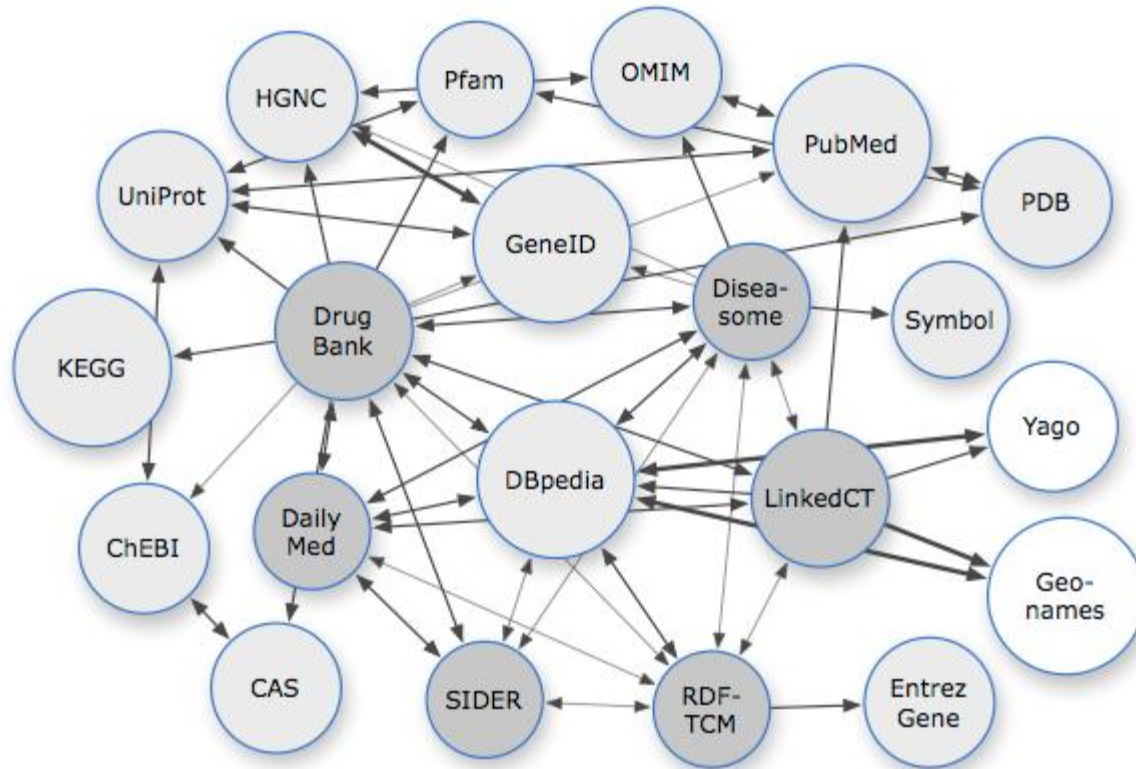
European Environment Agency in the Cloud

EUNIS biodiversity database

http://eunis.eea.europa.eu
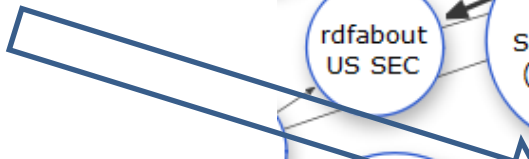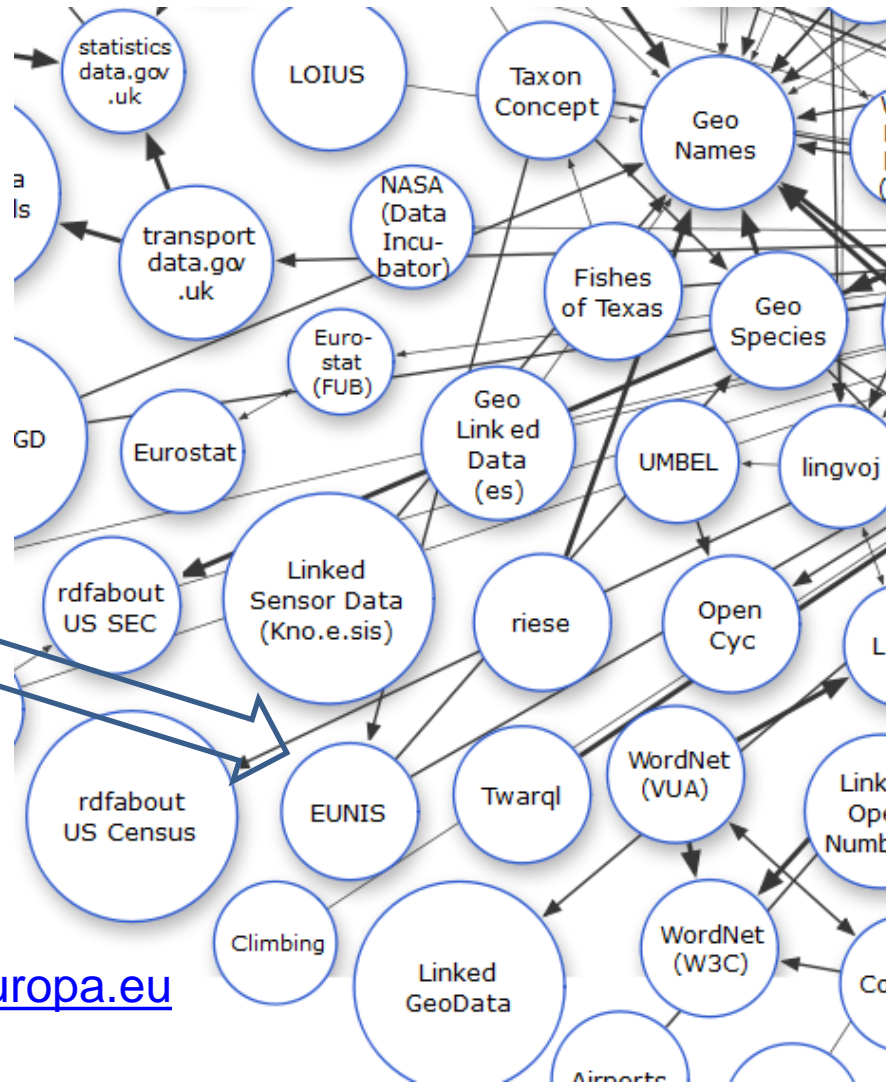
Wikipedia as a Reference Vocabulary
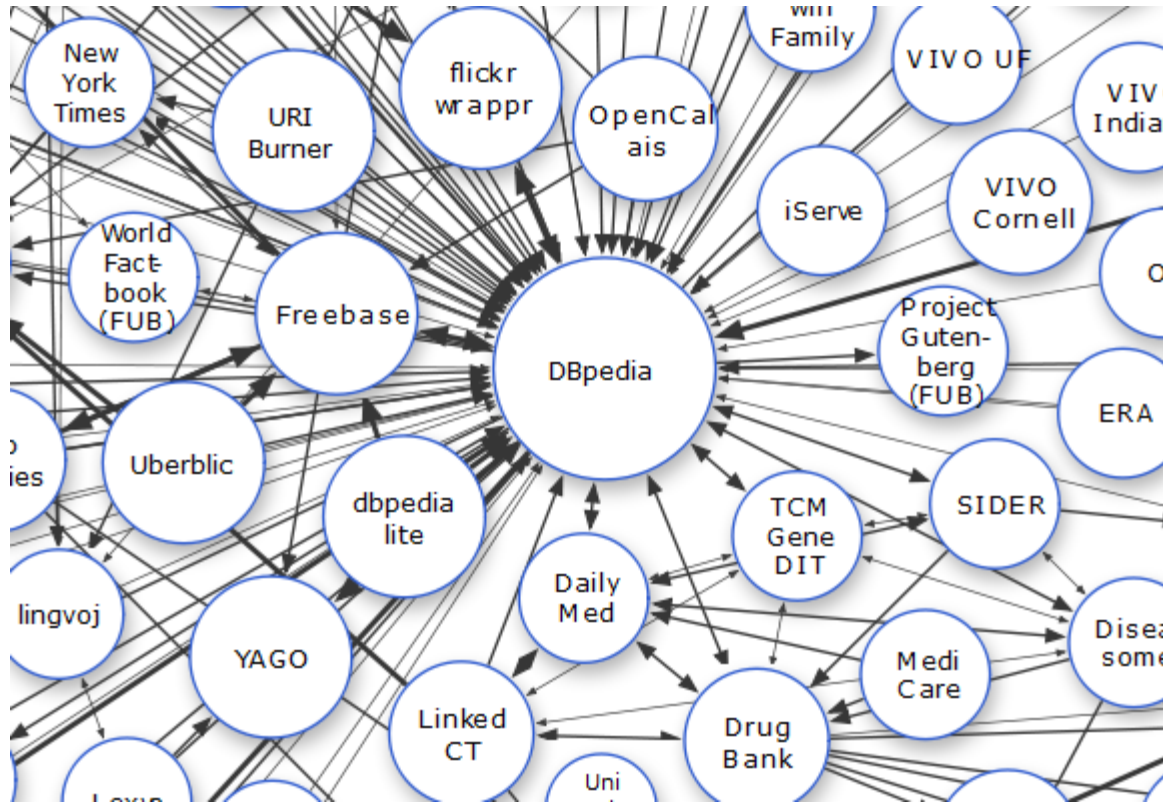
Linked Environment Data (vision@de 2010)

EEA

EUNIS

SENSE

GEMET

EARTh

AgroVoc

SNS

Location

Species

Event

Topic

Substance

UBA

more …

DioxinDB

Environment
Specimen Bank

I-ESB
Nordic Countries
Italy
UK
Spain
USA
Canada
Japan
South Corea

SENSE: Shared European National State of the Environment (EEA)
EARTh: Environmental Applications Reference Thesaurus (CNR)

Um
Bun
Amt
Für Mensch und Umwelt

innoQ
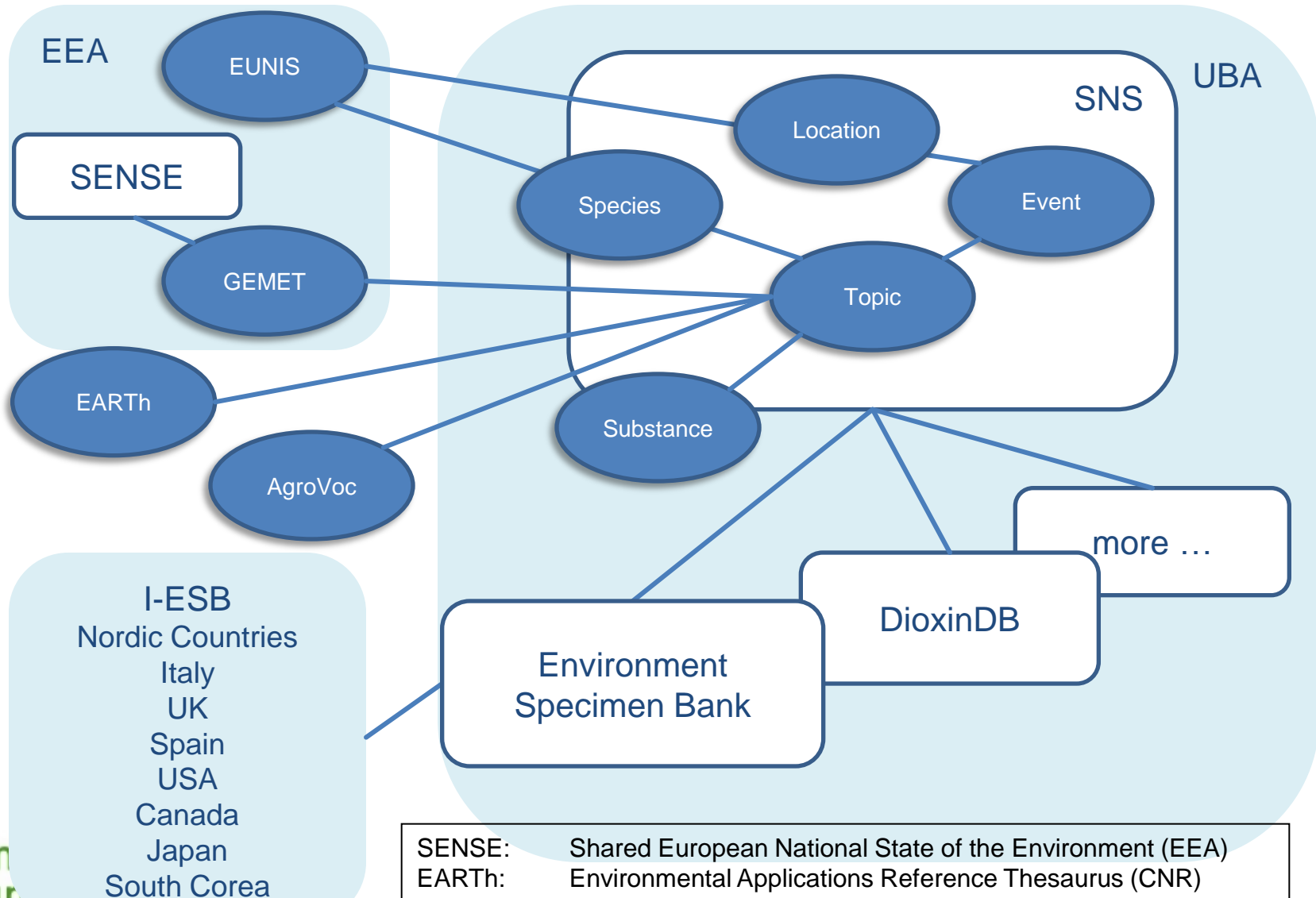
Short LED History

2009-03: The Prague eTerminology Workshop

http://www.e-envi2009.org/SummaryTerminologyW4.pdf

2009-10: 5th Ecoterm meeting at FAO, Rome

http://eea.eionet.europa.eu/Public/irc/envirowindows/jad/library?l=/ecoinformatics_indicator/ecoterm_5-6102009

2010:

- **GEMET** and **EUNIS** published as Linked Data by EEA

- **German Environment Specimen Bank** and **UMTHES** Thesaurus in test environment

- **EARTh** (CNR, Italy) in test environment

- LED at the W3C eGovernment Interest Group

  http://www.w3.org/egov/wiki/Linked_Environment_Data

- LED on Workshops and Conferences

Who is involved in Germany

- **Environment Specimen Bank**
  - holds data on exposure of humans and our environment to chemicals in different ecosystems since 1981. (de + en)
    http://www.umweltprobenbank.de

- **Semantic Network Service (SNS)**
  - Reference vocabulary service of environmental terms including common place names and events. (de + en)
    http://www.semantic-network.de

- **Species Catalog (built with iQvoc)**
  - mostly test organisms from the work of Division IV Chemical and Biological Safety

- **more to come next year**

UBA & innoQ developing iQvoc

- **open source SKOS vocabulary management tool**
  - **Web-based support for distributed editorial teams**
  - **Comfortable and safe editing with schema-based validation**
  - **Editorial workflow implementation**

- **Open Standards**
  - **W3C Linked Data**
  - **W3C SKOS Recommendation (incl. SKOS-XL)**

- **Technical platform**
  - **JRuby, JQuery, MySQL or Oracle, Virtuoso or BigOWLIM**

- **Availability**
  - **Open Source License (Apache 2.0)**
  - **iQvoc Version 1 demo (GEMET) 2008**
  - **iQvoc 2 release scheduled for end of 2010**

- **Applications: UMTHES, Species, GS-Soil-Thesaurus**

Umwelt Bundes Amt
Für Mensch und Umwelt

innoQ

Involved RDF Vocabularies

- **SKOS(XL) W3C rec** – thesauri, classifications
- **Dublin Core** – general metadata
- **Geonames** – for gazetteers
- **Linked Events Ontology** – for the chronicle
- **Darwin Core** – for species
- **SCOVO** – statistical and measuring data

- + local extensions

- SKOS is the only stable standard among them

About SPARQL endpoints

Exposing SPARQL endpoints on the Web means:

- unlimited query possibilities, but as well:

- invitation to **D**enial **o**f **S**ervice attacks
- (maybe by mistake)
- Would you expose an unlimited SQL SELECT endpoint?

- We need some authentication and trust mechanisms
  - FOAF+SSL vs. „walled gardens", WebID vs. OpenID
- + standard paging mechanisms of SPARQL endpoint implementations

Thank you for the attention!

Maria Rüther
Umweltbundesamt
Wörlitzer Platz 1, D-06844 Dessau-Roßlau
maria.ruether@uba.de

Thomas Bandholtz, Till Schulte-Coerne
innoQ Deutschland GmbH
Halskestr. 17, D-40880 Ratingen
thomas.bandholtz@innoq.com
till.schulte-coerne@innoq.com

What does Linked Data mean?

Linked Data Principles

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things.

Tim Berners-Lee, 2006-07-27
  http://www.w3.org/DesignIssues/LinkedData.html

Umwelt
Bundes
Amt
Für Mensch und Umwelt

inno Q

Wait – wasn't it „Linking *Open* Data"?

Linked Data is not necessarily „open".

- 2006: **Linked Data** Design Issues by Tim Berners-Lee, „open" not mentioned
http://www.w3.org/DesignIssues/LinkedData.html

- 2007: **Linking Open Data**
Linked Data + the Open Data Movement "aims at making data freely available to everyone"
http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

- 2008: **http://linkeddata.org/**
"a recommended best practice for **exposing, sharing, and connecting** pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

"Use URIs as names for things"

What kind of things do we have?

- Species, Substances, Locations, Events, Specimens, …

http://data.uba.de/samplingArea/10137
http://data.uba.de/specimen/t47098a_10220d1bc3e_4ee1
http://eunis.eea.europa.eu/species/9986

Are these literal names?

- nope, URIs are globally unique identifiers
- Standard: Uniform Resource Identifier (RFC 3986)
- literal names are mostly ambiguous
- they are provided as „labels" in the data

Umwelt
Bundes
Amt
Für Mensch und Umwelt

innoQ

„Use HTTP URIs …"

- Hypertext Transfer Protocol – the protocol of the Web
- an HTTP URI is a „Web address"
- Linked Data has a notion of „Cool URIs":

Tim Berners-Lee 1998
http://www.w3.org/Provider/Style/URI

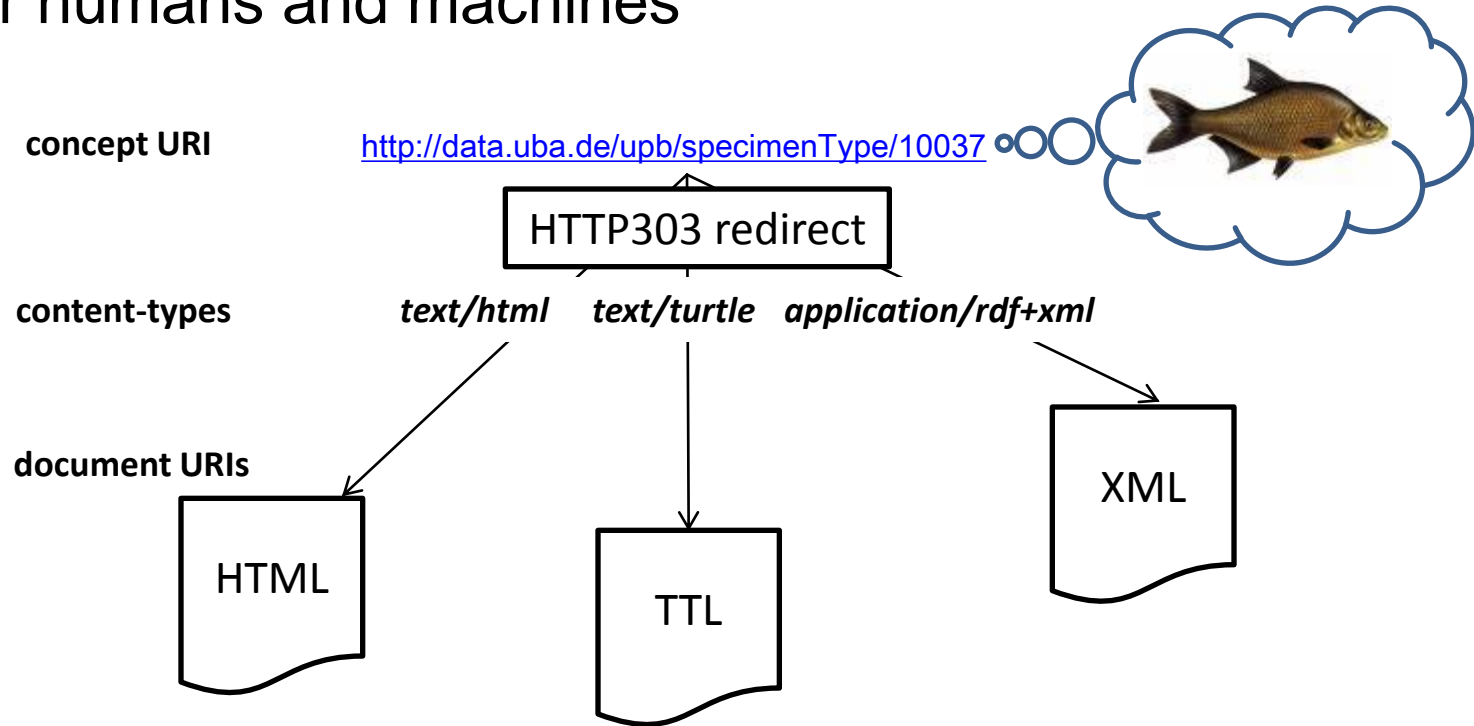## Cool URIs don't change

What makes a cool URI?
A cool URI is one which does not change.
What sorts of URI change?
URIs don't change: people change them.

„look up those names"- Content Negotiation

- for humans and machines

concept URI   http://data.uba.de/upb/specimenType/10037

HTTP303 redirect

content-types   *text/html*   *text/turtle*   *application/rdf+xml*

document URIs

HTML   TTL   XML

- 2008: Cool URIs for the Semantic Web
- http://www.w3.org/TR/cooluris/

„provide useful information" (HTML)
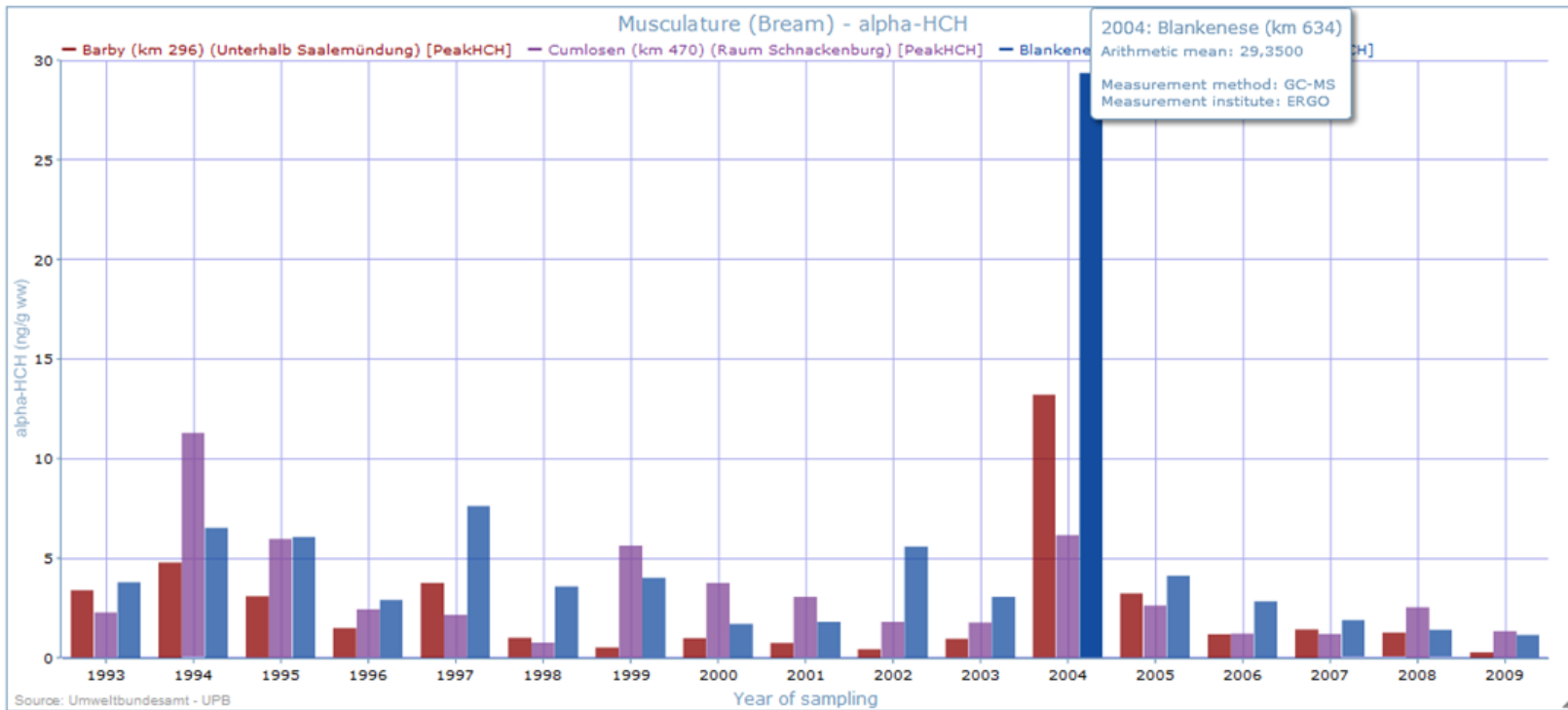
„provide useful information" (RDF)

```
esb:t47098a_10220d1bc3e_4ee1 a scv:Item ;
    rdf:value 13.2000;
    scv:dataset esb:chemicalExposure ;
    esb:specimenType esbd:10006 ;
    esb:samplingArea esbd:10137 :        # Blankenese (Unterelbe)
    esb:substance esbd:10053             # alpha-HCH ;
    esb:timeReference esbd:year2004 ;
    esb:summaryStat esbd:arithmeticMean ;
    esb:uom esbd:nggww ;
    sns:relatedEvent sns:t1d97d0d_102035cd5d4_-3384. # Elbehochwasser

esbd:10006 a esbd:SpecimenType ;
    skos:prefLabel "Brassenmuskulatur"@de ;
    skos:prefLabel "bream musculature"@en ;
    skos:broader esb:10037 .

esbd:10037 a esbd:SpecimenType ;
    skos:prefLabel "Brassen"@de ;
    dwct:scientificName "Abramis brama" ;
    owl:sameAs <http://eunis.eea.europa.eu/species/9986> .
```

„discover more things" (RDF)

```
esb:t47098a_10220d1bc3e_4ee1 a scv:Item ;
    rdf:value 13.2000;
    scv:dataset esb:chemicalExposure ;
    esb:specimenType esbd:10006 ;
    esb:samplingArea esbd:10137 :        # Blankenese (Unterelbe)
    esb:substance esbd:10053             # alpha-HCH ;
    esb:timeReference esbd:year2004 ;
    esb:summaryStat esbd:arithmeticMean ;
    esb:uom esbd:nggww ;
    sns:relatedEvent sns:t1d97d0d_102035cd5d4_-3384. # Elbehochwasser

esbd:10006 a esbd:SpecimenType ;
    skos:prefLabel "Brassenmuskulatur"@de ;
    skos:prefLabel "bream musculature"@en ;
    skos:broader esb:10037 .

esbd:10037 a esbd:SpecimenType ;
    skos:prefLabel "Brassen"@de ;
    dwct:scientificName "Abramis brama" ;
    owl:sameAs <http://eunis.eea.europa.eu/species/9986> .
```

„discover more things" (by SPARQL)

SPARQL **P**rotocol **a**nd **R**DF **Q**uery **L**anguage

time sequence of alpha-HCH in bream musculature from Blankenese.

```
SELECT DISTINCT ?year ?val WHERE {
  ?mes scv:dataset esb:chemicalExposure;
       esb:specimenType esbd:10006;
       esb:samplingArea esbd:10137;
       esb:substance esbd:10053;
       esb:summaryStat esbd:arithmeticMean;
       rdf:value ?val;
       esb:timeReference ?year.
}
```

Results in   …
                  esbd:year2005   04,12
                  esbd:year2004   29,35
                  esbd:year2003   03,46
                  …

Umwelt
Bundes
Amt
Für Mensch und Umwelt

innoQ

Technical LED Architecture in the UBA

- **Each contributing information system needs:**
  - Web (HTML) presentation of each „thing"
  - RDF rendering of the same data
  - Synchronize with an RDF database
- **One for all: Linked Data aware RDF DB (based on BigOWLIM)**
  - Content Negotiation (does not come out of the box!)
  - RDF representation in selectable RDF/XML syntax
  - SPARQL endpoint
- **Metadata**
  - Vocabulary of Interlinked Datasets (VOID) ?
    http://vocab.deri.ie/void/guide
  - Data Catalogue Vocabulary (DCAT) ?
    http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary
  - Comprehensive Knowledge Archive Network (CKAN) ?
    http://www.ckan.net/

Umwelt
Bundes
Amt
Für Mensch und Umwelt

innoQ

Summary & Outlook

Linked Data provides simple patterns for

- publishing raw data together with the Web presentation
- integration of heterogeneous databases by crosslinking
- referencing standard vocabularies from the data

Global W3C standards allow for

- open access with standard tools
- borderless integration inside & outside the agency
- unrestricted queries for everyone

Ongoing work

- first UBA contributions public by end of the year
- more UBA databases to follow
- international cross-agency networking
- join the network with *your* dataset!

Umwelt Bundes Amt
Für Mensch und Umwelt

innoQ