University of Reading

Earth Observation and the Web of Data - how can we join the dots?

Jon Blower
University of Reading

EO Science 2.0
ESA ESRIN, Frascati

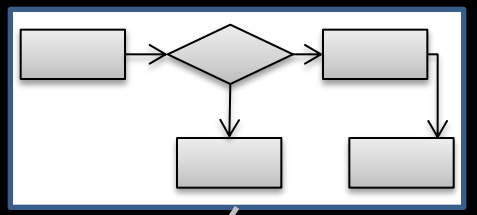
Introduction

- The Web is a huge VRE that we take for granted
  - **Decentralised publishing**
  - Low barrier to entry for users and publishers
  - Papers, people, data, code …
  - But little structure or machine readability
  - Organised chaos!
- Geospatial standards and EO infrastructures don't use the Web to best advantage
- How do we make better use of the Web in the EO community? What could we gain?

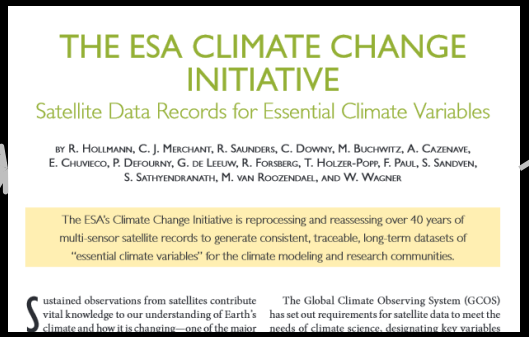
platform

instrument

algorithm



scientists



THE ESA CLIMATE CHANGE INITIATIVE

Satellite Data Records for Essential Climate Variables

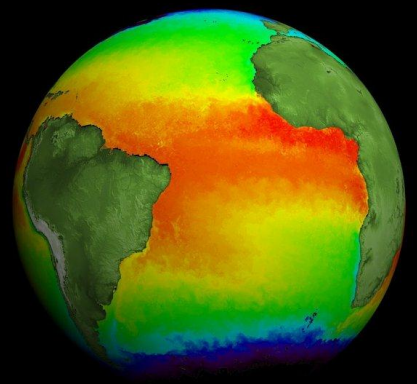
BY R. HOLLMANN, C. J. MERCHANT, R. SAUNDERS, C. DOWNY, M. BUCHWITZ, A. CAZENAVE, E. CHUVIECO, P. DEFOURNY, G. DE LEEUW, R. FORSBERG, T. HOLZER-POPP, F. PAUL, S. SANDVEN, S. SATHYENDRANATH, M. VAN ROOZENDAEL, AND W. WAGNER

The ESA's Climate Change Initiative is reprocessing and reassessing over 40 years of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.

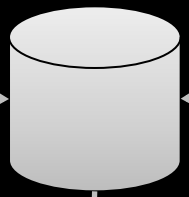
Sustained observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—one of the major

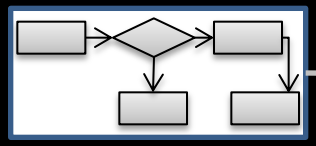
The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, designating key variables

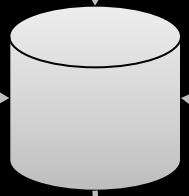
publication

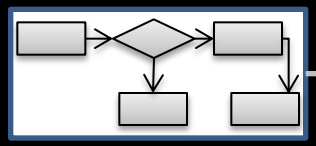


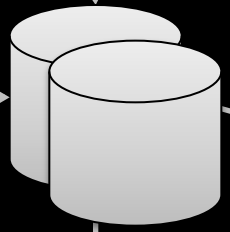
dataset

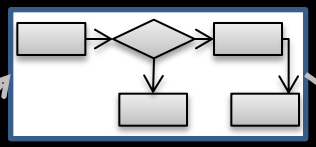


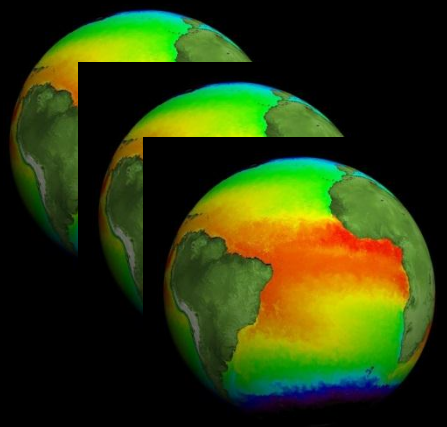




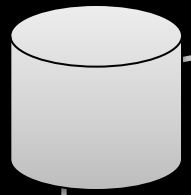


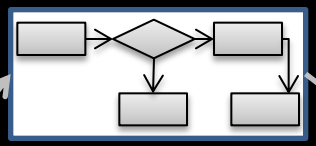


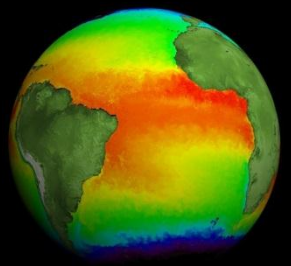





THE ESA CLIMATE CHANGE INITIATIVE
Satellite Data Records for Essential Climate Variables
BY R. HOLLMANN, C. J. MERCHANT, R. SAUNDERS, C. DOWNY, M. BUCHWITZ, A. CAZENAVE, E. CHUVIECO, P. DEFOURNY, G. DE LEEUW, R. FORSBERG, T. HOLZER-POPP, F. PAUL, S. SANDVEN, S. SATHYENDRANATH, M. VAN ROOZENDAEL, AND W. WAGNER
The ESA's Climate Change Initiative is reprocessing and reassessing over 40 years of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.
Sustained observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—one of the major
The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, designating key variables







THE ESA CLIMATE CHANGE INITIATIVE

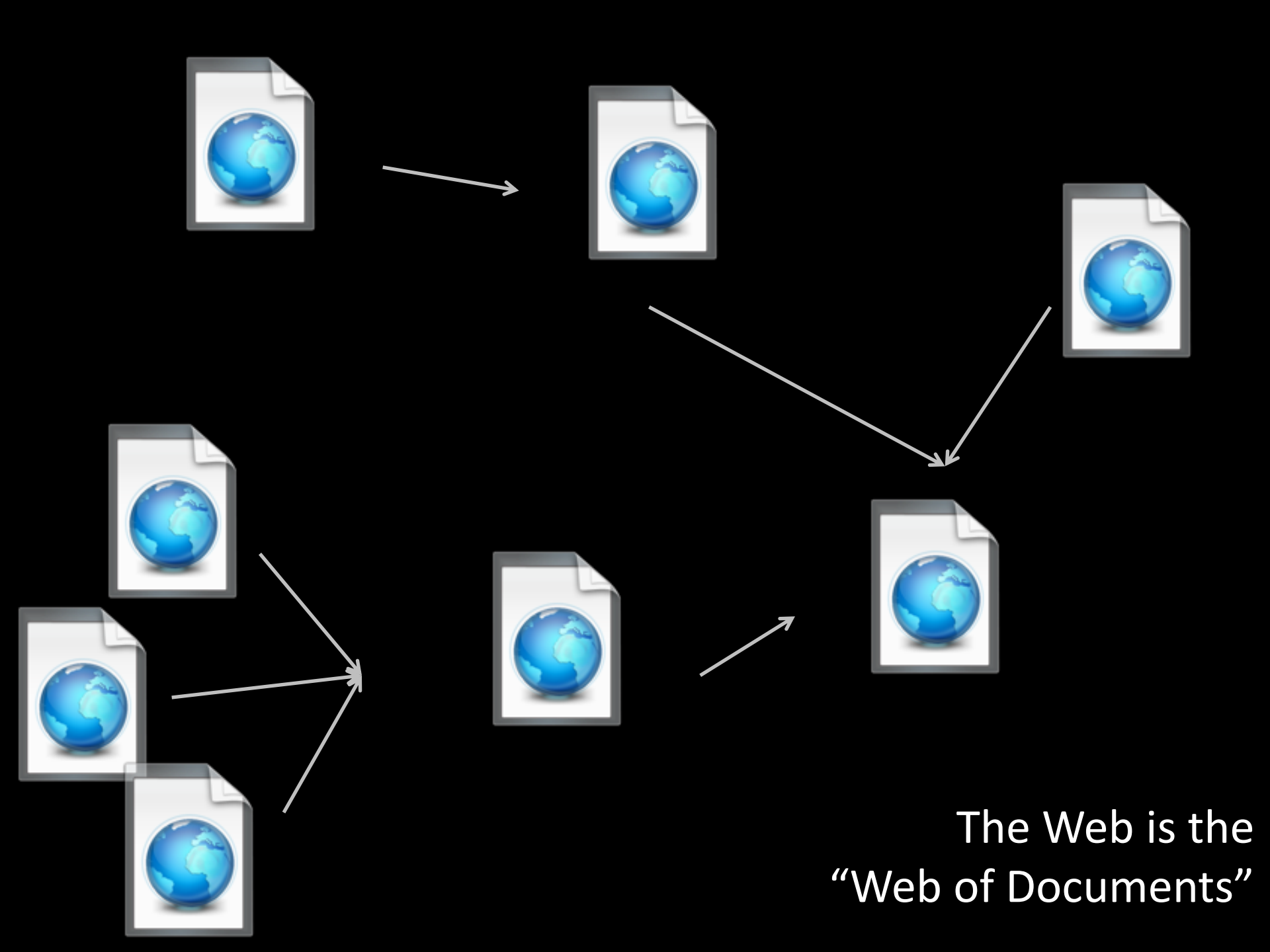
Satellite Data Records for Essential Climate Variables

BY R. HOLLMANN, C. J. MERCHANT, R. SAUNDERS, C. DOWNY, M. BUCHWITZ, A. CAZENAVE, E. CHUVIECO, P. DEFOURNY, G. DE LEEUW, R. FORSBERG, T. HOLZER-POPP, F. PAUL, S. SANDVEN, S. SATHYENDRANATH, M. VAN ROOZENDAEL, AND W. WAGNER

The ESA's Climate Change Initiative is reprocessing and reassessing over 40 years of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.

Sustained observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—one of the major

The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, designating key variables


The Web is the
“Web of Documents”

"If … the Web made all the online documents look like one huge book, [**Linked Data**] will make all the data in the world look like one huge database."

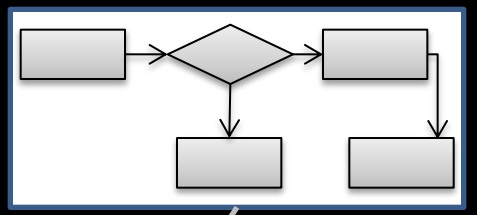
Sir Tim Berners-Lee

Photo by Susan Lesch, from Wikipedia

Linked Data principles

- Give things unique and persistent identifiers
  - Datasets, sensors, algorithms, publications, variables …
- Allow the identifier to be “looked up” on the web
  - i.e. the identifier should be an HTTP URL
  - e.g. http://dbpedia.org/resource/Frascati
- Provide a description of the thing in a standard format
  - Human readable (HTML)
  - Machine readable (RDF), using agreed vocabularies
- Use this description to link to other related things
  - And say *why* they are linked

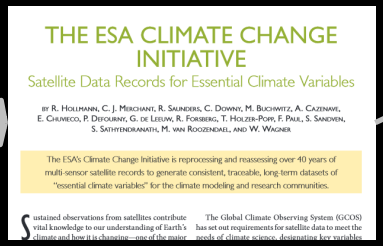
carries



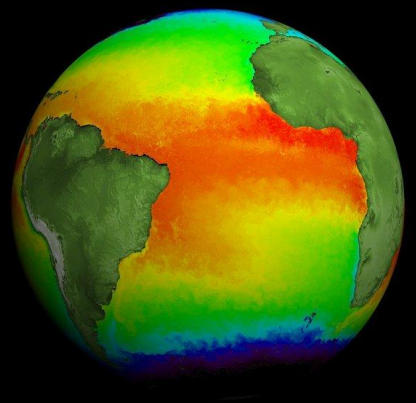
produced

produced

authored



describes /
uses /
queries …

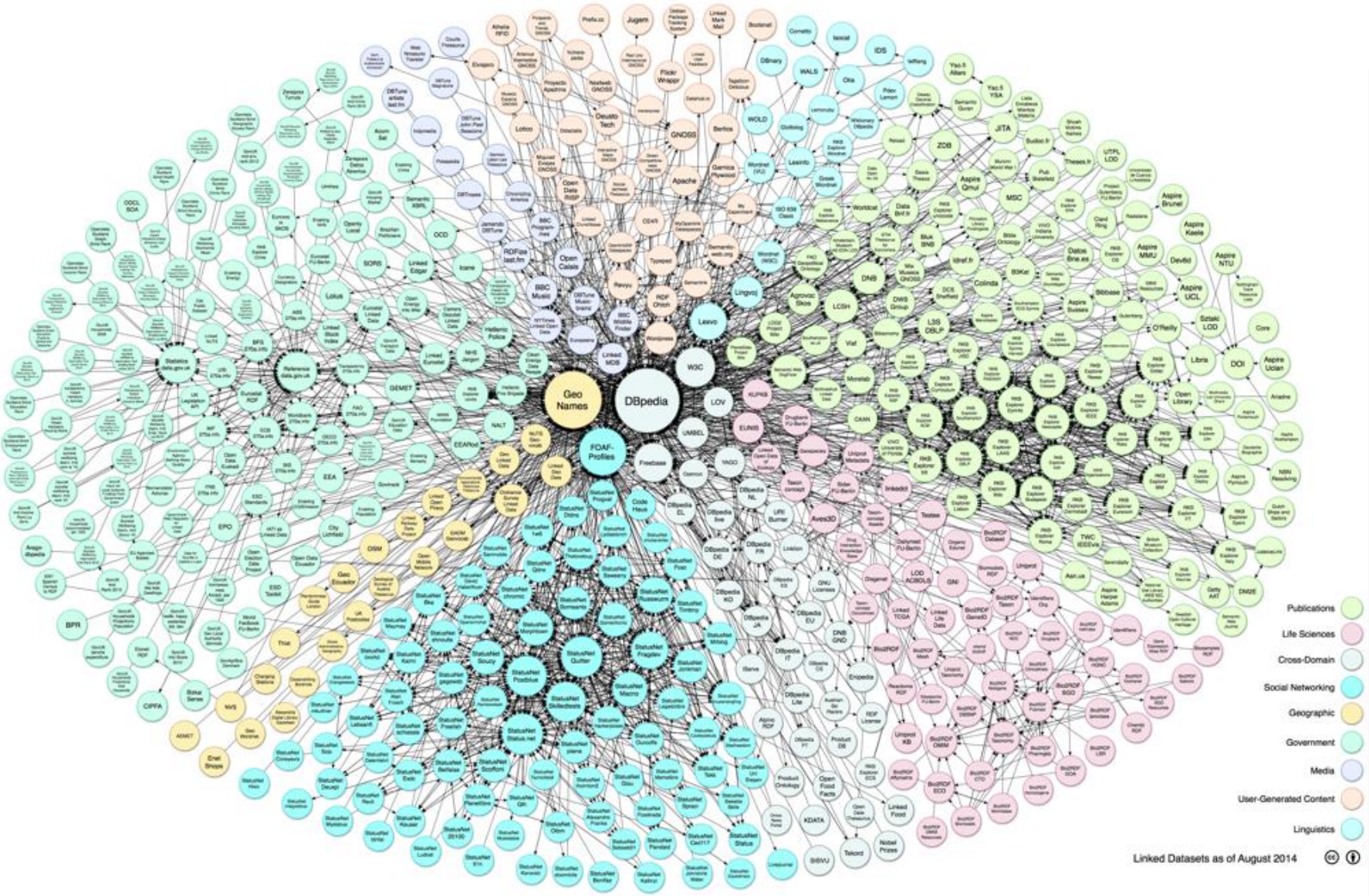


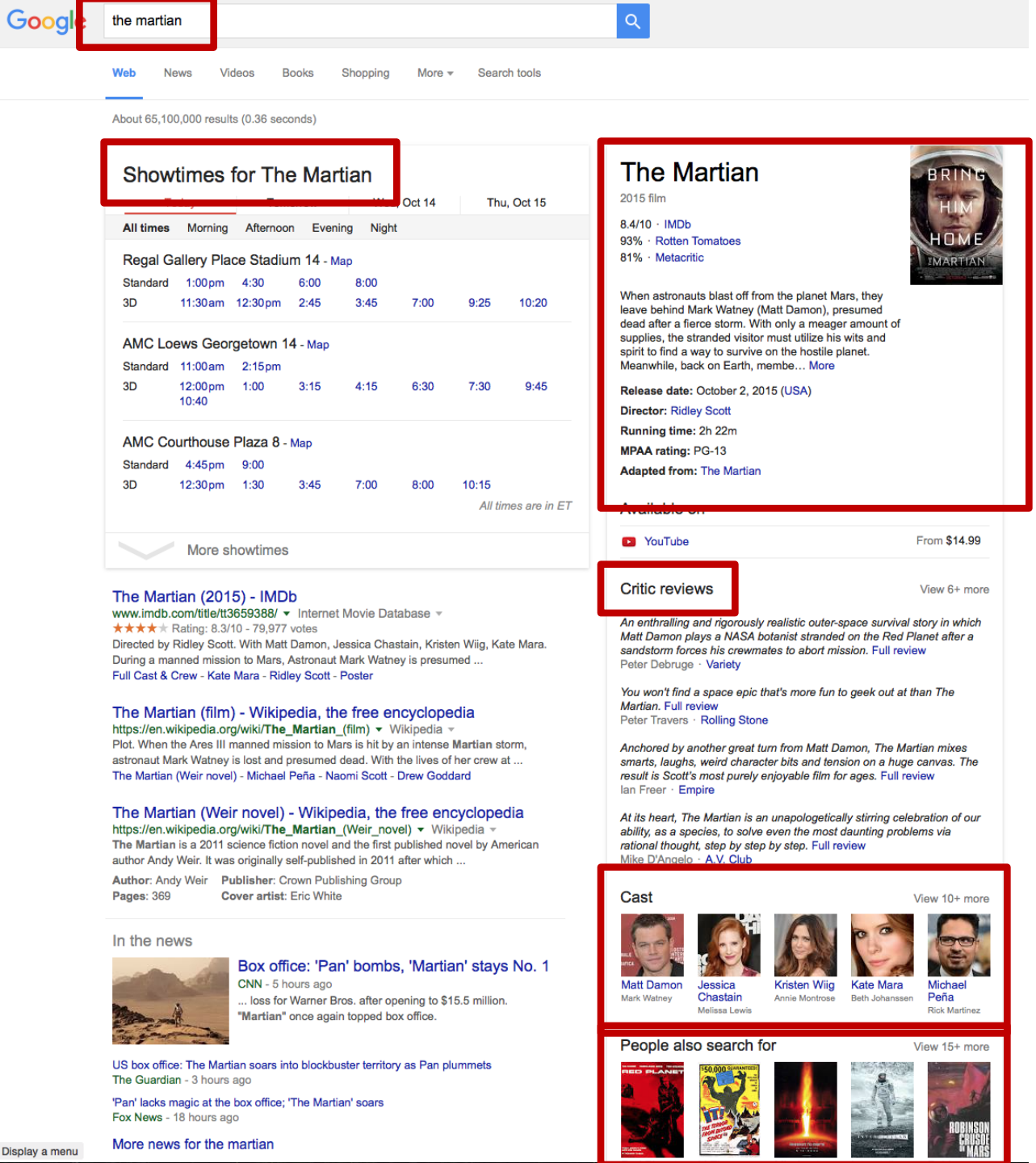
Linked Data techniques build the “Web of Data”

Linked Data standards
(community-independent)

- Data encoded in RDF **structure**
  - Resource Description Format
- **Semantics** encoded using agreed **vocabularies** of defined terms
- Data **querying** in SPARQL
  - RDF Query Language
  - cf. SQL
- Data can be **stored** in many ways
  - Native RDF stores (“triplestores”)
  - Relational databases
  - …

Linked Open Data Cloud (lod-cloud.net)





Search for “The Martian”, Google shows:

- Facts about the film
- Cast
- Showtimes
- Reviews
- Related films

This is Linked Data in action!

(powered by schema.org vocabulary)

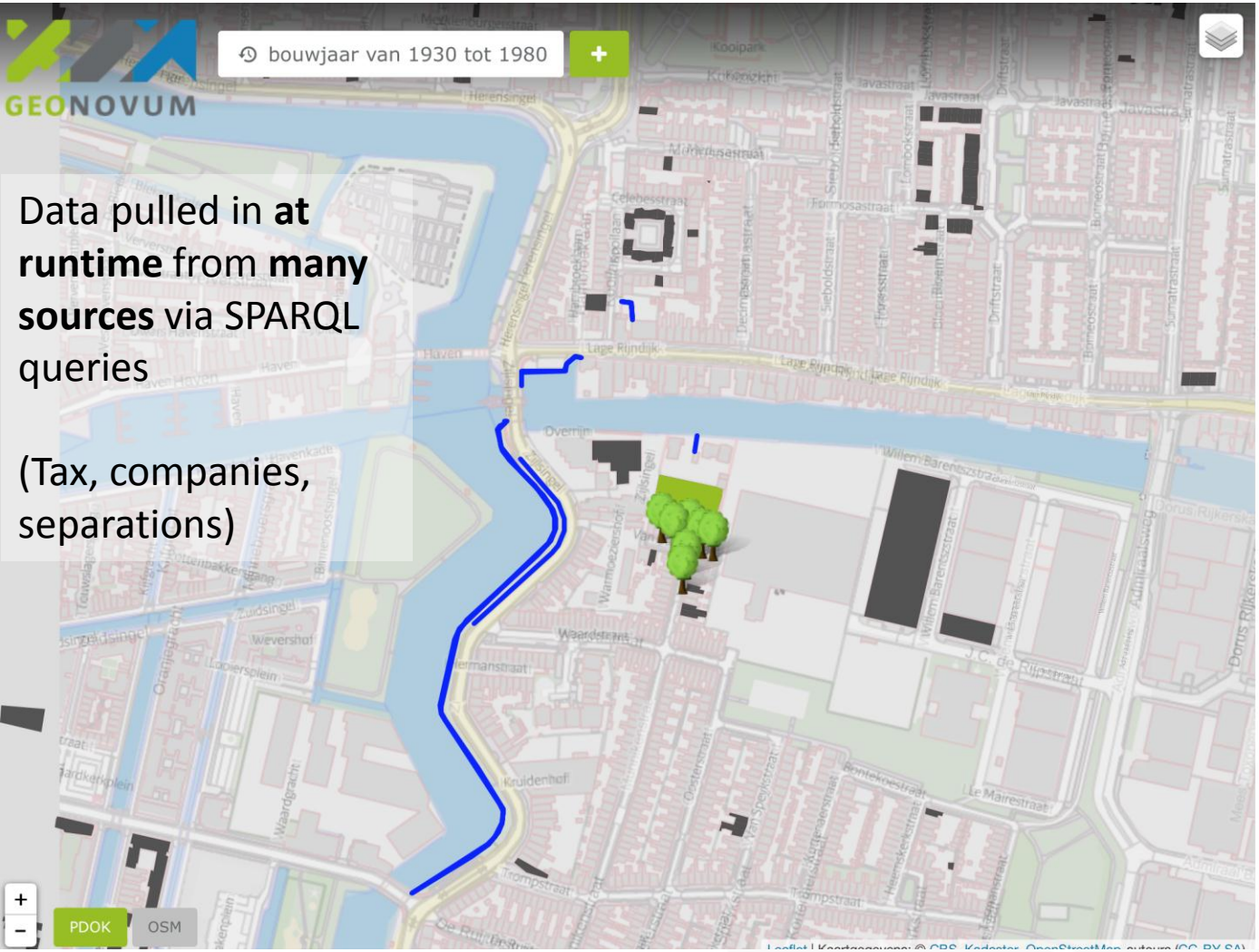
Why can’t we do the same for EO data?

What can we do with Linked Data?

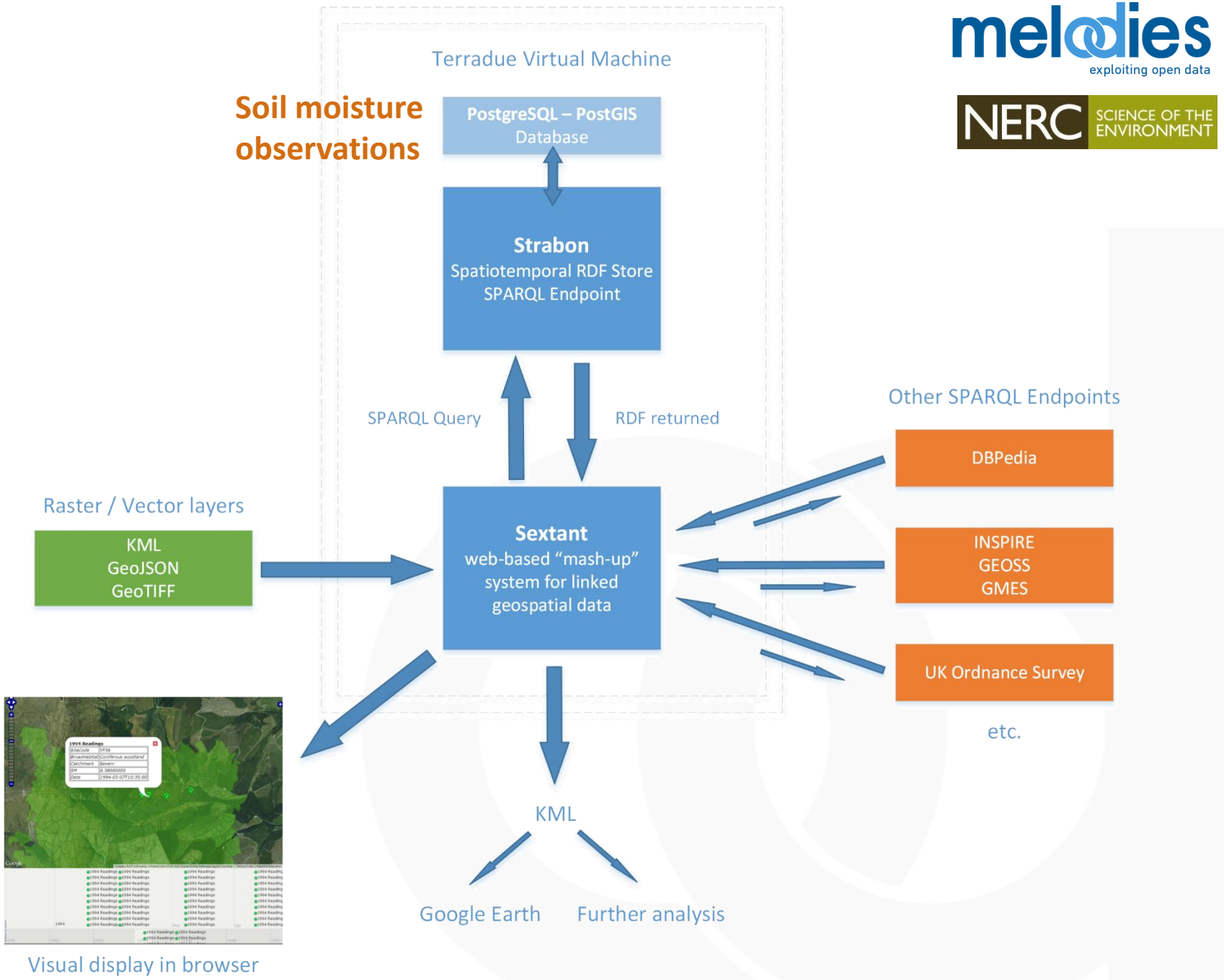
- Linking between datasets, instruments, algorithms
- Describing data more accurately (common vocabularies)
- Recording dataset provenance
- **Combining data from different sources**

Database of Dutch buildings (Geonovum)

http://almere.pilod.nl/bgtld/v2



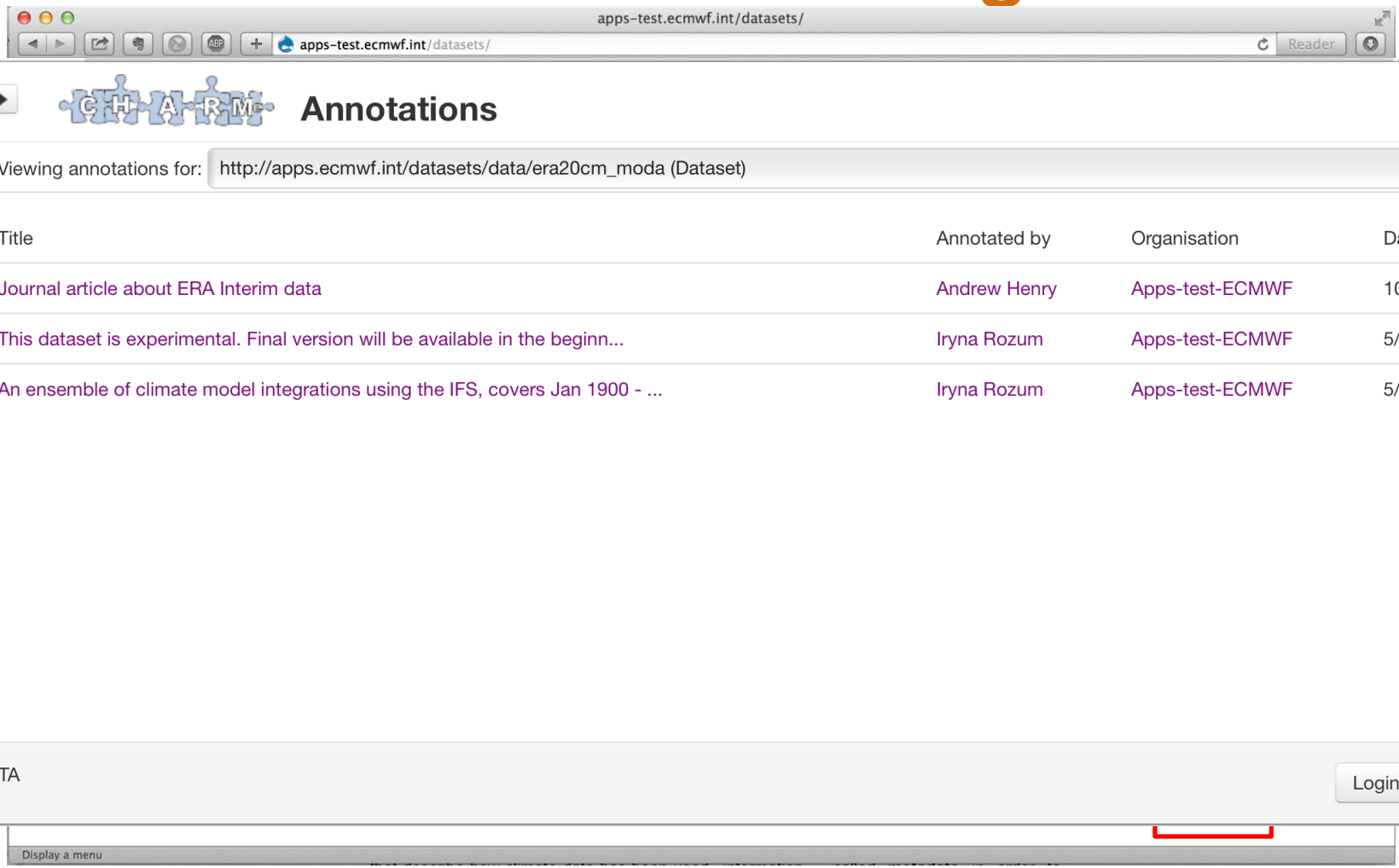
(thanks to Linda van den Brink)


Terradue Virtual Machine
Soil moisture observations
PostgreSQL – PostGIS
Database
Strabon
Spatiotemporal RDF Store
SPARQL Endpoint
SPARQL Query
RDF returned
Other SPARQL Endpoints
DBPedia
INSPIRE
GEOSS
GMES
UK Ordnance Survey
etc.
Raster / Vector layers
KML
GeoJSON
GeoTIFF
Sextant
web-based "mash-up" system for linked geospatial data
KML
Google Earth
Further analysis
Visual display in browser

What can we do with Linked Data?

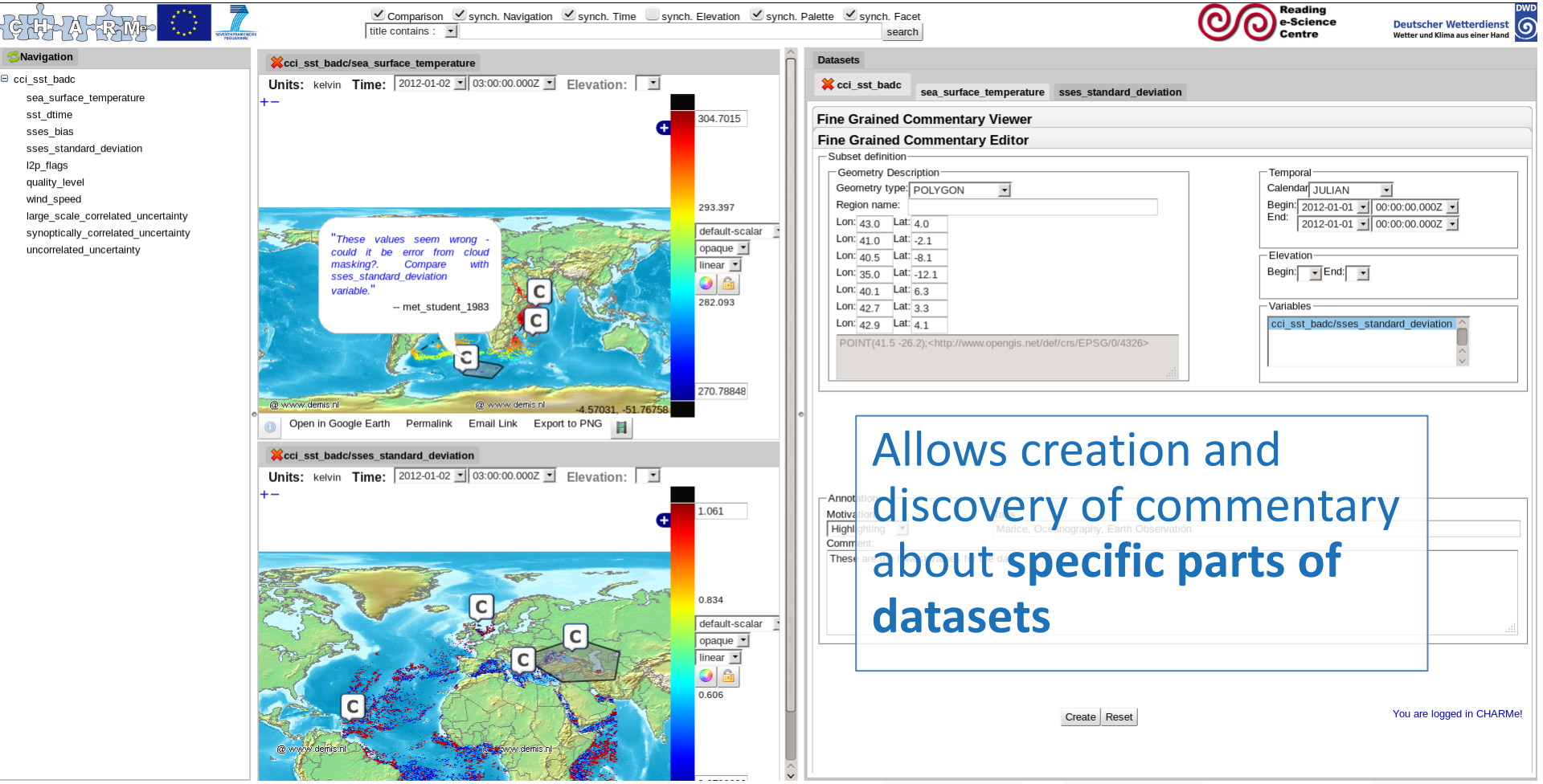
- Linking between datasets, instruments, algorithms
- Describing data more accurately (common vocabularies)
- Recording dataset provenance
- Dynamically combining data from different sources
- **User feedback and annotation**

User feedback on data using CHARMe



- http://www.charme.org.uk

“CHARMe Maps” tool



Allows creation and discovery of commentary about **specific parts of datasets**

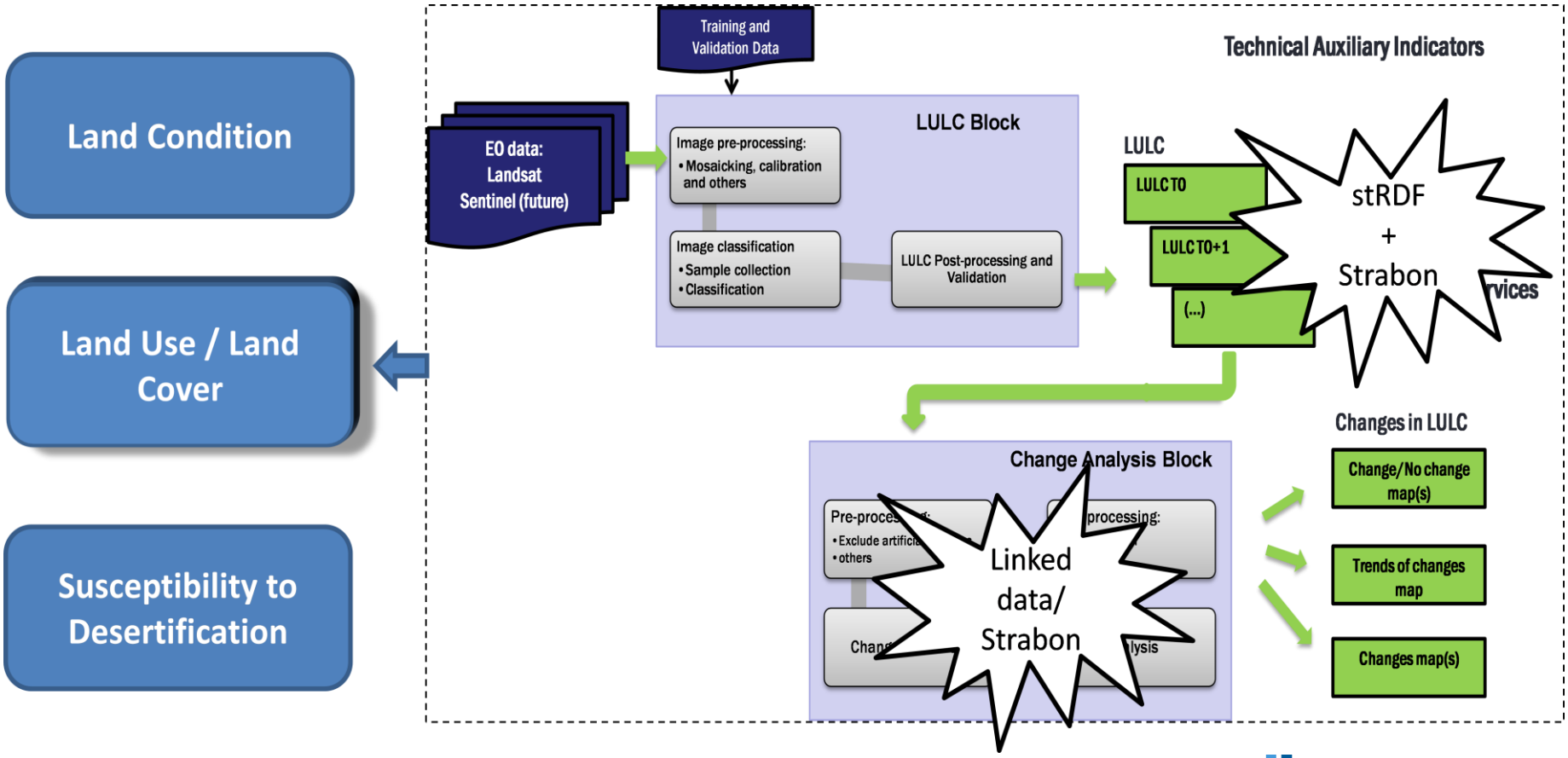
What can we do with Linked Data?

- Linking between datasets, instruments, algorithms
- Describing data more accurately (common vocabularies)
- Recording dataset provenance
- Dynamically combining data from different sources
- User feedback and annotation
- **Generating custom data products**

Enhancing land use change products

Part of previous processing chain replaced by Linked Data queries using SPARQL

- increased flexibility
- enables rapid experimentation and development



What can we do with Linked Data?

- Linking between datasets, instruments, algorithms
- Describing data more accurately (common vocabularies)
- Recording dataset provenance
- Dynamically combining data from different sources
- User feedback and annotation
- Generating custom data products
- **Enabling data discovery**
  - Follow-your-nose
  - mass-market search engines (Google screenshot)

Structured Data Testing Tool

Fetch URL Examples

```
1  <script type="app
2  {
3   "@context": "htt
4   "@type": "Datase
5   "@id": "http://w
6   "license": "http
7   "spatial": {
8     "geo": {
9       "box": "50 4
10    }
11  },
12  "datasetTimeInte
13  "about": {
14    "@type": "Thin
15    "@id": "http:/
16    "name": "Sea I
17  },
18  "datePublished":
19  "publisher": {
20      "@type": "Or
21    "name": "MELOD
22  },
23  "keywords": "sea
24  "hasPart": [
25    {
26      "@type": "Da
27      "@id": "http
28    },
29    {
30      "@type": "Dataset",
31      "@id": "http://www.example.com/datasets/1/2"
32    }
33  ]
34  }
35  </script>
```

Dataset (1)

1 Error

Dataset:	***http://www.example.com/datasets/1***
license:	https://creativecommons.org/licenses/by/4.0/
datasetTimeInterval:	2007-03-01
datePublished:	2010-01-01
keywords:	sea, ice
spatial [Place]:	
geo [StructuredValue]:	
box:	50 4 52 6
about [Thing]:	*http://www.example.com/sea_ice_fraction*
name:	Sea Ice Fraction
publisher [Organization]:	
name:	MELODIES
hasPart [Dataset]:	*http://www.example.com/datasets/1/1*
hasPart [Dataset]:	*http://www.example.com/datasets/1/2*

Google knows what a Dataset is!

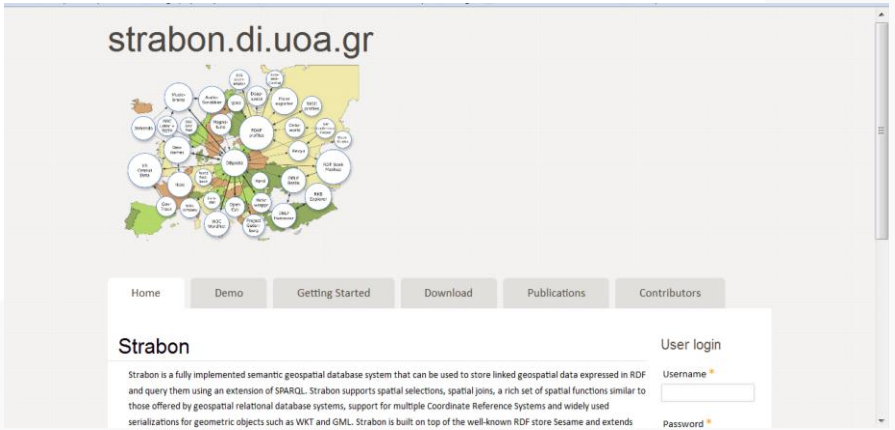
(again, uses schema.org vocab)

(Some of many) challenges

- Finding suitable vocabularies
  - Don’t invent your own unless you have to!
- **Handling geospatial data**

Geospatial Linked Data tools

- Strabon
  - spatiotemporal RDF store
  - GeoSPARQL / stSPARQL support
- GeoTriples
  - Convert geospatial data to RDF
- Ontop-spatial
  - “Wrap” existing geo-databases
- SILK
  - Discover links in datasets



HELLENIC REPUBLIC
National and Kapodistrian University of Athens

LEO

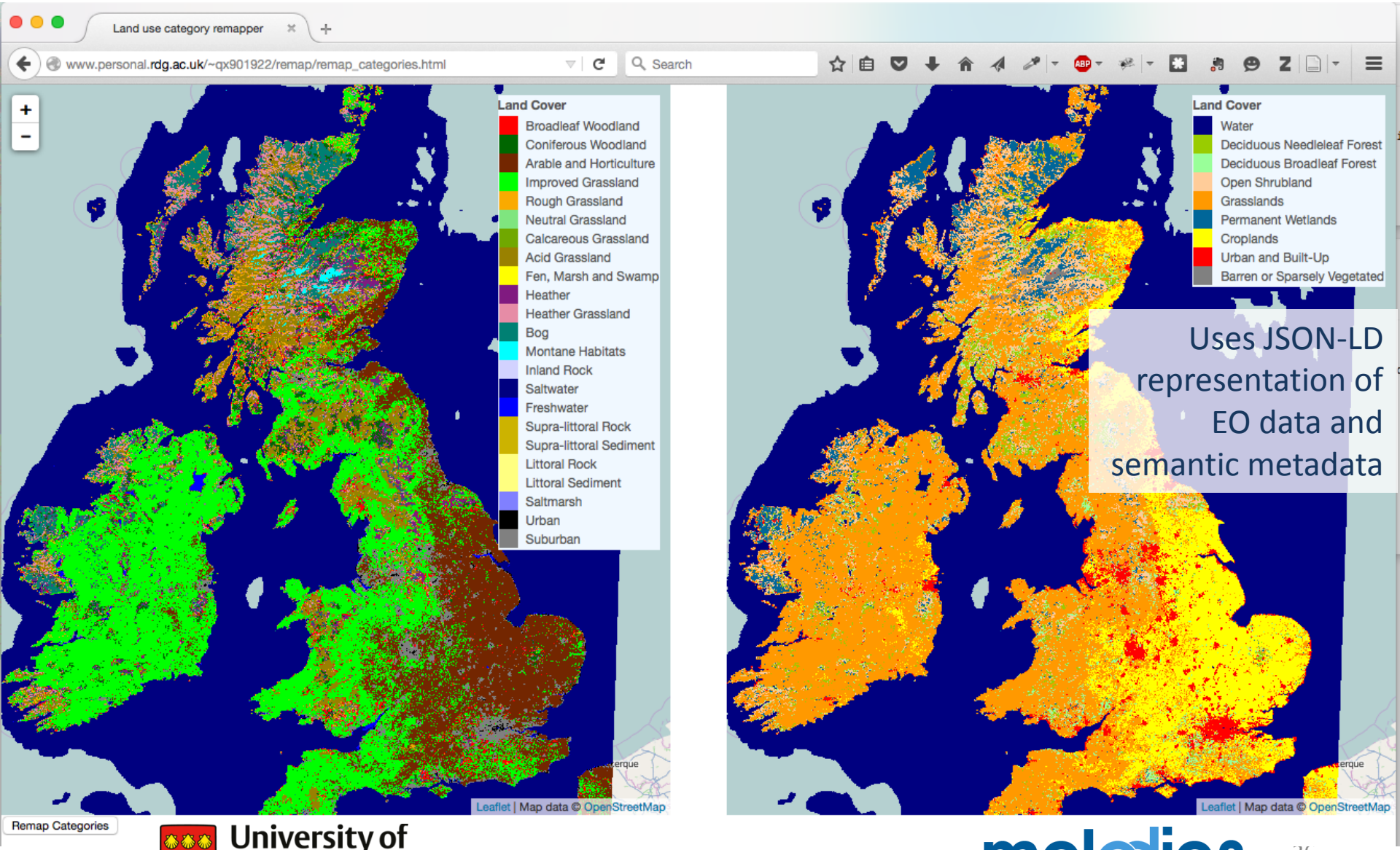
TELEIOS

melodies
exploiting open data

(Some of many) challenges

- Finding suitable vocabularies
  - Don’t invent your own unless you have to!
- Handling geospatial data
- **Handling big data** (data cubes)
  - Bridging between RDF and multidimensional arrays
  - (JSON-LD quite a good compromise format)

Interactive, in-browser mapping between land cover categories



Uses JSON-LD representation of EO data and semantic metadata

University of Reading

meldies exploiting open data

Conclusions

- Linked Data makes information **part of the web**, not just **on the web**
- Retains decentralised publication
  - “Small pieces, loosely coupled”
  - You don’t have to own everything!
- Rapidly maturing area, entering mainstream
- Breaks out of community-specific silos and architectures

- Good first actions:
  - Generate unique, persistent and resolvable identifiers for your “things”
  - Look for projects and systems that provide these identifiers already
  - Use microformats to embed semantic information in your website

melodies
exploiting open data

Thank you!

j.d.blower@reading.ac.uk
@Jon_Blower

http://melodiesproject.eu
@MelodiesProject

European Commission