

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?

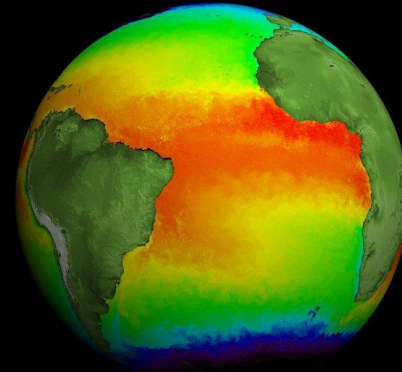
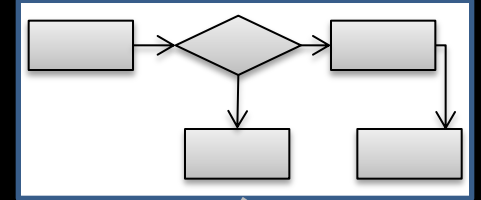


platform

instrument



algorithm



dataset

THE ESA CLIMATE CHANGE INITIATIVE
 Satellite Data Records for Essential Climate Variables

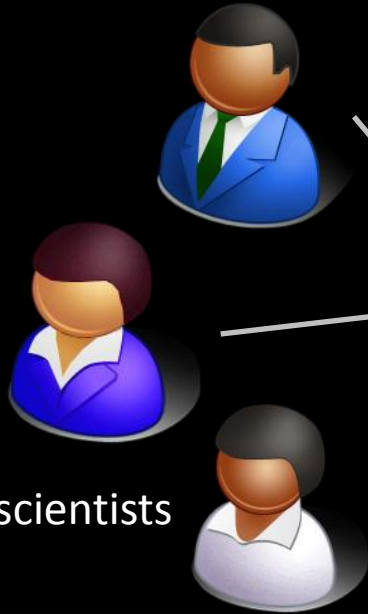
BY R. HOLLMANN, C. J. MERCHANT, R. SAUNDERS, C. DOWNY, M. BUCHWITZ, A. CAZENAVE,
 E. CHUVIECO, P. DEFOURNY, G. DE LEEUW, R. FORSBERG, T. HOLZER-POPP, F. PAUL, S. SANDVEN,
 S. SATHYENDRANATH, M. VAN ROOZENDAEL, AND W. WAGNER

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.

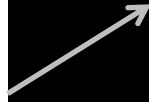
Sustained observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—one of the major needs of climate science, designating key variables

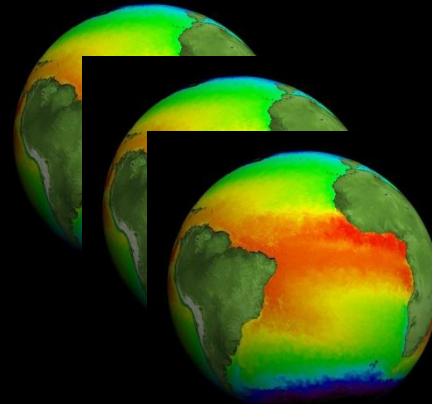
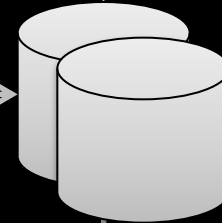
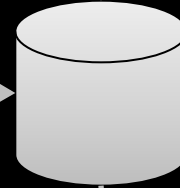
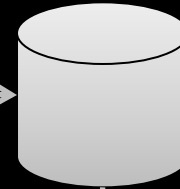
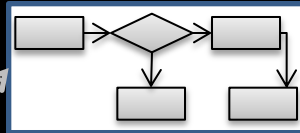
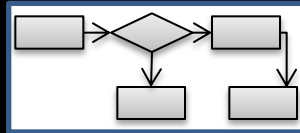
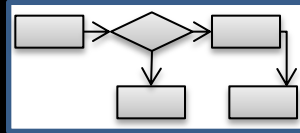
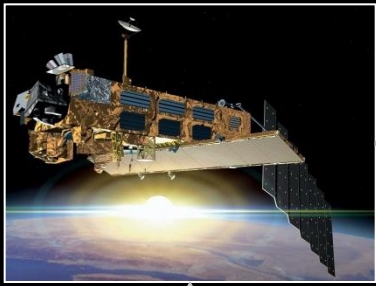
The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, designating key variables

publication



scientists



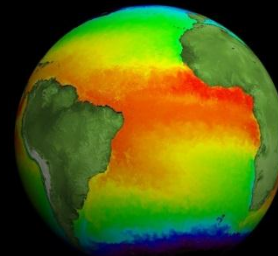
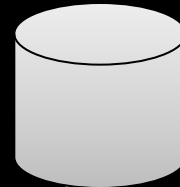
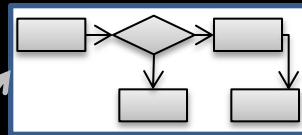


THE ESA CLIMATE CHANGE INITIATIVE
 Satellite Data Records for Essential Climate Variables

Dr R. HOLMANN, C. J. MICHANG, R. SANDERS, C. DONNO, M. BUCHWITZ, A. CAZZANI, E. CHIRICO, P. DIROUJAN, G. DE LEO, R. FORNARO, T. HÖRZNER, F. PIVA, S. SAKOCHI, S. SATTENBERGER, M. VAN ROSSUM, and W. WOLFF

The ESA's Climate Change Initiative is reprocessing and reanalysing 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.

Continued observations from satellites contribute vital knowledge to our understanding of Earth's environment and its changing state. The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science. Satellite law enables

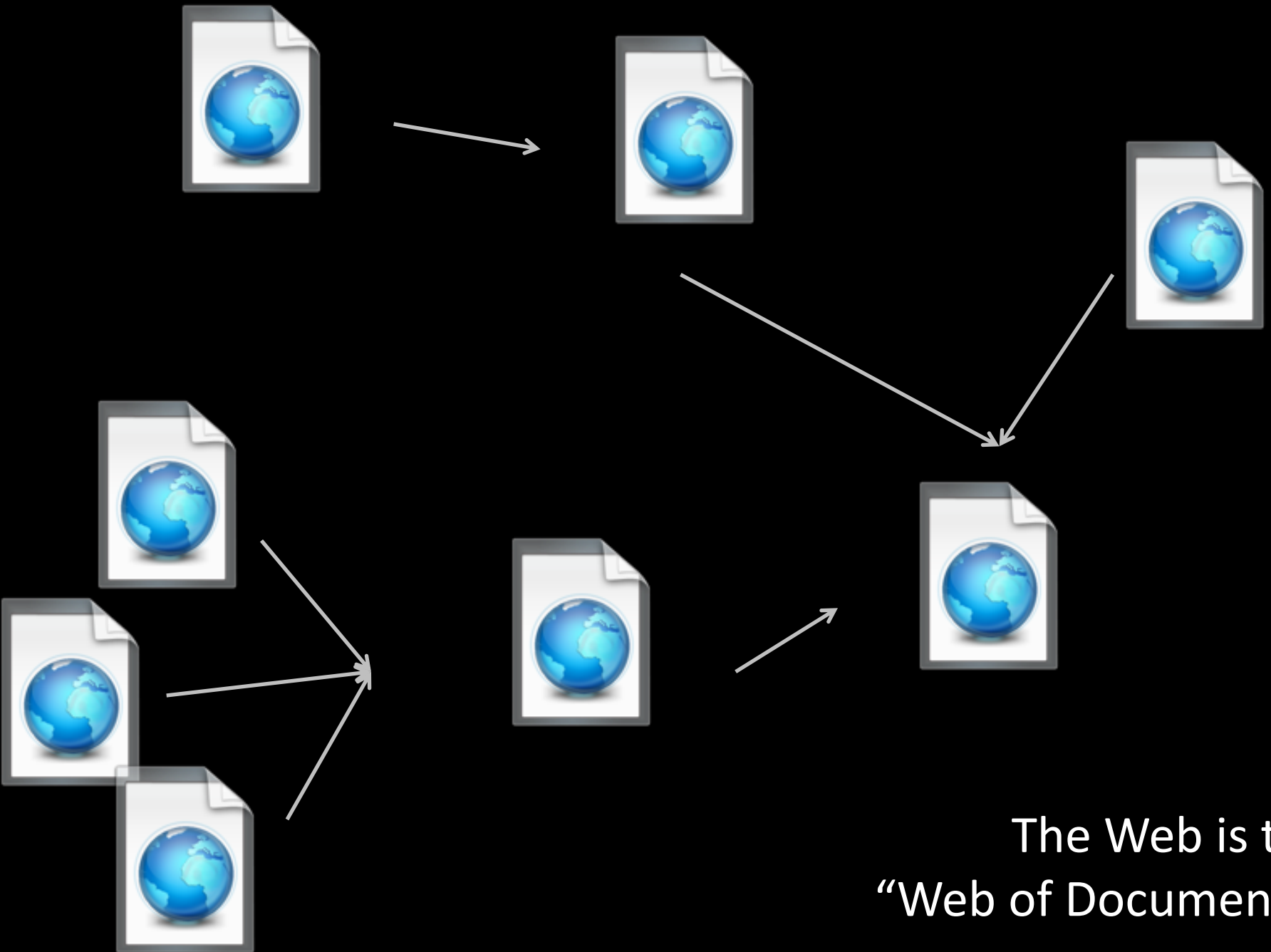


THE ESA CLIMATE CHANGE INITIATIVE
Satellite Data Records for Essential Climate Variables

BY R. HOLLMANN, C. J. MICHANGI, R. SANDERS, C. DONNO, M. BUCHWITZ, A. CAZZANI, E. CHIRICO, P. DIERCKMANN, G. DE LEEUW, R. FORBERG, T. HÖLZMANN, F. PAUL, S. SAKOCHI, S. SATTENBERGER, M. VAN ROSSUM, AND W. WOLFF

The ESA's Climate Change Initiative is reprocessing and reanalysing 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.

Continued observations from satellites contribute vital knowledge to our understanding of Earth's environment and its changing state. The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science. Cooperation between



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



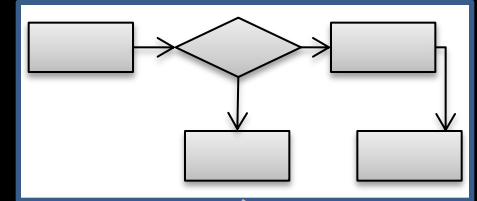
Photo by Susan Lesch, from Wikipedia

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

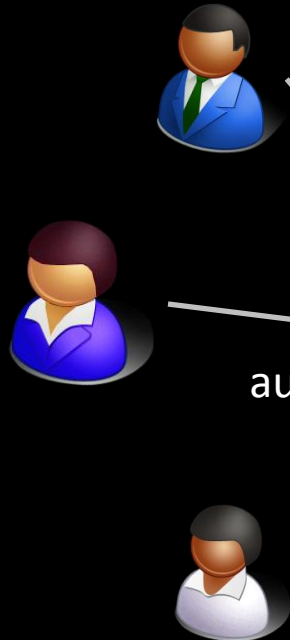
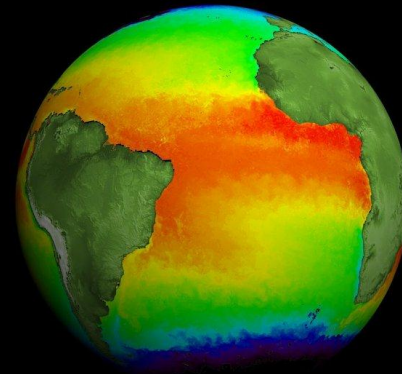


carries



produced

produced



authored

THE ESA CLIMATE CHANGE INITIATIVE
 Satellite Data Records for Essential Climate Variables

BY R. HOLLMANN, C. J. MICHANI, R. SHANDEL, C. DOWNY, M. BUCHWITZ, A. CAZZANI, E. CAHNECO, P. DEFOURNY, G. DE LEEUW, R. FORBERG, T. HOLZER-POPP, F. PAJ, S. SANDVEN, S. SATHYENDRANATH, M. VAN ROOZENEGEL, AND W. WAGNER

The ESA's Climate Change Initiative is reprocessing and reanalysing 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.

Satellite observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—some of the most important needs of climate science, decision-makers and the public.

The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, decision-makers and the public.

describes / uses / queries ...

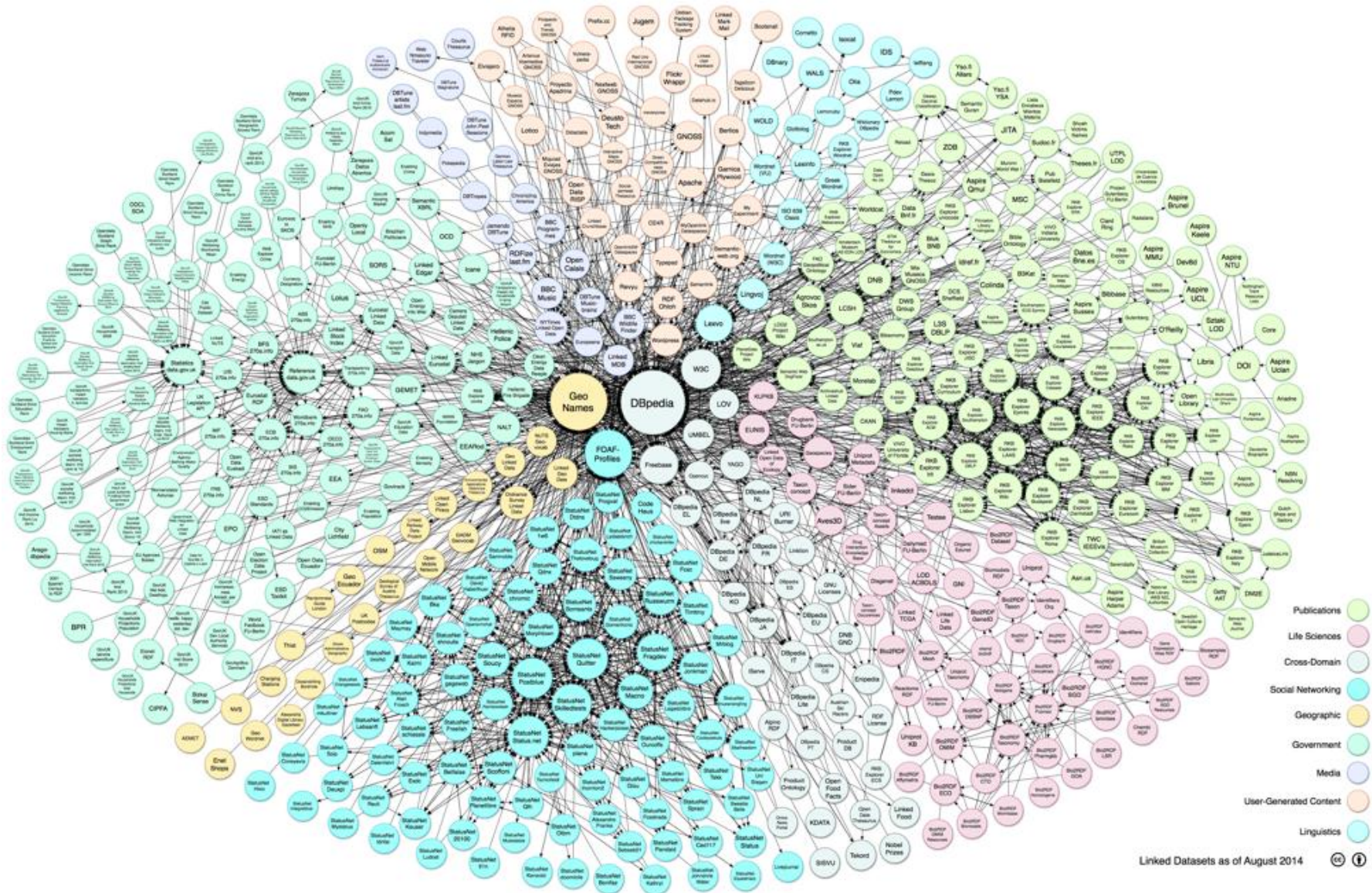
Linked Data techniques build the "Web of Data"

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)



About 65,100,000 results (0.36 seconds)

Showtimes for The Martian

Oct 14 Thu, Oct 15

All times	Morning	Afternoon	Evening	Night				
Regal Gallery Place Stadium 14 - Map								
Standard	1:00pm	4:30	6:00	8:00				
3D	11:30am	12:30pm	2:45	3:45	7:00	9:25	10:20	
AMC Loews Georgetown 14 - Map								
Standard	11:00am	2:15pm						
3D	12:00pm	1:00	3:15	4:15	6:30	7:30	9:45	
	10:40							
AMC Courthouse Plaza 8 - Map								
Standard	4:45pm	9:00						
3D	12:30pm	1:30	3:45	7:00	8:00	10:15		


All times are in ET

More showtimes

The Martian

2015 film

8.4/10 · IMDb
93% · Rotten Tomatoes
81% · Metacritic



When astronauts blast off from the planet Mars, they leave behind Mark Watney (Matt Damon), presumed dead after a fierce storm. With only a meager amount of supplies, the stranded visitor must utilize his wits and spirit to find a way to survive on the hostile planet. Meanwhile, back on Earth, membe... More

Release date: October 2, 2015 (USA)
Director: Ridley Scott
Running time: 2h 22m
MPAA rating: PG-13
Adapted from: The Martian

Critic reviews

View 6+ more

An enthralling and rigorously realistic outer-space survival story in which Matt Damon plays a NASA botanist stranded on the Red Planet after a sandstorm forces his crewmates to abort mission. Full review
Peter Debruge · Variety

You won't find a space epic that's more fun to geek out at than The Martian. Full review
Peter Travers · Rolling Stone

Anchored by another great turn from Matt Damon, The Martian mixes smarts, laughs, weird character bits and tension on a huge canvas. The result is Scott's most purely enjoyable film for ages. Full review
Ian Freer · Empire

At its heart, The Martian is an unapologetically stirring celebration of our ability, as a species, to solve even the most daunting problems via rational thought, step by step by step. Full review
Mike D'Angelo · A.V. Club

The Martian (2015) - IMDb
www.imdb.com/title/tt3659388/ · Internet Movie Database
★ ★ ★ ★ Rating: 8.3/10 - 79,977 votes
Directed by Ridley Scott. With Matt Damon, Jessica Chastain, Kristen Wiig, Kate Mara. During a manned mission to Mars, Astronaut Mark Watney is presumed ...
Full Cast & Crew - Kate Mara - Ridley Scott - Poster

The Martian (film) - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/The_Martian_(film) · Wikipedia
Plot. When the Ares III manned mission to Mars is hit by an intense Martian storm, astronaut Mark Watney is lost and presumed dead. With the lives of her crew at ...
The Martian (Weir novel) - Michael Peña - Naomi Scott - Drew Goddard

The Martian (Weir novel) - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/The_Martian_(Weir_novel) · Wikipedia
The Martian is a 2011 science fiction novel and the first published novel by American author Andy Weir. It was originally self-published in 2011 after which ...
Author: Andy Weir · Publisher: Crown Publishing Group
Pages: 369 · Cover artist: Eric White

In the news

Box office: 'Pan' bombs, 'Martian' stays No. 1
CNN - 5 hours ago
... loss for Warner Bros. after opening to \$15.5 million. "Martian" once again topped box office.






US box office: The Martian soars into blockbuster territory as Pan plummets
The Guardian - 3 hours ago

'Pan' lacks magic at the box office; 'The Martian' soars
Fox News - 18 hours ago

More news for the martian


Cast

View 10+ more

 Matt Damon Mark Watney	 Jessica Chastain Melissa Lewis	 Kristen Wiig Annie Montrose	 Kate Mara Beth Johansson	 Michael Peña Rick Martinez
--	--	--	--	--

People also search for

View 15+ more



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)

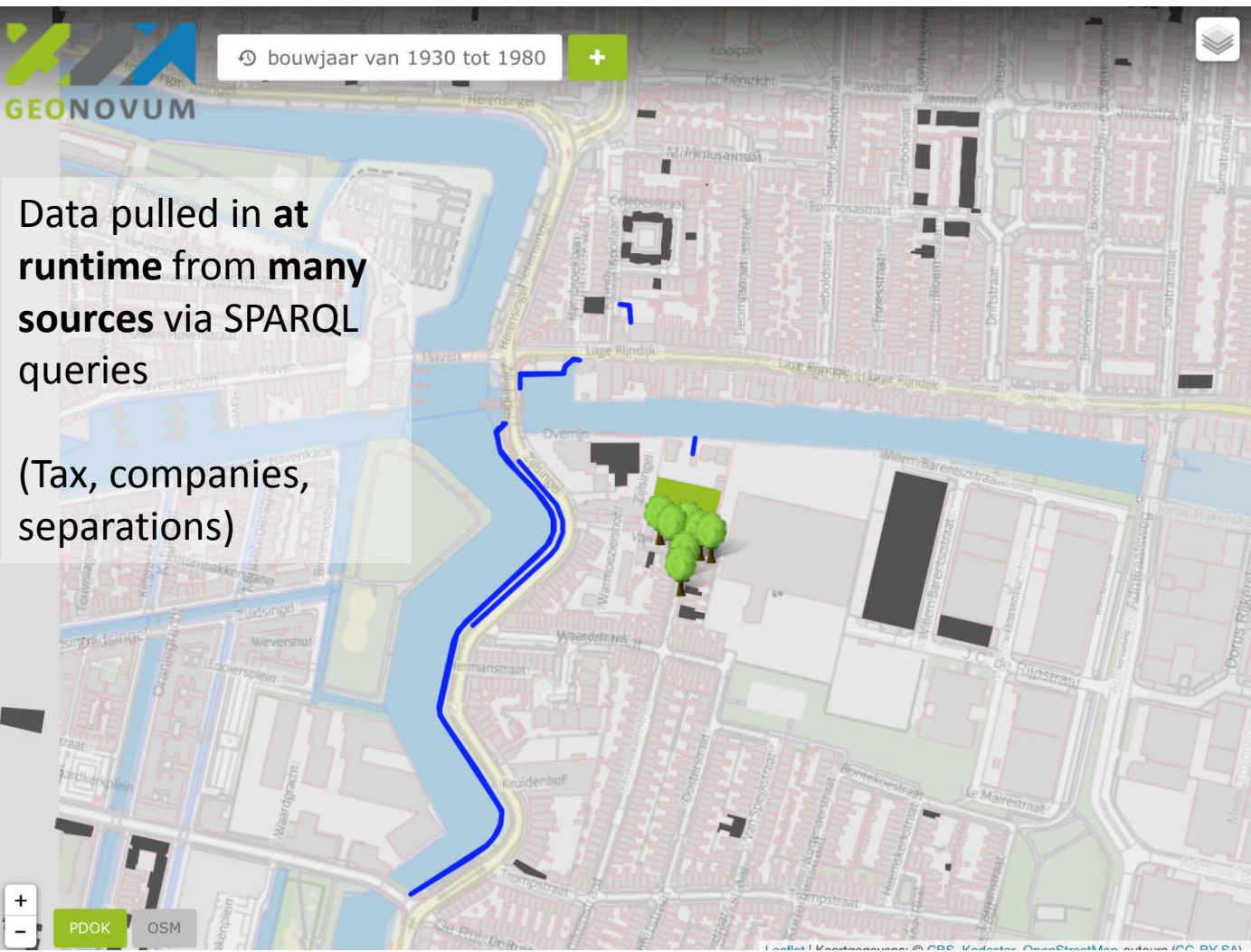
Why can't we do the same for EO data?

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>



Data pulled in at runtime from many sources via SPARQL queries

(Tax, companies, separations)

> Sluit venster

Gegevens uit de BAG

Adres	Zijlsingel 2
Postcode	2315KB
Stad	Leiden
Bouwjaar	1950
Status	Pand in gebruik

Winkel (1) + Zorginstelling (2)

Overig (6)

Gegevens uit de Basisregistratie WOZ

WOZ-objecten [↗](#)

In dit pand zijn 9 WOZ-objekten geregistreerd >

Gegevens uit het Nationaal Handels Register

Er zijn in dit pand geen bedrijven gevestigd.

Gegevens uit de BGT

Scheidingen

Om dit pand zijn 5 scheidingen gevonden >

(thanks to Linda van den Brink)

Soil moisture observations

Terradue Virtual Machine

PostgreSQL – PostGIS Database

Strabon
Spatiotemporal RDF Store
SPARQL Endpoint

SPARQL Query

RDF returned

Raster / Vector layers

KML
GeoJSON
GeoTIFF

Sextant
web-based “mash-up”
system for linked
geospatial data

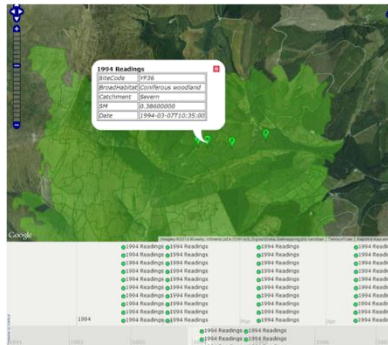
Other SPARQL Endpoints

DBPedia

INSPIRE
GEOS
GMES

UK Ordnance Survey

etc.



Visual display in browser

KML

Google Earth

Further analysis

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



Annotations

Viewing annotations for: http://apps.ecmwf.int/datasets/data/era20cm_moda (Dataset)

Title	Annotated by	Organisation	Date
Journal article about ERA Interim data	Andrew Henry	Apps-test-ECMWF	10/12/2014
This dataset is experimental. Final version will be available in the beginn...	Iryna Rozum	Apps-test-ECMWF	5/12/2014
An ensemble of climate model integrations using the IFS, covers Jan 1900 - ...	Iryna Rozum	Apps-test-ECMWF	5/12/2014

v1.0 BETA

Login

Cancel

Display a menu

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



“CHARMe Maps” tool



Comparison synch. Navigation synch. Time synch. Elevation synch. Palette synch. Facet

title contains : search

Navigation

- cci_sst_badc
- sea_surface_temperature
- sst_dtime
- sstes_bias
- sstes_standard_deviation
- l2p_flags
- quality_level
- wind_speed
- large_scale_correlated_uncertainty
- synoptically_correlated_uncertainty
- uncorrelated_uncertainty

cci_sst_badc/sea_surface_temperature

Units: kelvin Time: 2012-01-02 | 03:00:00.000Z Elevation:

304.7015
293.397
282.093
270.78848

default-scalar
opaque
linear

Open in Google Earth Permalink Email Link Export to PNG

cci_sst_badc/sstes_standard_deviation

Units: kelvin Time: 2012-01-02 | 03:00:00.000Z Elevation:

1.061
0.834
0.506

default-scalar
opaque
linear

Datasets

cci_sst_badc sea_surface_temperature sstes_standard_deviation

Fine Grained Commentary Viewer

Fine Grained Commentary Editor

Subset definition

Geometry Description

Geometry type: POLYGON

Region name:

Lon: 43.0 Lat: 4.0
Lon: 41.0 Lat: -2.1
Lon: 40.5 Lat: -8.1
Lon: 35.0 Lat: -12.1
Lon: 40.1 Lat: 6.3
Lon: 42.7 Lat: 3.3
Lon: 42.9 Lat: 4.1

POINT(41.5 -26.2);<http://www.opengis.net/def/crs/EPSSG/0/4326>

Temporal

Calendar: JULIAN

Begin: 2012-01-01 | 00:00:00.000Z
End: 2012-01-01 | 00:00:00.000Z

Elevation

Begin: End:

Variables

cci_sst_badc/sstes_standard_deviation

Annotation
Motivation
Highlighting
Comment
These are

Marine, Oceanography, Earth Observation

Create Reset

You are logged in CHARMe!

Allows creation and discovery of commentary about specific parts of datasets

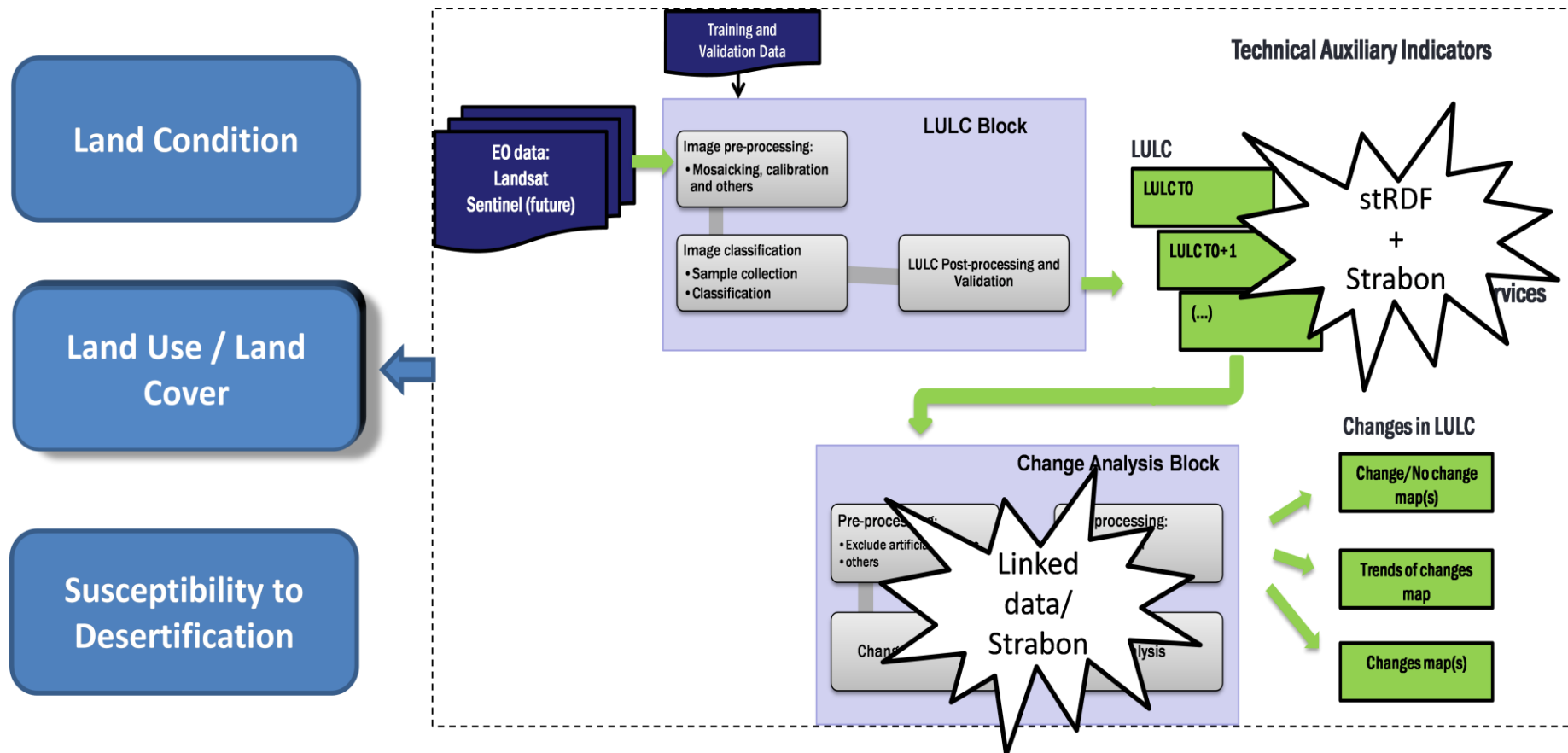
What can we do with Linked Data?

- 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)

Dataset (1)


1 Error 

Fetch URL Examples ▾

```

1 <script type="app
2 {
3   "@context": "http
4   "@type": "Dataset
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:	<i>http://www.example.com/datasets/1</i>
license:	<i>https://creativecommons.org/licenses/by/4.0/</i>
datasetTimeInterval:	2007-03-01
datePublished:	2010-01-01
keywords:	sea, ice
spatial [Place]:	
geo [StructuredValue]:	
 box:	50 4 52 6
about [Thing]:	<i>http://www.example.com/sea_ice_fraction</i>
name:	Sea Ice Fraction
publisher [Organization]:	
name:	MELODIES
hasPart [Dataset]:	<i>http://www.example.com/datasets/1/1</i>
hasPart [Dataset]:	<i>http://www.example.com/datasets/1/2</i>

Google knows what a Dataset is!

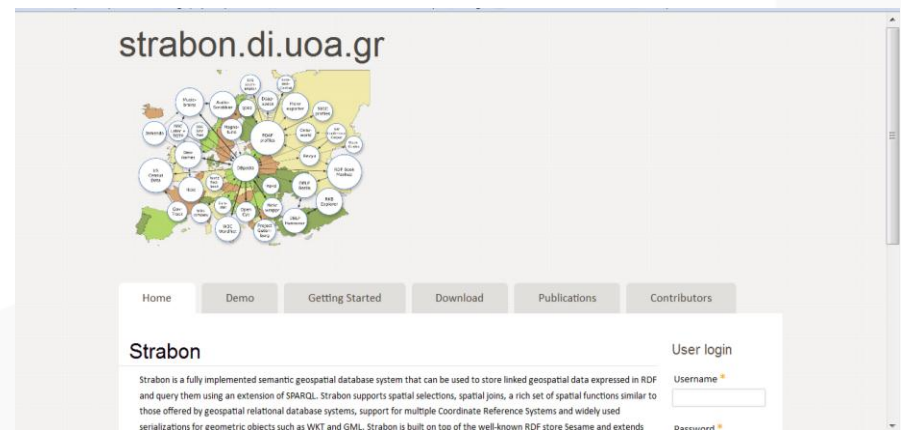
(again, uses schema.org vocab)

(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

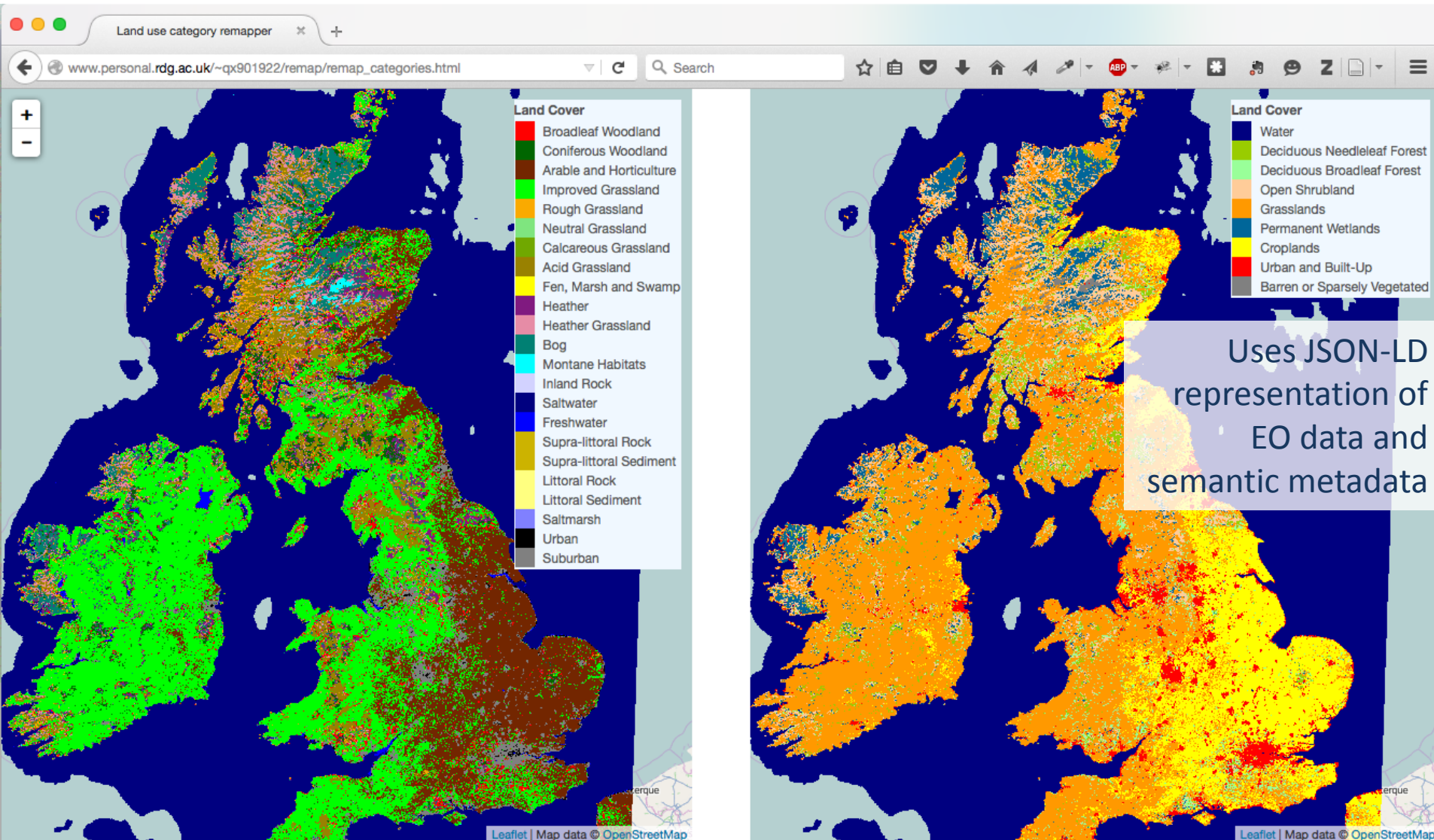


melodies
exploiting open data

(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



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



Thank you!

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

<http://melodiesproject.eu>
@MelodiesProject

