SeaDataCloud

Applying Linked Data principles for SeaDataNet Catalogues

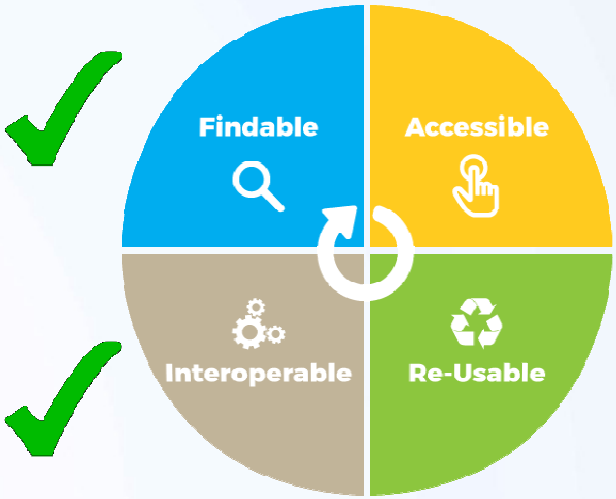
Rob Thomas and Adam Leadbetter
Marine Institute, Ireland

Second annual meeting, Barcelona, Spain - 8-9 November 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org

Why?

- In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable.
  - "H2020 Programme Guidelines on FAIR Data Management in Horizon 2020"

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (2016).



Reproducible?

Background

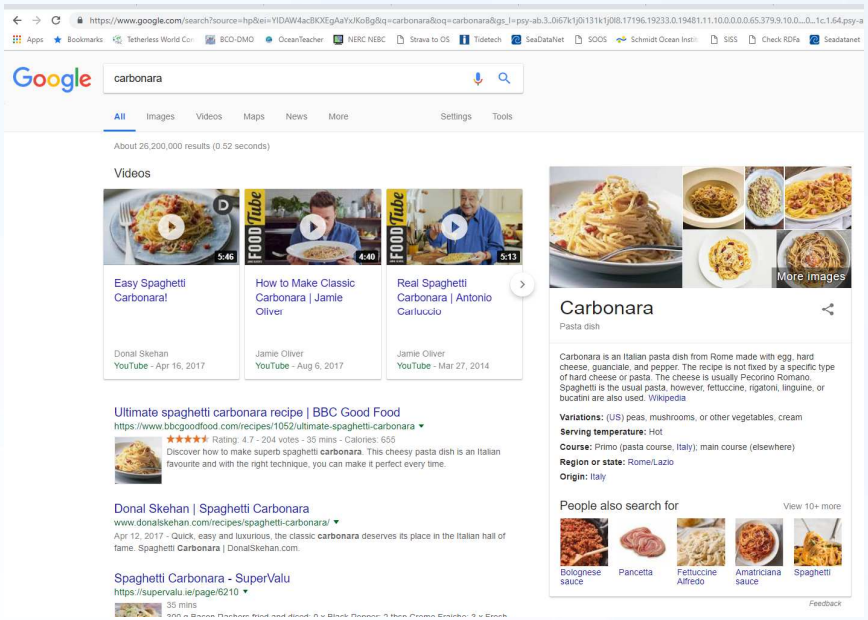
"The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With Linked Data, when you have some of it, you can find other, related, data."

– Sir Tim Berners-Lee, 2006

Background

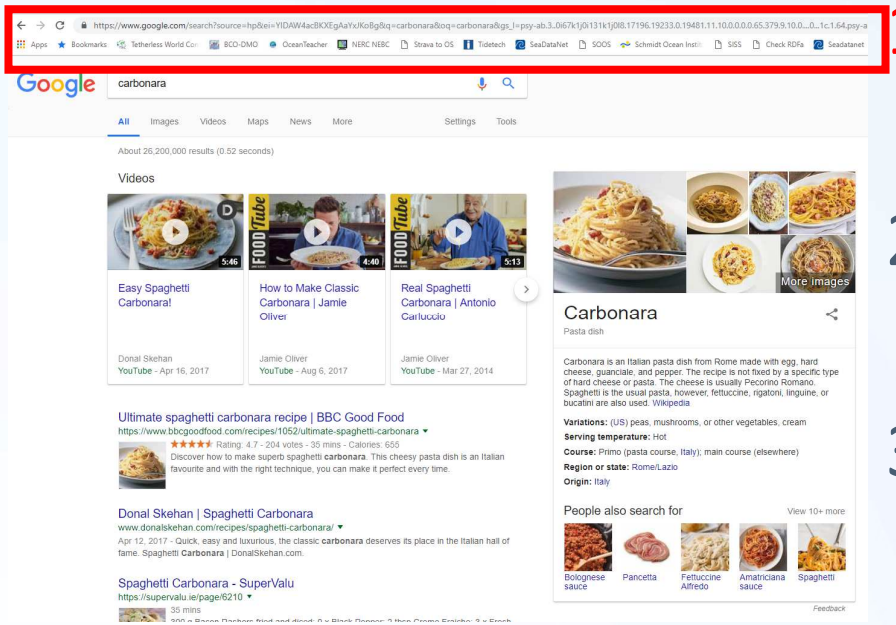
1. Use web addresses to name things
2. Allow those address to be looked up
3. Use web standards when the addresses are looked up
4. Include links to other web resources

Background



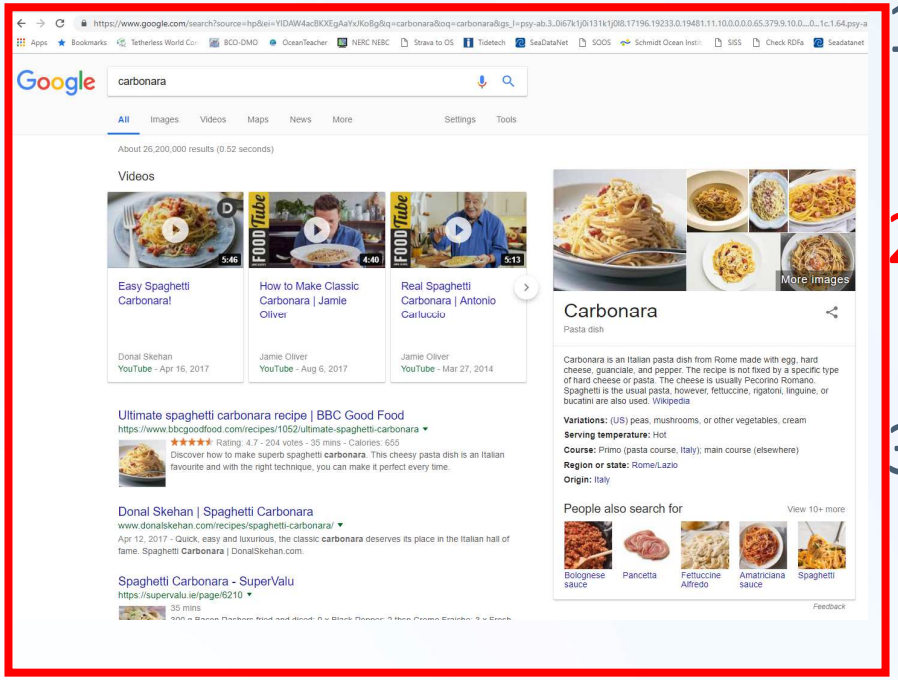
1. Use web addresses to name things
2. Allow those address to be looked up
3. Use web standards when the addresses are looked up
4. Include links to other web resources

Background



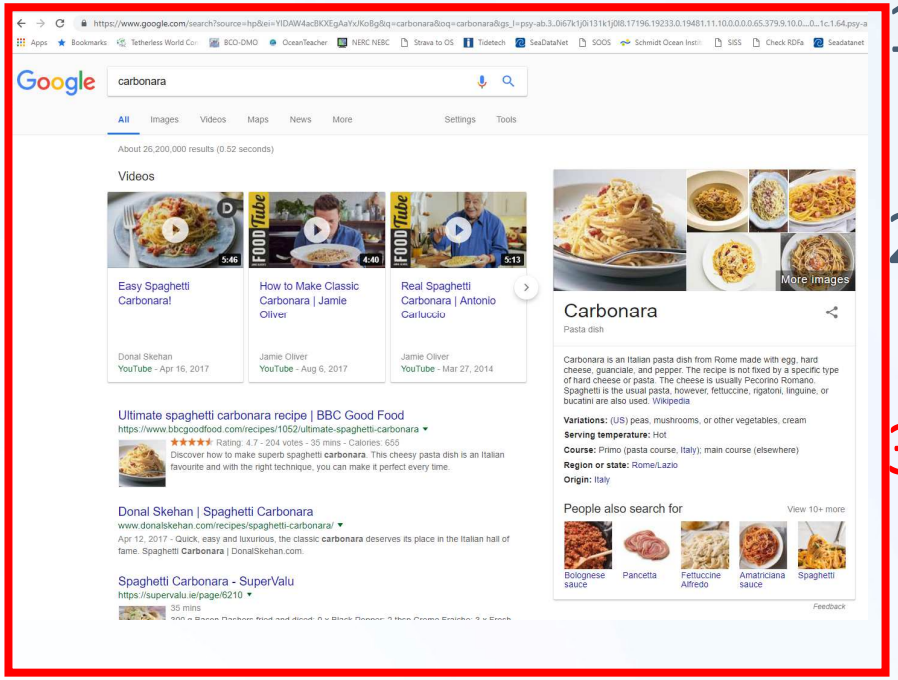
1. Use web addresses to name things
2. Allow those address to be looked up
3. Use web standards when the addresses are looked up
4. Include links to other web resources

Background



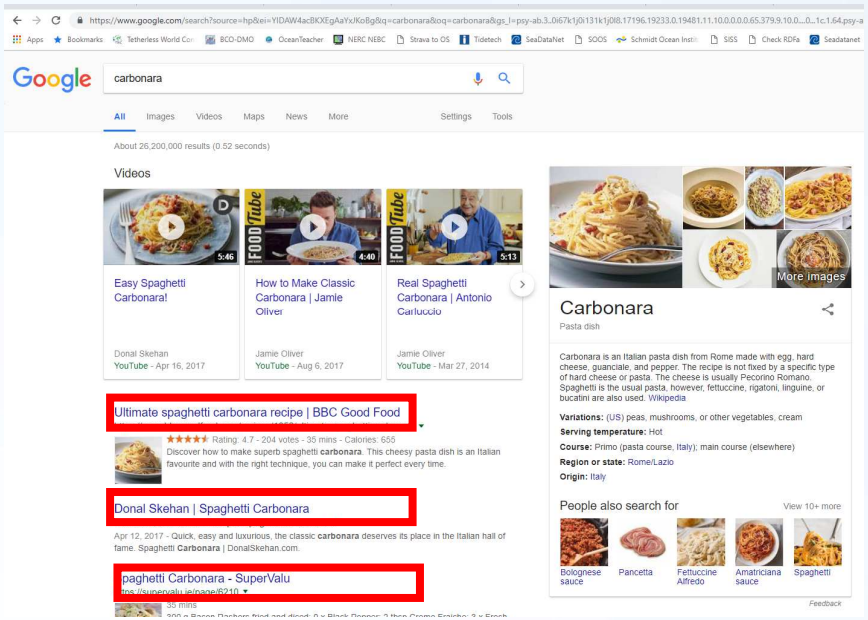
1. Use web addresses to name things
2. Allow those address to be looked up
3. Use web standards when the addresses are looked up
4. Include links to other web resources

Background



1. Use web addresses to name things
2. Allow those address to be looked up
3. Use web standards when the addresses are looked up
4. Include links to other web resources

Background



1. Use web addresses to name things
2. Allow those address to be looked up
3. Use web standards when the addresses are looked up
4. Include links to other web resources

Background

EDMERP

Projects (3 079)

EDMO

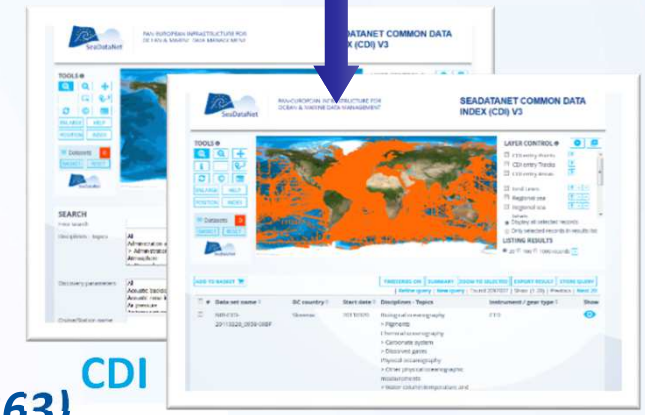
Organisations (4 523)

CSR

Research cruises (48 836)

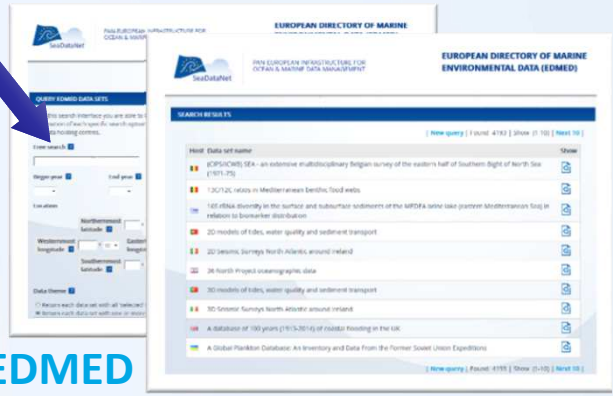
EDIOS

Observing programmes (363)



CDI

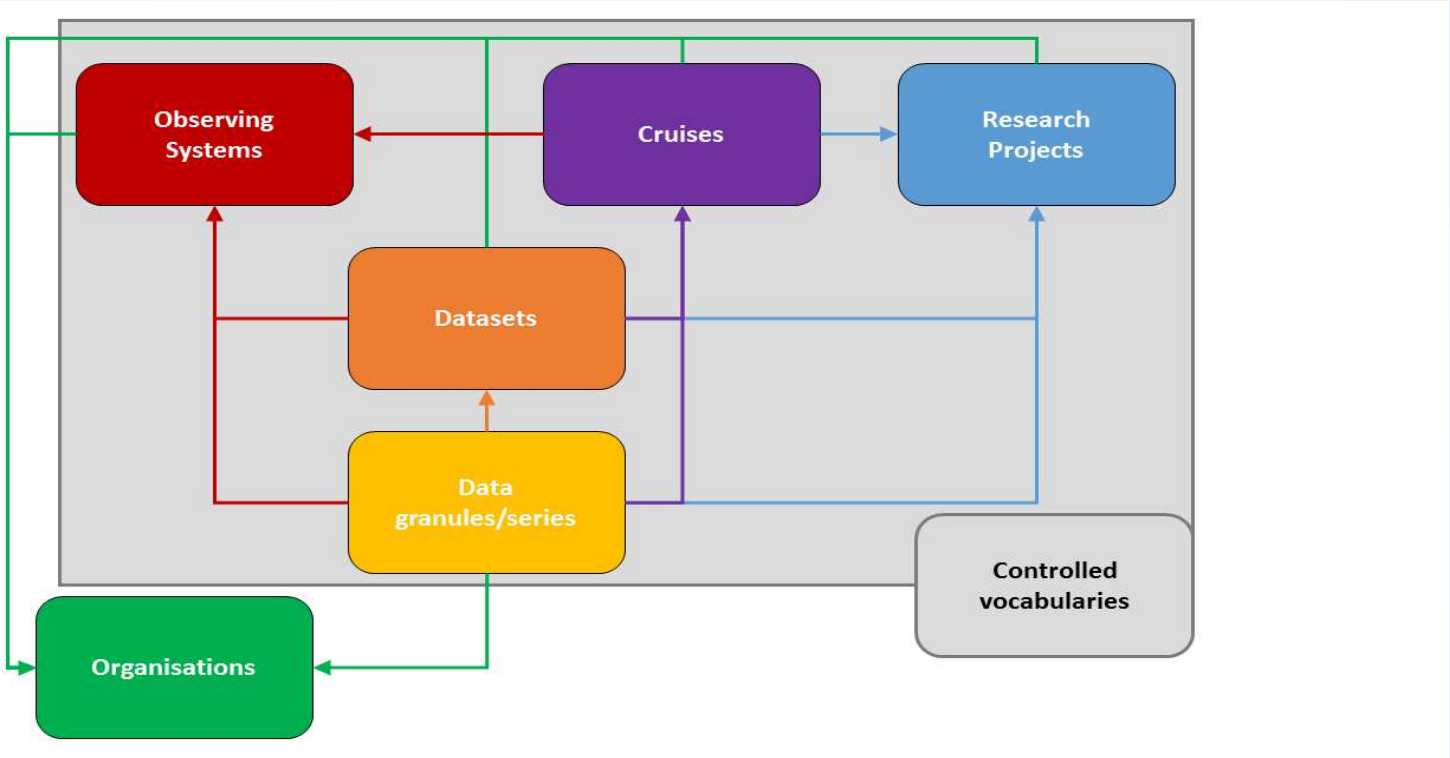
Data index (2 151 046)



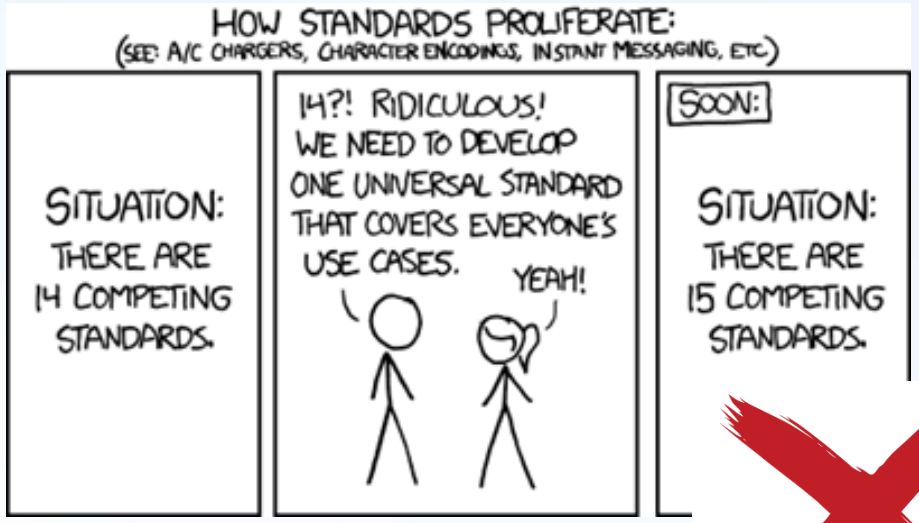
EDMED

Data sets (4 193)

Work so far


Observing Systems
Cruises
Research Projects
Datasets
Data granules/series
Controlled vocabularies
Organisations

Work so far


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.


No thanks!
We are too busy

Work so far

- Reusing existing patterns
  - Better understanding outside of SDN
  - Better interoperability with other organisations
  - Better INSPIRE compliance

Work so far

Catalogue	Entity	Exisiting Standard(s)	Host
EDMO	Organisation	W3C Organisation	Maris
EDMED	Data Set	W3C DCAT	BODC
EDMERP	Research Project	W3C Prov-O DBPedia Research Project	Maris
EDIOS	Observing System	INSPIRE Environmental Monitoring Facilities	BODC
CSR	Cruise	?... Publish SDC work	BSH
CDI	Data granule/series	W3C DCAT	Maris

Work so far

- Reusing existing patterns
  - Better understanding outside of SDN
  - Better interoperability with other organisations
  - Better INSPIRE compliance

  - Completed year 1 and included in:

  D8.3 Updated metadata formats and related XML schemas.

Implementation progress

- https://edmed.seadatanet.org/
  - https://edmed.seadatanet.org/search/
  - https://edmed.seadatanet.org/sparql/
  - https://edmed.seadatanet.org/report/<ID>
- https://edmo.seadatanet.org/
- https://edios.seadatanet.org/
- https://edmerp.seadatanet.org/
- https://cdi.seadatanet.org/

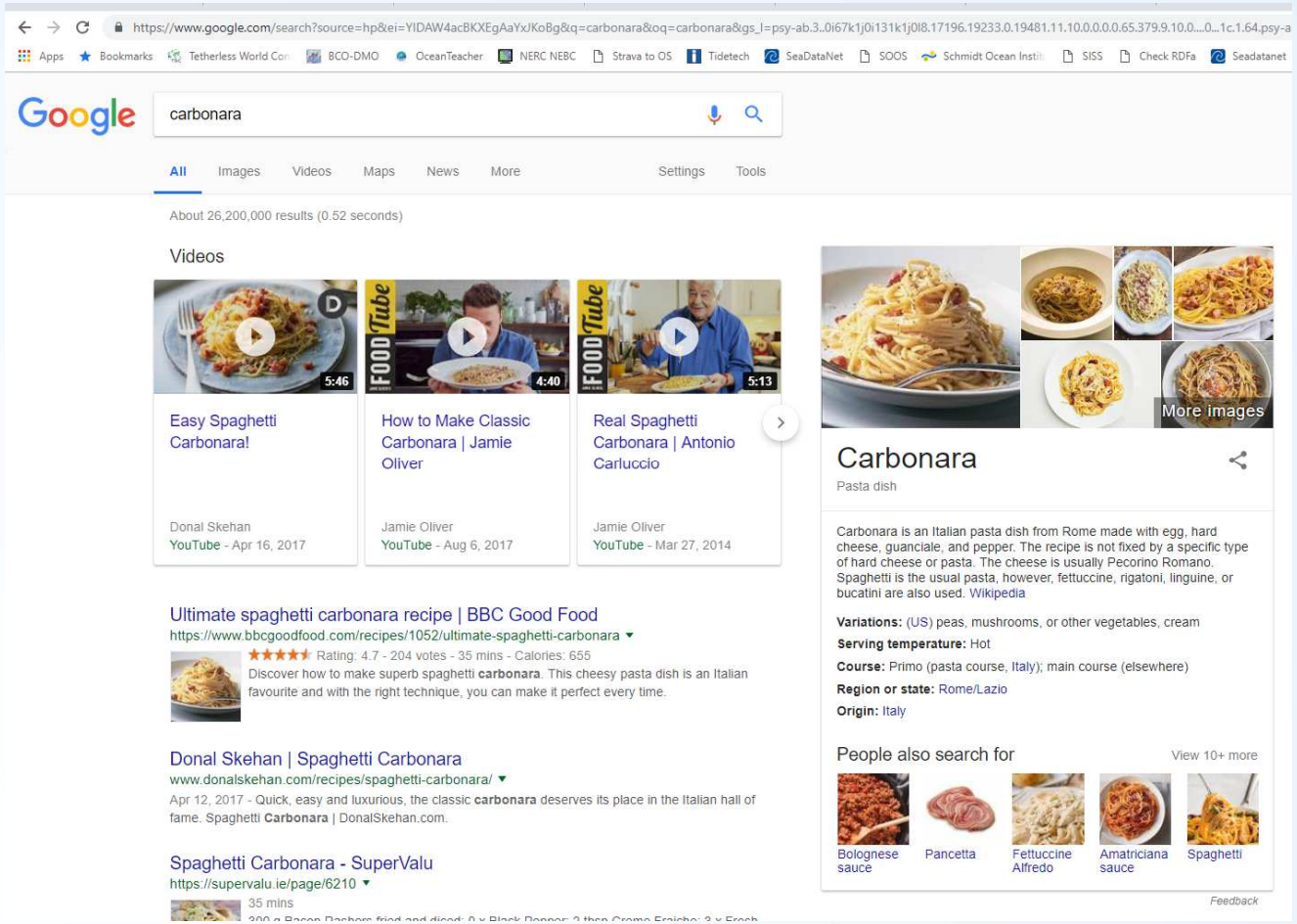
Implementation progress

Catalogue	Entity	Exisiting Standard(s)	Host	
EDMO	Organisation	W3C Organisation	Maris	✓
EDMED	Data Set	W3C DCAT	BODC	✓
EDMERP	Research Project	W3C Prov-O DBPedia Research Project	Maris	
EDIOS	Observing System	INSPIRE Environmental Monitoring Facilities	BODC	
CSR	Cruise	To be published	BSH	
CDI	Data granule/series	W3C DCAT	Maris	

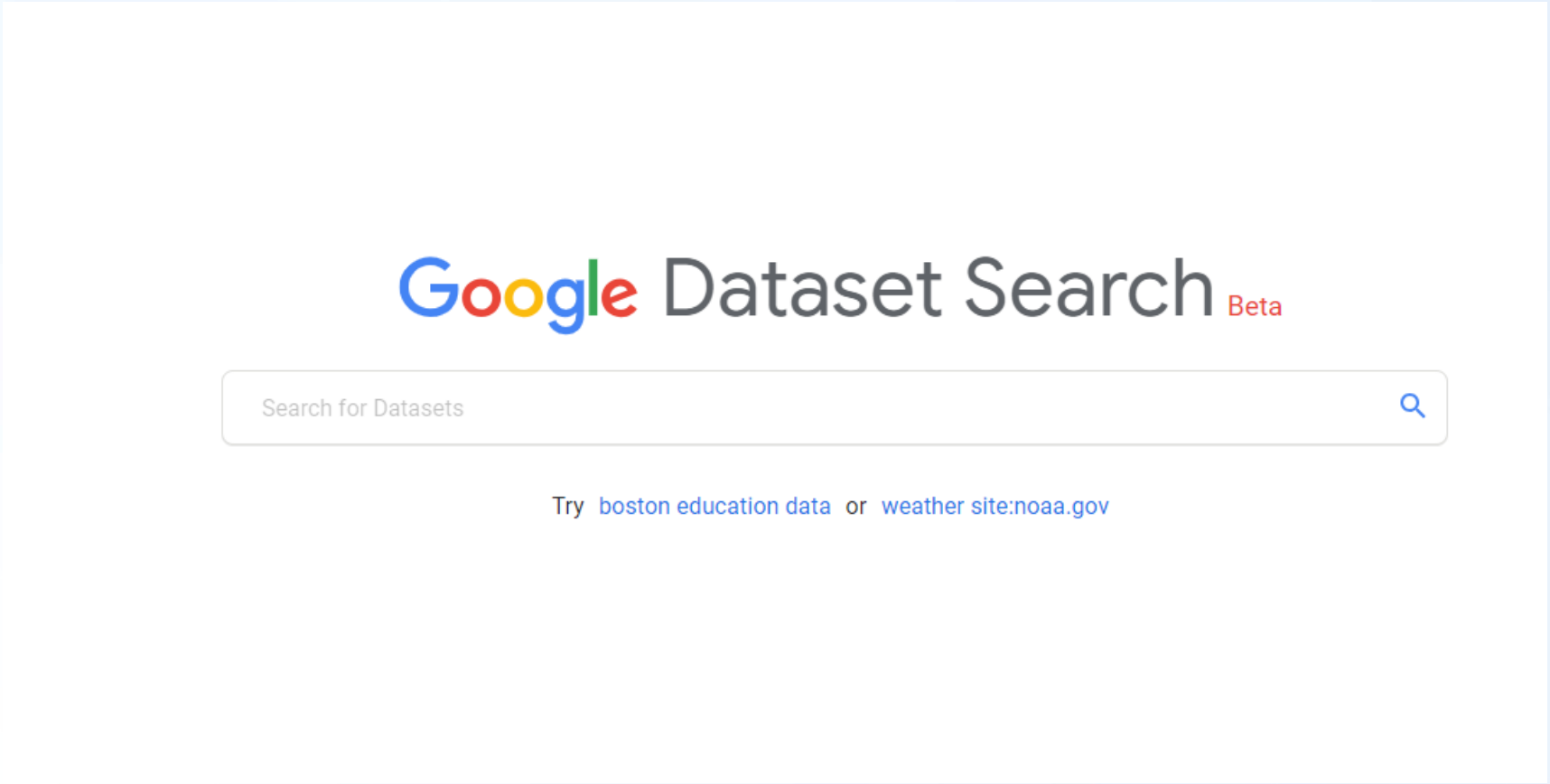
Work so far

- Reusing existing patterns
  - Better understanding outside of SDN
  - Better interoperability with other organisations
  - Better INSPIRE compliance
- Implementation
- Mapping to Schema.org
  - Added value work

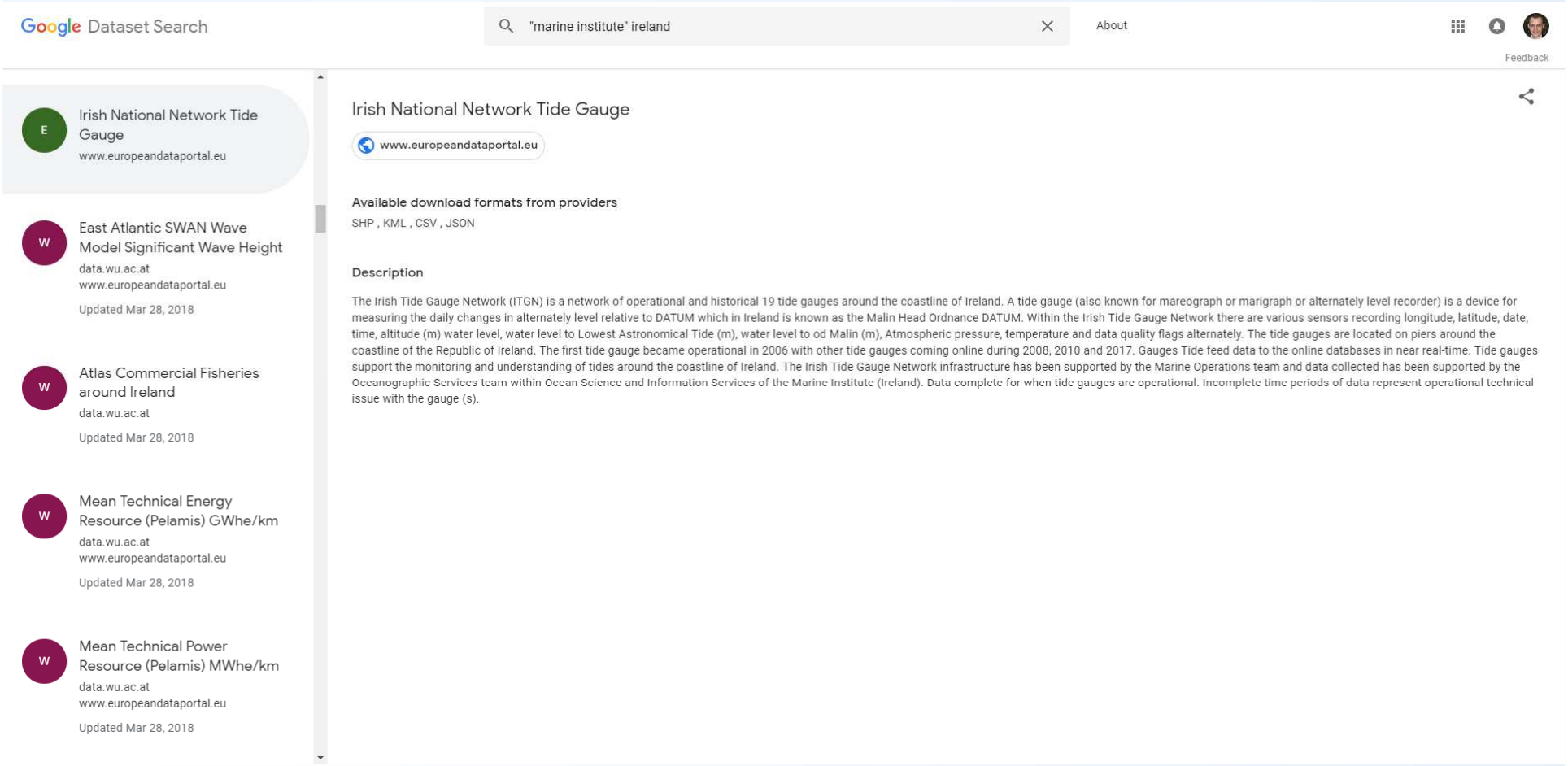
Work so far


carbonara
All
Images
Videos
Maps
News
More
Settings
Tools
About 26,200,000 results (0.52 seconds)
Videos
5:46
4:40
5:13
Easy Spaghetti Carbonara!
Donal Skehan
YouTube - Apr 16, 2017
How to Make Classic Carbonara | Jamie Oliver
Jamie Oliver
YouTube - Aug 6, 2017
Real Spaghetti Carbonara | Antonio Carluccio
Jamie Oliver
YouTube - Mar 27, 2014
Ultimate spaghetti carbonara recipe | BBC Good Food
https://www.bbcgoodfood.com/recipes/1052/ultimate-spaghetti-carbonara
Rating: 4.7 - 204 votes - 35 mins - Calories: 655
Discover how to make superb spaghetti carbonara. This cheesy pasta dish is an Italian favourite and with the right technique, you can make it perfect every time.
Donal Skehan | Spaghetti Carbonara
www.donalskehan.com/recipes/spaghetti-carbonara/
Apr 12, 2017 - Quick, easy and luxurious, the classic carbonara deserves its place in the Italian hall of fame. Spaghetti Carbonara | DonalSkehan.com.
Spaghetti Carbonara - SuperValu
https://supervalu.ie/page/6210
35 mins
More images
Carbonara
Pasta dish
Carbonara is an Italian pasta dish from Rome made with egg, hard cheese, guanciale, and pepper. The recipe is not fixed by a specific type of hard cheese or pasta. The cheese is usually Pecorino Romano. Spaghetti is the usual pasta, however, fettuccine, rigatoni, linguine, or bucatini are also used. Wikipedia
Variations: (US) peas, mushrooms, or other vegetables, cream
Serving temperature: Hot
Course: Primo (pasta course, Italy); main course (elsewhere)
Region or state: Rome/Lazio
Origin: Italy
People also search for
View 10+ more
Bolognese sauce
Pancetta
Fettuccine Alfredo
Amatriciana sauce
Spaghetti
Feedback

Work so far


Google Dataset Search Beta
Search for Datasets
Try boston education data or weather site:noaa.gov

Work so far


Google Dataset Search
"marine institute" ireland
About
Feedback
Irish National Network Tide Gauge
www.europeandataportal.eu
East Atlantic SWAN Wave Model Significant Wave Height
data.wu.ac.at
www.europeandataportal.eu
Updated Mar 28, 2018
Atlas Commercial Fisheries around Ireland
data.wu.ac.at
Updated Mar 28, 2018
Mean Technical Energy Resource (Pelamis) GWhe/km
data.wu.ac.at
www.europeandataportal.eu
Updated Mar 28, 2018
Mean Technical Power Resource (Pelamis) MWhe/km
data.wu.ac.at
www.europeandataportal.eu
Updated Mar 28, 2018
Irish National Network Tide Gauge
www.europeandataportal.eu
Available download formats from providers
SHP , KML , CSV , JSON
Description
The Irish Tide Gauge Network (ITGN) is a network of operational and historical 19 tide gauges around the coastline of Ireland. A tide gauge (also known for mareograph or marigraph or alternately level recorder) is a device for measuring the daily changes in alternately level relative to DATUM which in Ireland is known as the Malin Head Ordnance DATUM. Within the Irish Tide Gauge Network there are various sensors recording longitude, latitude, date, time, altitude (m) water level, water level to Lowest Astronomical Tide (m), water level to od Malin (m), Atmospheric pressure, temperature and data quality flags alternately. The tide gauges are located on piers around the coastline of the Republic of Ireland. The first tide gauge became operational in 2006 with other tide gauges coming online during 2008, 2010 and 2017. Gauges Tide feed data to the online databases in near real-time. Tide gauges support the monitoring and understanding of tides around the coastline of Ireland. The Irish Tide Gauge Network infrastructure has been supported by the Marine Operations team and data collected has been supported by the Oceanographic Services team within Ocean Science and Information Services of the Marine Institute (Ireland). Data complete for when tide gauges are operational. Incomplete time periods of data represent operational technical issue with the gauge (s).

Future directions?

Future directions?

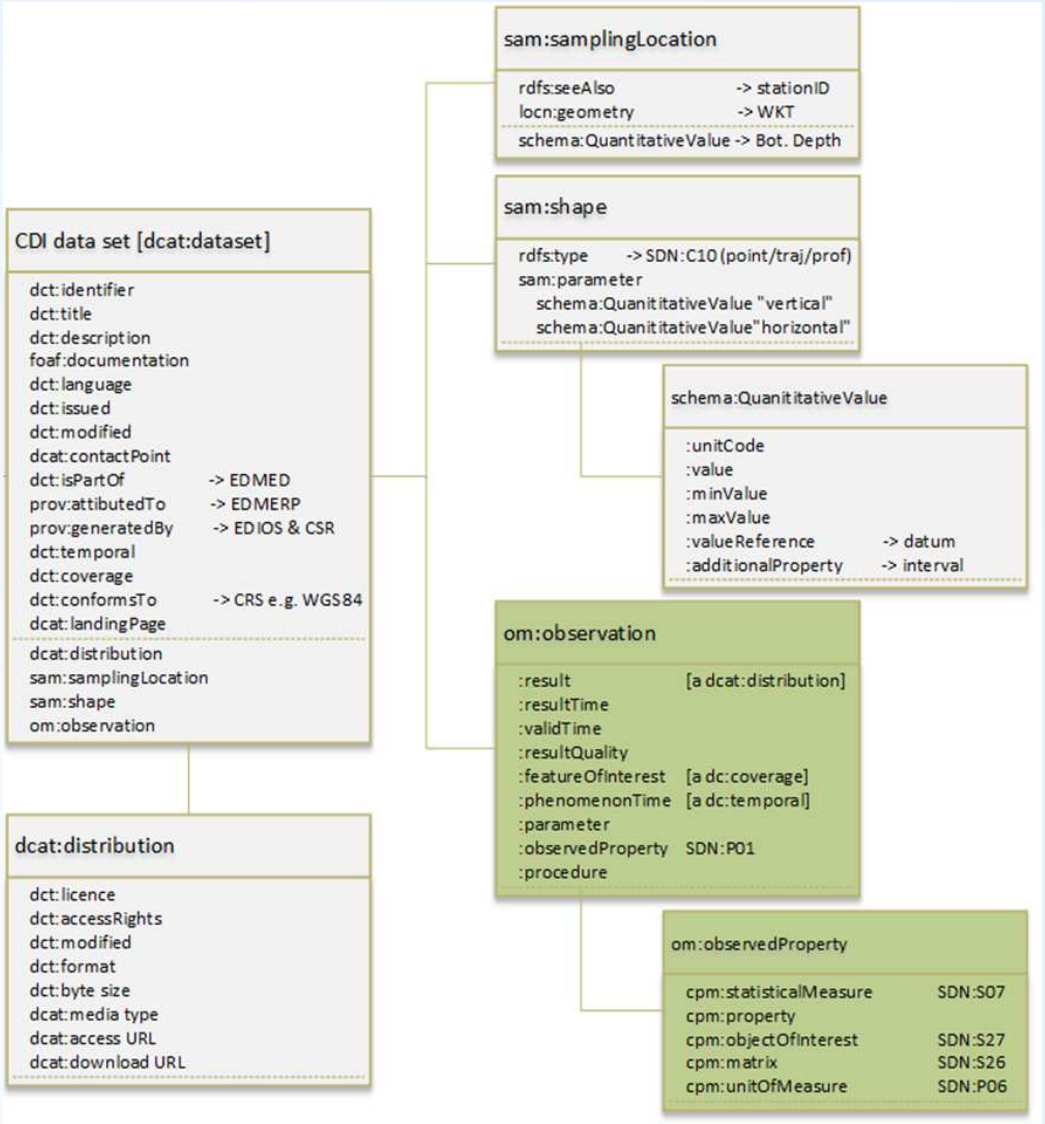
Future directions?

Future directions?

Catalogue	Linked Data Standard
Organisations (EDMO)	World Wide Web Consortium
Datasets (EDMED)	World Wide Web Consortium As used in European Data Portal
Projects (EDMERP)	World Wide Web Consortium / DBPedia
Common Data Inventory	World Wide Web Consortium
Cruise Summary Reports	*Liaised with US-NSF Rolling Deck to Repository & with Australia (through ODIP/SDC)*
Observing Systems (EDIOS)	INSPIRE Spatial Data Infrastructure
ODV Header	*Open Geospatial Consortium / ISO*

```
//SDN_parameter_mapping
//<subject>SDN:LOCAL:Depth</subject><object>SDN:P01::DEPHPR01</object><units>SDN:P06::ULAA</units>
//<subject>SDN:LOCAL:Pressure</subject><object>SDN:P01::PRESPR01</object><units>SDN:P06::UPDB</units>
//<subject>SDN:LOCAL:Salinity</subject><object>SDN:P01::PSALST01</object><units>SDN:P06::UUUU</units><instrument>SDN:L22::TOOL0058</instrument>
//<subject>SDN:LOCAL:Temperature</subject><object>SDN:P01::TEMPST01</object><units>SDN:P06::UPAA</units><instrument>SDN:L22::TOOL0058</instrument>
```

Future directions?



Future directions?

INSPIRE - Infrastructure for Spatial Information in Europe

CDI data set [dcat:dataset]

- dct:identifier
- dct:title
- dct:description
- foaf:documentation
- dct:language
- dct:issued
- dct:modified
- dcat:contactPoint
- dct:isPartOf -> EDMED
- prov:attibutedTo -> EDMERP
- prov:generatedBy -> EDIOS & CSR
- dct:temporal
- dct:coverage
- dct:conformsTo -> CRS e.g. WGS84
- dcat:landingPage
- dcat:distribution
- sam:samplingLocation
- sam:shape
- om:observation

dcat:distribution

- dct:licence
- dct:accessRights
- dct:modified
- dct:format
- dct:byte size
- dcat:media type
- dcat:access URL
- dcat:download URL

sam:samplingLocation

- rdfs:seeAlso -> stationID
- locn:geometry -> WKT
- schema:QuantitativeValue -> Bot. Depth

sam:shape

- rdfs:type -> SDN:C10 (point/traj/prof)
- sam:parameter
  - schema:QuanititativeValue "vertical"
  - schema:QuanititativeValue"horizontal"

schema:QuanititativeValue

- :unitCode
- :value
- :minValue
- :maxValue
- :valueReference -> datum
- :additionalProperty -> interval

om:observation

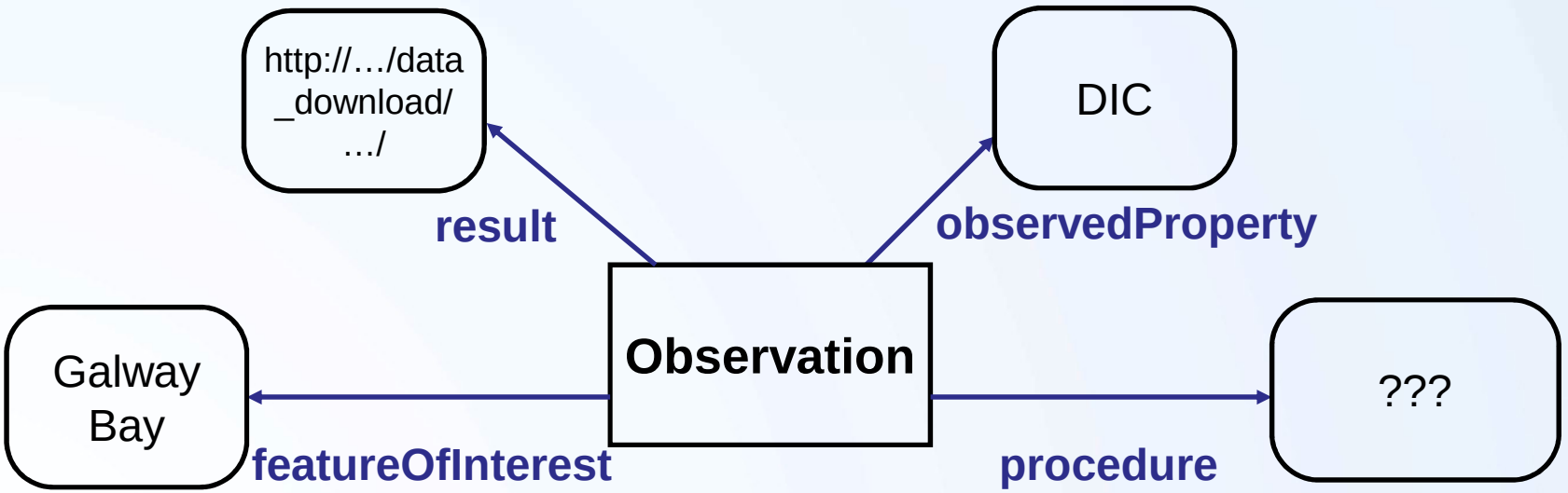
- :result [a dcat:distribution]
- :resultTime
- :validTime
- :resultQuality
- :featureOfInterest [a dc:coverage]
- :phenomenonTime [a dc:temporal]
- :parameter
- :observedProperty SDN:P01
- :procedure

om:observedProperty

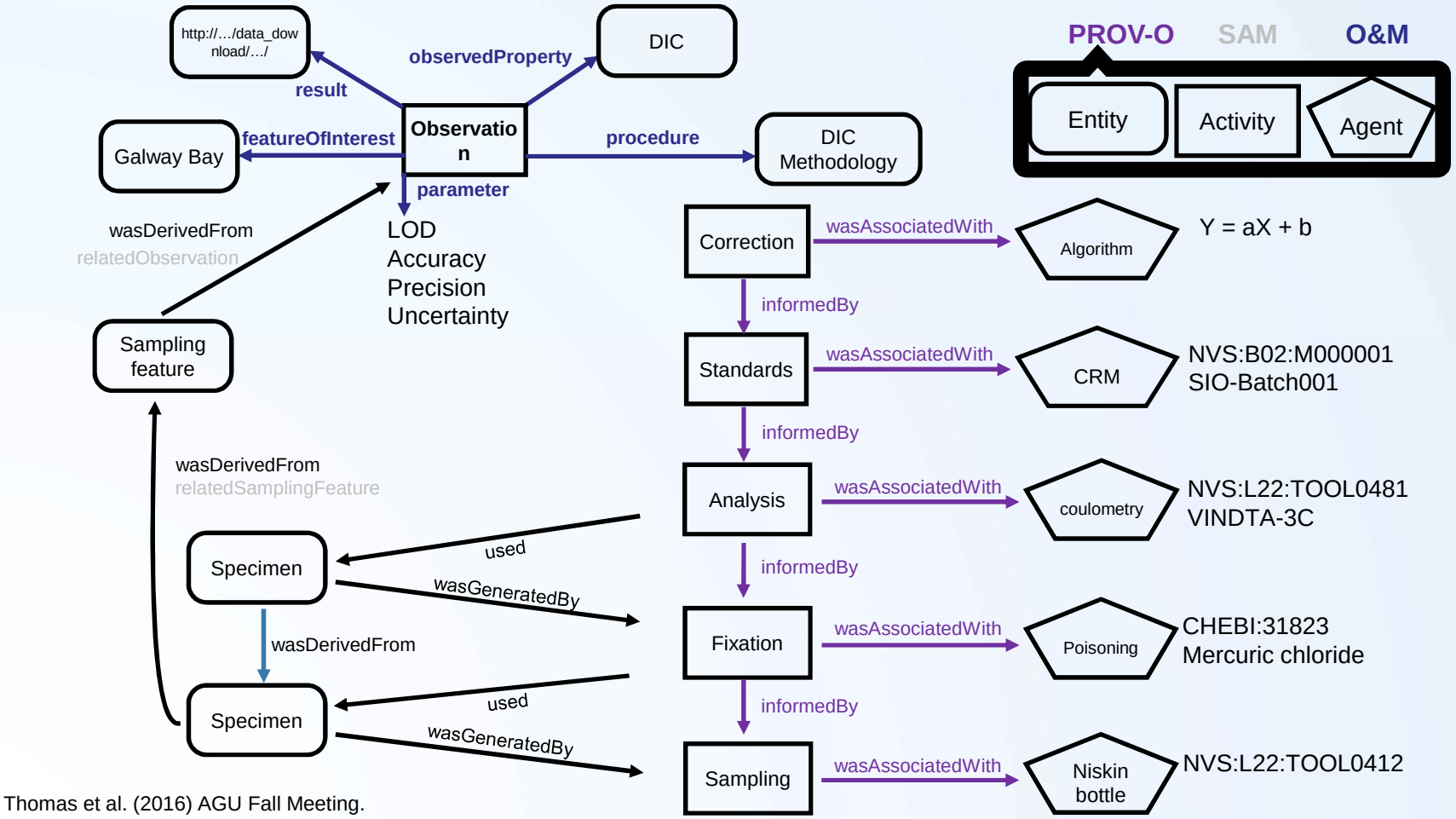
- cpm:statisticalMeasure SDN:S07
- cpm:property
- cpm:objectOfInterest SDN:S27
- cpm:matrix SDN:S26
- cpm:unitOfMeasure SDN:P06

Future directions?

Enrich datasets with usage metadata



Extend Observation and Measurements (ISO 19156) for the procedure



Thomas et al. (2016) AGU Fall Meeting.

Future directions?

Everything above is “metadata”…

Linked ‘meta’data

.

.

.

Linked ‘observational’ data

Future directions?

A tabular dataset is one organized primarily in terms of a grid of rows and columns. For pages that embed tabular datasets, you can also create more explicit markup, building on the basic approach described above. At this time we understand a variation of CSVW ("CSV on the Web"), provided in parallel to user-oriented tabular content on the HTML page.

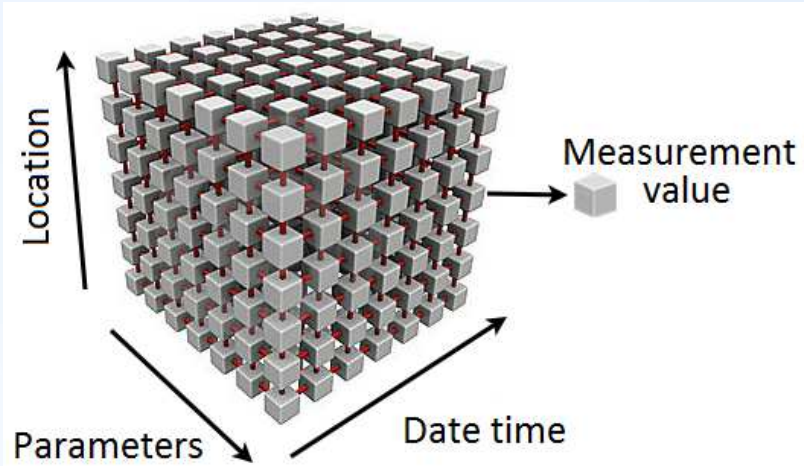
https://developers.google.com/search/docs/data-types/dataset

Future directions?

A tabular dataset is one organized primarily in terms of a grid of rows and columns. For pages that embed tabular datasets, you can also create more explicit markup, building on the basic approach described above. At this time we understand a variation of CSVW ("CSV on the Web"), provided in parallel to user-oriented tabular content on the HTML page.

https://developers.google.com/search/docs/data-types/dataset

Future directions?



W3C Recommendation

W3C

The RDF Data Cube Vocabulary

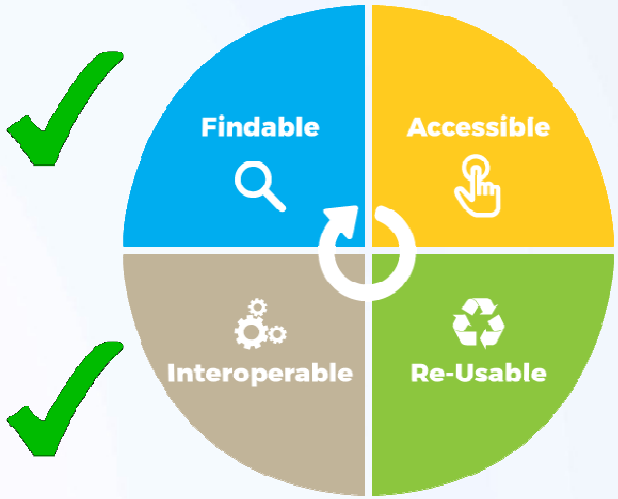
W3C Recommendation 16 January 2014

opengovintelligence

Why?

- In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable.
  - "H2020 Programme Guidelines on FAIR Data Management in Horizon 2020"

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (2016).



CSVW
RDF DataCubes

Enriched usage metadata

SeaDataCloud

Applying Linked Data principles for SeaDataNet Catalogues

Rob Thomas and Adam Leadbetter
Marine Institute, Ireland

SeaDataNet Plenary, Barcelona, 8-9 November 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org