Next frontier in open earth science: Interoperable information about observed environments

Markus Stocker

PANGAEA
  Data Publisher for Earth & Environmental Science
MARUM
  Center for Marine Environmental Sciences
University of Bremen
Germany

Twitter: @envinf
Email: mstocker@marum.de
ORCID: orcid.org/0000-0001-5492-3212

Open Science

- … is more than Open Data [EGU 2017]
- Principles, open ...
  - Access and Data
  - Source and Methodology [Kraker et al.]
  - Peer Review and Educational Resources [Open Science as a Practice]
  - Reproducible Research, Science Evaluation, Science Policies, Science Tools [FOSTER]
- The Open Definition: defines "openness" in relation to data and content
- Freely used, modified, and shared by anyone for any purpose

EGU 2017, Is Open Science the way to go?
Kraker et al. (2011), https://doi.org/10.1504/IJTEL.2011.045454
Open Science as a Practice, http://openscienceasap.org/open-science/
Facilitate Open Science Training for European Research (FOSTER), https://www.fosteropenscience.eu
Open Knowledge International, http://opendefinition.org

Open Data

… and Open Methodology

Open Data is great

but

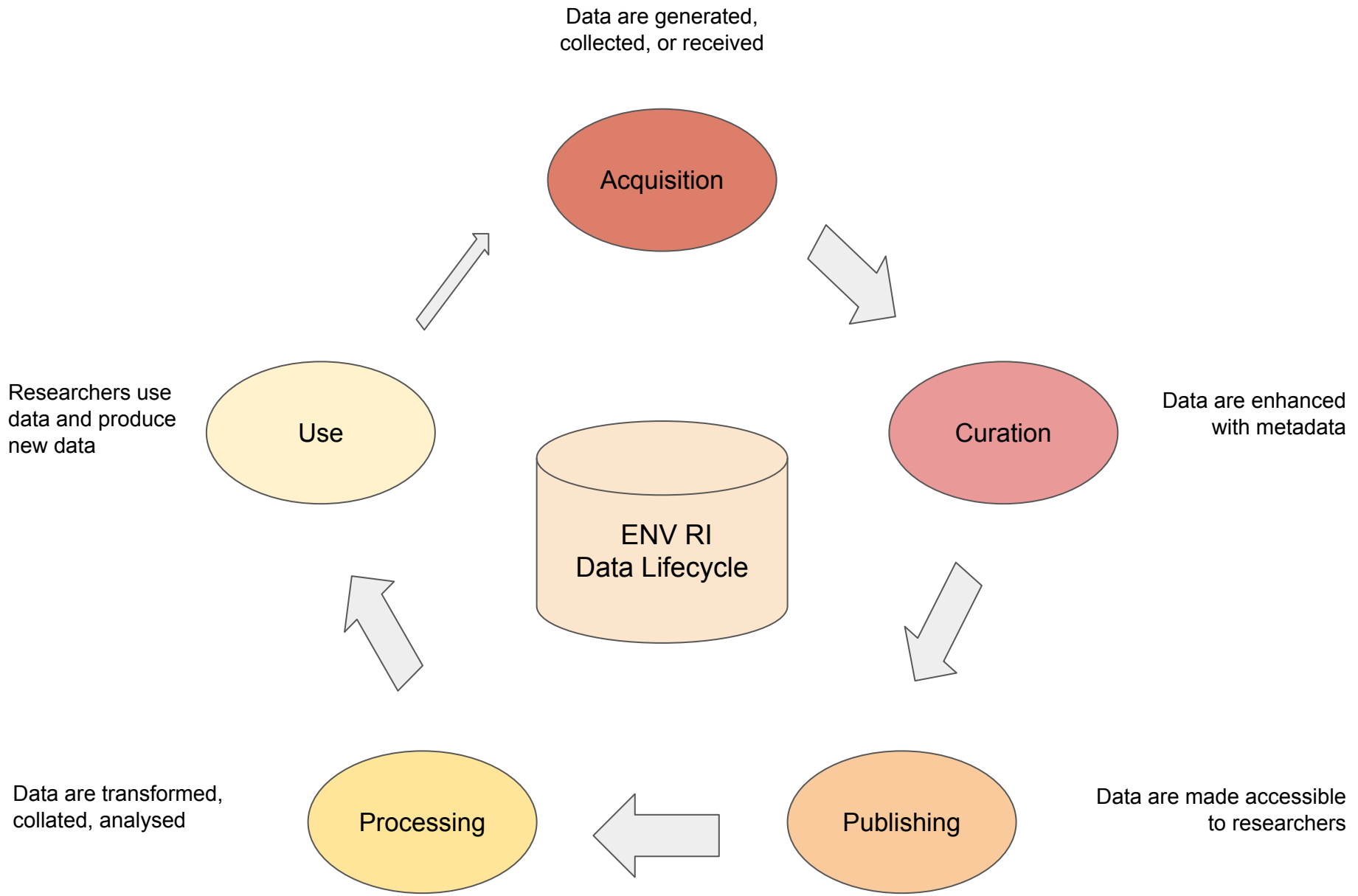we need to take a closer look at information

734546 0 1 0 0 0 0 0 0 0 0 0 1
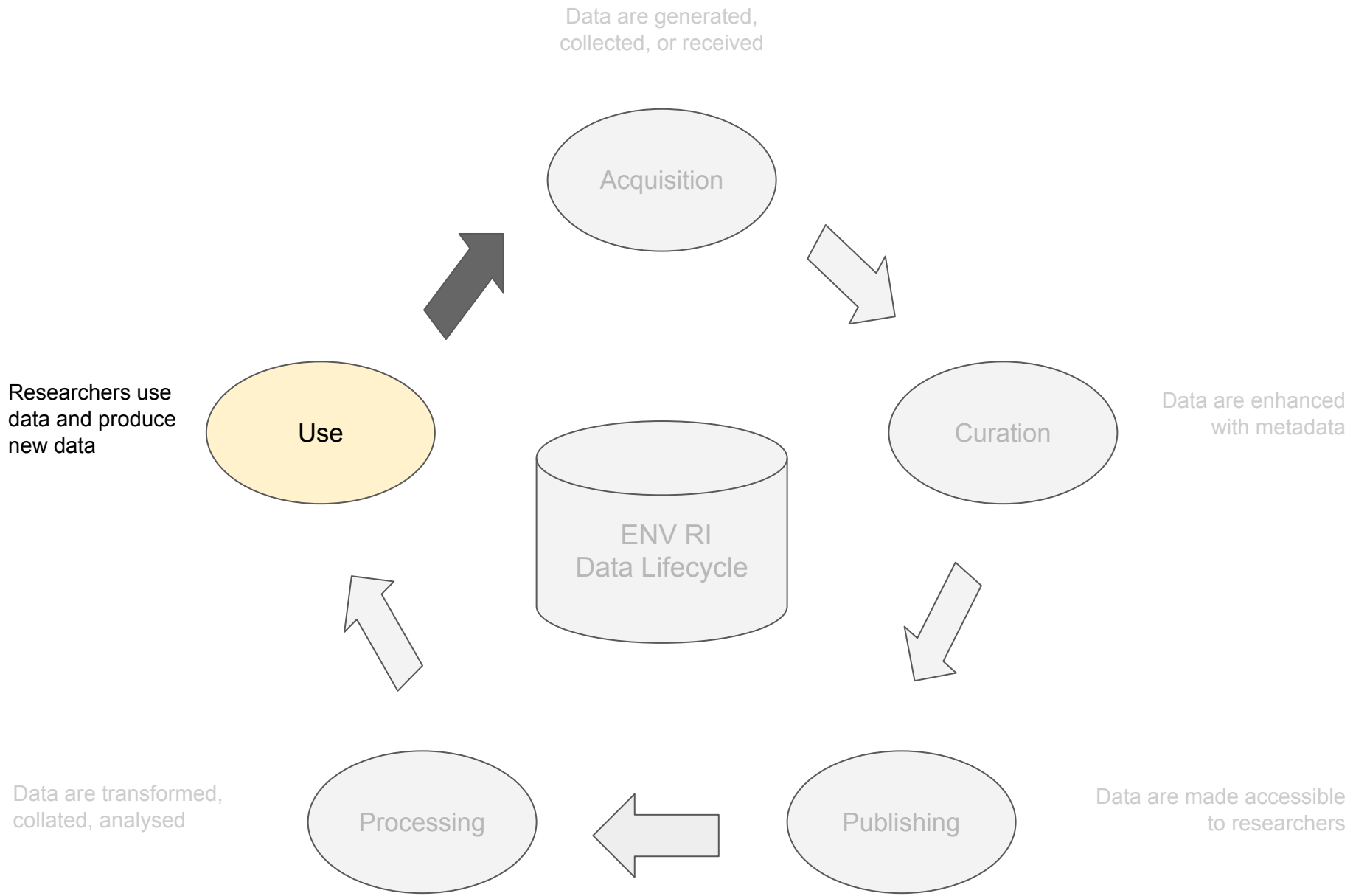
```
[] a lode:Event ;
    lode:atTime [
        time:inXSDDateTime "2011-02-12"
    ] ;
    smear:hasClassification [
        rdfs:label "Class Ib" ;
    ]
] .
```

Open Semantic Content

or

Meaningful structured data

Floridi, L. (2011). The Philosophy of Information. Oxford University Press.

Semantic content
about the observed environment
derived from primary observational data
by research communities

Data are generated,
collected, or received

Acquisition

Data are enhanced
with metadata

Curation

ENV RI
Data Lifecycle

Researchers use
data and produce
new data

Use

Data are transformed,
collated, analysed

Processing

Publishing

Data are made accessible
to researchers

ENVRI Reference Model, http://envri.eu/rm

Data are generated, collected, or received

Acquisition

Researchers use data and produce new data

Use

ENV RI Data Lifecycle

Curation

Data are enhanced with metadata

Data are transformed, collated, analysed

Processing

Publishing

Data are made accessible to researchers

ENVRI Reference Model, http://envri.eu/rm

Data use

- Scientists interpret observational data
- For their meaning in the context of research investigations
- Result is semantic information about the observed environment
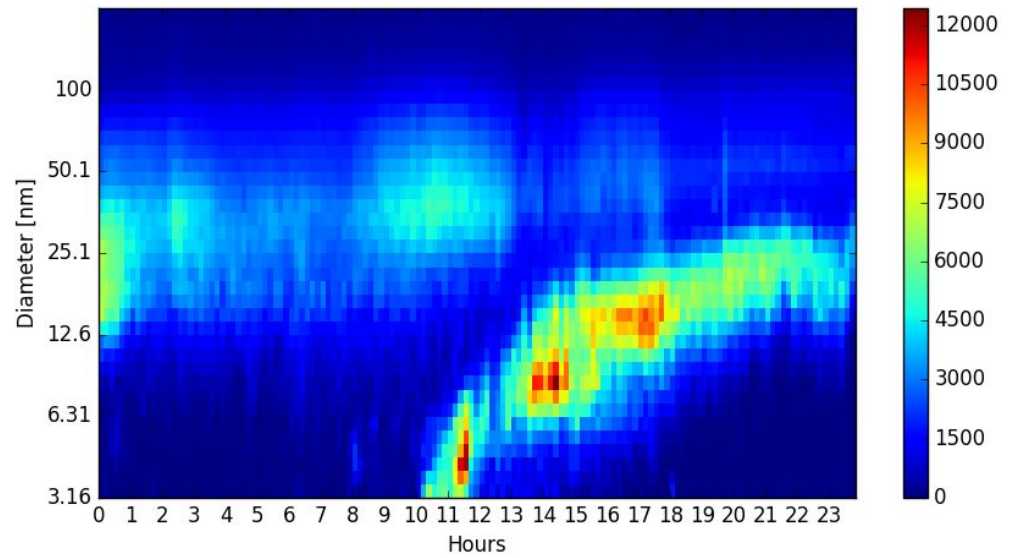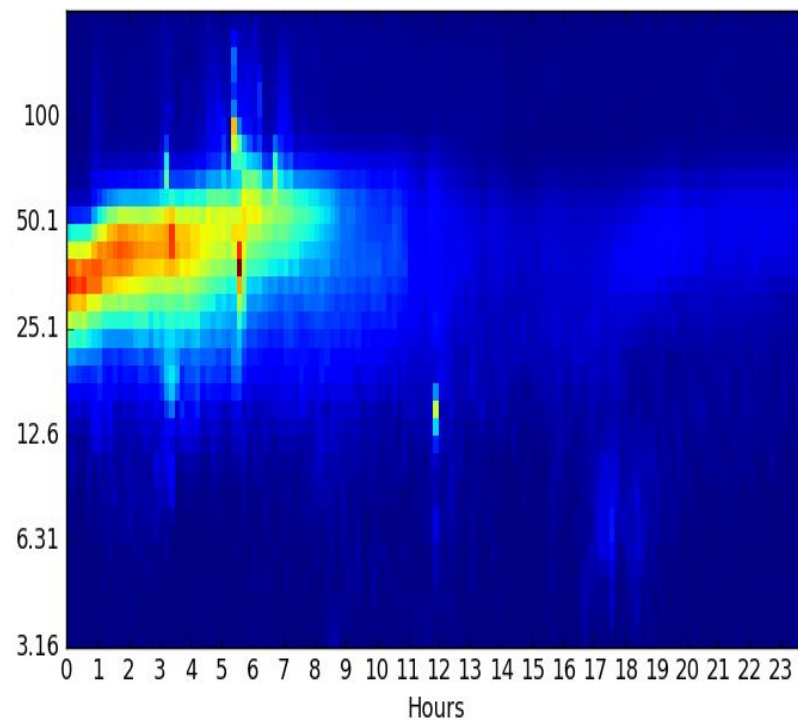- Semantic information is *meaningful* and truthful structured data

Floridi, L. (2011). The Philosophy of Information. Oxford University Press.
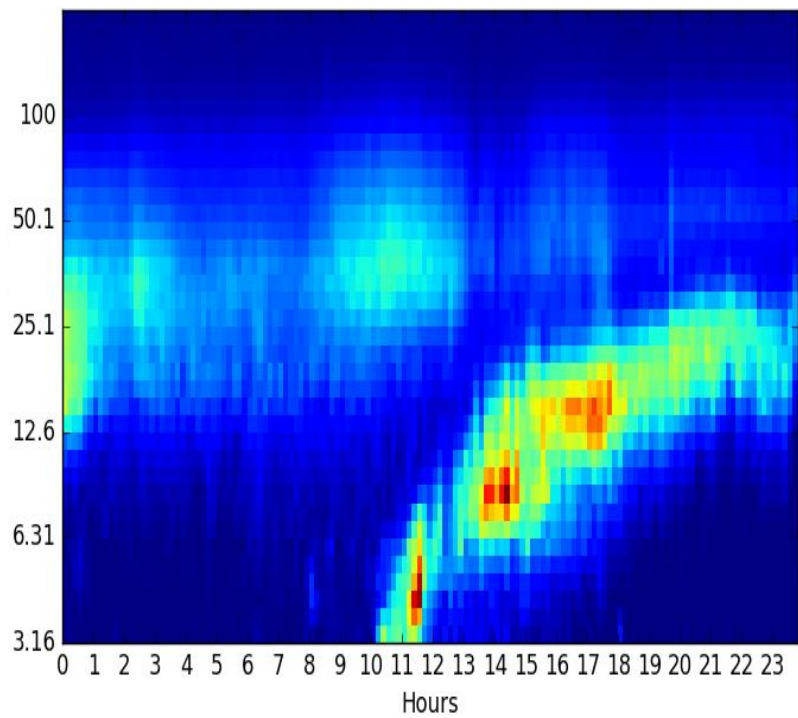
Concern I

Meaning gained in interpretation is lost
when information is recorded

Concern II

Lack of curation in interpretation means
research is not reproducible

Use case

Diameter [nm]

100
50.1
25.1
12.6
6.31
3.16

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Hours

12000
10500
9000
7500
6000
4500
3000
1500
0

734544 0 0 0 0 0 0 1 0 0 1
734545 0 0 0 0 0 0 0 1 0 0 1
734546 0 1 0 0 0 0 0 0 0 0 1
734547 0 0 0 0 0 0 0 1 0 0 1
734548 0 0 0 0 0 0 1 0 0 1
734549 0 0 0 0 0 0 1 0 0 1
734550 0 0 0 0 0 0 0 1 0 0 1
734551 0 0 0 0 0 0 0 1 0 0 1
734552 0 0 1 0 0 0 0 0 0 0 1
734553 0 0 0 0 0 0 0 1 0 0 1
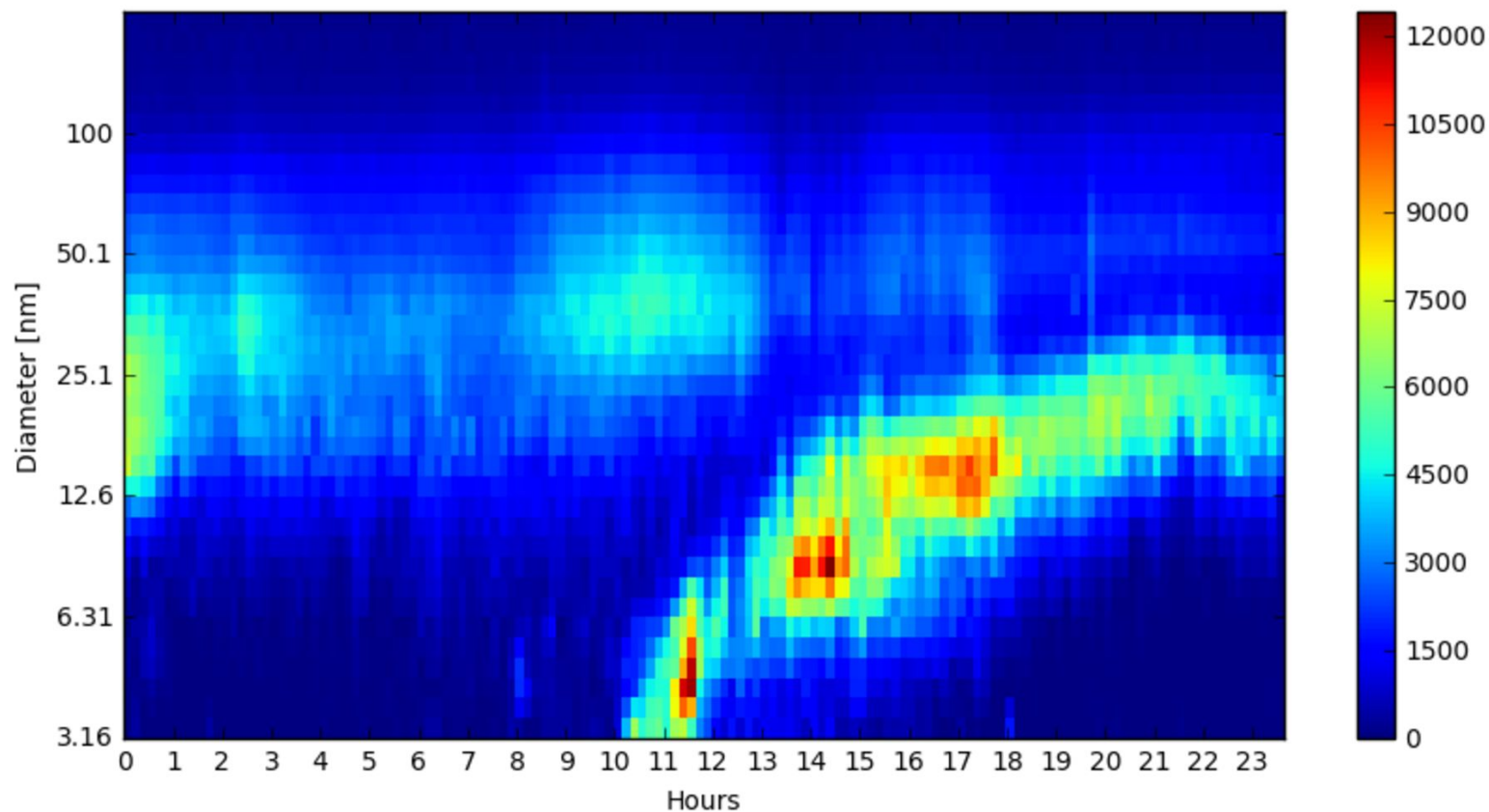734554 0 0 0 0 0 0 1 0 0 1
734555 0 0 0 0 0 0 0 1 0 0 1

Record *meaning*
gained in interpretation

```python
from smear.datafetcher import fetchdata
from smear.dataplotter import plotdata

# Fetch and plot concentration data for the given time and location
# from SmartSMEAR, https://avaa.tdata.fi/web/smart
plotdata(fetchdata('2013-04-04', 'Hyytiälä'))
```
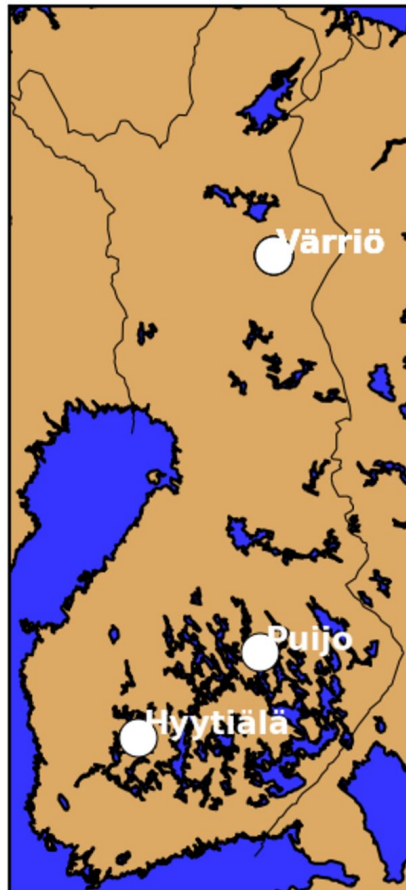
```python
from factory import record, event

# Record information about the new particle formation event
record(event('2013-04-04', 'Hyytiälä', '11:00', '19:00', 'Class Ia'))
```

```
from processing.visualization import map
from factory import events

# Visualize the locations of events on a map
map(events())
```

```python
from processing.description import describe
from factory import events

# Describe an event in plain English text
describe(events('2013-04-04', 'Hyytiälä'), format='text')
```

A Class Ia event occurred at Hyytiälä (FI)
[http://www.geonames.org/656888/hyytiaelae.html]
on 2013-04-04 starting at 11:00 and ending at 19:00.

```python
from factory import events
from processing.description import describe

# Describe an event with information in machine readable format
describe(events('2013-04-04', 'Hyytiälä'), format='rdf')
```

```
<http://pangaea.de/eb1ad69f11aecd2449f6d5741c3b8ac3> a lode:Event ;
    smear:hasClassification smear:ClassIa ;
    lode:atPlace <http://sws.geonames.org/656888/> ;
    lode:atTime <http://pangaea.de/92be5465a05cc56156422d6cdb4603e1> ;
    lode:inSpace <http://pangaea.de/7f885190eb43154e01c97f814b287a4b> .


<http://pangaea.de/3ca2edb9bb13c998a0ea94082aa592dd> a time:Instant ;
    time:inXSDDateTime "2013-04-04T11:00:00+03:00"^^xsd:dateTime .


<http://pangaea.de/7f885190eb43154e01c97f814b287a4b> a sf:Point,
        wgs84:SpatialThing ;
    geosparql:asWKT "POINT (24.29077 61.84562)"^^geosparql:wktLiteral .


<http://pangaea.de/92be5465a05cc56156422d6cdb4603e1> a time:Interval ;
    time:hasBeginning <http://pangaea.de/3ca2edb9bb13c998a0ea94082aa592dd> ;
    time:hasEnd <http://pangaea.de/cc87885836375b64bf023a6bcd5ad814> .


<http://pangaea.de/cc87885836375b64bf023a6bcd5ad814> a time:Instant ;
    time:inXSDDateTime "2013-04-04T19:00:00+03:00"^^xsd:dateTime .


<http://sws.geonames.org/656888/> a gn:Feature,
        DUL:Place ;
    gn:countryCode "FI"^^xsd:string ;
    gn:locationMap <http://www.geonames.org/656888/hyytiaelae.html> ;
    gn:name "Hyytiälä"^^xsd:string ;
    wgs84:lat "61.84562" ;
    wgs84:long "24.29077" .
```
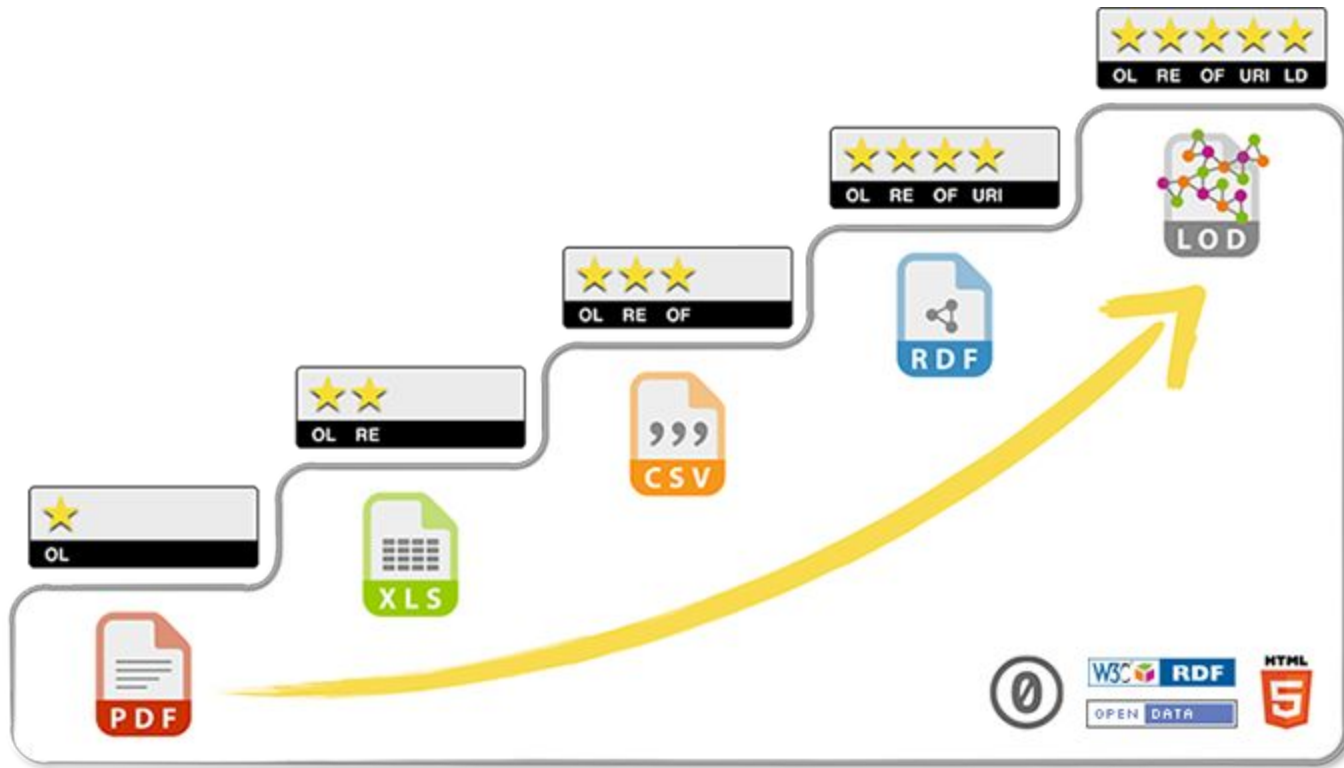
http://5stardata.info/en/

European Research Infrastructure for the observation of Aerosol, Clouds, and Trace gases

http://www.actris.eu

@ACTRISRI

Establishing WG, From Observational Data to Information (D2I WG)

Take aways

- We need Open *Meaningful* Data, semantic content
- In particular about the observed environment
- Make sure meaning gained in interpretation is explicit
- Use semantic technologies for formal representation
- Jupyter notebooks for open methodology