

Agenda

1. **(13:00) Update on project admin (contracts, CA etc) [15mins]**
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for BADC/new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

Project admin

- Finalised project docs need sending to JISC
 - Project plan
 - Work packages
 - Project budget
- Consortium Agreement
 - Drafted by UEA, agreed by STFC
 - Now just needs to be signed
 - UEA holds this at present

Project plan

- Covers:
 - Background, aims & objectives, overall approach
 - Project outputs & outcomes
 - Stakeholders, risk analysis
 - Use of standards
 - Technical development approach
 - IPR
 - Partners, project management
 - Budget, workpackages
 - {Evaluation, Quality, Dissemination, Exit & sustainability} plans

Aims & objectives

- develop an information architecture addressing some of the scientific data workflows in climate research
- deploy infrastructure to capture relevant metadata for climate research data, software, and workflows
- develop a 'linked-data' approach to publishing and citing climate research data
- prototype our approach using four high-profile climate research datasets: CRUTEM, CRU TS, tree-ring chronologies, and HadCET

etc...

- Stakeholders:
 - BADC, Met Office, UK Location Programme, Linked-data/data.gov.uk, academic publishers
- Biggest risks:
 - low take-up of results
 - loss of staff
 - too ambitious
- Standards:
 - Linked data (RDF, SPARQL), OGC, OAI-ORE, NetCDF, CSML, ...
- Partners:
 - UEA: lead partner, data holder
 - STFC: tech lead, project mgt
 - Met Office: interested party

Work packages

WORKPACKAGES	Month	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
1: Project management													
2: Requirements analysis													
3: Research data management													
4: Data linking, citing, integrating													
5: Prototype													
6: Engagement													

Budget

- Budget covers
 - roughly 0.5FTE Arif and Colin over 12 months
 - thus almost full-time for remainder of project
 - roughly 0.1FTE STFC project manager
 - STFC ‘in-kind’ contribution of ~0.5FTE (GeoTOD)
 - £1k each T&S
- Late start!
- No STFC budget spent as yet.... 😊

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
- 2. (13:15) Project handover to BADC [15mins]**
3. Project planning [30mins]
4. ACRID overview for BADC/new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

Project handover to BADC

- Sarah Callaghan new BADC project manager
 - excellent experience in data citation (previous CLADDIER and OJIMS projects) and project management
- CA/(sub)contract implications?
- Practical aspects ESC \leftrightarrow BADC

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
- 3. (13:30) Project planning [30mins]**
4. ACRID overview for BADC/new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

Project planning

WORKPACKAGE 1: Project Management <u>Objective:</u> This work package will ensure the smooth running of the project, and that project milestones and deliverables are achieved. It will coordinate the other work packages and follow JISC project management guidelines.					
1. <i>Project plan:</i> A detailed project plan, work plan and budget will be developed in the first month of the project. A consortium agreement will be established.	1-Aug-2010	15-Dec-2010	<ul style="list-style-type: none"> D1.1a (15-Dec-2010): Detailed project plan, work plan, budget D1.1b (15-Dec-2010): consortium agreement. 	1	AW
1. <i>Project meetings:</i> Overall project meetings will be held at least once per month over the duration of the project, with smaller working meetings held as needed.	1-Aug-2010	31-Jul-2011			TO/AW
1. <i>Project website:</i> A project website will be established and maintained throughout the project. This will provide the primary broad dissemination and engagement point.	1-Aug-2010	31-Jul-2011	<ul style="list-style-type: none"> D1.2 (1-Oct-2010): Initial project website 	2	TO
1. <i>Reporting:</i> This task is a combination of all project reporting, including mid-term progress and final reports.	1-Aug-2010	31-Jul-2011	<ul style="list-style-type: none"> D1.3 (1-Feb-2011): Mid-term report D1.4 (31-Jul-2011): Final project/budget reports 	9	AW

Project planning

WORKPACKAGE 2: Requirements analysis <u>Objective:</u> The first work of the project will analyse existing workflows around the target climate research datasets to understand requirements for collecting and linking data provenance, software configuration, and other metadata.					
1. Analyse workflows: Scientific workflows associated with the prototype climate research datasets will be analysed in order to identify the requirements for capturing software and dataset metadata (including versioning, provenance and configuration). General patterns of information flow for climate research data will be identified to ensure generality of the information architecture developed and applicability to other HEIs.	1-Dec-2010	1-Feb-2011	<ul style="list-style-type: none"> D2.1 (1-Feb-2011): Description of scientific workflows 		AW/TO
1. Develop information architecture: An information architecture will be developed, addressing the following: <ul style="list-style-type: none"> required data management infrastructure to capture versioning information and metadata, interaction with external data centres (e.g. BADC), etc. an information model (including RDF vocabularies) for the climate research datasets deployment architecture (indicating software components that need to be developed) 	1-Jan-2011	1-Feb-2011	<ul style="list-style-type: none"> D2.2 (1-Feb-2011): Information architecture 	3	AW

Project planning

WORKPACKAGE 3: Research data management <u>Objective:</u> This work package will design and develop the required data management infrastructure to ensure capture of relevant metadata for climate research data, software, and workflows.					
1. <i>Software management:</i> Tooling will be developed and deployed to ensure adequate management of the software artefacts involved in processing climate data - including software versioning, configuration, publishing, and run-time metadata capture.	1-Jan-2011	1-Jun-2011	<ul style="list-style-type: none"> D3.1 (1-Jun-2011): Deployed tools for managing <i>climate research software</i> 	5	TO
1. <i>Data management:</i> This task will focus on the climate data themselves, developing tools for managing update and versioning of datasets, capturing the workflow associated with their generation, and enabling metadata export.	1-Jan-2011	1-Jun-2011	<ul style="list-style-type: none"> D3.2 (1-Jun-2011): Deployed tools for managing <i>climate research data</i> 	5	TO

Project planning

WORKPACKAGE 4: Data linking, citing, integrating <u>Objective:</u> This work package is concerned with publishing climate research data in citeable linked-data form, enabling integration with other datasets through use of common data models. CSML explicitly enables integration of the following elements across distributed climate datasets: instrument/station metadata, observed parameters (e.g. using vocabulary services), observation timeseries.					
1. <i>RDF vocabularies:</i> A relevant data model is needed for both climate research data, and the aggregation of digital objects (raw and processed data, software, metadata) associated with their generation and scholarly publication. This work package will adopt existing models (OAI-ORE, CSML) in a linked-data framework to develop a best practice approach for climate research data, enabling integration with other distributed datasets and tools.	1-Feb-2011	1-Apr-2011	<ul style="list-style-type: none"> D4.1 (1-Apr-2011): RDF-based climate research data model 	4	AW
1. <i>Linked-data server:</i> A linked-data server will be configured to provide access to the web of data objects used in climate research. It will implement content-negotiation and support SPARQL querying.	1-Mar-2011	1-Jul-2011	<ul style="list-style-type: none"> D4.2 (1-Jul-2011): Deployed linked-data server 	6	AW
1. <i>Data representations:</i> Software will be developed to transform climate research data to a number of different representations (RDF, HTML, KML, netCDF) for the linked-data server.	1-Mar-2010	1-Jul-2011	<ul style="list-style-type: none"> D4.3 (1-Jul-2011): Software for rendering data in different forms 	6	AW/TO
1. <i>Citation infrastructure:</i> A mechanism will be developed to enable citation of climate research data objects within the context of scholarly communication. This task will adopt approaches developed within the previous CLADDIER and OJIMS projects.	1-May-2011	1-Aug-2011	<ul style="list-style-type: none"> D4.4 (1-Aug-2011): Citation infrastructure 	7	AW

Project planning

WORKPACKAGE 5: Prototype Objective: The approaches and tools developed in WP2-4 will be applied in this work package to demonstrate the full chain of climate research data citation, linking, and integration for a number of target datasets. This will validate the approach and demonstrate the added value of managing and exposing the full set of data-related artefacts (e.g. software configuration, workflow metadata, provenance) in the context of scholarly publication.					
1. <i>CRUTEM</i> : This reference dataset of monthly global gridded land temperatures is generated by processing station observation data, and is described in a series of highly-cited papers (e.g. Brohan et. al. (2006)). We will collate raw data and metadata around its generation, aggregating the components using OAI-ORE, exposing as linked-data, and integrating using data models that are interoperable with other climate research datasets and tools (i.e. using CSML and NetCDF where appropriate and feasible).	1-Apr-2011	1-Aug-2011	• D5.1 (1-Aug-2011): CRUTEM dataset exposed	8	TO
1. <i>CRU TS</i> : The CRU TS dataset includes multiple climate variables interpolated at a relatively high spatial resolution, and updated twice per year. A complex chain of processing is performed on a large number of raw observational datasets. A research publication is currently in preparation, and will provide a valuable opportunity to validate the entire framework developed in this project.	1-Apr-2011	1-Aug-2011	• D5.2 (1-Aug-2011): CRU TS dataset exposed	8	TO
1. <i>Tree-ring chronologies</i> : Tree-rings are a widely used proxy in paleo-climate reconstruction, but the relevant chronologies developed are highly dependent on an empirical 'standardisation' procedure. We will capture the process associated with production of tree-ring chronologies and represent the entire workflow (including processing software and parameters) in a linked, citable, and integrated form.	1-Apr-2011	1-Aug-2011	• D5.3 (1-Aug-2011): Tree-ring data exposed	8	TO
1. <i>HadCET</i> : The Met Office Hadley Centre's 'Central England Temperature' research dataset is the longest instrumental record of temperature in the world. We will generate a linked-data representation of this important dataset, providing a timely exploration of linked-data principles applied to Met Office research data.	1-Apr-2011	1-Aug-2011	• D5.4 (1-Aug-2011): HadCET dataset exposed	8	AW/Met Office

Project planning

WORKPACKAGE 6: Engagement Objective: A major element of this project will be engagement with JISC and key stakeholder communities in climate research, linked-data, and open public data. This will ensure maximum information exchange and impact of project outcomes. The main result of the project will be a prototype information architecture for citing, linking, and integrating climate research data - this will provide a landmark exemplar that may be widely exploited by climate researchers across the academic and government sectors.					
1. <i>Academic/research community:</i> The project will engage with the JISC 'Managing Research Data Programme', fully participating in relevant activities. As well, project outcomes will be presented at relevant research meetings of the academic community (e.g. the UK e-Science All-Hands Meeting). Climate research falls within the remit of the Natural Environment Research Council, and project results are likely to be relevant to the NERC Science Information Strategy. We will engage with the SIS Architecture Project to promote the project results.	1-Dec-2010	1-Aug-2011	• M6.1.x (<i>various dates</i>): Participate in JISCMRD activities		AW/TO
1. <i>Public data community:</i> Since climate research is today a matter of significant public policy interest, the results of this project will be relevant to the UK's public data initiative. We will engage with The National Archives and the UK Location Programme, jointly taking responsibility for developing the UK approach to linked public geospatial data. We will also work with the Met Office on developing approaches to exposing public climate datasets.	1-Dec-2010	1-Aug-2011	• M6.2.x (<i>various dates</i>): Specific knowledge exchange meetings held with UK LP/TNA; also with Met Office. Meetings will be arranged as opportunities arise.		AW
1. <i>Climate research community:</i> The European Geosciences Union recently established a new scientific division on 'Earth and Space Science Informatics', recognising the growing importance to Earth science of advanced data and computational techniques. We will promote project outcomes to the climate research informatics community.	1-Dec-2010	1-Aug-2011	• M6.3.x (<i>various dates</i>): Results presented at meetings/workshops of the climate research community. Key events will be identified as opportunities arise.		TO
1. <i>Linked-data community:</i> The semantic web and linked-data community has focussed primarily on relatively 'static' datasets. Climate timeseries add a time dimension which may be relevant to other areas. RDF vocabularies and data models developed in this project may be of interest to this community, and so we will engage through online media and relevant meetings/workshops to ensure project outcomes are relevant and exposed.	1-Dec-2010	1-Aug-2011	• M6.4.x (<i>various dates</i>): Results presented at meetings of the linked-data community. Key events will be identified as opportunities arise.		AW

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
- 4. (14:00) ACRID overview for new people [15mins]**
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

JISC 14/09

- *Citing*: “Agreed conventions for data citation and for data description are important for research data discovery. Persistent identification is required...”
- *Linking*: “A recent position paper written for JISC ... makes a case for the benefits of linking research data using semantic or linked data technology ... data on which a journal article is based are bi-directionally linked to other data, resources, articles and people.”
- *Integrating*: “Integrating heterogeneous data across distributed sources can enable effective and innovative reuse”

Data citation: DOI

- JISC 14/09 refers to *DataCite* initiative
 - International consortium, incl. British Library, assigning DOIs to datasets
- Parsons and Duerr, EOS **91**(34), 297-304
 - “Digital Object Identifiers (DOIs) are increasingly being used as a way to precisely indicate which particular data set was used and to enable traceability to the original source.”
- Wilson et. al. (AGU abstract submission): “Enhancing The Recognition, Reusability, And Transparency Of Scientific Data Using Digital Object Identifiers”
- UNESCO (2010): “SCOR/IODE/MBLWHOI Library Workshop on Data Publication”
 - Focussed on DOI
- Earth System Science Data journal: provides DOIs for data publications

But: what does a DOI point to?

- Parsons and Duerr: “What is the citable unit with a DOI? A file? A collection of files? How many?”
- Wilson et. al.: “Currently, the DOI resolves to document that describes the data set and provides a link to the data set itself...”
- UNESCO: “The Meeting considered the question “what lies at the end of a DOI” and identified the following answers: (i) URL of the homepage of a Web data delivery system; (ii) Deep URL into a Web data delivery system; (iii) URL of a zip file; (iv) URL of a CF-compliant NetCDF file; and (vi) Archival Information Package (AIP) as specified in the Open Archival Information System (OAIS) model.”
- *Answer: linked-data?*

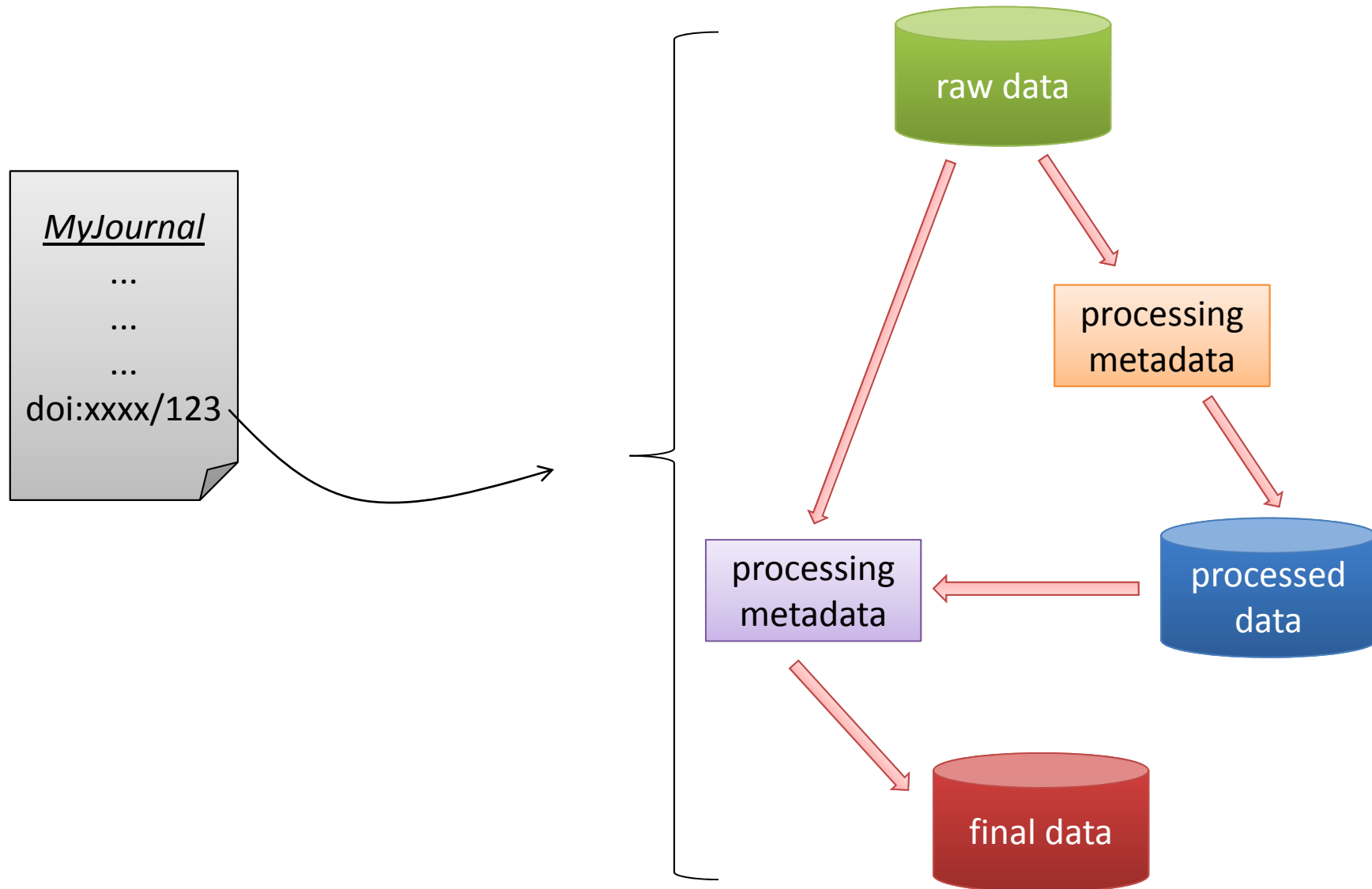
The 'linked-data cloud'

- Bizer, Heath, Berners-Lee (Linked Data – The Story So Far):
 - “Linked Data refers to a set of best practices for publishing and connecting structured data on the Web”
- [Linkeddata.org](http://linkeddata.org) 'linked data cloud'

data.gov.uk

- HMG (Dec. 2009) (*Putting the frontline first – smarter government*):
 - “we will aim for the majority of government-published information to be reusable, linked data by June 2011”
<http://www.hmg.gov.uk/media/52788/smarter-government-final.pdf>
- data.gov.uk provides entry point

Linked-data for ACRID



The web: *documents* to *data*

- Keys to success of the *document web*:
 - Ability to *identify* documents
 - <http://www.cru.uea.ac.uk>
 - Ability to *link* documents
 - `Climatic Research Unit`
- Need the same capability for data

The web: *documents* to *data*

- Document web is based on HTML

```
<html>
  <head>
    <title>My nice webpage</title>
  </head>
  <body>
    <h1>Introduction</h1>
    Here is my great webpage
  </body>
</html>
```

- Data web is based on RDF...

Resource Description Framework

- W3C language for “representing information about resources”
 - Identifies things with URIs (nodes)
 - Describes them with properties and property values (arcs)
- Example:
 - "there is a Person whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr."



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

(<http://www.w3.org/TR/rdf-primer/>)

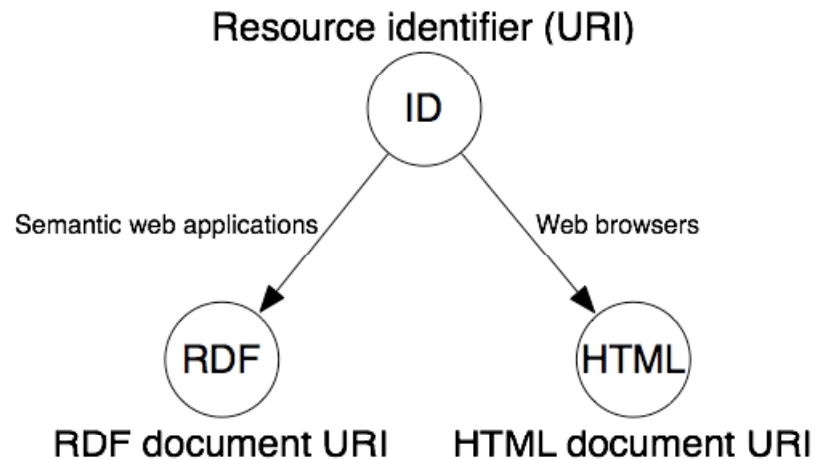
Linked-data principles (TBL)

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

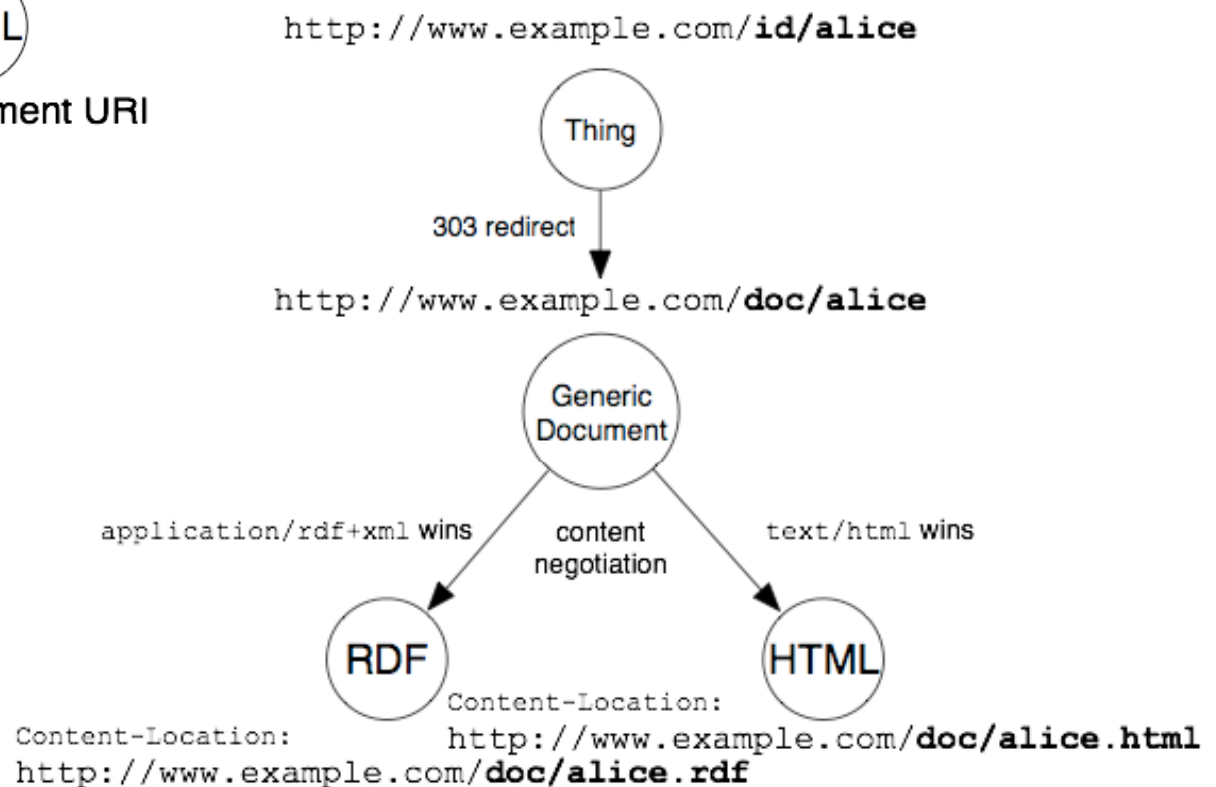
Designing 'URI sets'

- Distinguish between *real-world thing*, and a *description* of it
- May also have different *representations* of a description
- (Ref: 'Cool URIs for the semantic web', W3C)
- Cabinet Office (2009) "Designing URI Sets for the Public Sector"
 - Identifier URI: <http://education.data.gov.uk/id/school/121242>
 - Document URI: <http://education.data.gov.uk/doc/school/121242>
 - Representation URI:
<http://education.data.gov.uk/doc/school/121242.rdf>
 - Definition URI: <http://education.data.gov.uk/def/school/>

How it works...



(<http://www.w3.org/TR/cooluris/>)



Example

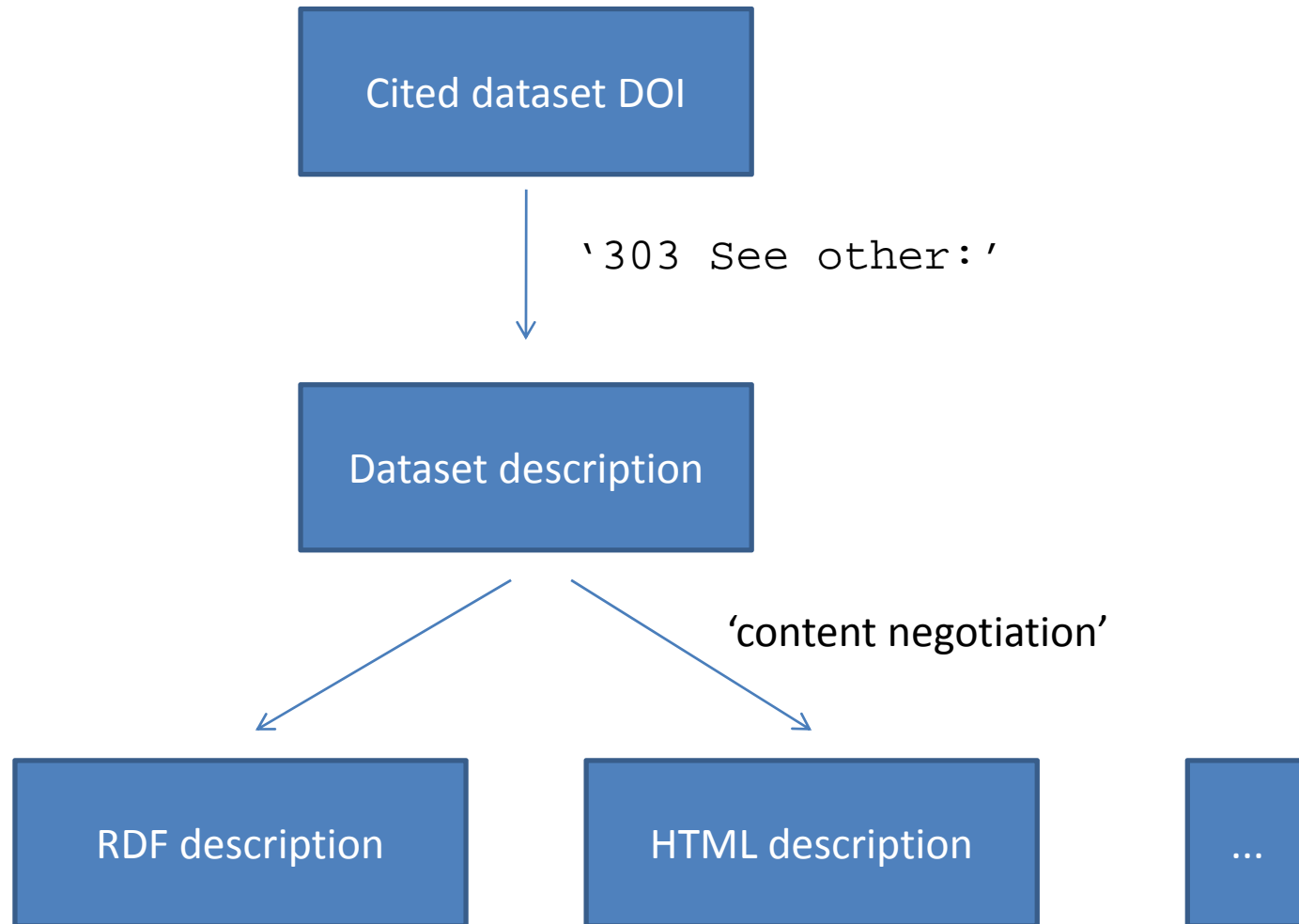
- `wget --max-redirect=0`
`"http://education.data.gov.uk/id/school/121242"`
- `wget -O - --header="Accept: text/html"`
`"http://education.data.gov.uk/doc/school/121242" | more`
- `wget -O - --header="Accept: application/rdf+xml"`
`"http://education.data.gov.uk/doc/school/121242" | more`
- Or direct representation URIs:
 - <http://education.data.gov.uk/doc/school/121242.rdf>
 - <http://education.data.gov.uk/doc/school/121242.json>
 - <http://education.data.gov.uk/doc/school/121242.html>
- <http://education.data.gov.uk/def/school/IndependentSchoolTypeMusic.rdf>
- [http://dbpedia.org/resource/Climatic Research Unit](http://dbpedia.org/resource/Climatic_Research_Unit)

Linked-data querying: SPARQL

- <http://data.gov.uk/sparql>

```
prefix sch-ont: <http://education.data.gov.uk/def/school/>
SELECT ?name WHERE {
  ?school a sch-ont:School;
    sch-ont:establishmentName ?name;
    sch-ont:districtAdministrative
      <http://statistics.data.gov.uk/id/local-authority-district/33UG> ;
}
ORDER BY ?name
```

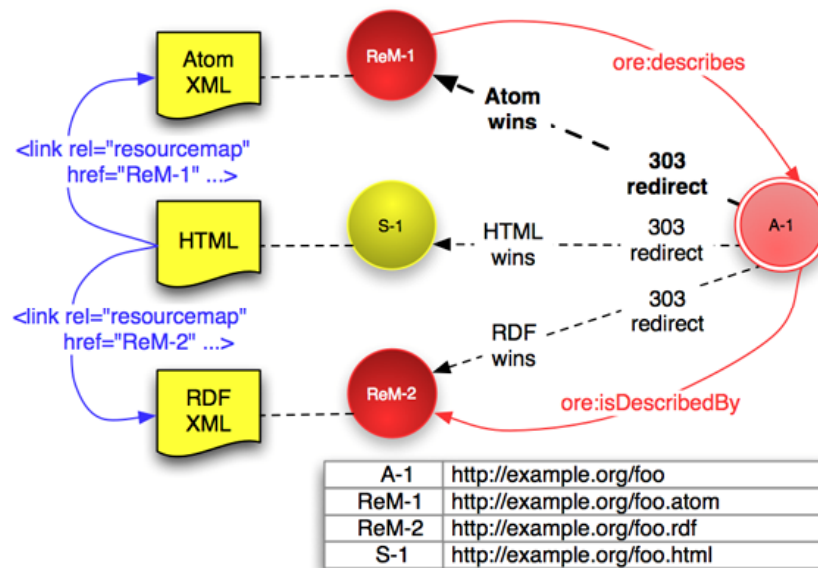
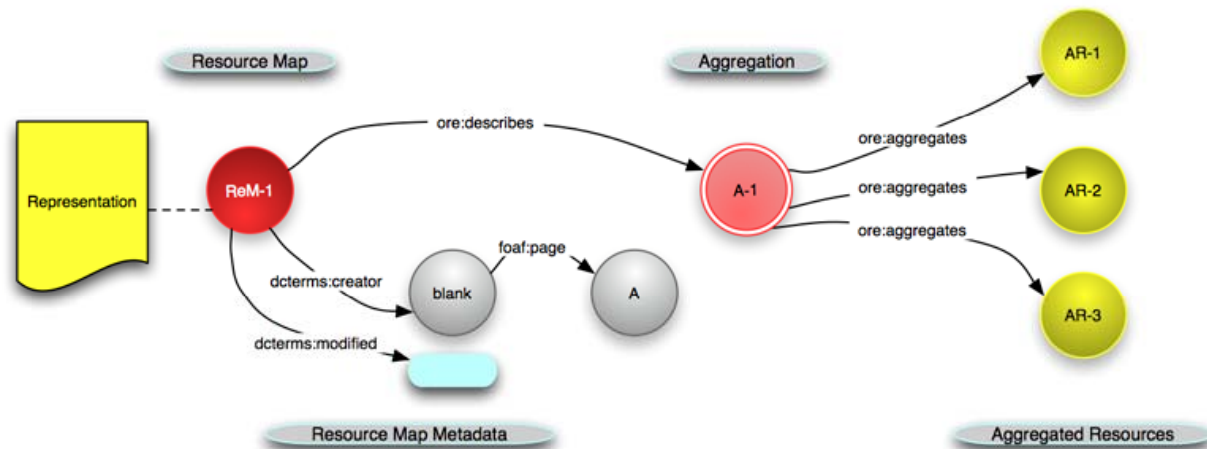
For ACRID



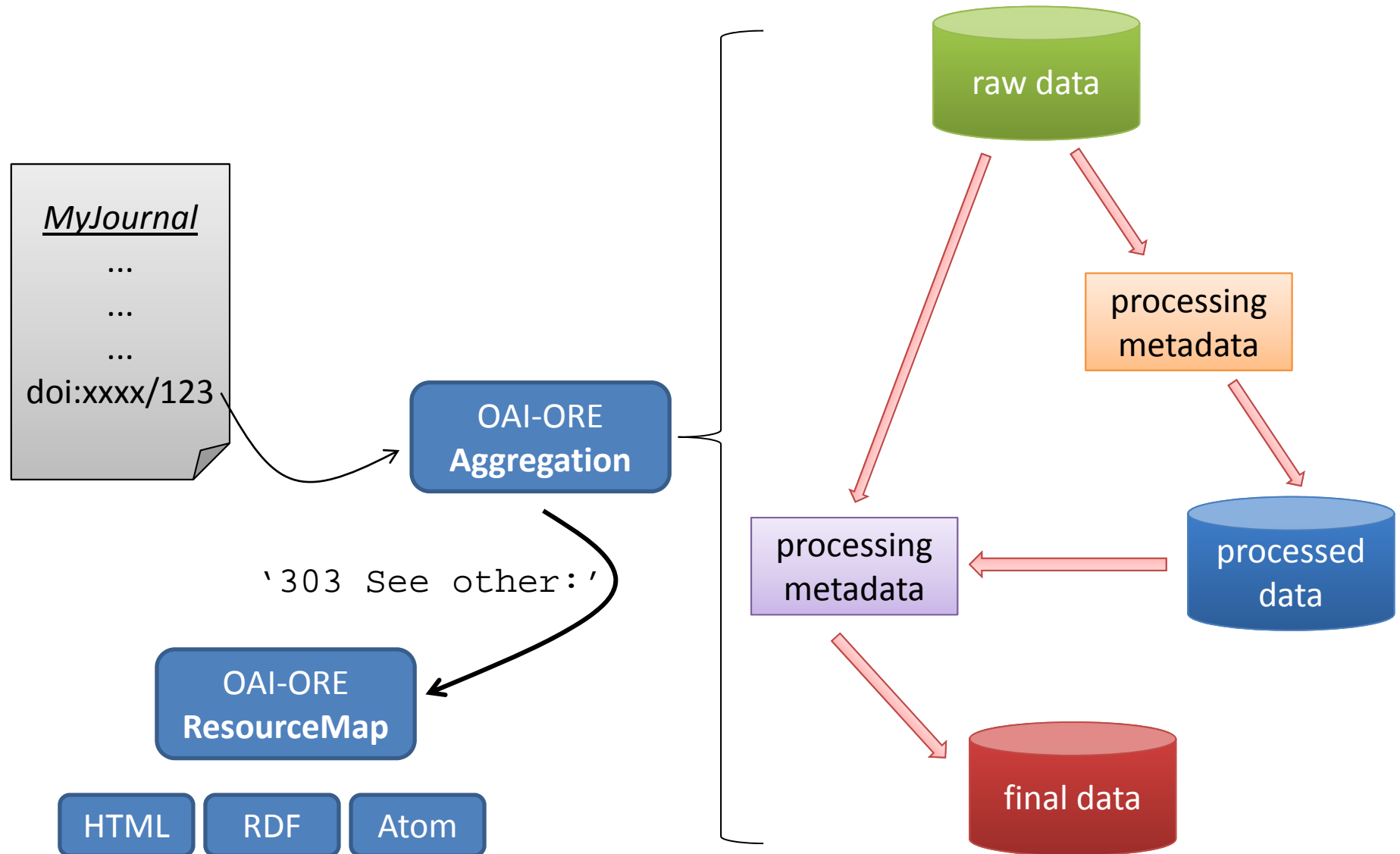
RDF vocabulary / ontology

- Open Archives Initiative Object Reuse and Exchange (OAI-ORE)
 - description and exchange of aggregations of Web resources
 - <http://www.openarchives.org/ore/>

OAI-ORE



Linked-data for ACRID – v2



Information model

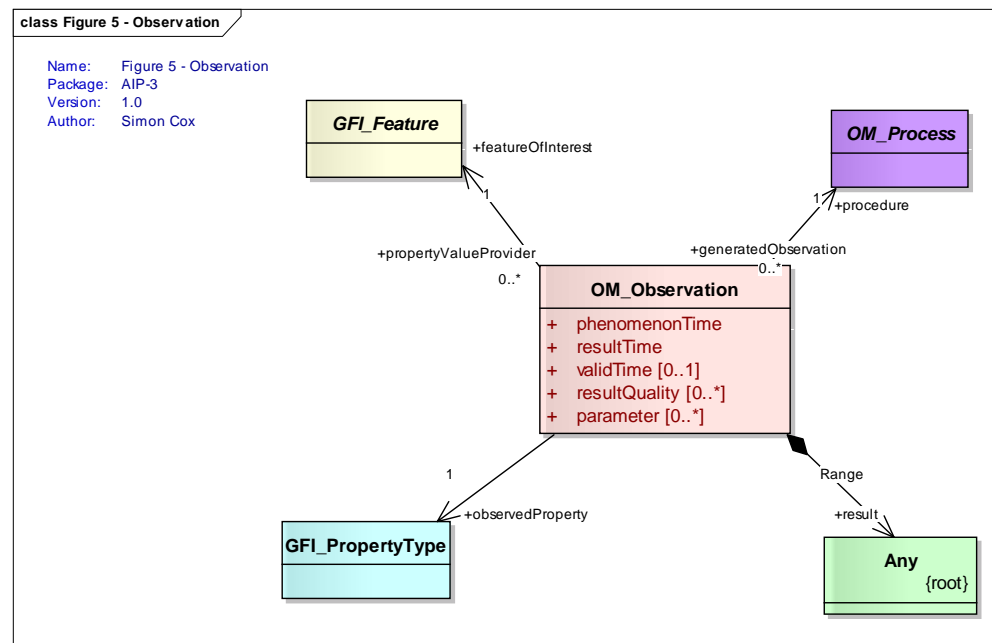
- The information model is key!
- Need a model that captures key aspects of data workflow
 - What are the ‘things’ we want to expose?
 - Are there general patterns that could be re-used (between our ‘test’ datasets, other domains)?

Issues

- Issues:
 - Information model itself
 - Dynamic/evolving data
 - “Many datasets are being updated on a rolling basis, adding new data as and when received. ... A citation, however, is supposed to link a reader back to the same publication which the citing author read. In the case of a dynamic dataset, linking back to the dataset as it was when an author used it to write a paper is clearly impossible. This poses a significant challenge.” (OECD, 2009)
 - Data subsets
 - ““Deep citation" or references to subsets of data sets, are analogous to page references in printed matter. ... Devising a simple standard for describing the chain of evidence from the data set to the subset would be highly valuable.”
 - Versioning
 - ...

'Integrating'


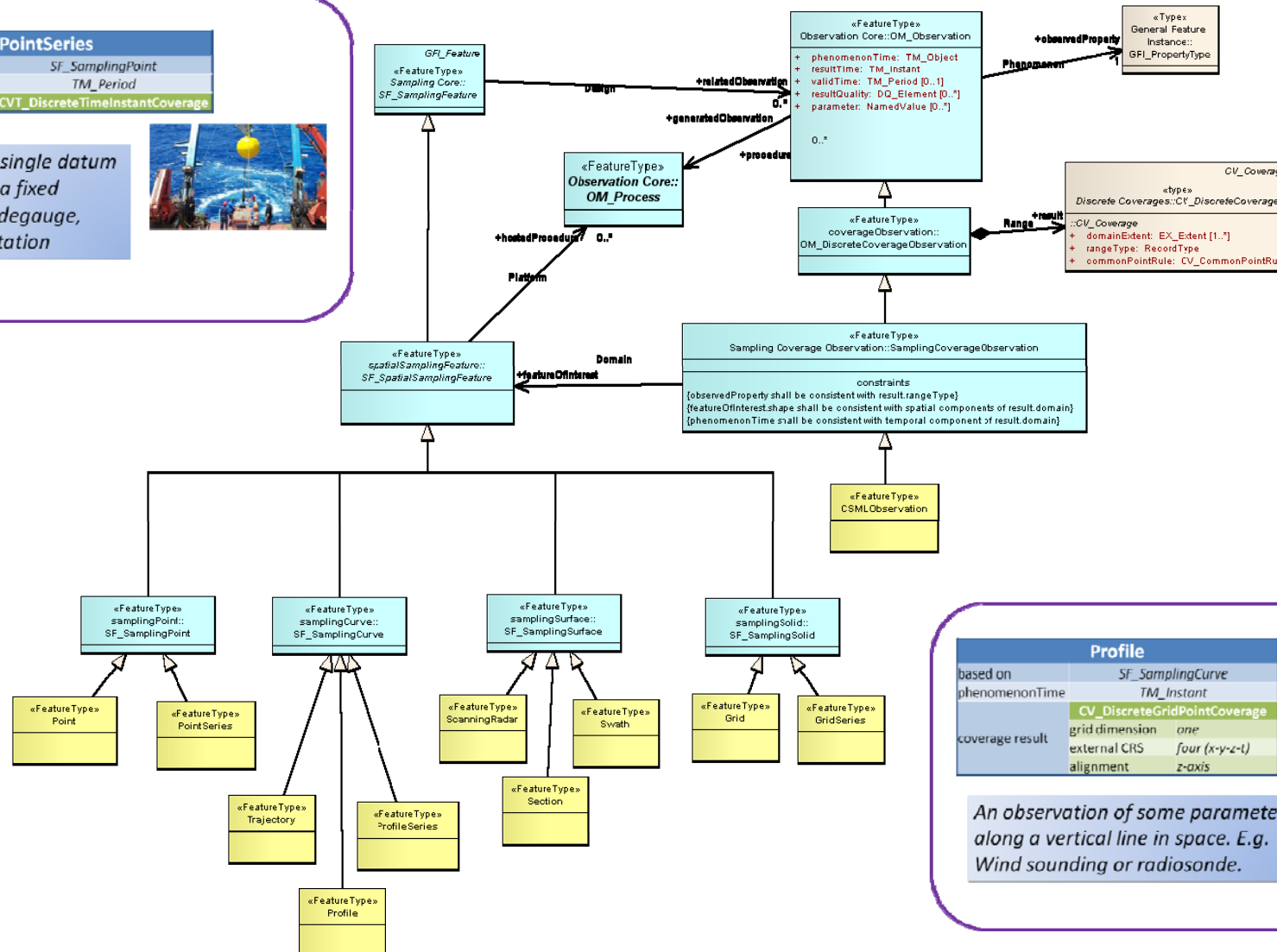
- An example information model for 'Observations and Measurements' (ISO/DIS 19156)
 - An **observation** is an event that estimates an **observed property** of a **feature of interest**, using a **procedure**, and generating a **result**



Climate Science Modelling Language (CSML)


PointSeries	
based on	SF_SamplingPoint
phenomenonTime	TM_Period
coverage result	CVT_DiscreteTimeInstantCoverage

A time-series of single datum observations at a fixed location. E.g. Tidegauge, buoy, weather station

Profile	
based on	SF_SamplingCurve
phenomenonTime	TM_Instant
coverage result	CV_DiscreteGridPointCoverage
grid dimension	one
external CRS	four (x-y-z-t)
alignment	z-axis

An observation of some parameter along a vertical line in space. E.g. Wind sounding or radiosonde.

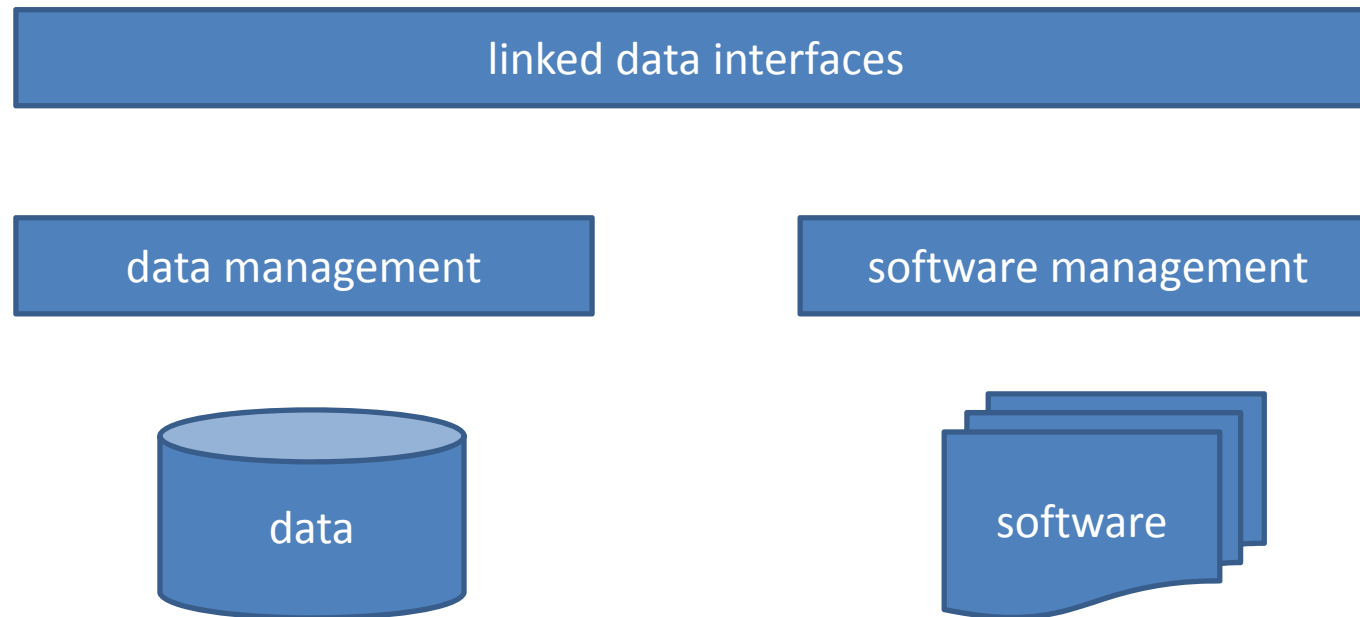


Formats

- netCDF
 - CF conventions
 - Controlled terms for physical parameter
- Could be provided as a specific 'representation' in linked-data

Overall architecture

- (High level)



Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for new people [15mins]
5. **(14:15) GeoTOD demo [15mins]**
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

GeoTOD

- ‘Geospatial Transformation with OGSA-DAI’
- Developing a linked-data server with OGSA-DAI
 - implementing Cabinet Office guidelines on URI structure for spatial data (id/doc/so/def URI types, scoped with INSPIRE two-letter theme code)
 - prototyped with Ordnance Survey open data (‘Strategi’ dataset)
 - based on ontologies generated from INSPIRE UML conceptual models
- <http://tiger.dl.ac.uk:8080/geotodls>

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for new people [15mins]
5. GeoTOD demo [15mins]
- 6. (14:30) Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]**
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

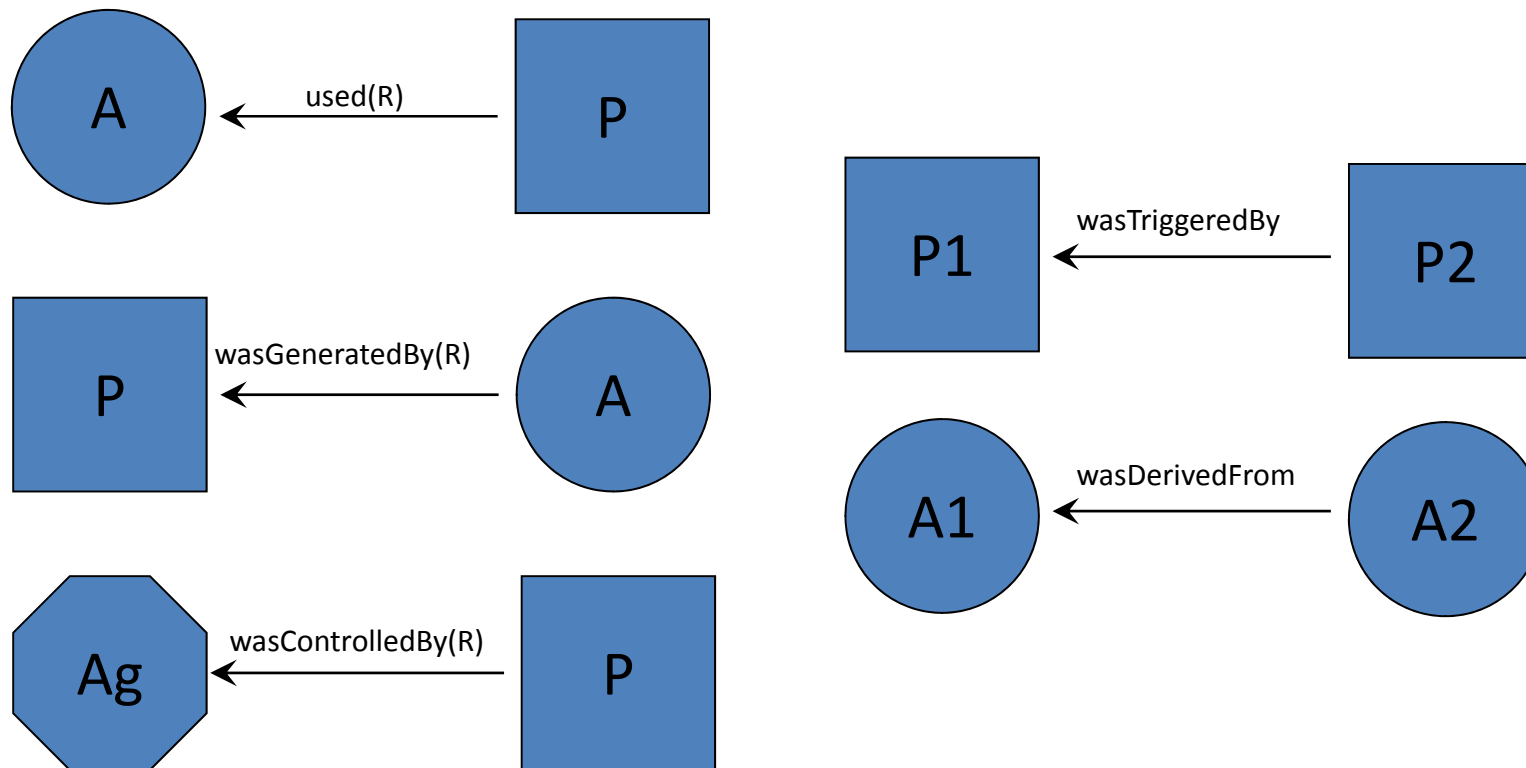
Digital Object Identifier (DOI®)

- Actionable, interoperable, persistent links for (digital) objects
- Resolution based on the Handle System®
- >98% DOIs are for scholarly articles
- DataCite (incl. BL, ANDS, etc.) developing DOIs for research data
- e.g. [doi:10.1594/WDCC/dphase_mpeps](https://doi.org/10.1594/WDCC/dphase_mpeps)

Open Provenance Model (OPM)

- A widely-adopted generic model that
 - enables digital representation of the provenance information about any digital or physical object.
 - enables exchange of provenance information between computer systems.
- OPM consists of three different “Nodes” (notions): Artefact (A), Process (P) and Agent(Ag).

Open Provenance Model - Edges

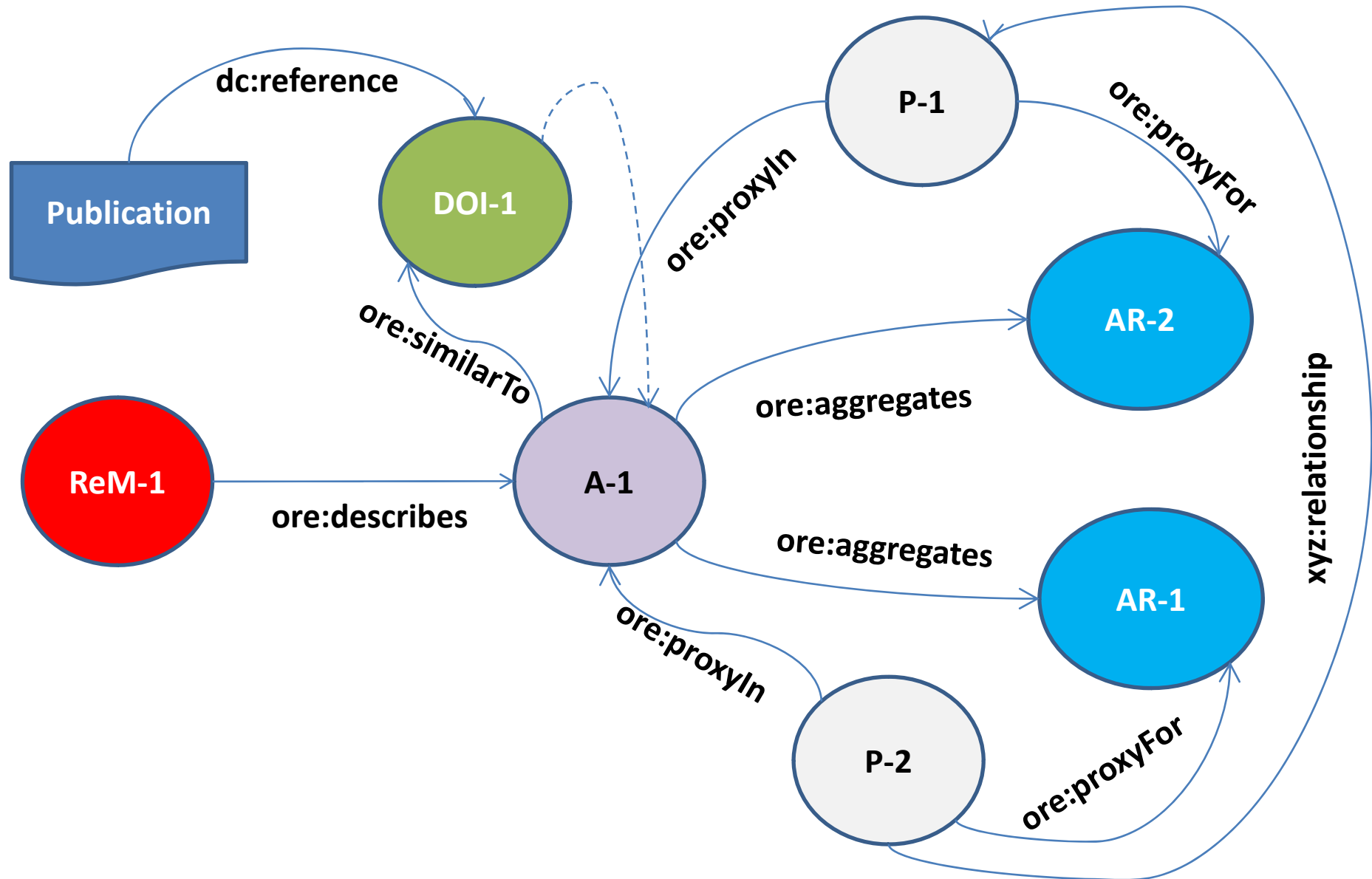


Source: <http://openprovenance.org/tutorial/slides/2-opm-semantics.pptx>

Open Archives Initiative Object Reuse and Exchange (OAI-ORE)

- OAI-ORE *defines standards for the description and exchange of aggregations of Web resources.*
- Leverages the RDF and Linked Data concepts.
- Consists of the following notions:
 - **Aggregation (A)**: a set of Web-based Resources.
 - **Aggregated Resource (AR)**: a Resource that is a constituent of an Aggregation.
 - **Resource Map(ReM)**: describes an Aggregation.
 - **Proxy (P)**: used in an assertion specific to an Aggregated Resource (e.g. relationship with another aggregated resource) in the context of a specific Aggregation.

OAI-ORE (An Example)



Metadata Objects for Linking in Environmental Science (MOLES)

- Based on O&M (ISO 19156):
 - An **observation** measures an **observed property** of a **feature of interest** using a **procedure** and generating a **result**
- Specialises **procedure** for
 - ***acquisition*** (*operation of an instrument on a platform*)
 - ***computation*** (*source data, algorithm, etc.*)

References

- <http://www.openarchives.org/ore/1.0/datamodel.html>
- <http://openprovenance.org>
- <http://www.datacite.org>
- <http://www.doi.org>

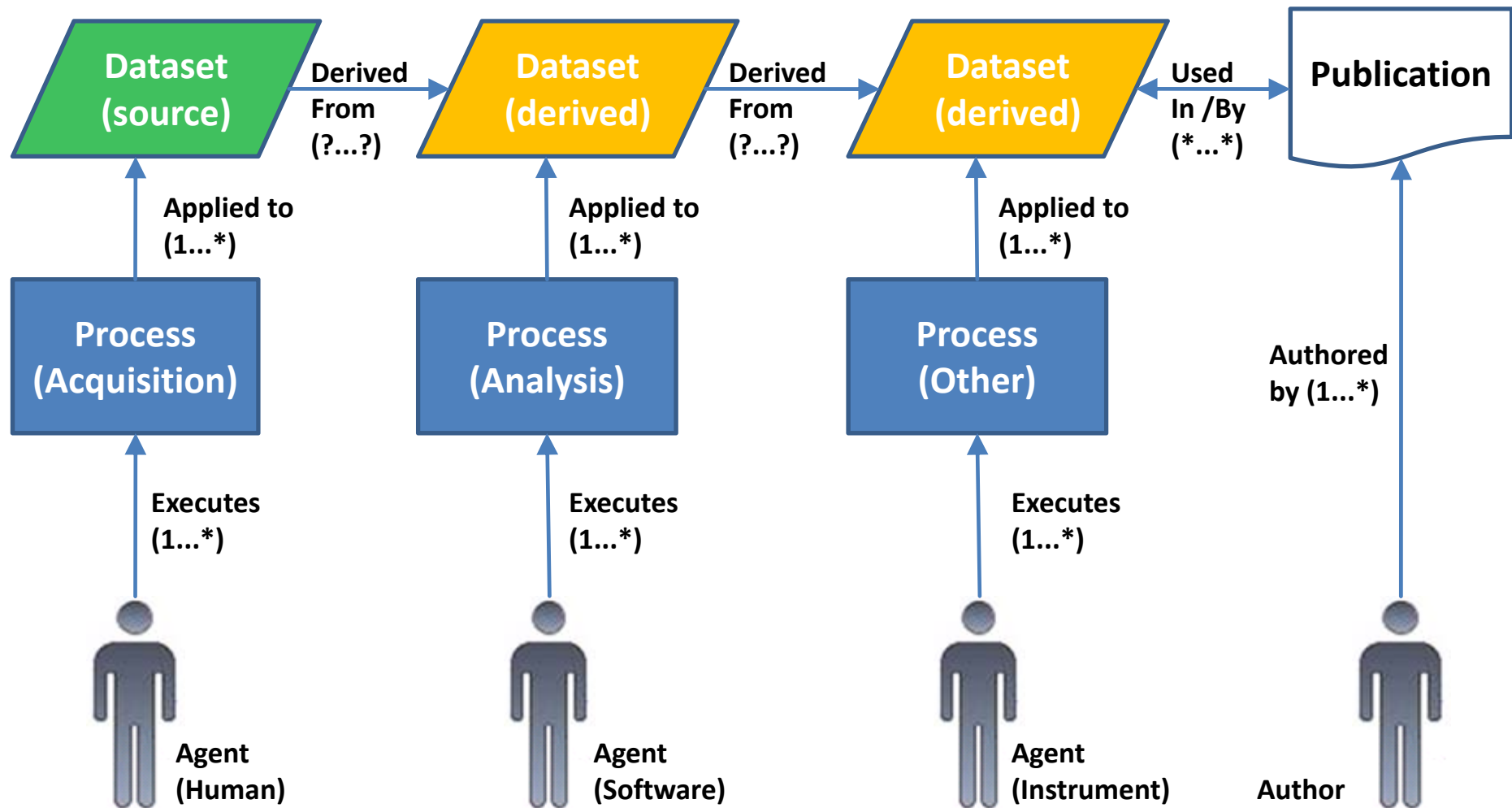
Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
- 7. (14:45) STFC (meta)data modelling [15mins]**
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
10. AoB [30mins]

ACRID Provenance Model (APM) - Rationale

- A basic model to
 - capture detailed information about the workflows behind publishing various UAE CRU datasets
 - thus, enable re-enactment of the workflows
 - facilitate traceability of the provenance of published data
 - provide a platform for developing a more comprehensive model

ACRID Provenance Model (APM) - Overview



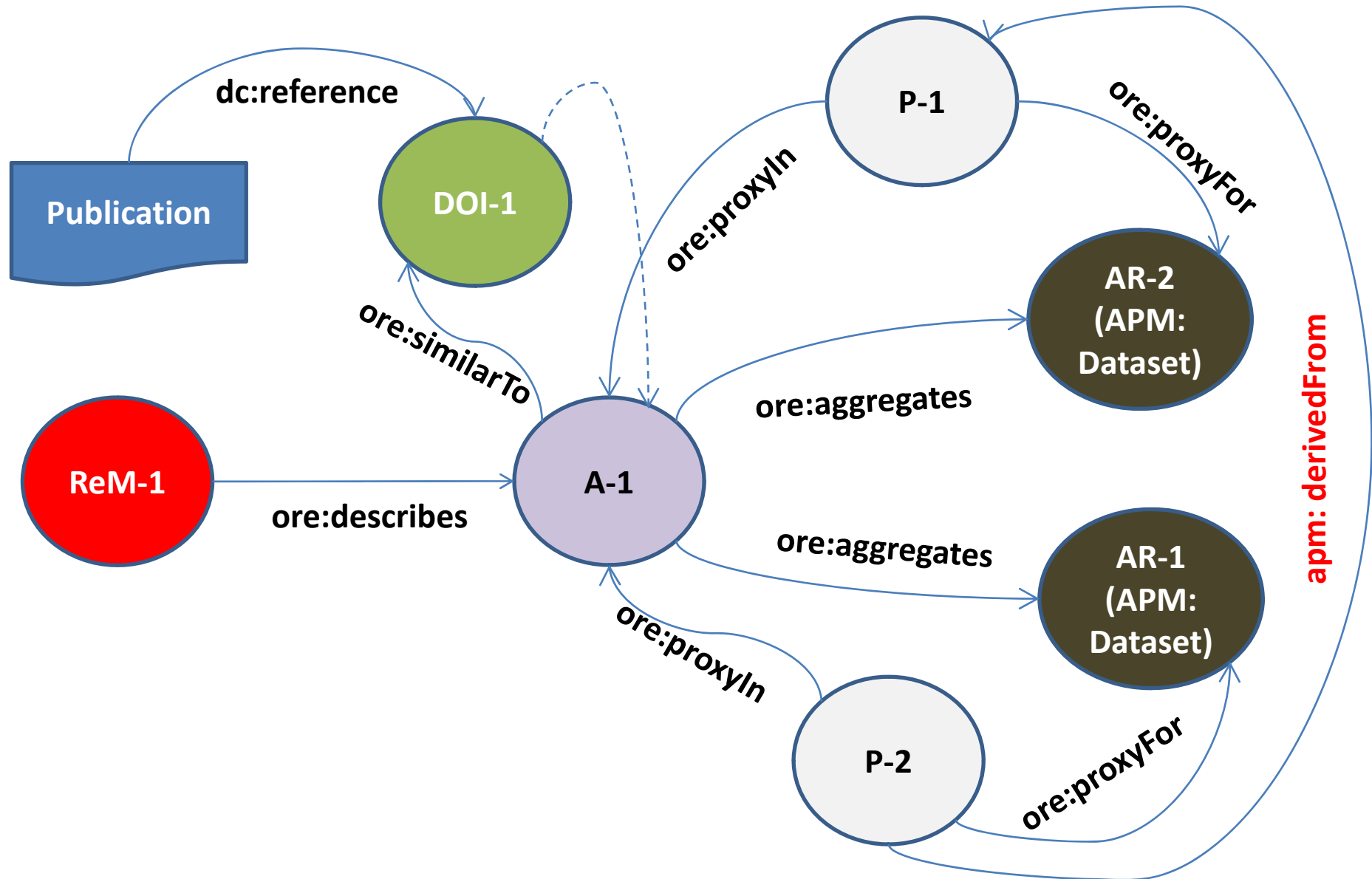
ACRID Provenance Model (APM) – Principal Concepts (1)

- **Dataset** – a data object (the provenance of which is being captured) in a particular state.
 - A dataset (derived) could be derived from one or more other datasets (source).
 - A dataset has a reference to the publication(s) it is used in.
 - Analogous to “Artefact” in OPM.
- **Process**– an action performed (as standalone or as part of a series of other actions) to either acquire a dataset (source) or produce a dataset (derived) from one or more other datasets (source or derived).
 - Analogous to the “Process” concept in OPM.
 - Typical geospatial processes include acquisition, observation and analysis.

ACRID Provenance Model (APM) – Principal Concepts (2)

- **Agent**– an entity (alone or along with other entities) responsible for performing one or more processes to either acquire a dataset or produce a dataset from one or more other datasets.
 - An agent could be a human, computer software or any type of hardware, such as weather observation instrument .
 - Comparable to the “Agent” concept in OPM.
- **Publication**– a copy of a published work that is available to public and is based on one or more datasets.
 - Produced by one or more **Authors** .
 - “Persistently” and uniquely identified using DOI.
 - Similar to the “Artefact” concept in OPM

APM in OAI-ORE



ACRID Provenance Model (APM) – Examples of Questions

- Relationship between datasets - unidirectional or bidirectional?
- Record change of ownership/custodianship of datasets?
- Record Metadata about APM instances, e.g. who created it, who maintains it, etc.?
- Capture annotations made to the datasets?

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
- 8. (15:00) Update on CRU work [30mins]**
9. Met Office interests [30mins]
10. AoB [30mins]

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
- 9. (15:30) Met Office interests [30mins]**
10. AoB [30mins]

Agenda

1. Update on project admin (contracts, CA etc) [15mins]
2. Project handover to BADC [15mins]
3. Project planning [30mins]
4. ACRID overview for new people [15mins]
5. GeoTOD demo [15mins]
6. Summary of: DOI, OAI-ORE, Open Provenance Model, MOLES [15mins]
7. STFC (meta)data modelling [15mins]
8. Update on CRU work [30mins]
9. Met Office interests [30mins]
- 10. (16:00) AoB [30mins]**