

EOSCpilot data interoperability

technical workshop

Data catalogues and datasets in the European Open Science Cloud

4-5 October

Genome Campus, Hinxton, UK

EOSCpilot data interoperability technical workshop

Data catalogues and datasets in the European Open Science Cloud

4-5 October, 2017. Genome Campus, Hinxton, UK

[Abstract](#)

[Objectives](#)

[Day 1](#)

[Day 2](#)

[Agenda](#)

[Shared folder](#)

[Attendees](#)

[Logistics](#)

[Venue](#)

[Getting here](#)

[Traveling and accommodation](#)

[Online access](#)

[Dinner](#)

Abstract

The European Open Science Cloud (EOSC) aims to provide by 2020 a virtual repository to access data from publicly funded research. The EOSCpilot is a two year project to support the first phase of development of EOSC. The EOSCpilot data interoperability task aims to establish principles, propose recommendations and demonstrate how FAIR data hosted by domain specific data repositories can be exposed to EOSC to be used and reused by EOSC services and users. In this workshop we will review data models used to describe datasets and data catalogues as well as requirements of infrastructure services dependent on datasets. The workshop will be focus on understanding, testing and evaluating existing technologies to expose and facilitate the integration of dataset metadata.

Objectives

Day 1

- Get an update of the EOSCpilot project and the work of the data interoperability task
- Review use cases and metadata models describing dataset catalogues and datasets
- Review e-infrastructure requirements for services working with datasets
- Propose key minimum properties to expose and find dataset metadata

Day 2

- Evaluate methods and technologies to expose data catalogues and datasets metadata

Agenda

The first day we will have updates and discussion about relevant activities in the context of EOSC. At the end of the first day and in the second day we will have more hands on activities with data providers and technology experts to explore technologies and metadata models to expose data into EOSC.

Please contribute collecting notes and asking questions in this document. We will process this document later to build the workshop report.

Shared folder

Please if you are a presenter:

1. Copy your presentation in the [google drive shared folder](#)

2. Link the title of your talk to your presentation in this agenda

Day 1 - Wednesday 4 October (Garden Room - EBI Main Building)			
09:00 - 09:05	Welcome and introduction	5'	Carole Goble ELIXIR - UMAN
09:05 - 10:00	1.1 - EOSC updates 10' presentation per speaker		<i>Chaired by Carole Goble</i>
	<ul style="list-style-type: none">• EOSC and EOSCpilot	10'	Brian Matthews STFC
	<ul style="list-style-type: none">• EOSCpilot services	10'	Massimiliano Assante CNR
	<ul style="list-style-type: none">• EOSCpilot interoperability	10'	Volker Beckmann CNRS
	<ul style="list-style-type: none">• EOSCpilot data interoperability	10'	Rafael C Jimenez ELIXIR
	<ul style="list-style-type: none">• Questions and answers	15'	All

Notes

- Vocabulary: what is a service provider? A glossary?
 - A Service Provider is an organisation willing to offer a set of (Open Science) services, e.g. OpenAIRE, EGI, EUDAT.

Questions

- Can you give a concrete example of a EOSC service using some data catalogue metadata. Is AAI the only one?
 - There is no “official” list of EOSC services, so far we talked about class of services. However many services can be built by relying on data catalogue metadata. For instance it is possible to develop a recommender service i.e. a service that automatically suggests users on existing items of potential interest (by having a certain knowledge of the user). Another example might be a VRE generator, this service exploits the information that is stored into the catalogue and provide a VRE designer with a list of possible datasets to be automatically made available by the VRE.

REPORT

- Can we really assume that most data are already FAIR? For example, in particle physics data sets are simply too big to be shared and used in an EOSC context. In nuclear physics data are often taken in experiments which do not follow the FAIR principle (just to give some examples, certainly there are more). Is a dedicated effort necessary to help communities to make their data FAIR?
 - Agree we can't assume they are FAIR - but even massive data or private data (e.g. health care) **can** be identified, attributed, described, etc. Accessible does not mean it has to be a ZIP file shared with http on figshare, but open protocols must be used, and only with authentication when actually required. So perhaps very important here are tutorials and documentation on how to "go FAIR". (Stian).
 - Rafa: There is already a group working on how to make data FAIR funded by the European Commission and a group working on FAIRmetrics. We will rely on this work ... <https://github.com/FAIR-Data-EG/consultation>
 - The idea at least for this project is to work on how to make existing FAIR data resources available into the EOSC.
 - There is also the FAIRmetrics.org soon releasing a strawman metrics set
- Motivation of EOSC to break barriers among disciplines. Use cases does not look like support this motivation.
 - Demonstrators are stretching their research infrastructure using the horizontal infrastructures like EUDAT, EGI, ...
 - The challenge is to identify the commonalities among demonstrators
 - EGA is part of the demonstrators and has a reproducibility use case which is transversal to other domains
- As an outcome EOSC should provide a list of data catalogues and services part of EOSC
- Please help participating in the Data Catalogue survey:
<https://tinyurl.com/eosc-cat-survey>

10:00 - 10:20	Coffee break		
10:20 - 11:00	1.2 - EOSC review: datasets and dataset catalogues 10' presentation per speaker	<i>Chaired by Rafael C Jimenez</i>	
	<ul style="list-style-type: none">• Metadata catalogues and data repositories	10'	Massimiliano Assante CNR
	<ul style="list-style-type: none">• Datasets	10'	Carole Goble ELIXIR - UMAN
	<ul style="list-style-type: none">• e-infrastructure service requirements	10'	Rafael C Jimenez on behalf of Matthew Dovey JISC

	<ul style="list-style-type: none"> • Questions and answers 	10'	All
--	---	-----	-----

Notes

- We will have a hands-on session this afternoon (1.7 Proposal of key minimum properties to expose and find dataset metadata). One of the groups will work on the survey. If you are a data catalogue and if you are interested you can join this group to complete the survey.

Questions

- What about services that are built on top of those services?
 - Basic requirements from the service that are consuming the content

11:00 - 11:40	1.3 - Datasets in EOSCpilot demonstrators Flash talks - 5'		<i>Chaired by Dario Vianello and Erik Van Den Bergh</i>
	<ul style="list-style-type: none"> • Digital humanities and Cultural Heritage - Datasets in TEXTCROWD 	5'	Franco Niccolucci PIN
	<ul style="list-style-type: none"> • Life Sciences - Pan-Cancer genomics datasets 	5'	Dario Vianello EMBL-EBI
	<ul style="list-style-type: none"> • Cryo-Electron microscopy datasets 	5'	Carlos Oscar Sorzano (online CSIC
	<ul style="list-style-type: none"> • EGA datasets 	5'	Jordi Rambla EGA
	<ul style="list-style-type: none"> • Questions and answers 	20'	All

Notes

- Let's not reinvent another standard
- Look for what is common

Questions

- For SciPion, could it make sense to use [Common Workflow Language](#) for describing the workflow - we have already got the [CWL Viewer](#) for visualization (Developed by Manchester/BioExcel) and CWL workflows can be executed by [multiple engines/platforms](#) (including Airflow/Butler which was just mentioned by Dario)

Carlos: "CWL seems command line focused - but seems less natural for our streaming approach"

- For all: Are you aware if any data catalogue is indexing your datasets?
- For all: Is it clear for your data repositories what a dataset is? do you have a standard way to represent datasets?
- (Susheel) What is FAIR or unFAIR? Do we need FAIRness Levels?

11:40 - 12:15	1.4 - Research Infrastructure dataset catalogues 1/2 Flash talks - 5'		<i>Chaired by Carole Goble</i>
---------------	---	--	--------------------------------

REPORT

	<ul style="list-style-type: none">• OMICsDI, a catalogue of datasets in life sciences - ELIXIR	5'	Yasset Perez-Riverol EMBL-EBI
	<ul style="list-style-type: none">• Dataset catalogues in astronomy	5'	Andrea Bignamini INAF / ASTERICS
	<ul style="list-style-type: none">• Dataset catalogues in and Cultural Heritage	5'	Franco Niccolucci PIN
	<ul style="list-style-type: none">• Dataset catalogues in Environment sciences and Particle Physics	5'	Eileen Kühn (online) KIT
	<ul style="list-style-type: none">• Questions and answers	15'	All

Notes

- OmicsDi
 - Has information about replicas and the location of datasets
 - It links (re)analysis to the datasets
 - All the functionality is provided by APIs
- Asterics
 - Data repository and catalogue of metadata
- Parthenos
 - Survey of registries
 - 15 catalogues
 - Mapping their schemas
 - One data model

Questions

- OmicsDi
 - About (re)analysis of the datasets, how can you know this information? Who tells you that a dataset was used for analysis?
 - Data provider provides that information
 - OmicsDI provides the link to the original dataset files. What about if the information is provided by a web service? Beside the url (location), does the metadata say which type of interface (FTP, Aspera, Webservice, API, ...)?
 - Not yet, providing ID. Looking at expanding this kind of metadata.
 - Could OmicsDI host sensitive/controlled access datasets?
 - Working with EGA and will possibly work towards providing data access/license metadata for OmicsDI
 - Use the DUO (Data Usage Ontology)
- Parthenos
- General
 - Are the datasets coming from the presented demonstrators indexed by the presented data catalogues?

REPORT

- Action: Test how datasets from demonstrators are indexed in our data catalogues and test our how recommendations.

12:15 - 12:45	Lunch break		
12:45 - 13:25	1.4 - Research Infrastructure dataset catalogues 2/2 Flash talks - 5'	<i>Chaired by Carole Goble</i>	
	<ul style="list-style-type: none">Dataset catalogue in Earth Environmental science - EPOS	5'	Garry R. Baker BCS/NERG
	<ul style="list-style-type: none">Dataset catalogue in Language Resources - CLARIN-EL	5'	Stelios Piperidis Athena Research Center
	<ul style="list-style-type: none">Dataset catalogue in BlueBRIDGE	5'	Massimiliano Assante CNR
	<ul style="list-style-type: none">Dataset catalogue in the Integrated Carbon Observation System - ICOS	5'	Ari Asmi ICOS
	<ul style="list-style-type: none">FAIRDOM Catalogue for Systems Biology Investigations	5'	Carole Goble FAIRDOM Association University of Manchester
	<ul style="list-style-type: none">Questions and answers	15'	All

Notes

- ...

Questions

- ...

13:25 - 14:10	1.5 - Dataset models Flash talks - 10'	<i>Chaired by Carole Goble</i>	
	<ul style="list-style-type: none">Developing a core metadata profile for the UK Research Data Discovery Service. Service requirements.	10'	Dale Robertson JISC
	<ul style="list-style-type: none">DataMed and DATS, for discovery and description of Biomedical datasets - NIH	10'	Philippe Rocca-Serra & Alejandra Gonzalez-Beltran Oxford
	<ul style="list-style-type: none">Data catalogues and datasets in schema.org and bioschemas	10'	Alasdair Gray HW

REPORT

	<ul style="list-style-type: none">• Preliminary analysis of minimum properties based use cases and requirements for finding data	10'	Eileen Kühn KIT
	<ul style="list-style-type: none">• Questions and answers	5'	All

Notes

- ...

Questions

- How do all these activities sit with the W3C Data Exchange Working Group ([DXWG](#)) that is currently revising DCAT?
 - DATS group and linked into this activity
- Important aspects
 - Geospatial: e.g. longitude and latitude (very strong community: [OGC](#) – also connected in the DXWG)
 - Sky coordinates: right ascension and declination
 - Samples
 - Observations
- Questions relating to the Jisc RDDS can be addressed to dom.fripp@jisc.ac.uk .

14:10 - 15:00	1.6 - e-infrastructure services working with datasets Flash talks - 10'		<i>Chaired by Rafa</i>
	<ul style="list-style-type: none">• Data Set Distribution Service (DSDS)	10'	Jinny Chien & Delisa Simonovic EMBL-EBI
	<ul style="list-style-type: none">• Bringing datasets to Amazon and some recommendations	10'	Cyrus Vahid Amazon
	<ul style="list-style-type: none">• BioMAJ	5'	Rafael Jimenez on behalf of Olivier Sallou
	<ul style="list-style-type: none">• Questions and answers	15'	

Notes

- ...

Questions

- DSDS: Relate to BdBags paper [I'll Take That to Go: Big Data Bags and Minimal Identifiers for Exchange of Large, Complex Datasets](#) which also uses Globus GridFTP and MinId for large site-independent data transfers based on BagIt
- Lets not get bogged down in whether we should use AWS or public solutions and instead focus on the metadata needed to drive all the service providers so can avoid lock-in and support hybrid and migrations solutions.
- Security in ALL cloud resources needs to be transparently compliant.

15:00 - 15:30	Coffee break		
15:30 - 17:00	1.7 - Proposal of key minimum properties to expose and find dataset metadata Hands-on Group session	90'	<i>Chaired by Rafael C Jimenez</i>

Notes

- **UK RDDS mapping spreadsheet:**
<https://docs.google.com/spreadsheets/d/1XTrJgcb0OEEdqSrsrfU-38BztBKm59TLDTkZ33MLViSI/edit#gid=1554679326>
- **User story spreadsheet:**
<https://docs.google.com/spreadsheets/d/1dtHpbp5cVaooVdqhvDjLHKM5Y8IfC-iRSU6OA6BLSUg/edit#gid=0>
- **Link to the Researchers break-out group copy of the spreadsheet:**
https://docs.google.com/spreadsheets/d/1887xFpzXEnfLH3Bgf4TzRLuobjciWC_4gzAQ5CeDct4/edit#gid=0

REPORT

- Data catalogue
 - 10 new mappings
 - 5 more responses to the survey
 - Issues during the mapping
 - Need to revise mapping data
 - Mappings are not complete
 - Give some days to complete (10 more days, **deadline is Fri 13th October**)
 - Need to look at the level of granularity
 - There are missing properties
 - Eg. Temporal evolution
 - Date and date type could solve that problem?
 - Bias of viewpoint / use cases of the descriptions. E.g. time is critical but missing.
 - ACTION:
 - Massimiliano: Send around responses from survey
- Researchers
 - Feeling that some of the use cases were requirements of the interface or the search function interrogating the metadata, rather than of the metadata schema itself
 - Difficulty of interpreting the use cases given - sometimes the group weren't sure what the requirement meant
 - Probably there are some missing use cases. One which was identified was *time*.
 - Broad agreement with uses cases except row 2, where the group didn't reach a conclusion. General trend was towards higher cardinality (i.e. more mandatory fields) due to individual disciplines considering particular fields to be of very high

REPORT

importance for their work. But different disciplines considered different fields to be critical for their work!

- Column D

- Services

- <https://docs.google.com/spreadsheets/d/1dtHbbp5cVaooVdqhvDjLHKM5Y8IfC-iRSU6OA6BLSUg/edit#gid=0>
- New tab
- Legal contact for licence

17:00 - 18:00	1.8 - Methods and technologies to expose metadata for dataset catalogues Flash talks - 5'		<i>Chaired by Rafael Jimenez with cheering on by Carole Goble</i>
	<ul style="list-style-type: none">● FAIRification of datasets and data catalogues	5'	Andra Waagmeester GO-FAIR
	<ul style="list-style-type: none">● OpenAIRE dataset exchange format	5'	Natalia Manola OpenAIRE
	<ul style="list-style-type: none">● Dataset indexing at EUDAT-B2FIND	5'	Heinrich Widmann EUDAT (DKRZ)
	<ul style="list-style-type: none">● Questions and answers	15'	
	<ul style="list-style-type: none">● RDA Metadata Standards Catalog and Metadata Element Set	5'	Alex Ball University of Bath, RDA
	<ul style="list-style-type: none">● DataCite description of datasets	5'	Kristian Garza DataCite
	<ul style="list-style-type: none">● FAIRsharing - mapping the landscape of standards, data repositories and data policies	5'	Peter McQuilton and Susanna Sansone University of Oxford, RDA, FORCE11
	<ul style="list-style-type: none">● Mechanisms of International Virtual Observatory Alliance (IVOA) to guarantee FAIRness to astronomical data.	5'	Andrea Bignamini INAF / ASTERICS
	<ul style="list-style-type: none">● Questions and answers	10'	

Notes

- A OpenAIRE use case is EU policy making
- OpenAire guidelines
 - How to expose data to OpenAire

- Version 4.0 Based on DataCite Schema 4.0

Questions

- OpenAire
 - Consume data from data repositories. What about data catalogues?
 - OpenAIRE harvests aggregators (i.e., data catalogues) and ensures that correct provenance chain is maintained and exposed (for transparency reasons).
 - What is the researcher's use case other than deposit and supplementary materials - is there any de novo find?
 - Action: Clarify context and scope of catalogues
 - Use cases are clear - the scopes of these resources are the long tail and for contextualised outcomes for funders etc
 - Linked to publications, funding, facilities, research communities, etc.
 - In the deliverable the scope of these dataset catalogues and that the "one universal catalogue" vision of the EC is looney-tunes.
 - Depends on the scope. If the scope is to have a catalogue with 5-6 common attributes, as an entry point to other catalogues, yes this is possible. But we need to ask ourselves (or the users) what would be the benefit of such a catalogue, other than keeping track of numbers. See the comment below (General).
- B2FIND
 - Comment. Different models for collecting data: Push or pull. DataCite presented as a harvesting service. I think I does not pull information but users push to them the information.
 - That's right : DataCite does not harvesting data, but it provides a service to assign a DOI to your data. I want only show DataCite as an example of a cross-discipline metadata provider from where B2FIND harvests.
 - As a user I think it would make sense for domain specific data catalogues to push their metadata to data catalogues like b2find. But what would be the benefit for the domain specific data catalogues? Somehow they might consider it as a threat since they might lose users?
 - No, I think the opposite is the case : If you are visible as a domain specific catalogue in B2FIND you get visible in a wide, cross-disciplinary scope. And some users will often not found the domain specific dataset they search for in EUDAT-B2FIND, but get linked instead to the domain specific catalogue/repository and continue the search there. I see the different catalogue not as competing but as complementary
 - Harvesting for harvesting sake without smart tools - what's the point?
 - The main purpose of B2FIND is the faceted search functionality over a wide spread cross-disciplinary scope. E.g. we provide the possibility to filter for 'Discipline' or 'Temporal Coverage' - I don't know about lot of cross-discipline search portals that have these features. You can argue

that these tools are not (very) smart. But this depends (as well) on your 'search use cases' and your expectations

- General
 - Need an ecosystem of *interoperable* data catalogues (interoperability at the minimum level of Findability - PIDs/Accessibility - URL/access point of the catalogue entry), so we can build the smart tools/service which will in turn crawl this ecosystem of catalogues and access the datasets and respond to researcher questions. A side effect: this crawling mechanism may be used to enrich these catalogues or make them more interoperable.
 - The paradox we discussed in Athens was generic catalogues and metadata are broad, shallow and useful as filing cabinets; specific catalogues are deeper, narrower and useful to domains but end towards silos only accessible to those "in the know" ..
 - What is also important is to have validation services (very light certification process), that will allow us to verify the right content/values even of the minimal attributes. OpenAIRE has used one for OAI-PMH for years and we find that such a service is most useful in order to make repository managers aware of the issues and, using a rating system, provide the incentives for them to comply to the minimal schema.
- RDA - Metadata standards catalogue (MSC), based on the DCC
 - Er....isn't this already done by FAIRsharing.org?
 - The MSC is the latest phase of work began with DCC Disciplinary Metadata Catalogue, which at launch in 2013 was the only actively maintained cross-disciplinary catalogue of metadata standards. FAIRsharing became cross-disciplinary in 2017.
 - Different scope
 - DRA: Migration and comparison among standards
 - Both the [Metadata Standards Directory](#) and FAIRsharing are RDA recommended resources
 - Programmatic interface
- FAIRsharing.org
 - Quality curation - by in house knowledge engineers working with the maintainers of the resources themselves
 - Lots of features including collections and recommendations (which are collections based on a policy document)
 - Is the FAIRsharing API open?
 - Yes (we provide an API key for tracking). Contact Pete (peter.mcquilton@oerc.ox.ac.uk) for more information
- General
 - Which one to use for recommending standards within DMPs?
 - Both can be used (e.g. both will be represented on the DMPTool/DMPOnline)

REPORT

18:00 - 18:10	Introduction to the hands on workshop on day 2	10'	Rafael C Jimenez ELIXIR
---------------	---	-----	-----------------------------------

Notes

- ...

Questions

- ...

18:30 - 19:30	Dinner
---------------	------------------------

Day 2 - Thursday 5 October - Hands-on workshop (WGC Conference Centre James Watson room)			
08:45 - 9:00	Coffee	15'	
09:00 - 09:10	Recap and introduction Hands on workshop with data providers from different domains and technology experts to evaluate methods and technologies to expose data (dataset) catalogues metadata.	10'	Rafael C Jimenez ELIXIR-EMBL
09:10 - 11:20	Mapping to the RDA MIG Metadata Element Set 10:30 - 11:00 Exercise: Complete documents in folder 11:00 - 11:20 Reporting	90'	Alex Ball University of Bath, RDA

Notes

- ...

Questions

- RDA Metadata
 - Stian: PUUID is confusing name if it does not have anything to do with [UUID](#)
 - Many of these metadata properties can be decomposed. Do the work of RDA also elaborate on more specific sub-properties?
 - Absolutely. This is part of what we mean by unpacking the elements!
 - Many of the metadata properties have open questions. I guess the aim will be to transform them into proper descriptions and also recommendations?
 - Recommendations could be generic or tailored to a domain specific research area. Eg. a recommendation for identifiers could be slightly different for humanities than life sciences.
 - The recommendations should be accompanied by a discussion that explains the rationale, which may help illuminate how strongly it applies to a given case.
 - We should have ideally an alignment of recommendation from this RDA group and this EOSCpilot group.

REPORT

- I assume RDA would want to reuse existing standards for each of these concepts (recomending how to use them), rather than define yet another mega-metadata-schema. At least that should be the way to go for EOSCPilot.
 - The idea of the set is to aid interoperability. In the long term we want to be able to express existing standards in terms of this metadata model, in order to facilitate conversion to/from other schemes.
- When this RDA working group aims to have a formal recommendation for the Metadata Element Set?
 - We do not have a fixed deadline to work to, but we aim to have a draft model ready for the next RDA Plenary in March 2018. After that, it depends on how long it takes to achieve consensus.
- What do you think about incorporating operational metadata like the metadata properties proposed yesterday?
 - `This work is based on addressing use cases, so if new use cases transpire that need more operational metadata elements, then I see no reason why the set could not be expanded to accommodate them.
- Report
 - Data sources, properties and in some cases values for properties are available in the mapping spreadsheet
 - But the mapping spreadsheet does not have all the concepts we are unpacking in the Metadata Element Set. We need examples of how the high level elements are unpacked within the property docs to justify the eventual decomposition.
 - Documents for each property could use a template with some structure to guide the collection of feedback
 - We should have a reference section in each property doc with guidelines already propose for a specific property
 - We could link the mapping spreadsheet with the property docs and vice versa. The docs will be useful for collecting feedback the spreadsheet to map the data sources properties.
 - Not really. See comment above about the spreadsheet.
 - Missing properties
 - Mainly operational properties we discussed yesterday
- General
 - Action
 - Get all the data catalogues engage and providing feedback
 - Have a deadline for providing feedback

10:30 - 11:00	Coffee break		
11:20 - 12:20	Bioschemas: Using Schema.org for describing scientific information	90'	Leyla J Garcia EMBL-EBI Sarala Wimalaratne

REPORT

			EMBL-EBI Alasdair Gray Heriot-Watt University, Bioschemas
12:10 - 12:40	Lunch break		
12:40 - 13:10	continue		

Notes

- Need to add guidelines for specific properties like identifiers
 - Bioschemas should identify which properties require recommendations. Eg identifiers, citation, etc. This plays well with RDA effort to describe and provide recommendations for properties.
- Types include properties to describe and try to model something, profiles try to define what properties in a type are minimum. Profile also define cardinality and vocabularies that should use to describe properties.
- It is not just properties for search by properties that help to discriminate the results of the search
- Generic model could be easily used for other sciences who could be creating domain specific profiles on top of the generic types.
- ACTION
 - Brian: Explore have to have researchSchemas as a pilot/demonstrator in EOSCpilot
 - Engage data catalogues
 - Describe the benefits of marking up with researchschemas

Questions

- What is bio in bioschemas and what is generic and applicable to any other scientific domain? How portable would this be for other scientific domains? eg. astronomy, earth science, humanities, ... For instance what would I need to do to describe a planet. How easy would be to mimic bioschemas for earth science (earthschemas)?
- As a data provider why should I use schema.org markup in HTML pages instead of a proper API or Webservice?
 - Just half of the data catalogues have an API or WS interface
 - Best practice for HTML content representation and to be indexed by search indexes
 - Common interface
- Some properties should be at a higher level than dataset. Eg. "measureTechnique". Properties could be associated at the level of the investigation, experiment, study, ...
- Issue
 - Where is the "protein" profile described within the description?
 - Idea: We could add a context "bioschemas" and show the protein profile.
 - Is there a conforms to property?

REPORT

13:10 - 13:40	Exposing metadata to EUDAT-B2FIND	30'	Heinrich Widmann EUDAT (DKRZ)
---------------	-----------------------------------	-----	---

Notes

- Data provider exposing metadata with OAI-PMH
 - Good
 - Simple, low-barrier, standard API
 - Based on XML
 - Data model
 - DublinCore
 - Your own model
 - Domain agnostic
 - Bad
 - Inefficient/slow

Questions

- ...

13:40 - 14:10	GO-FAIR implementations to expose scientific metadata	30'	TBC GO-FAIR
---------------	---	-----	----------------

Notes

- hands-on tutorial example: <https://tinyurl.com/fair-hands-on>
- FAIR Data Port paper: <https://peerj.com/articles/cs-110/>
- github repository <https://github.com/DTL-FAIRData/>
 - FAIRifier: <https://github.com/DTL-FAIRData/FAIRifier>
 - FAIR Data Point: <https://github.com/DTL-FAIRData/FAIRDataPoint>
- FAIRification example with Friend of a friend:
 - <https://www.dropbox.com/sh/5gg3m46rdk4d1mb/AACHqICbJHXIawhX5od9GaTWa?dl=0>
- FAIR metrics working group <http://fairmetrics.org/>

Notes

- FAIRifier extension of openRefine

Questions

- ...

14:10 - 15:00	Wrap-up and conclusions	20'	All
---------------	-------------------------	-----	-----

Notes

- ...

Questions

- ...

Attendees list (Private):

https://docs.google.com/document/d/11nHn7AZKs9WlcvNEPDoCOxh60OA1reTEwdtS8_hfCRQ/edit

Contact person: Rafael Jimenez

Logistics

Venue

[Wellcome Genome Campus](#), Hinxton, Cambridgeshire, CB10 1SD, UK

- Day 1: Garden Room at the EBI-EBI building.
- Day 2: [James Watson](#) at the Wellcome Genome Campus conference centre.



Getting here

To get to the campus please follow [these steps](#). once you are at the campus please follow this steps:

1. Go to the **security building** and let the security team know you are coming to this workshop. They will have your name and will have a badge ready for you to come in as a visitor.
2. Walk into the campus and following the map look for the EMBL-EBI **Main building** and its main entry.

3. From the EMBL-EBI reception you will see signs that will guide you to the Garden room (ground floor).

Traveling and accommodation

Suggestions about traveling and accommodation can be found at the ELIXIR website:

<https://www.elixir-europe.org/about-us/who-we-are/hub/visiting>

NOTE: Unfortunately, due to many events, the accommodation in Hinxton has been already fully booked. We recommend to book the hotel nearby which can be found at our [website](#) or you can check the availability at booking.com

Please, put your hotel to the registration table (above), it can help to find the company for the shared taxi from your hotel.

Online access

If you would like to present remotely please contact Rafael <rafael.jimenez@elixir-europe.org>

If you would like to attend the meeting online please follow this link:

<https://global.gotomeeting.com/join/193025037>

You can also dial in using your phone.

United Kingdom: +44 330 221 0086

Access Code: 193-025-037

More phone numbers

United States: +1 (571) 317-3129

Australia: +61 2 8355 1020

Austria: +43 7 2088 3962

Belgium: +32 28 93 7018

Canada: +1 (647) 497-9350

Denmark: +45 32 72 03 82

Finland: +358 942 45 0453

France: +33 184 880 733

Germany: +49 692 5736 7211

Ireland: +353 15 360 734

Italy: +39 0 694 80 34 40

Netherlands: +31 207 941 377

New Zealand: +64 9 280 6302

Norway: +47 21 93 37 51

Spain: +34 932 20 0596

Sweden: +46 853 527 836

Switzerland: +41 225 4599 78

Joining from a video-conferencing room or system?

Dial: 67.217.95.2##193025037

Cisco devices: 193025037@67.217.95.2

First GoToMeeting? Try a test session: <http://link.gotomeeting.com/email-welcome>

Dinner

The dinner are not covered by the meeting. Still we think it would be nice to have a dinner together so we have reserve a table in the [Red Lion](#). Please fill this [spreadsheet](#) you would like to join us **before 1pm on 2nd October**.

[Directions](#) to the Red Lion

