

ALL HANDS 2020 WORKSHOP 2

Interoperability

Bioschemas as a flagship activity of the Interoperability Platform

Date: 8 June 2020, 13:00-15:00 BST

Connection details: [Link](#)

In-session note/question collection sheet: [Link](#)

Chair: Alasdair Gray (ELIXIR-UK), Sirarat Sarntivijai (Hub)

Coordinators: Kathi & Hannah

Abstract:

The Interoperability Platform will focus on the Bioschemas including the Strategic Implementation Study (SIS) and other [Bioschemas' community](#) works. The session aims to enable progress to be made towards the goals of the SIS, and identify other areas of collaborations with the Bioschemas' community at large. Leaders and participants of all use cases in this study are invited to contribute. Engagement with OpenAIRE via Bioschemas scraped data for the OpenAIRE knowledge graph is a subject of interest that is highlighted in the opening remark. Recent discussions at the [EIPvF2F-Bioschemas](#) have emphasized on refining the [phenotype profile](#) which may have overlaps in the Plant Informatics and the Rare Disease use cases with case-specific differences to be identified.

Note: [Bioschemas Monthly Call](#) - becoming the community member

13.00 - 13.10	Introduction - Bioschemas, a Flagship activity of the Interoperability Platform Speaker: Carole Goble (ELIXIR-UK)
13.10-13.25	Title: OpenAIRE Knowledge Graph and Bioschemas Speaker: Paolo Manghi (OpenAIRE - CNR)
13.25-13.45	Discussion: Moderator: Sirarat Sarntivijai
13.45-13.55	Title: The Bioschemas Community Reboot • How we are reorganising Bioschemas • The ELIXIR Bioschemas Strategic Implementation Study



	Bioschemas Monthly Call - becoming the community member Speaker: Alasdair Gray (ELIXIR-UK)
13.55-14.00	Title: Overview of breakout sessions Speaker: Alasdair Gray
14.00-14.10	Coffee break (move into breakout sessions)
14.10-14.45	Breakout 1: Phenotypes for plants Leader: Cyril Pommier, Daniel and Bert
	Breakout 2: Phenotypes for (rare) diseases Leader: Marc, Raja, Marco Roos
	Breakout 3: Site maps and crawling concerns Leader: Petros Papadopoulos, Egon Willighagen
14.45-15.00	Reporting: (please add content to breakout slides) Plant phenotype: Disease phenotype: Sitemap:

Outcomes from Bioschemas Interoperability Workshop

OpenAIRE

Action: Follow up with OpenAIRE on

- Using Bioschemas markup to populate knowledge graph
- Validating markup

Phenotypes for plants

Plant Search Use Case

Alice is looking for experiments about a taxon (down to the variety level, maybe germplasm). She may want to filter by:

- Experimental factor (drought, nitrogen, etc),
- Growth facility (field, greenhouse, growth chamber),
- Trait (e.g. plant height, flowering time)
- Phenotype proper

She can find relevant experiments in plant phenotyping databases through google/FAIDARE and ideally related publications

Adjustments to Bioschemas Specifications

Phenotype type/profile: Need to capture the development stage

New types/profiles:

- Trait: distinct from phenotype, want to search by both trait and phenotype
- Experimental Factor: likely to be of interest beyond plant use case



- Growth facility: specific to plants but crucial to a wide range of experiments

Follow up activities

- Plant community markupathon: 3 July
 - Extend invite to more members of the community
- Distinguish what makes a trait different from a phenotype and develop profiles/types
- Develop ExperimentalFactor profile/type
 - Decide if growth facility is an experimental factor

Phenotypes for (rare) diseases

Rare Diseases Use Cases

1. As a researcher of a specific RD phenotype, I am looking for resources with data on that specific phenotype.
2. A clinician doing research on one specific RD, I am searching for resources with data from the same RD
3. As a catalogue curator, I am looking for resources that have data on any rare disease phenotype.
4. As a data resource provider I want to expose the phenotypes associated with the datasets in the data resource
5. As a researcher, I am interested in the supporting evidence related to the association between a phenotype and a disease of a BioChemEntity
6. As a patient with a disease phenotype I am looking for researcher and/or medical providers experts on that phenotype

Adjustments to Bioschemas Specifications

Need to capture:

- Patient/sample dataset
- Diagnostic criterion
- Pathognomonic sign
- Exclusion criteria
- Supporting evidence

Follow up Activities

- Distinction between bioschemas:Phenotype, schema:MedicalCondition, and schema:MedicalSignOrSymptom
- Schedule a Rare Disease Markupathon
 - Need to identify which use case we will focus on
 - Ensure that we get the data providers for the specific use case

Sitemaps and crawling concerns

Concerns

- Cost for hosts of having your site crawled: both by search engines and community crawler
 - Include rules for crawling in the sitemap to limit page rate, etc
- Challenge of maintaining a sitemap: several web hosting frameworks automate the process
 - Some sites have specific issues they are encountering, e.g. solr index
- Cost for consumers of crawling sites, particularly dynamic sites
- Time required to crawl an entire site



Follow up Activities

- Provide how to guide for including a sitemap for your resource
 - Could be for bioschemas use only, i.e. registered directly with bioschemas scraper
 - Google search console already permits the registering of a sitemap with them
 - Still need to have markup on pages to enable search engines to discover it
- Investigate the feasibility of Bioschemas data dumps
 - Avoids hosting and crawling costs
 - Makes data immediately available from a single location
 - Data would not be usable by search engines: anti-spam measure

Outcomes from Other All Hands Meeting Sessions

FAIRification for Knowledge Management (Monday 9:00–10:30)

Session notes:

<https://docs.google.com/document/d/19qYk-7eq4ctmlqec7JL9SHH-82sIDhm3irw0gfpnKS8/edit>

FAIRsharing (and Identifiers.org) markup only uses FAIRsharing ids/urls. No way of connecting their markup to the data sources that they are providing a record about.

Biodiversity Mini-symposium (Monday 11:00–12:30)

Session document:

<https://docs.google.com/document/d/11NDVKT5baCmFfA0bUdIxCauIKhfQofQPjXkJsVAbP0/edit?usp=sharing>

Guy Cochrane suggested the embedding of Bioschemas markup in the COVID-19 data portal.

Action: Follow up with Guy (Done)

Plant Sciences Mini-symposium (Tuesday 8:30–12:00)

Session notes:

<https://docs.google.com/document/d/1SAkuH-YFoALE7UBHSrBprndaa-71f5vfZYjaExVz4wc/edit>

e!DAL data repository (Daniel Arend (IPK)) embeds schema.org markup in their resources. This is DataCatalog/Dataset markup. It is appropriate that this does not go down to the Bioschemas level. See for example this [record](#) on Metabolite profiling of red and blue potatoes. e!DAL is a generic framework and could be used for any type of data.

Useful search use cases presented in the talk “[Data findability and integration : FAIDARE Elixir data search service](#)” by Cyril Pommier

FONDUE - How to FAIRify and link plant genotyping data using ELIXIR Platforms
Sebastian Beier

- Use as an implementation verification of the BioSample profile

General discussion section

- Need to have a community registry of data sources (long-tail of science)
 - Not only for plant community but for other communities as well
- Ideal use case for Bioschemas
 - Source only need to embed markup in their site and register sitemap



- Community registry scrapes markup and maintains record
- Connect with FAIRsharing collection?

Short Talks Session (Wednesday 9:00–11:00)

- Alban Gaignard (ELIXIR France): How FAIR is my resource? Monitoring progress in FAIRification through automatic FAIR maturity testing
 - Nice validation approach, based on SHACL
 - Could we reuse to provide a Bioschemas validation service?
- Alasdair Gray (ELIXIR UK): Exploiting Bioschemas Markup to Support ELIXIR Communities
 - Australian BioCommons would like to follow up on using Bioschemas

Metabolomics Mini-symposium

Session notes:

https://docs.google.com/document/d/1I1eoy37dGSDtYCmKEWHozsLiIrDUt_VPcxxgzeRA3xI/edit#

Good discussion between Egon, Steffen, and Alasdair: full transcript in session notes

- Need to identify the consumption use case for fluxomics. Types of data
 - Reactions
 - Studies
 - Enzyme, may be Protein
- Need to break the task into manageable chunks
 - Metabolites plus simple chemical equation plus dynamic reactions with kinetics/rates
 - Is a metabolite just a MolecularEntity?
- Egon developed an example Metabolights [markup](#).
 - Needs to be updated to 0.5-RELEASE version

Tools Workshop

Session notes:

???

Bioschemas profile used by Workflow Hub.

