The BioSharing Registry: connecting data policies, standards and databases in the life sciences

The products of the WG:

- **principles** for linking information about databases, content standards and journal and funder policies in the life sciences
- these will be implemented in an existing registry
Community-developed **content standards**

- To **structure, enrich** and **report** the description of the datasets and the experimental context under which they were produced
- To facilitate **discovery, sharing, understanding** and **reuse** of datasets

- Including **conceptual model, conceptual schema** from which an exchange format is derived to allow data to flow from one system to another
- Including **controlled vocabularies, taxonomies, thesauri, ontologies** etc. to use the same word and refer to the same ‘thing’
- Including **minimum information reporting requirements**, or **checklists** to report the same core, essential information
In the life sciences alone.....almost 600!
A web-based, curated and searchable registry ensuring that standards and databases are registered, informative and discoverable; also monitoring the development and evolution of standards, their use in databases and the adoption of both in data policies.
Researchers, developers and curators lack support and guidance on how to best navigate and select content standards, understand their maturity, or find databases that implement them;

Funders, journals and librarians do not have enough information to make informed decisions on which content standards or database to recommended in policies, or funded or implemented...
Core functionalities:

- search and filtering, e.g. by funder
- submissions forms to add new records
- “claim” functionality of existing records
- person’s profile (as maintainer of records) associated to the ORCID profile (for credit, as incentive)
- visualization and views of content
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Publications</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Taxa defined</td>
<td>No taxa defined.</td>
<td>No taxa defined.</td>
<td>1 Taxa types, including:</td>
<td>No taxa defined.</td>
</tr>
<tr>
<td>Data types defined</td>
<td>No data types defined.</td>
<td>No data types defined.</td>
<td>2 Data types, including:</td>
<td>No data types defined.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Countries</td>
<td>DW: Darwin Core Terminology Artifact</td>
<td>FLOW: Global Biodiversity Information Facility</td>
<td>GUID and LSID Applicability Statements</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standards</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Publications</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Taxa defined</td>
<td>No taxa defined.</td>
<td></td>
<td>4 Taxa types, including:</td>
<td>No taxa defined.</td>
</tr>
<tr>
<td>Data types defined</td>
<td>No data types defined.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxonomies</td>
<td>ALL</td>
<td></td>
<td>1 Data types, including:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANIMALIA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BACTERIA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FUNGI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintainers</td>
<td></td>
<td></td>
<td>1 Data types, including:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIOSHARING ID</td>
<td>STANDARD NAME</td>
<td>FULL NAME</td>
<td>TYPE</td>
<td>FOUNDRY</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>------------------------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>BSG-000202</td>
<td>NCD</td>
<td>Natural Collections Descriptions</td>
<td>exchange format</td>
<td></td>
</tr>
<tr>
<td>BSG-000204</td>
<td>ABCDDNA</td>
<td>Access to Biological Collection Data DNA extension</td>
<td>exchange format</td>
<td></td>
</tr>
<tr>
<td>BSG-000201</td>
<td>TDWG SDS</td>
<td>TDWG Standards Documentation Specification</td>
<td>reporting guideline</td>
<td></td>
</tr>
<tr>
<td>BSG-000199</td>
<td>TCS</td>
<td>Taxonomic Concept transfer Schema</td>
<td>exchange format</td>
<td></td>
</tr>
</tbody>
</table>
Standards and databases cross-linked

ISA-Tab - Investigation Study Assay Tabular

General Information
General-purpose ISA-Tab file format - an extensible, hierarchical structure that focuses on the description of the experimental metadata file, sample characteristics, technology and measurement types, sample-to-data relationships
Developed in: UK
Database scope and data types
Support
documentation: https://data.broadinstitute.org/omo/ISA-
MA/release古代雇佣
Email: support@isa.org
Contact: Biocurated Services
Homepage: http://isa.broadinstitute.org

Implementing Databases
MetaboLights
MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-referenced, cross-referenced and overset metabolite structures and their reference spectra as well as their biological roles, functions and experimental data from metabolite experiments. We offer open submission and have strong reporting capabilities. We will use ISA and further develop ISA-standards to standard formats where various components are encapsulated, such as the encoded spectral and chromatic data, and associated Information about the chemical standard, as well as metadata describing studies and the study as a whole. We are dedicated to collaborating closely with major companies in the metabolomics community, such as the Metabolomics Society and the associated Metabolomics Standards Initiative (MSI).
Advisory Board and Working Group - core members and adopters

Jessica Tenenbaum (Duke University, USA; Chair Elect for the American Medical Informatics Association's Genomics WG), Co-Chair of the Advisory Board
Melissa Haendel (OHSU, USA; Force11; Monarch Initiative), Co-Chair of the Advisory Board
Susanna-Assunta Sansone (University of Oxford, NPG Scientific Data), Co-Chair of the RDA WG

Simon Hodson (CODATA), Co-Chair of the RDA WG
Rebecca Lawrence (F1000), Co-Chair of the RDA WG
Michael Ball (BBSRC)
Theo Bloom (BMJ)
Jennifer Boyd (OUP)
Scott Edmunds (GigaScience, BGI)
Rafael Jimenez (ELIXIR)
Amye Kenall (BMC)
Varsha Khodiyar (NPG)
Thomas Lemberger (EMBO Press)
Jennifer Lin (PLoS)
Jonathan Tedds (Ubiquity Press Open Health Data journal; University of Leicester, UK)
Todd Vision (UNC Chapel Hill & NESCents, USA; Dryad)
Michael Witt (re3data; Purdue University Libraries)
Anita de Waard (Elsevier)

Operational Team

Susanna-Assunta Sansone
Principal Investigator and Associate Director

Philippe Rocca-Serra
Senior Research Lecturer

Alejandra Gonzalez-Beltran
Research Lecturer

Milo Thurston
Senior Bioinformatician

Allyson Lister
Knowledge Engineer

Ajlfe Abdul-Rahman
Research Software Engineer

Eamonn Maguire
Contractor