BoF: Using schema.org and enriched metadata to enable/boost FAIRness on research resources

Leyla Garcia, John Graybeal, Josef Hardi, Nick Juty, Siri Jodha Khalsa, Fotis Psomopoulos, Adam Shepherd, Mingfang Wu

research data sharing without barriers
rd-alliance.org

3 April, 2019, RDA 13th Plenary Meeting, Philadelphia
Session notes:  https://goo.gl/q7d1te

Session page:

(please sign in)
Motivation & goals

- Schema.org well-established
  - Leverages on Schema.org
    - Simple to adopt, moderately expressive
    - Extendable
    - Indexed (discoverability)
- Becoming a popular mechanism
  - DataCite DOIs: schema.org (JSON-LD) through content negotiation
  - DataVerse: citation metadata to dataset landing pages
- Other research domains (Survey)
  - Will have their own ‘best practice’
    - commonalities across domains
    - Key differences expressing domain relevant information
- ResearchSchemas
Motivation & goals

● DDP IG TF: "Using schema.org for research data discovery"
● Propose creation of a focused and formal RDA working group to establish ‘ResearchSchemas’ as a community

● Initial landscaping survey to determine & refine objectives
  ■ Which metadata schemas are used across research areas
    ▪ Not necessarily schema.org
  ■ Mapping across equivalent properties (crosswalk)
    ▪ Facilitate interoperability & MI recommendations
  ■ Identify gaps
    ▪ Proposing new types, properties or relations

● Metadata & standards => improve FAIR
Agenda

- Welcome, goal of the session
- Current practices in working on schemas to describe datasets
  - Bioschemas (https://github.com/bioschemas)
  - Science-on-schema (https://github.com/ESIPFed/science-on-schema.org)
- Tools to facilitate schema converting, validation, etc
  - From existing metadata schemas to schema.org
- Report on the survey
- Process for setting up the research schemas WG
- Discussion
  - WG case statement, Schemas and extensions, Tooling, Guidelines, FAIR alignment in research schemas, ..
- Summary & next steps
• Collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond
Background

Types

What we are talking about

Properties

What we can say about it

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>additionalName</td>
<td>Text</td>
<td>An additional name for a Person, can be used for a middle name.</td>
</tr>
<tr>
<td>address</td>
<td>PostalAddress or Text</td>
<td>Physical address of the item.</td>
</tr>
<tr>
<td>affiliation</td>
<td>Organization</td>
<td>An organization that this person is affiliated with. For example, a school, university, a club, or a team.</td>
</tr>
<tr>
<td>alumniOf</td>
<td>EducationalOrganization or Organization</td>
<td>An organization that the person is an alumni of. Inverse property alumni.</td>
</tr>
<tr>
<td>award</td>
<td>Text</td>
<td>An award won by or for this item. Supersedes awards.</td>
</tr>
<tr>
<td>birthDate</td>
<td>Date</td>
<td>Date of birth.</td>
</tr>
<tr>
<td>birthPlace</td>
<td>Place</td>
<td>The place where the person was born.</td>
</tr>
<tr>
<td>brand</td>
<td>Brand or Organization</td>
<td>The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.</td>
</tr>
<tr>
<td>children</td>
<td>Person</td>
<td>A child of the person.</td>
</tr>
<tr>
<td>colleague</td>
<td>Person or URL</td>
<td>A colleague of the person. Supersedes colleagues.</td>
</tr>
<tr>
<td>contactPoint</td>
<td>ContactPoint</td>
<td>A contact point for a person or organization. Supersedes contactPoints.</td>
</tr>
<tr>
<td>deathDate</td>
<td>Date</td>
<td>Date of death.</td>
</tr>
<tr>
<td>deathPlace</td>
<td>Place</td>
<td>The place where the person died.</td>
</tr>
<tr>
<td>duns</td>
<td>Text</td>
<td>The Dun &amp; Bradstreet DUNS number for identifying an organization or business person.</td>
</tr>
<tr>
<td>email</td>
<td>Text</td>
<td>Email address.</td>
</tr>
<tr>
<td>familyName</td>
<td>Text</td>
<td>Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the name property.</td>
</tr>
<tr>
<td>faxNumber</td>
<td>Text</td>
<td>The fax number.</td>
</tr>
<tr>
<td>follower</td>
<td>Person</td>
<td>The most generic uni-directional social relation. A person or organization that monitors or maintains.</td>
</tr>
</tbody>
</table>
Bioschemas

• Community initiative built on top of schema.org

• Aim
  • Improve data discoverability and interoperability in Life Sciences

• How
  • Adding Life Science types to schema.org
  • Providing usage guidelines, examples and tools
Motivation

• Researcher looking for help to
  • sequence a cactus → people, expertise
  • find similar sequences → datasets
  • compare and visualize → tools
Use cases

• Generic
  • Search engines
  • Google dataset specialized search

• Life sciences
  • Data depositions can harvest markup from small providers → small providers can benefit back from data depositions
  • Tess (training and events registry) → Automatic population → trainers add markup
  • Common/rare terms → Resource index
It is all about context
Profiles: types in context

Dataset Profile

Version: 0.2 (25 February 2018)

Bioschemas specification for describing a dataset in the life-science.

Schema.org hierarchy
This Profile fits into the schema.org hierarchy as follows:

Thing > CreativeWork > Dataset

Description
A guide for how to describe datasets in the life-sciences using Schema.org-like annotation.
Contribution to FAIRability

• Unique identifiers (included in metadata)
• Descriptive metadata
• Indexed and available

- Identifier → minimum → always part of the metadata
- Metadata focused on use cases
- Indexed by Google, some specialized resources (bio.tools, TeSS)
- Future → Bioschemas specialized indexes
- Open protocol
- Authorization-based protocols
- Metadata always accessible

- Mainly HTTP & HTTPS
- Authorization protocols could be express via metadata (but not included yet)
- Longevity via specialized registries (not always and mostly no plan if registries disappear)
• Metadata on structured language
• Use of FAIR vocabularies
• Qualified references to other resources

- Schema.org + profiles → JSON-LD
- Recommendations on controlled vocabularies (but FAIR vocabularies are not always available/suitable)
- Not supported, mainly generic relations to others
- Rich metadata

- License

- Provenance

- Community standards

- Partially supported → Minimum, recommended and optional

- License (but not always as minimum)

- Some provenance covered → Dataset provider, publication authors, structure data creator

- Community involvement since conception
Summary

• Focus on key properties prioritized as Minimum, Recommended and Optional based on community agreements and common practices

• Additional recommendations regarding properties cardinality

• Customization on schema.org types to better supports needs on the life sciences community

• Terms reused from well-known ontologies thus avoiding reinventing the wheel
Themes Can include:
Text mining, Structured metadata, Identifiers, Data distribution, Data integration, Data Validation, Tools, Containers, Tools discovery, and Training materials.
Opportunities for companies to submit hacking topics.
Project submission closes 7th April.

Contact: Jen.Harrow@elixir-europe.org
Science-on-schema

Adam Shepherd

research data sharing without barriers
rd-alliance.org
NSF EarthCube Repository Working Group

PROBLEM
- Lots of websites describing repositories
- Repository managers duplicating effort

SOLUTIONS
- Repositories self-publish using schema.org + extensions
- Based on re3data.org
- Produced Guidelines
- Dozen NSF data repositories
NSF EarthCube Project 418 (P418)

- Registry of NSF-GEO funded datasets

- Expanded Repository Working Group:
  - **Guidelines** for publishing schema.org for Datasets
  - schema.org harvesting & indexing **software**
  - **prototype UI** and API
A point, or coordinate, would be defined in this way:

```json
{
  "@context": {
    "@vocab": "http://schema.org/",
    "datacite": "http://purl.org/spar/datacite/"
  },
  "@type": "Dataset",
  "name": "Removal of organic carbon by natural bacterioplankton",
  "spatialCoverage": {
    "@type": "Place",
    "geo": {
      "@type": "GeoCoordinates",
      "latitude": 39.3280,
      "longitude": 120.1633
    }
  }
}
```

All other shapes are defined using the `schema:GeoShape`:

```json
"spatialCoverage": {
  "@type": "Place",
  "geo": {
    "@type": "GeoShape",
    "line": "39.3280,120.1633 40.445,123.7878"
  }
}
```

http://www.opengis.net:def/crs/OGC/1.3/CRS84

dbpedia:Spatial_reference_system
Decisions
document:
  • status
  • context
  • consequences

follows
Architectural Decision Records (ADRs)
Extending existing schema.org properties

- Technique blessed by Dan Brickley, Google

Temporal Coverage

The OWL Time vocabulary defines useful ways for describing the temporal coverage of some schema:CreativeWork such as a Dataset.

Example

```json
{
    "@context": [
        {
            "@vocab": "http://schema.org/",
            "time": "http://www.w3.org/2006/time#"
        },
        "https://geoschemas.org/context/temporal.jsonld"
    ],
"@type": "Dataset",
"temporalCoverage": {
    "@type": "time:TemporalEntity",
    "time:hasBeginning": {
        "@type": "time:Instant",
        "time:inXSDDateTimeStamp": "2015-11-01T17:58:16.102Z"
    }
}
```

1. Make your `@context` an array.
2. Make sure the first context is schema.org
3. Add the temporal coverage context file to the `@context`
4. Now you can begin to use OWL Time classes as a valid `schema:temporalCoverage`
1,113,210 entities
7,087,380 triples

47,650 Dataset
54,665 DataDownload
599,960 PropertyValue
~ 35k Identifiers
~560k Dataset Variables
<table>
<thead>
<tr>
<th>Dataset Properties</th>
<th>Google Requires / Recommends</th>
<th>Provider Usage</th>
<th>Dataset Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Implemented</td>
</tr>
<tr>
<td>@context</td>
<td><strong>Required.</strong> Set @context to &quot;<a href="http://schema.org/">http://schema.org/</a>&quot;</td>
<td>80%</td>
<td>omitted ending slash: '<a href="http://schema.org">http://schema.org</a>'</td>
</tr>
<tr>
<td>@type</td>
<td><strong>Required.</strong> Set @type to &quot;Dataset&quot;</td>
<td>100%</td>
<td>47,650 datasets</td>
</tr>
<tr>
<td>name</td>
<td><strong>Required.</strong> A descriptive name</td>
<td>80%</td>
<td>99.9%</td>
</tr>
<tr>
<td>description</td>
<td><strong>Required.</strong> A short summary</td>
<td>70%</td>
<td>97%</td>
</tr>
<tr>
<td>url</td>
<td><strong>Recommended.</strong></td>
<td>70%</td>
<td>100%</td>
</tr>
<tr>
<td>citation</td>
<td><strong>Recommended.</strong></td>
<td>60%</td>
<td>100%</td>
</tr>
<tr>
<td>keywords</td>
<td><strong>Recommended.</strong></td>
<td>70%</td>
<td>99.9%</td>
</tr>
<tr>
<td>spatialCoverage</td>
<td><strong>Recommended.</strong></td>
<td>80%</td>
<td>92%</td>
</tr>
<tr>
<td>temporalCoverage</td>
<td><strong>Recommended.</strong></td>
<td>10%</td>
<td>15%</td>
</tr>
<tr>
<td>variableMeasured</td>
<td><strong>Recommended.</strong></td>
<td>30%</td>
<td>83%</td>
</tr>
<tr>
<td>version</td>
<td><strong>Recommended.</strong></td>
<td>40%</td>
<td>95%</td>
</tr>
<tr>
<td>sameAs</td>
<td><strong>Recommended.</strong> Same data, different URL.</td>
<td>10%</td>
<td>100%</td>
</tr>
</tbody>
</table>

[https://developers.google.com/search/docs/data-types/dataset](https://developers.google.com/search/docs/data-types/dataset)
## Use of Guidelines cont'd

<table>
<thead>
<tr>
<th>Dataset Properties</th>
<th>Provider Usage</th>
<th>Dataset Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Implemented</td>
</tr>
<tr>
<td><strong>identifier</strong></td>
<td>30%</td>
<td>10,556 datasets</td>
</tr>
<tr>
<td><strong>author/creator/contributor</strong></td>
<td>80%</td>
<td>28,765 datasets</td>
</tr>
<tr>
<td><strong>funder (not awards)</strong></td>
<td>30%</td>
<td>4,069 datasets</td>
</tr>
<tr>
<td><strong>distribution</strong></td>
<td>60%</td>
<td>45,221 datasets</td>
</tr>
<tr>
<td><strong>license</strong></td>
<td>70%</td>
<td>42,523 datasets</td>
</tr>
<tr>
<td><strong>hasPart</strong></td>
<td>10%</td>
<td>122 datasets</td>
</tr>
</tbody>
</table>

*ex: linking PhysicalSamples to Datasets*
Project 418 Tools: Four main tools developments

- **Gleaner**: The access and process tool.
  - Go based, deployed as a Docker image
  - A Docker Compose/Stack file can deploy Gleaner and all its dependencies
  - Accesses SDO data graphs and processes them into indexes

- **Geodex**
  - A simple web site used to provide a test UI for the generated indexes from Gleaner. NOT a production site. Only for testing.
  - Also a set of test APIs to explore leveraging the index

- **SHACL**
  - Not a service, though we have developed a RESTful service around the TopQuadrant library. This service allows easy access to SHACL processing.

- **Fence**
  - In development, a inspection site/tool along the lines of JSON-LD playground or Google SDTT but more focused on the needs of the NSF data facilities
Project 418 Tools: Gleaner

Two components

- **Summoner**
  
  Accesses files via sitemap.xml and extracts JSON-LD (validates JSON-LD form)

- **Miller**
  
  A simple framework to support multiple processing pipelines on the summoned JSON-LD. These include; Graph, Spatial, Prov, Full text indexing (multiple options), SHACL validation, more

- Go based, Dockerized, GitHub: [https://github.com/earthcubearchitecture-project418/gleaner](https://github.com/earthcubearchitecture-project418/gleaner)
Project 418 Tools: Geodex, SHACL, Fence

Geodex.org
A simple testing site with web interface and services (RESTful). Allows us to leverage the indexes to see how they perform. (other clients too)

SHACL
Developing a set of “shape files” to test data graphs against. Developing as services to use in Gleaner and other tool chains. Allows providers to validate their data graphs.

Fence
Under development. A website to allow inspection of data graphs. Test their use via web components, validate with SHACL, route to other tools.
A mapping tool from metadata to schema.org

Josef Hardi, John Graybeal
Stanford BMIR
About the tool

- We have developed an ETL tool that uses a *system-independent mapping language* that works well with source data in open-standard formats to produce schema.org-compliant instances.

- The idea of using a high-level language for data mapping is not new and many off-the-shelf ETL tools have it. We re-implemented the idea as an open source library where we then built a tool that could make the whole data transformation processes automatic and transparent to users.

- Currently, our tool supports *XML* and *RDF* source formats.

- We have transformed over 550,000 public scientific metadata from [ClinicalTrials.gov](https://clinicaltrials.gov), [PubMed](https://pubmed.ncbi.nlm.nih.gov) and [DrugBank](https://www.drugbank.ca) by using the tool.
The tool requires an input of a data source document and a data mapping. In the batch mode, the source documents can be many and a single mapping will process them. Other than that, the ETL processing details are hidden from the users.
Choosing the mapping language

Users can choose from two language options. Both work equally well but they have pros and cons as described below:

1) **RML / “RDF Mapping Language”**
   - A well-known mapping language in the semantic web community
   - Rich language features
   - Too verbose

2) **CAML / “CAML is Another Mapping Language”**
   - Less verbose
   - An experimental language developed for the tool
   - Less rich language features

A map requires a **target field** and a **source field** as a reference for extracting the value.

<table>
<thead>
<tr>
<th>Target:</th>
<th>Source:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCHEMA.ORG FIELDS</strong></td>
<td><strong>VALUE FIELDS</strong></td>
</tr>
<tr>
<td>@type:</td>
<td>'MedicalTrial'</td>
</tr>
<tr>
<td>name:</td>
<td>/clinical_study/official_title</td>
</tr>
<tr>
<td>identifier:</td>
<td>/clinical_study/id_info/nct_id</td>
</tr>
<tr>
<td>status:</td>
<td>/clinical_study/overall_status</td>
</tr>
<tr>
<td>description:</td>
<td>/clinical_study/detailed_description/textblock</td>
</tr>
<tr>
<td>studySubject:</td>
<td>/clinical_study/condition</td>
</tr>
<tr>
<td>phase:</td>
<td>/clinical_study/phase</td>
</tr>
<tr>
<td>code:</td>
<td>/clinical_study/condition/browser/mesh_term</td>
</tr>
<tr>
<td>codeValue:</td>
<td>'MedicalCode'</td>
</tr>
<tr>
<td>studyLocation:</td>
<td>/</td>
</tr>
<tr>
<td>@type:</td>
<td>/clinical_study/location/facility</td>
</tr>
<tr>
<td>name:</td>
<td>'AdministrativeArea'</td>
</tr>
<tr>
<td>address:</td>
<td>/name</td>
</tr>
<tr>
<td>@type:</td>
<td>/address</td>
</tr>
<tr>
<td>addressLocality:</td>
<td>'PostalAddress'</td>
</tr>
<tr>
<td>addressRegion:</td>
<td>/city</td>
</tr>
<tr>
<td>postalCode:</td>
<td>/state</td>
</tr>
<tr>
<td>addressCountry:</td>
<td>/zip</td>
</tr>
</tbody>
</table>

**Figure 2:** Data mapping in CAML
Automatic and transparent processing

The tool will automatically create a data transformation program based on the chosen data format. At the end of the pipeline, it will produce a schema.org instance in JSON-LD format. Two possible paths are available:

- If XML is chosen then a data transformation program based on XSLT will be auto-created,
- If RDF is chosen then a data transformation program based on SPARQL will be auto-created.

**Figure 3**: The auto-creation from a data mapping to an XSLT transformation specification
Please visit [https://schemaorg.metadatacentor.org/playground](https://schemaorg.metadatacentor.org/playground) and try some of the data transformation examples to get a better sense about the mapping and the tool itself.

All code resources are available freely at [https://github.com/metadatacentor/schemaorg-pipeline](https://github.com/metadatacentor/schemaorg-pipeline)
Data Discovery Paradigms IG
Using schema.org for research data discovery TF

Mingfang Wu / Australia Research Data Commons
Data discoverability via web search engines

- Common way to describe metadata across many resources
- Structured metadata even in resources without an API
Extensions to schema.org

Extensions

As schema.org has grown, we have developed mechanisms for community extension as a way of adding more detailed descriptive vocabulary that builds on the schema.org core.

Hosted extensions are managed and published as part of the schema.org project, with their design often led by one of more dedicated community groups.

External extensions live elsewhere in the Web, typically managed by other organizations with their own processes and collaboration mechanisms. Please consult external documentation for full details of their vocabulary, versioning system and release history.

Hosted Extensions

Specialized terms from hosted extensions can be used alongside core schema.org terms like Event and Person. For example in the auto extension there is a property for emissionsCO2, and in the bibliographic extension we have a property publisherImprint.

Using the extension mechanism the core vocabulary is extended by the following hosted extensions:

- auto.schema.org
- bib.schema.org
- health-lifesci.schema.org
Task force objectives

- **Common elements across research domains**
  - Objective 1 - Define research schemas types and minimum information guidelines for discoverability and accessibility
  - Objective 2 - Crosswalk and gap analysis evaluating existing standards and guidelines

- **Domain specific elements**
  - Objective 3 - Review existing efforts working on Schemas to describe research types
  - Objective 4 - Engagement and communication strategy; collaboration and with existing efforts
This survey will gather information on existing work involving schemas to describe research data and related resources.

Analysis of the survey results will help repositories and the proposed working group understand current practices, identify commonalities, gaps and barriers in using schemas for describing and discovering research datasets.

It is envisaged that the survey results can inform the work group in planning its objectives and deliverables, along with sharing practices between data repositories.
Survey: current practices in using Schemas to describe research datasets

- Repository/Catalogue profile
  - Organisation name, URL of catalogue, domain covered, metadata schema(s)

- Current status of applying schema.org
  - Mapping from/to schema.org, mapping between other (non schema.org) schemas, the way schema.org is being applied

- Issue identification
  - Missing resource type, property, or relation property

- Suggestions to the research schemas ‘working group’
Survey: Catalogue profile

20 participations, dated on 25/03/2019

Domains covered by participating catalogues (20/20)

- Thermal Engineering/Process Engineering, 1
- Medicine, 2
- Social and Behavioural Sciences, 1
- Materials Sciences and Engineering, 1
- Life Sciences and Biomedical, 1
- Geosciences, 2
- Biology, 2
- Agriculture; Forestry; Horticulture; Veterinary Medicine, 1
- Other: Biogeochemical Dynamics, Culture Heritage
- All domains, 7

Types of metadata schemas (19/20)

- DCAT-AP, DCAT DataCite Schema.org ISO2146 DataCite + Dublin Core
- Customised - standard compliant (6)
- General purpose (10)
- Domain specific (3)
- EML DATS DDI
Survey: Current status of applying schema.org

Mapping from/to schema.org (15/19)
- EML -> schema.org
- B2FIND <-> schema.org
- ISO2146/RIF-CS -> schema.org
- Dataverse -> schema.org
- CATS -> schema.org
- HCLS -> schema.org
- DCAT-AP -> schema.org

Ways of applying schema.org (16/19)
- Mark up of landing page in JSON-LD (8)
- Metadata schema (6)
- Possibly complementing (3)
### Resource type
- Scientific measurement
- Environmental entities
- Data services / APIs
- Tissue samples
- Data access arrangements
- Data reuse conditions/consent
- Data Controller (legal frameworks)
- Performances
- Digital artefacts
- Some from DataCite ResourceTypeGeneral, e.g. DataPaper, Model, Workflow

### Relation property
- Dataset -> FundingAward
- Dataset -> Cruise (Event)
- Dataset -> Funder
- Study -> Study design
- Many from DataCite <relationType>, e.g. IsCitedBy, HasVersion, IsNewVersionOf, ...

### Issues:
- Mapping multiple relation types into one
- Not sure if predicates (e.g. in the OBO Foundry Relation Ontology (RO), EnvO, and SWEET) are expressible

### Property
- Keyword -> external vocabulary (e.g. DefinedTerm, CategoryCode)
- Controlled vocabulary from DataCite <dateType>, e.g. Accepted, Available, Copyrighted, Updated, etc.
- Some semantic difference, e.g. schema:Dataset:name, DataCite:Author:name
- Specific term to generic term, e.g. dct:provenance to schema:description
Unfettered schema heterogeneity will hinder interoperability, so some mention of harmonisation strategies to be provided in the (still-to-be-developed) RDA Guidelines would be useful if we are hoping for: close data compatibility if not integratability.

Schema.org (for metadata) rarely uses or allows for the use of controlled terminology (i.e. semantics) in the data values, and it does not make any effort to establish constraints, which result in a large set of metadata not actionable, validatable, or interoperable.

Schema.org is not supporting some basic metadata information which is common across domains. This includes controlled vocabularies/thesauri/code lists.
Would like to add Arts & Humanities in addition to scientific types.

The ‘problem’ of sensitive data and managed access datasets needs to be tackled by any group attempting to provide a comprehensive metadata framework.

Developing/describing use cases and examples

One valuable contribution would be to either provide a list of well-tested software for “collecting” semantic assets (vocabularies, schemas, ontologies …) or host such a repository themselves.

Being able to document consistently “semantic assets” would be fundamental for enabling interoperability via re-use.
Survey: Participants and Organisations - Thank you!

- NSF DataONE, NSF Arctic Data Center, university of California, Santa Barbara
- Scientific Computing Department, STFC Daresbury Laboratory
- Heinrich Widmann (European Datainfrastructure (EUDAT), hosted at and maintained by Deutsches Klimarechenzentrum)
- Joel Benn (Australian Research Data Commons)
- Douglas Fils (EarthCube Science Support Office)
- Adam Shepherd (Biological and Chemical Oceanography Office (BCO-DMO))
- Matt Styles (UK Clinical Research Council - Tissue Directory and Coordination Centre)
- Julian Gautier (Harvard Dataverse)
- Europeana
- Josef Hardi, John Graybeal (Stanford Center for Biomedical Informatics Research, Stanford University)
- Bruce Wilson (ORNL Distributed Active Archive Center for Biogeochemical Dynamics)
- Steve Canham, Christian Ohmann (European Clinical Research Infrastructure Network)
- Leopold Talirz (Swiss National Centre of Competence in Research for the Computational Design and Discovery of New Materials))
- Andrea Perego (European Commission, Joint Research Centre (JRC))
- Philippe Rocca-Serra, Susanna-Assunta Sansone (Oxford e-Research Centre, University of Oxford, UK)
- Chris Hunter (GigaScience Database, China National Gene Bank (CNGB))
- DataDryad (Version re-launch this summer)
- Kerrin Borschewski (Consortium of European Social Science Data Archives, European Research Infrastructure Consortium)
- Institut national de la recherche agronomique (INRA)
- The British Library
Survey instrument design
- Leyla Garcia (Elixir Hub)
- Nick Juty (ELIXIR-UK)
- Fotis Psomopoulos (INAB|CERTH)
- Siri Jodha Khalsa (NSIDC)
- Mingfang Wu (ARDC)

Provide feedback to the instrument
- Doug Fils (Ocean Leadership)
- Joel Benn (ARDC)
- Simon Cox (CSIRO)
- Kathleen Gregory (DANS)

And to those who helped to promote the survey!!

Survey: https://goo.gl/iLunug
Setting up a RDA WG

Mingfang Wu
Case statement

- Case statement content (https://goo.gl/jCin6h)
  - WG Charter
  - Value Proposition
  - Engagement with editing work in the area
  - Work plan
  - Adoption plan
  - Initial membership

The group page: https://www.rd-alliance.org/groups/research-schemas-wg
RDA WG case statement review process

1. WG Case Statement Development
2. Community Review
3. TAB Review
4. Council Review
5. Group Endorsement

https://rd-alliance.org/groups/creating-and-managing-rda-groups/creating-or-joining-rda-working-group.html
WG review criteria

- Fit with the overall RDA vision and mission
- International membership spanning, ideally, three or more continents
- 2-4 co-chairs leading the initiative
- Measurable outcomes
- Outcomes will foster data sharing and/or exchange, and be taken up by the intended community
- Proposed work, outcomes /deliverables, and Action Plan described in the Case Statement can be accomplished in 12-18 months
- Appropriate scope of the WG
- The effort adds value over and above what is currently being done within the community.

https://rd-alliance.org/groups/creating-and-managing-rda-groups/creating-or-joining-rda-working-group.html
Group Discussion

research data sharing without barriers
rd-alliance.org
Discussion

- WG Case statement
- Schemas and extensions
- Tooling (to consume, to generate)
- Guidelines
- FAIR alignment in research schemas
- ...
Action items and next steps

research data sharing without barriers
rd-alliance.org
1. Review of Actions coming out of this meeting
   - Action 1 (responsible person)
   - Action 2 (responsible person)

2. Next Steps