

SWAT4LS 2018 Hackathon proposals

[Theme: Bioschemas Markup and Support Tools](#)

[Goals](#)

[Prerequisites](#)

[Expected Results](#)

[Theme: Metadata Interoperability - Mapping Metadata from Health Care, Clinical Research and EHR utilizing ISO 21526](#)

[Goals](#)

[Prerequisites](#)

[Expected Results](#)

[Theme: Data2services, converting your data to a standard data model.](#)

[Goals](#)

[Prerequisites](#)

[Expected Results](#)

[Theme: Schemas and Vocabularies for Agriculture and Plant Biology](#)

[Goals](#)

[Prerequisites](#)

[Expected Results](#)

[Please add your proposal here](#)

Theme: Bioschemas Markup and Support Tools

Correspond: Alasdair Gray (A.J.G.Gray@hw.ac.uk)

[Bioschemas](#) aims to make life sciences resources on <mailto:A.J.G.Gray@hw.ac.uk> web more Findable. It does this by recommending a small set of schema.org properties to embed in the resource. Where such properties do not exist, e.g. properties to describe a gene, the Bioschemas community are proposing extensions to schema.org.

Goals

- Generate markup for life sciences resources
- Enable validation of generated markup

Prerequisites

- Markup: none, although an awareness of JSON-LD and schema.org is an advantage
- Validation: ShEx and Javascript

Expected Results

- Deployment of markup in web resources
-

Theme: Metadata Interoperability - Mapping Metadata from Health Care, Clinical Research and EHR utilizing ISO 21526

Correspond: Matthias Löbe (matthias.loebe@imise.uni-leipzig.de)

Data integration and data sharing are ongoing fields of research in health care. As part of this year's "Medical Informatics Initiative" in Germany, all university hospitals are to make their health care data available to researchers within a few years. In addition to technical and regulatory issues, this poses a major challenge for semantic interoperability, as different standards for information modelling compete with each other, e.g.:

- HL7 CDA, HL7/IHE Value Sets, CIMI
- FHIR Resources
- CDISC ODM/SDTM/CDASH
- openEHR Archetypes
- Common Data Models (OMOP, PCORnet)
- Semantic Web/Linked Data

Goals

The aim of the hackathon is to **map exemplary data elements** from clinical research and care, which are available in various information models, into a common metamodel. The metamodel is represented by relevant ISO standards:

- ISO 21526 Healthcare Informatics - Metadata Repository Requirements (Working Draft 2018-03-12) which itself is based on
 - ISO 11179-3 Metadata registry
 - ISO 19763-12 Metamodel interoperability

ISO 21526 is currently under Community development TC 215 Health Informatics.

The information architectures of the various consortia of the Medical Informatics Initiative will be based (and will be based) on different information models. But within two or three years, an overarching interoperability for data queries by external researchers will have to be ensured. So, a way of mapping between the information models will be required. One step in this direction is a mapping to a common standard/metamodel. In addition, current approaches of data sharing, e.g. aspects of data stewardship (data quality, provenance and accessibility), FAIR data and metadata, Data Sharing and annotation with medical terminology (LOINC, SNOMED CT) will be examined.

Prerequisites

- Basic knowledge of ISO 11179-3, ISO 19763, ISO 21526 is required
- Some knowledge of the Health Care domain and data modeling in information systems
- Knowledge of at least one standards for information modelling mentioned above

Expected Results

- To find and state weaknesses in the ISO 21526 draft
 - To get a deeper insight of existing standards for medical information modeling and their modeling power
 - To develop best practices to translate mappings into a common metamodel
 - To align with medical terminologies and research data principles (provenance etc.)
-

Theme: Data2services, converting your data to a standard data model.

Correspond: Vincent Emonet (vincent.emonet@maastrichtuniversity.nl), Alexander Malic (alexander.malic@maastrichtuniversity.nl)

Today an increasingly amount of data is available on the Web, but this data usually comes in a myriad of formats (XML, CSV, RDB...) with no inherent semantic representation. Inspired by the FAIR data principles (Findable, Accessible, Interoperable, Reusable), we propose Data2services, a framework built with scalability in mind to convert any type of data to a standard data model that follows the Semantic Web standards.

This data is then accessible as RDF through a SPARQL endpoint. We are also working on automatically generating web services based on the data model for simplified access.

Goals

- Execute the [data2services pipeline](#) on your machine using Docker to convert your data to a generic RDF, where the data representation is based on the input data structure.
- Craft SPARQL queries to map the relevant data from the generated generic RDF to the target data model of your choice.
- Draft and start to develop web services that consume and expose the refined RDF to simplify its access.

Prerequisites

- Docker already installed, or admin rights on your machine to install Docker
- Bring data you want to transform (relational database and/or XML-, CSV-, TSV-, PSV-files)
- Provide a target data model to convert your data to, or we can propose you one

Expected Results

- Data2services running on your hardware using Docker

- Your data now complying to a standard data model, and accessible from a SPARQL endpoint and auto-generated API.
 - A draft or a prototype of a web service that expose data2services RDF
-

Theme: Schemas and Vocabularies for Agriculture and Plant Biology

Correspond: Marco Brandizi (marco.brandizi@rothamsted.ac.uk)

As machine-readable data, APIs and linked data are growing popularity in the fields of agriculture, plant research and food, we have an increasing need to define common schemas and terminologies.

In fact, these are fundamental tools in an area that ranges from research to business, consumers and public policies.

schema.org, bioschemas and fairsharing.org are good examples of projects and resources that could be leveraged to cover the specific needs in agrifood.

Goals

The main aim of the hackathon is to outline simple and practical extensions of existing schemas/ontologies, like schema.org and bioschemas.org, to cover the typical entities used in the agriculture, plant biology and food field. Examples of such entities are:

- Organism Descriptors
 - Specie taxonomies
 - Pedigrees
 - Phenotypes
- Soil descriptors
 - (plot size, composition, treatments, fertilisers/pesticides, irrigation)
- Weather Data
 - (variables like temperatures, precipitations, wind speed, descriptors like station geo-location)
- Experimental Data Reporting
 - Field Trial Experiments
- Molecules (Proteins, Genes, etc)
- Molecular Interactions (pathways, reactions, gene expression)
- Publications
- Metadata descriptors for data sets
- Relevant links between research, farming and food products, e.g., link food ingredients to plant species taxons, as well as health-related facts (e.g., allergens).

Prerequisites

- Knowledge of the application domain, and/or data modelling would be useful

Expected Results

- A first informal document collecting main entity types and their relations
- Ideas on how to leverage existing frameworks like schema.org and bioschemas
- Possibly first implementation examples.
- Deployment of a validation tool

Please add your proposal here