The **EDAM** ontology and its integration into Galaxy

**Hervé Ménager**, Jon Ison, Matúš Kalaš, Veit Schwämmle and EDAM Contributors

**Overview**

EDAM is an ontology of well established, familiar concepts that are prevalent within bioinformatics, including types of data and data identifiers, data formats, operations, and topics. EDAM has a simple structure, and comprises a set of concepts with terms, synonyms, definitions, relations, links, and some additional information (especially for data formats).

8 consecutive stable versions of EDAM have been released since July 2015 (version 1.10), with version 1.17 being the current one at the time of the abstract submission. EDAM is developed in a participatory and transparent fashion, with a growing community of contributors. EDAM is used by multiple bioinformatics projects, including ELIXIR bio.tools, Debian Med and the Common Workflow Language.

**Concepts** - All concepts have a name (term or label) and a definition. Further, a concept may have relations (see below) to other EDAM concepts, as well intrinsic properties, e.g. an identifier may have a regular expression defining its syntax.

**Hierarchy** - Every concept (excluding top-level concepts) is related to one or more other concepts within the same branch by an is a (specialisation) relation. Hence EDAM has 4 primary hierarchies (for Data, Format, Operation, and Topic).

**Relations** - Concepts are related by defined relation types (see figure), which reflect well established or self-evident principles, and are used primarily to define internal consistency of EDAM.

**Datatypes annotation**

The core Galaxy datatypes have been mapped to EDAM. This mapping is defined in the Galaxy source code, where datatype classes are annotated with EDAM data and format terms. We strongly encourage the authors of new datatypes to include these mappings.

**Tools annotations**

Tools can now be annotated with EDAM topics and operations using the specific tags `<edam_topic>` and `<edam_operation>`.

**Integration with bio.tools**

The annotation of Galaxy datatypes and tools is currently used for the integration with the ELIXIR bio.tools registry:

- **ReGaTE** (github.com/c3bi-pasteur-fr/regate) is a utility that enables the automated registration of Galaxy tools available on a given server in the bio.tools (bio.tools) registry.
- Conversely, **ToolDog** (github.com/bio-tools/ToolDog, see Poster 10) is a tool that uses the bio.tools description of a command line tool to generate templates for Galaxy XML and CWL tool descriptions.

**Architecture**

**Datatypes**

- **Data**
  - Sequence
  - Image
  - Table
  - Structure
  - DNA
  - Protein
  - Data
  - Data identifier
  - Data format
  - Data type
  - Data category
  - Data class
  - Data type
- **Data formats**
  - XML
  - YAML
  - CSV
  - JSON
  - FASTA
  - VCF
  - BED
- **Operations**
  - Sequence alignment
  - Sequence comparison
  - Sequence analysis
  - Sequence identification
  - Sequence classification
  - Sequence annotation
- **Topics**
  - **Biology**
  - **Chemistry**
  - **Computer science**
  - **Engineering**
  - **Healthcare**

**Contact & Information**

Enquiries: edam@elixir-dk.org

Documentation & downloads:

 github.com/edamontology/edamontology
 edamontologydocs.readthedocs.io

**Browsing**

EDAM @ OLS
elcbi.univ-angers.fr/edam

EDAM @ BioPortal
bioportal.bioontology.org/ontologies/EDAM

**Developers**

Jon Ison (DTU, DK) - lead developer
Matúš Kalaš (UoB, NO)
Veit Schwämmle (SDU, DK)

**Acknowledgements**

We are deeply thankful to the Galaxy community for their contributions in enabling the integration of Galaxy and EDAM, especially John Chilton, Björn Grüning, Eric Rasche, Nicola Soranzo and Marius Van Den Beek.