1. MOTIVATION
Researchers demand simple and powerful means to organise, find, compare, use and connect an increasingly large and complex set of tool and data resources. These tasks depend on consistent, machine-understandable resource descriptions. There is an urgent need for an ontology that unifies semantically the common bioinformatics and biomedical concepts and provides a controlled vocabulary for the annotator.

2. EDAM ONTOLOGY
EDAM (Figure 1) includes 5 sub-ontologies (Table 1) within the scope of bioinformatics and biomedical resource (tool and data) description. There are 5 types of relationships (Table 2) which relate concepts from different branches.

EDAM includes:
- A hierarchical set of coarse topics for categorising any bioinformatics resource
- Comprehensive sets of terms for describing common types of data and operations
- Comprehensive catalogues of common data formats and types of data identifiers

EDAM provides a starting point for nomenclature and is ready for use in production environments. The goal is to provide, at a coarse level at least, a controlled vocabulary for all the topics, data types, formats, identifiers and operations in general use.

3. EDAM CONCEPTS
EDAM concepts are well established and familiar within bioinformatics and biomedical science. Each concept corresponds to a category of thing. The current version includes over 2000 concepts with names (terms), synonyms, definitions and other useful information, for example regular expressions for validating identifier values, or references to formal definitions of file formats.

Table 1 EDAM sub-ontologies
EDAM includes 5 sub-ontologies which collectively define the scope.

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>topic</td>
<td>A common topic in bioinformatics or biomedical research, including subjects, disciplines, general fields of analysis or techniques etc. e.g. “Sequence alignment”, “Chemoinformatics”, “Phylogenetics”</td>
</tr>
<tr>
<td>data</td>
<td>A type of data in common use e.g. “Sequence alignment”, “Sequence record”, “Image”</td>
</tr>
<tr>
<td>format</td>
<td>A commonly used data format e.g. “FASTA”, “SAM”, “MALE-ML”</td>
</tr>
<tr>
<td>identifier</td>
<td>A label that identifies (typically uniquely) a biological or computational entity or resource e.g. “Ensembl ID”, “EC number”, “Gene symbol”</td>
</tr>
<tr>
<td>operation</td>
<td>A specific, unique function performed by a tool or other resource what is done, but not typically have or in what context e.g. “Sequence database search”, “Data submission”</td>
</tr>
</tbody>
</table>

Table 2 EDAM relations
EDAM includes 5 relations (in addition to is_a) which relate concepts from one branch (in quotes, e.g. ‘data’) to another.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>is_identifier_of</td>
<td>A data ‘identifier’ is an identifier of a certain type of ‘data’ e.g. “EMBL accession is_identifier_of ‘Sequence record’”</td>
</tr>
<tr>
<td>has_output</td>
<td>An ‘operation’ produces a certain type of ‘data’ e.g. “Sequence alignment has_output Sequence”</td>
</tr>
<tr>
<td>has_input</td>
<td>An ‘operation’ consumes a certain type of ‘data’ e.g. “Sequence alignment has_input Sequence”</td>
</tr>
<tr>
<td>is_format_of</td>
<td>A data ‘format’ is a format of a certain type of ‘data’ e.g. “FASTA is_format_of Sequence”</td>
</tr>
<tr>
<td>in_topic</td>
<td>A concept (‘data’ or ‘operation’) is within scope of a ‘topic’ e.g. “Sequence alignment” in_topic “Sequence analysis”</td>
</tr>
<tr>
<td>is_a</td>
<td>A child concept is a specialisation of its parent e.g. “Parsite sequence alignment is_a Sequence alignment”</td>
</tr>
</tbody>
</table>

Applicability
EDAM provides a controlled vocabulary for semantic description in diverse contexts:
- Application software
- Tool collections and packages
- Web services e.g. WSDL files or REST API docs
- Resource catalogues and registries
- Workflows and workflow software
- Databases and ontologies
- XSD data schema and file formats
- Web portals and pages
- Documents, e.g. scientific publications
- and more !

Downloads and interfaces
EDAM in OWL format: http://edamontology.org/projects/edamontology/files/
Bioportal browser: https://bioportal.bioontology.org/ontologies/EDAM
EDAM Ontology Lookup Service: https://www.ebi.ac.uk/ontology-lookup/

Further information
EDAM documentation: http://edamontology.org/ Please subscribe to one or both mailing lists: http://lists.sanger.ac.uk/mailman/listinfo/edamontology-users http://lists.sanger.ac.uk/mailman/listinfo/edamontology-developers

Contact the developer:
Jon Ison
jison@ebi.ac.uk

Thanks!
- Herve Menager (Manchester University, UK)
- Nathalie Conte, Simon Pettifer (CBS EBI, UK)
- Jonassen Inge Jonassen, Carole Goble, (Manchester University, UK)
- Service Manager (European Bioinformatics Institute, Hinxton, Cambridge, UK)
- Nathalie Conte, Carole Goble, Simon Pettifer, Simon Roper, Herve Menager (Table 1, Column 1, Cell 3, Row 4)

doi: 10.1093/bioinformatics/btt113