MONASH University

MONASH
BIOINFORMATICS
PLATFORM

Using Standards for greater reuse of training material

Dr. Sonika Tyagi

Monash Bioinformatics Platform Manager

Training Coordinator EMBL-ABR

Chair Standards Committee GOBLET

Member Education sub-committee ABACBS

EMBL Australia Bioinformatics Resource

GOBLET Global Organisation for Bioinformatics Learning, Education & Training

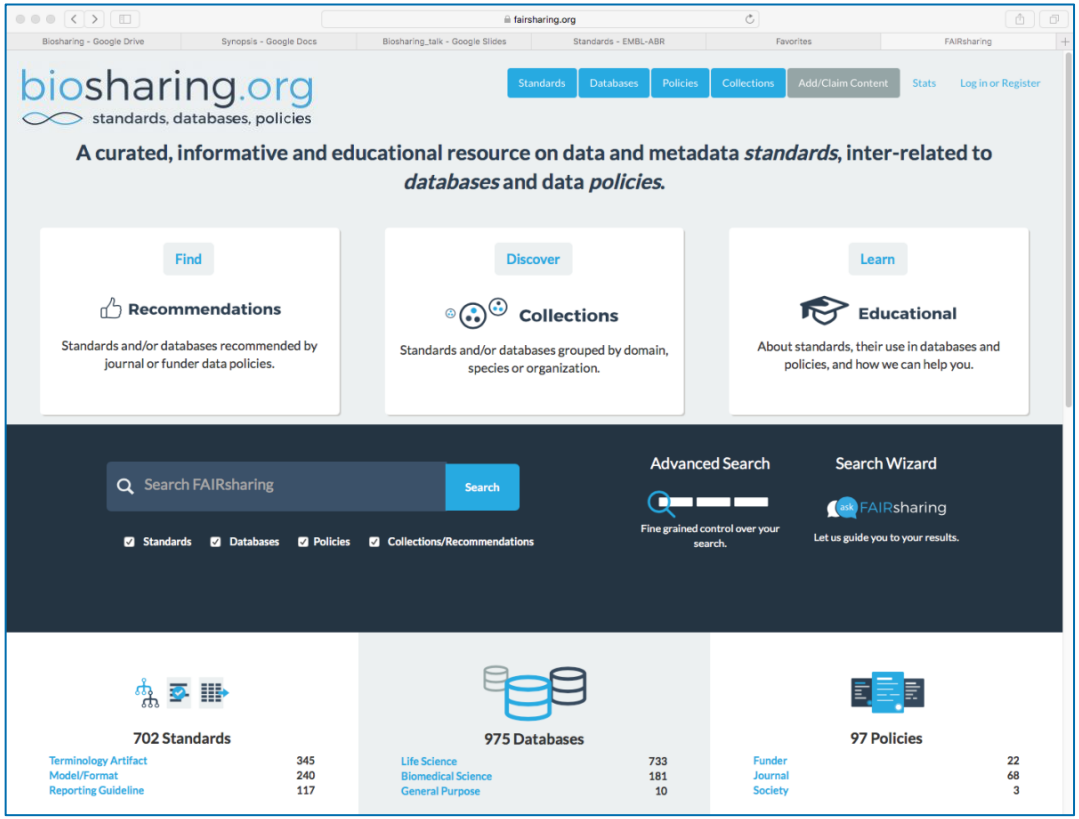
GROUP OF EIGHT AUSTRALIA

Why Standards ?

- In a digital world, **standards** of many types are essential for underpinning **interoperability** between data, computational tools/software that act on the data, and the underlying infrastructure that these tools/software use.

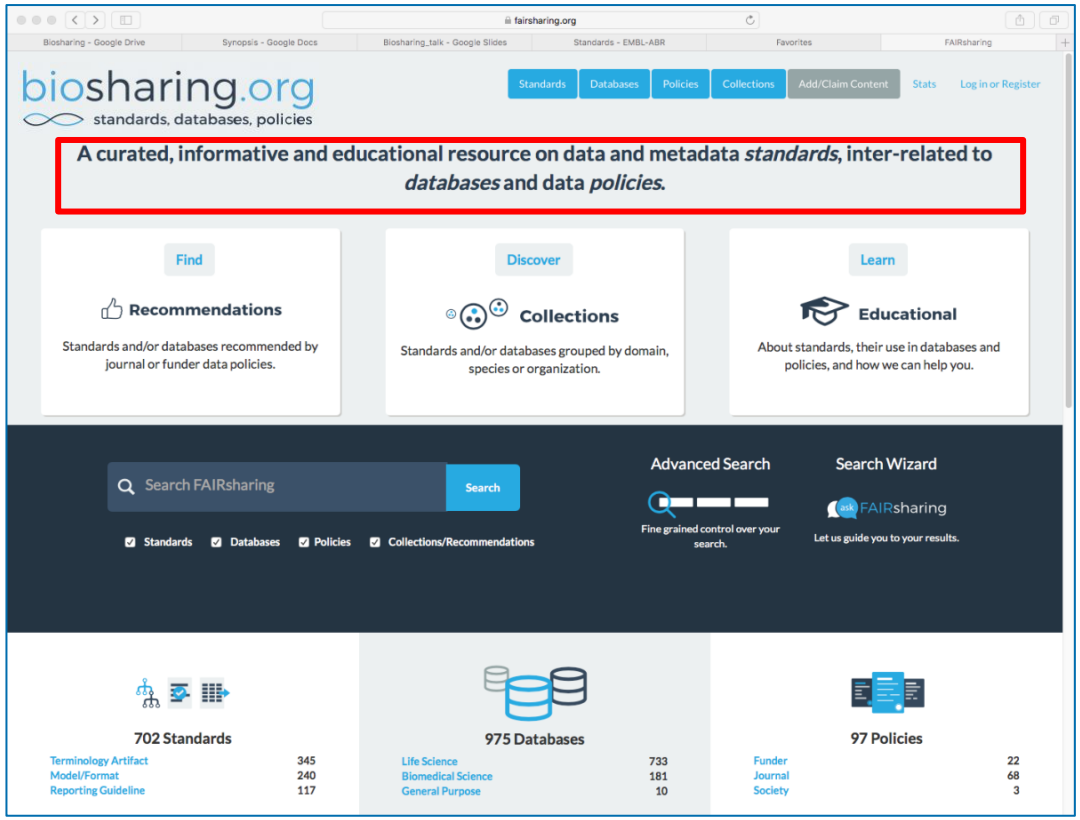
- Standards allow automated efficient exchange between tools and environment.

biosharing.org



Oxford e-Research Centre & University of Oxford

biosharing.org : What it is?



Oxford e-Research Centre & University of Oxford

biosharing.org : Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

biosharing.org : Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

For developers seeking to make their resource more findable

- To register or update an standard and/or database description to make them more discoverable to others and get credit for it.

biosharing.org : Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

For developers seeking to make their resource more findable

- To register or update an standard and/or database description to make them more discoverable to others and get credit for it.

For journal editors/librarians creating guidelines

- To create a view on an interrelated set of standards and/or databases, a simple way to complement data guidelines and assist users.

biosharing.org : Why was it developed ?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

For developers seeking to make their resource more findable

- To register or update an standard and/or database description to make them more discoverable to others and get credit for it.

For journal editors/librarians creating guidelines

- To create a view on an interrelated set of standards and/or databases, a simple way to complement data guidelines and assist users.

For funders developing data policies

- To help refine policy by discovering which standards and databases are inter-related, more used and mature, or are funded by the organisation

biosharing.org : What is in it ?

STANDARDS

The standards in biosharing are manually curated from a variety of sources, including:

- BioPortal (repository of biomedical ontologies),
- MIBBI (repository of Minimum Information specifications for Biological and Biomedical Investigations)
- Equator Network (database of health research reporting guidelines)

Current content - 701 standards:

- Terminologies (ontology, controlled vocabulary, staging systems, etc) 344 (7 Aus)
- Model/Format (exchange formats, markup languages, RIF-CS, etc) 240 (11 Aus)
- Reporting Guideline (diagnostic reporting guidelines, minimum info specs, etc) 117 (10 Aus)

biosharing.org : What is in it ?

DATABASES

A catalogue of databases, described according to the BioDBcore guidelines (a community-defined, uniform, generic description of the core attributes of biological databases), along with the standards used within them; partly compiled with the support of Oxford University Press (NAR Database Issue and *DATABASE* Journal).

Current content - 975 databases:

- Life Science (non-human) 733 (20 Aus)
- Biomedical Science (human) 181 (5 Aus)
- General Purpose (incl. figshare, OSF, zenodo, wikidata, RDA, etc) 10 (1 Aus)

biosharing.org : What is in it ?

POLICIES

A catalogue of data preservation, management and sharing policies from international funding agencies, regulators and journals.

Current content - 97 Policies

- Funder 22 (1 Aus)
- Journal 68 (0 Aus)
- Society 3 (1 Aus)

biosharing.org : Who can contribute ?

Anyone can register to contribute content, or claim ownership of records created by a 3rd party

biosharing.org : Who can contribute ?

Anyone can register to contribute content, or claim ownership of records created by a 3rd party

Recently (2017) biosharing has become discipline agnostic:

biosharing.org : Researcher attitude

biosharing.org : Researcher attitude

2015/16 survey

BioSharing Survey - Summary

Peter McQuilton[1], Pascale Gaudet[2] and Susanna-Assunta Sansone[1].

1. Oxford e-Research Centre, University of Oxford, UK
2. CALIPHO, Swiss Institute of Bioinformatics, CH

EXECUTIVE SUMMARY

A 10-question survey was conducted from the 18th December 2015 to 22nd February 2016 to gather users' views on which features and content they need to make informed decisions, e.g. on how to best select standards and understand their maturity, or to find the databases that implement them. A link to the questions can be found here: https://bd2kccc.org/2016/01/15/biosharing-standards-registry-survey.

https://figshare.com/articles/New_draft_item/3795810

biosharing.org : Researcher attitude

2015/16 survey - 3 months, 533 respondents

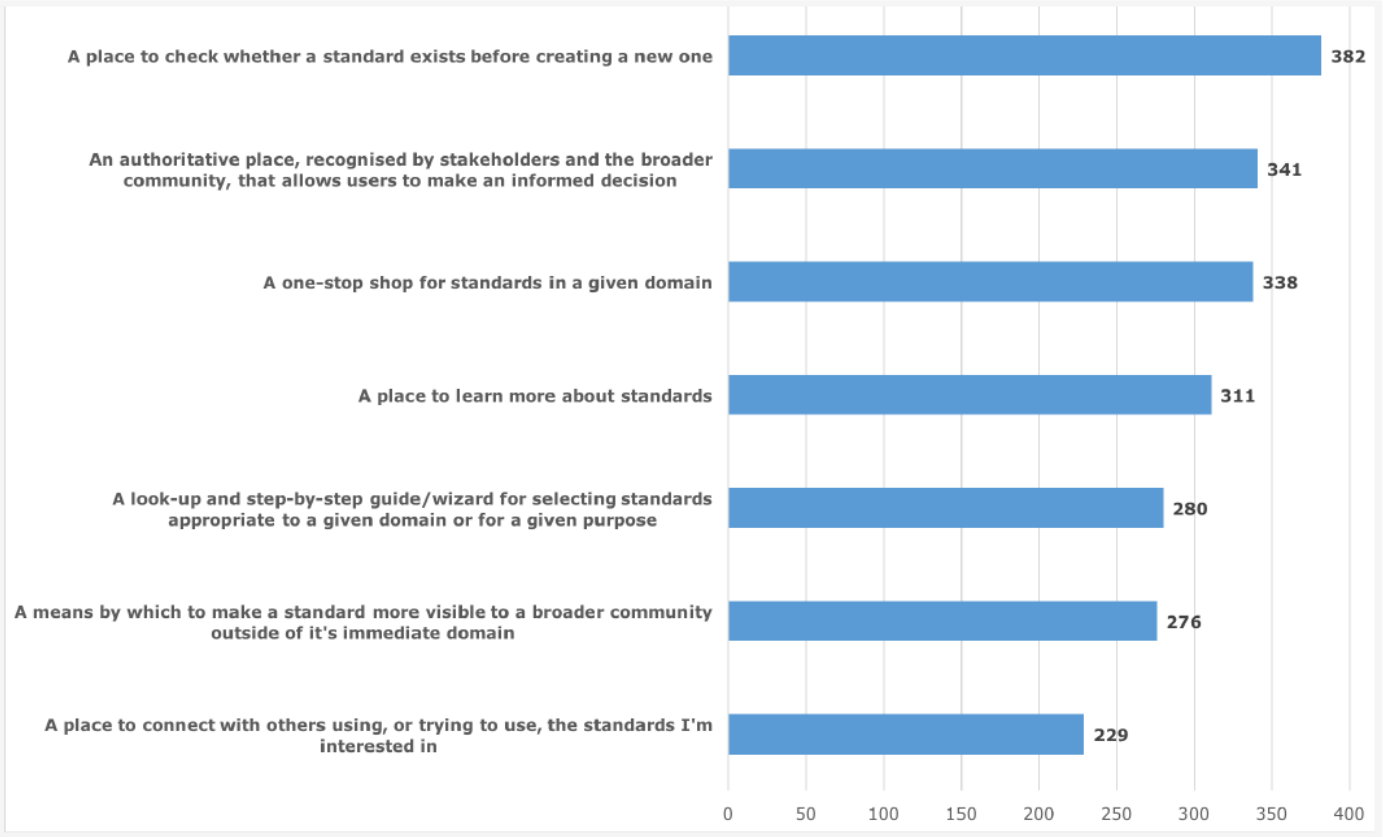
Researcher	323
Tool/database developer	274
Standard Developer/maintainer	206
Data Curator	151
Data Manager	150
Journal Publisher/Editor	31
Librarian	20
Funding Agency	20

- 25% assoc. with Elixir (EU)
- 21% assoc. with NIH BD2K (US)

- 65% aware of biosharing
- 30% had used biosharing

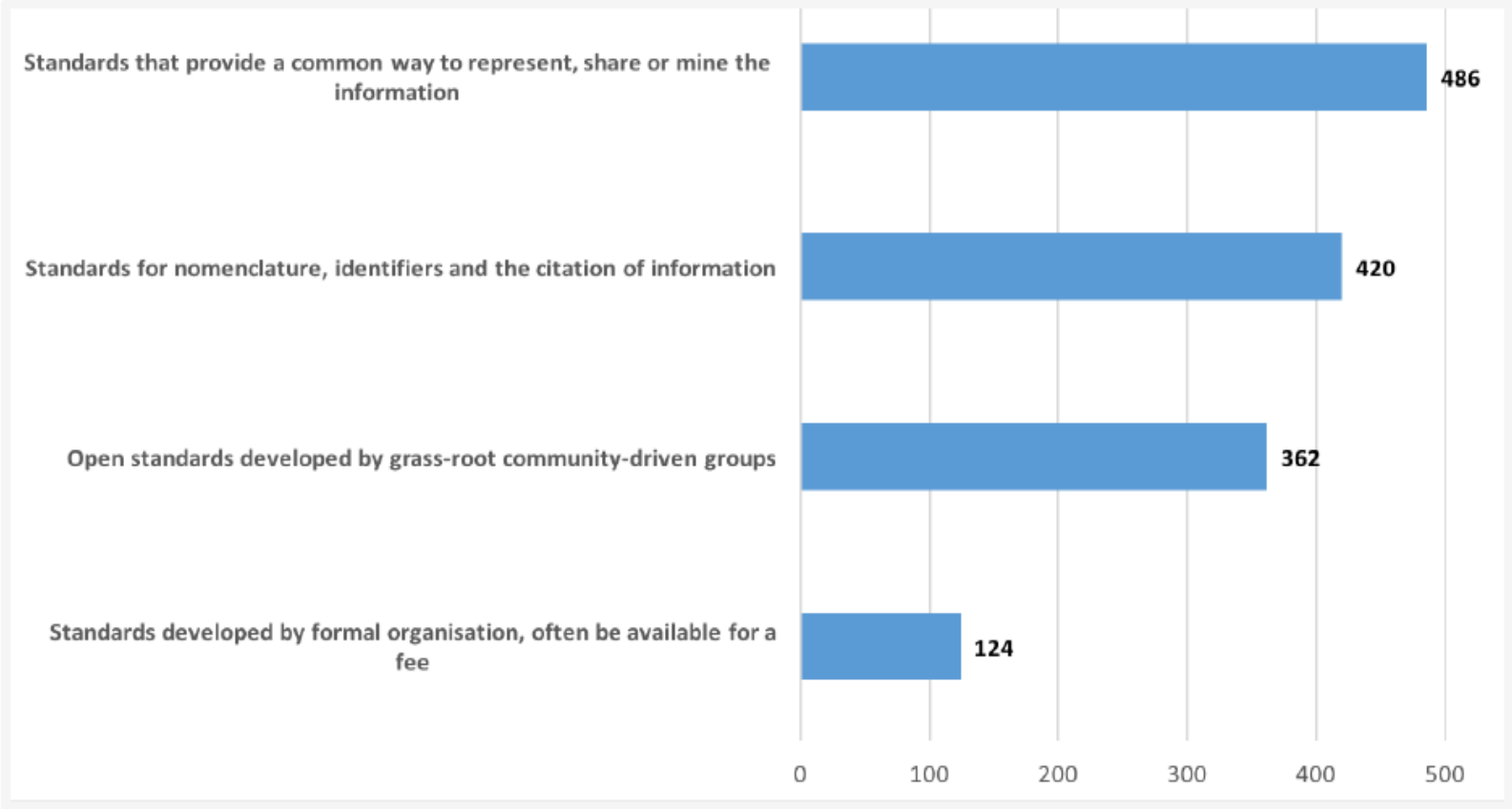
biosharing.org : Researcher attitude

2015/16 survey : What do you need from a standards registry?



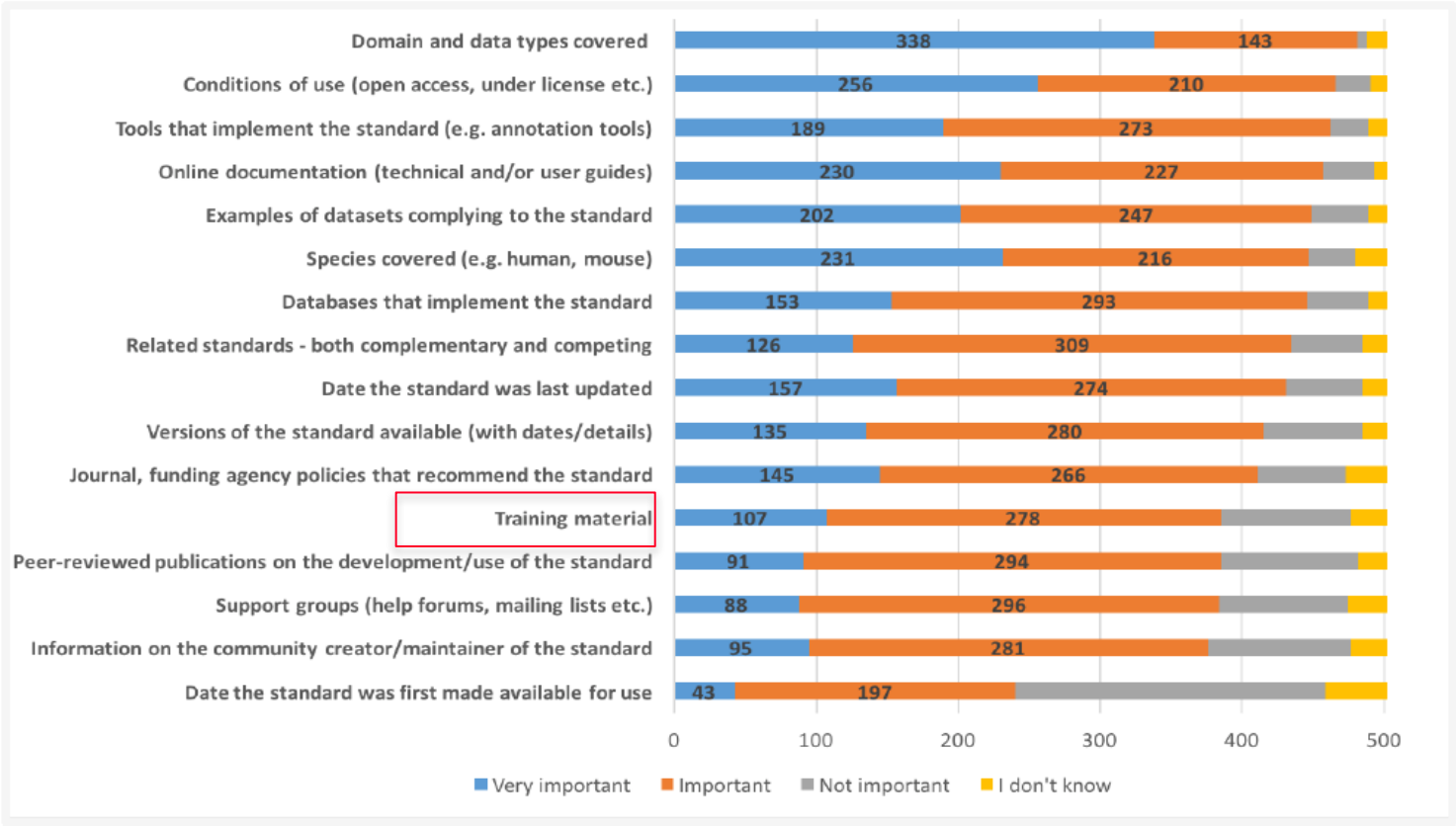
biosharing.org : Researcher attitude

2015/16 survey: What types of standards would you like to see in the registry?



biosharing.org : Researcher attitude

2015/16 survey: What information should a standards registry capture about a standard?



@MonashBioinfo
Bioinformatics.Platform@monash.edu
http://bioinformatics.erc.monash.edu

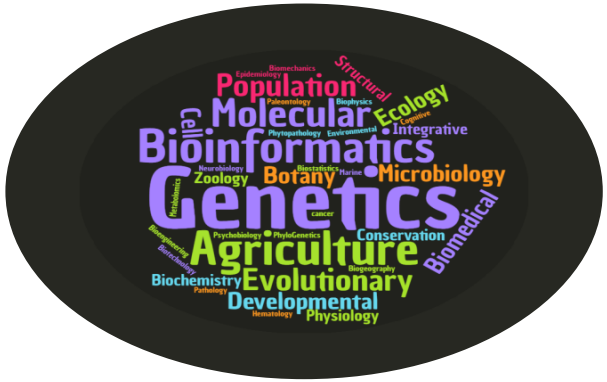
biosharing.org: Awareness in Australian Bioscience community

biosharing.org: Awareness in Australian Bioscience community

How aware is Australia's bioscience community of Standards?

EMBL-ABR + GOBLET Standards Key Area: 2017 Survey into Standards

- Aim: to collect information to direct EMBL-ABR node efforts and resources for the maximum impact for the needs of the Australian research community when it comes to standards across key areas in bioinformatics including data, tools, workflows and training
- Promotion: via each node (email lists, social media, etc.)
- Response rate: extremely low (*n*=17) *



* previous data and tool surveys have previously attracted ~100 responses

How aware is Australia's bioscience community of Standards?

EMBL-ABR + GOBLET Standards Key Area: 2017 Survey into Standards

- Is interest in and/or understanding of standards by biology and bioinformatics researchers in Australia therefore low?
- Awareness and Education is required ?

Overall, there is confusion between file formats and standards for metadata and annotation. Most respondents are unable to name an ontology or are unaware of how to find one. As yet, we have received limited responses relating to training material standards; however, this was listed amongst the top priority standards to be developed

Standards for Training material ?

Bioschemas

- to improve **data interoperability** in life sciences.
  - By using **schema.org** to add semantic markup to webpages for consistently structured information.
  - makes it easier to discover, collate and analyze distributed data.
- a collection of specifications that provide guidelines to facilitate a more consistent adoption of **schema.org** markup within the life sciences.

The IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata

"Reusability of learning objects, to aid discoverability, and to facilitate their interoperability, usually in the context of online learning management systems

Standards for Training material ?

Bioschemas

- to improve **data interoperability** in life sciences.
  - By using **schema.org** to add semantic markup to webpages for consistently structured information.
  - makes it easier to discover, collate and analyze distributed data.
- a collection of specifications that provide guidelines to facilitate a more consistent adoption of **schema.org** markup within the life sciences.

The term "**schema**" refers to the organization of data as a blueprint of how the **database** is constructed.

The schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy. e.g. 'Event' is a type that has properties like 'startDate', 'endDate' and 'description'.

Standards for Training material ?

Bioschemas

GOBLET
Global Organisation for Bioinformatics Learning, Education & Training

Training specification is under development. It is being written by a multi-institutional team from ELIXIR, Pistoia Alliance, GOBLET, TeSS, BioSharing, EMBL-ABR and BBMRI.

You can find more about the project and similar projects on the bioschemas GitHub pages.

Standards for Training material ?

GOBLET
Global Organisation for Bioinformatics Learning, Education & Training

Bioschemas

Standards committee:

- started to test how to make easier sharing bioinformatics events and training materials using schema.org
- minimum information and vocabularies which are important in our community.
  - Minimum descriptors to define training material
    - **http://tiny.cc/TrainingStandards**
  - Bioschema mapping
    - **http://tiny.cc/bioschemaMapping**

Acknowledgements

Jeff Christiansen (EMBL-ABR Data Coordinator)

EMBL Australia Bioinformatics Resource

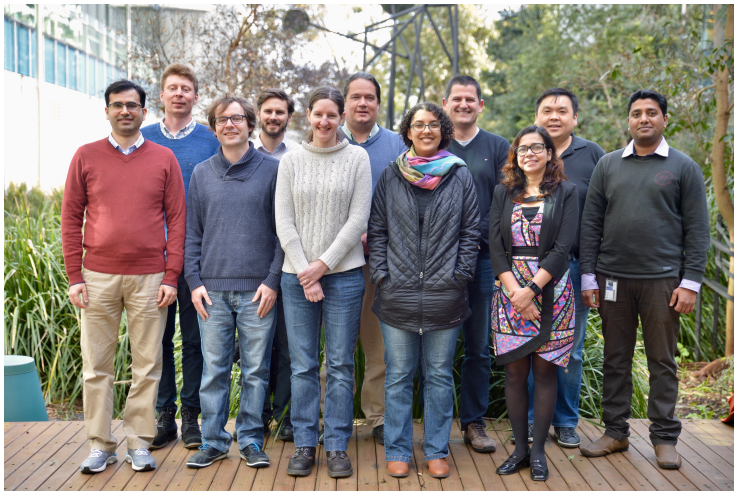
GOBLET
Global Organisation for Bioinformatics Learning, Education & Training

Bioschemas

Standards Committee

- Sonika Tyagi (Acting chair) (Monash, Aus)
- Sarah Morgan(EBI, UK)
- Michael Charston (UTAS, Aus)
- Patricia Palagi (Swiss Institute of Bioinformatics)
- Anette McGrath (CSIRO, Aus)
- Gabriella Rustici (Cambridge University, UK)
- Judit Kumuthini (H3BioNet South Africa)

- Saravanan Dayalan (EMBL-ABR Standards Coordinator)



Monash Bioinformatics Platform team

Group exercises

1. Go to https://fairsharing.org/standards/ and find out if there are existing standards for training material?

2. How many standards are there for the term ‘expression’ ? What different sub-types exist?

3. Go to http://schema.org/docs/schemas.html and look at ’CreativeWork’ type and see if you can use some of the properties to ‘training material’ ? [hint: look at examples here http://schema.org/docs/gs.html]

Registrations are extended till the ABACBS17 week!

We are hosting BioinfoSummer 2017