Metadata for Interoperable Bioscience

Alejandra González-Beltrán
Oxford e-Research Centre, University of Oxford

DDI Metadata Sprint

October 19-23 2015

isa-tools.org

stato-ontology.org
statistics ontology

biosharing.org

LIFE, NATURAL & BIOMEDICAL SCIENCES

NIH Big Data to Knowledge (BD2K)

Roche

elixir UNITED KINGDOM

SEVENTH FRAMEWORK PROGRAMME

bbsrc biotechnology and biological sciences research council

MRC Medical Research Council

NATURAL ENVIRONMENT RESEARCH COUNCIL

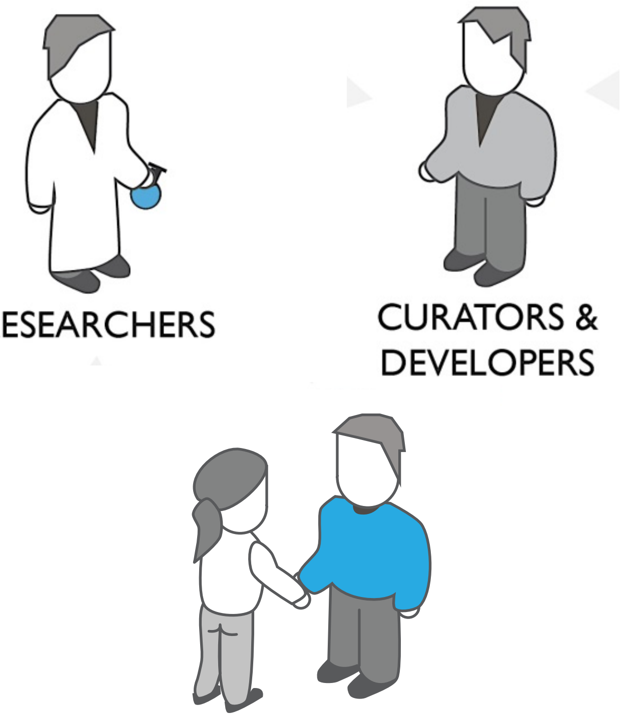
UNIVERSITY OF OXFORD

Oxford e-Research Centre

Our areas of activity:

- Data capture and curation
- Data (nano)publication
- Data provenance
- Open, community ontologies and standards
- Semantic web
- Software development
- Training

Communities we work with/for:



RESEARCHERS

CURATORS & DEVELOPERS

FUNDERS, JOURNAL EDITORS & LIBRARIANS

As part of:

- UK, European and international consortia
- Pre-competitive informatics public-private partnerships
- Standardization initiatives

Some of the groups we engage with incl.:

EMBL-EBI

nature publishing group npg

Imperial College London

UNIVERSITY OF CALIFORNIA SAN DIEGO

MANCHESTER 1824 The University of Manchester

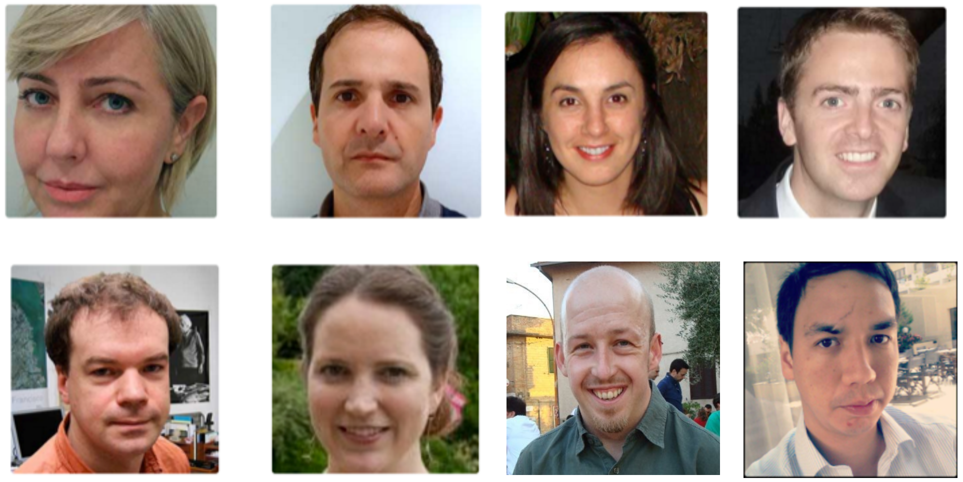
STANFORD SCHOOL OF MEDICINE

HARVARD SCHOOL OF PUBLIC HEALTH

DRYAD

RDA RESEARCH DATA ALLIANCE

elixir Data for Life



Overview

biosharing.org

isa-tools.org

isacreator configurator isacreator bii converter Risa linkedisa

NIH NIH Big Data to Knowledge (BD2K)

CEDAR
CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL

bio CADDIE

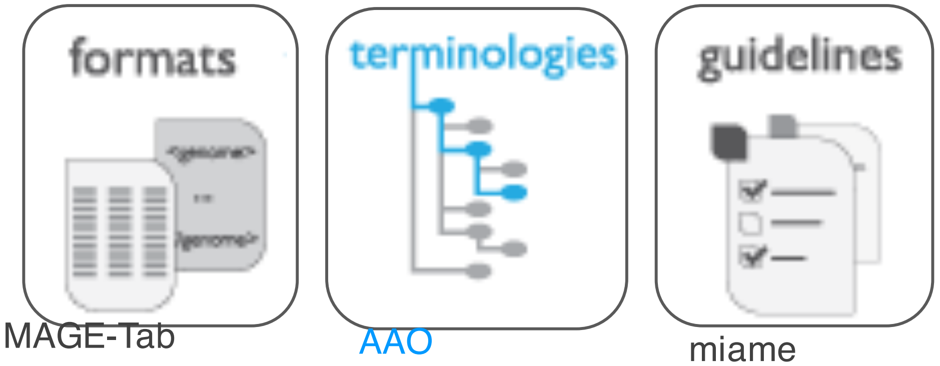
stato-ontology.org
statistics ontology

biosharing.org
Information Resources

Mapping the landscape of **standards**, **databases** and **data policies** in the life sciences (including biological, environmental and biomedical sciences)

data policies

over 600 standards



MAGE-Tab
GCDML
SRAxml
SOFT
FASTA
CML
DICOM
GELML
SBRML
MITAB
MzML
ISA-Tab
SEDML...

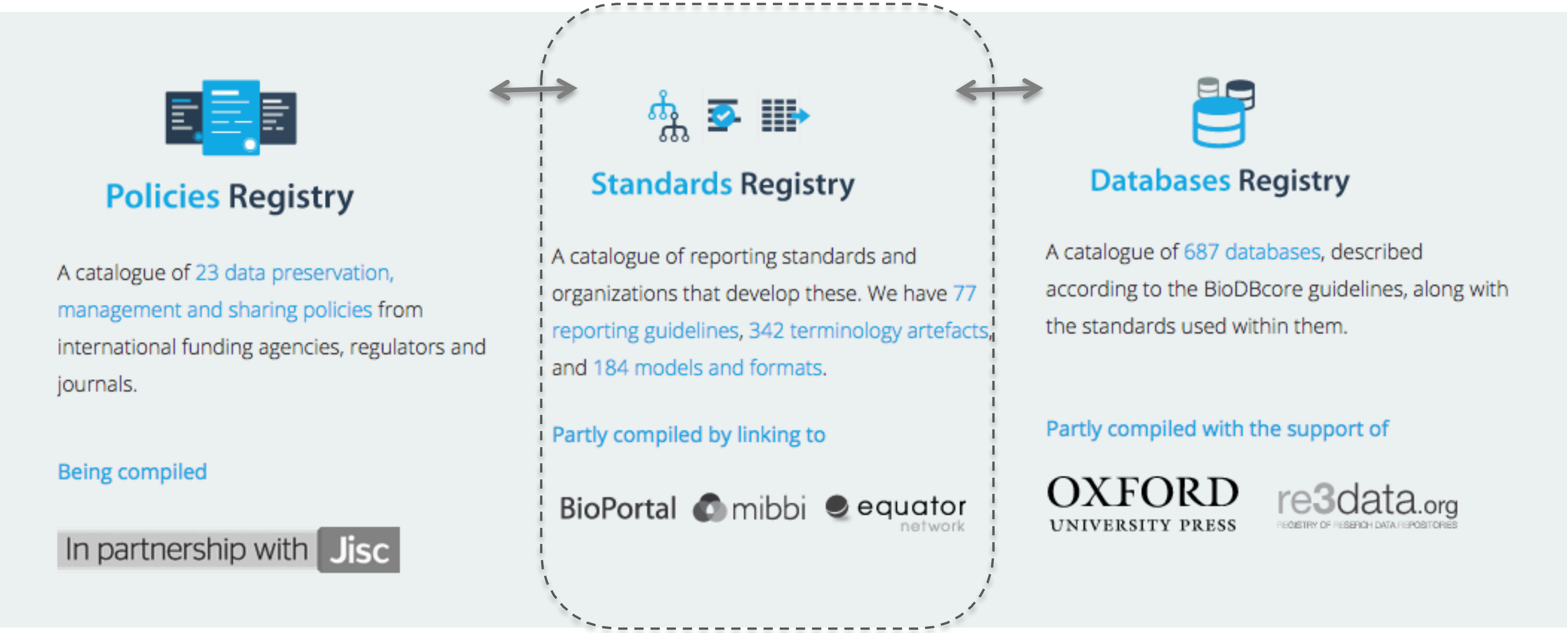
AAO
CHEBI
OBI
VO
PATO
ENVO
MOD
TEDDY
XAO
BTO
DO
PRO
IDO...

miame
MIAPA
MIRIAM
MIQAS
MIX
REMARK
MIGEN
MIQE
MIAPE
ARRIVE
CONSORT
MIASE
MISFISHIE....

databases & training material

biosharing.org

Information Resources



Policies Registry

A catalogue of 23 data preservation, management and sharing policies from international funding agencies, regulators and journals.

Being compiled

In partnership with Jisc

Standards Registry

A catalogue of reporting standards and organizations that develop these. We have 77 reporting guidelines, 342 terminology artefacts, and 184 models and formats.

Partly compiled by linking to

BioPortal mibbi equator network

Databases Registry

A catalogue of 687 databases, described according to the BioDBcore guidelines, along with the standards used within them.

Partly compiled with the support of

OXFORD UNIVERSITY PRESS re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

A **web-based**, **curated** and **searchable registry** ensuring that biological standards and databases are *registered*, *informative* and *discoverable*; also monitoring the **development** and **evolution** of standards, their **usage** in databases and the **adoption** of both in data policies.

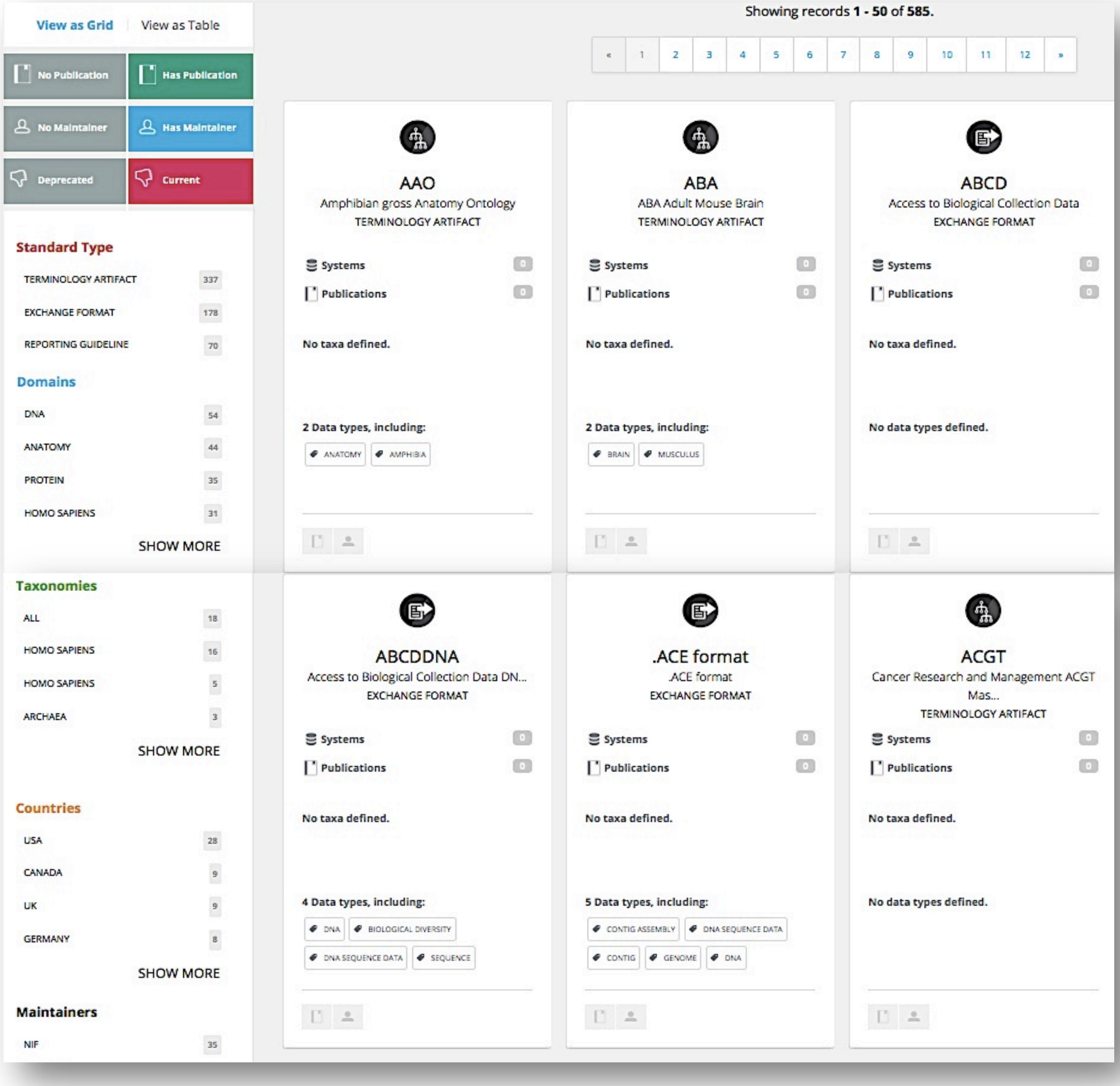
Search filter, refine

Core functionalities:

- search and filtering, e.g. by funder, domain, type of standard
- Refine by publication, maintainer etc.
- add new records, edit existing records
- “claim” records
- person’s profile (as maintainer of records) associated to the ORCID profile (for credit)
- visualization and views of content and linking

Annotation Sources:

- 4 axes: (material, process, quality, information)
- NIF,OBI,CL,GO,IAO,EDAM



View as Grid | View as Table

No Publication | Has Publication

No Maintainer | Has Maintainer

Deprecated | Current

Standard Type

TERMINOLOGY ARTIFACT	337
EXCHANGE FORMAT	178
REPORTING GUIDELINE	70

Domains

DNA	54
ANATOMY	44
PROTEIN	35
HOMO SAPIENS	31

SHOW MORE

Taxonomies

ALL	18
HOMO SAPIENS	16
HOMO SAPIENS	5
ARCHAEA	3

SHOW MORE

Countries

USA	28
CANADA	9
UK	9
GERMANY	8

SHOW MORE

Maintainers

NIF	35

Showing records **1 - 50** of **585**.

« 1 2 3 4 5 6 7 8 9 10 11 12 »

AAO
Amphibian gross Anatomy Ontology
TERMINOLOGY ARTIFACT
Systems 0
Publications 0
No taxa defined.
2 Data types, including:
ANATOMY | AMPHIBIA

ABA
ABA Adult Mouse Brain
TERMINOLOGY ARTIFACT
Systems 0
Publications 0
No taxa defined.
2 Data types, including:
BRAIN | MUSCULUS

ABCD
Access to Biological Collection Data
EXCHANGE FORMAT
Systems 0
Publications 0
No taxa defined.
No data types defined.

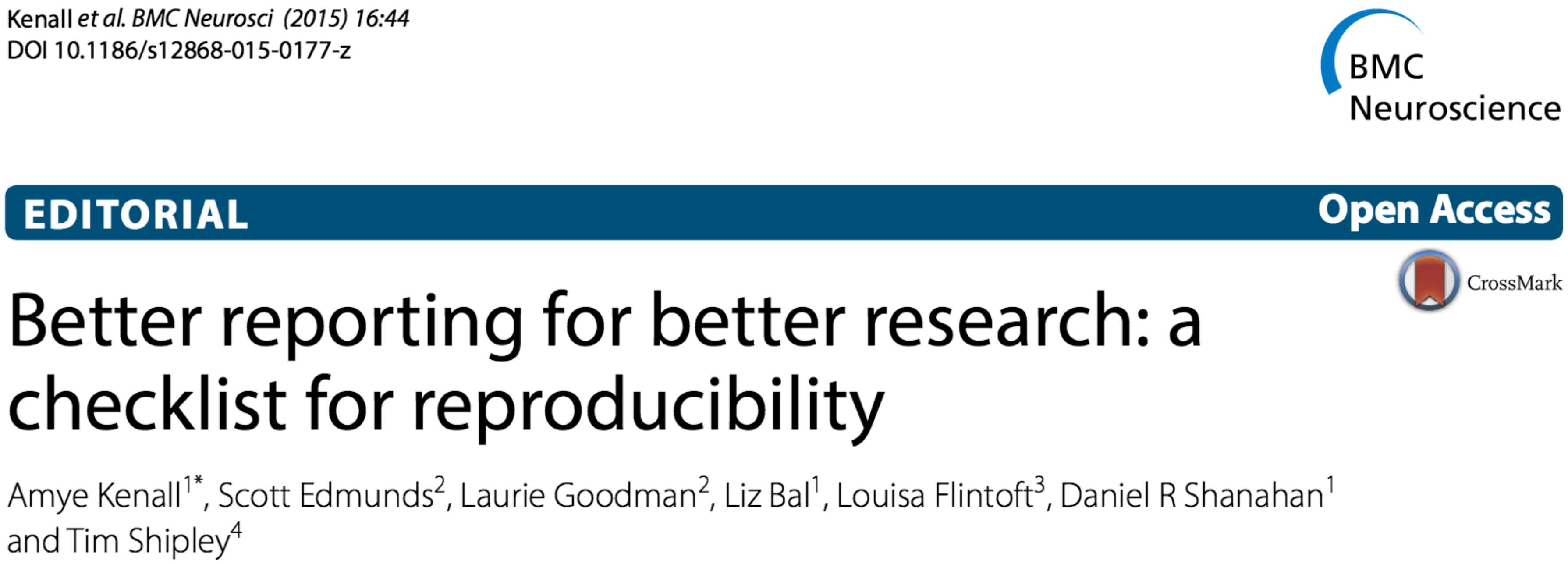
ABCDDNA
Access to Biological Collection Data DN...
EXCHANGE FORMAT
Systems 0
Publications 0
No taxa defined.
4 Data types, including:
DNA | BIOLOGICAL DIVERSITY | DNA SEQUENCE DATA | SEQUENCE

.ACE format
.ACE format
EXCHANGE FORMAT
Systems 0
Publications 0
No taxa defined.
5 Data types, including:
CONTIG ASSEMBLY | DNA SEQUENCE DATA | CONTIG | GENOME | DNA

ACGT
Cancer Research and Management ACGT Mas...
TERMINOLOGY ARTIFACT
Systems 0
Publications 0
No taxa defined.
No data types defined.

biosharing.org
Information Resources

Collections



Kenall *et al. BMC Neurosci* (2015) 16:44
DOI 10.1186/s12868-015-0177-z

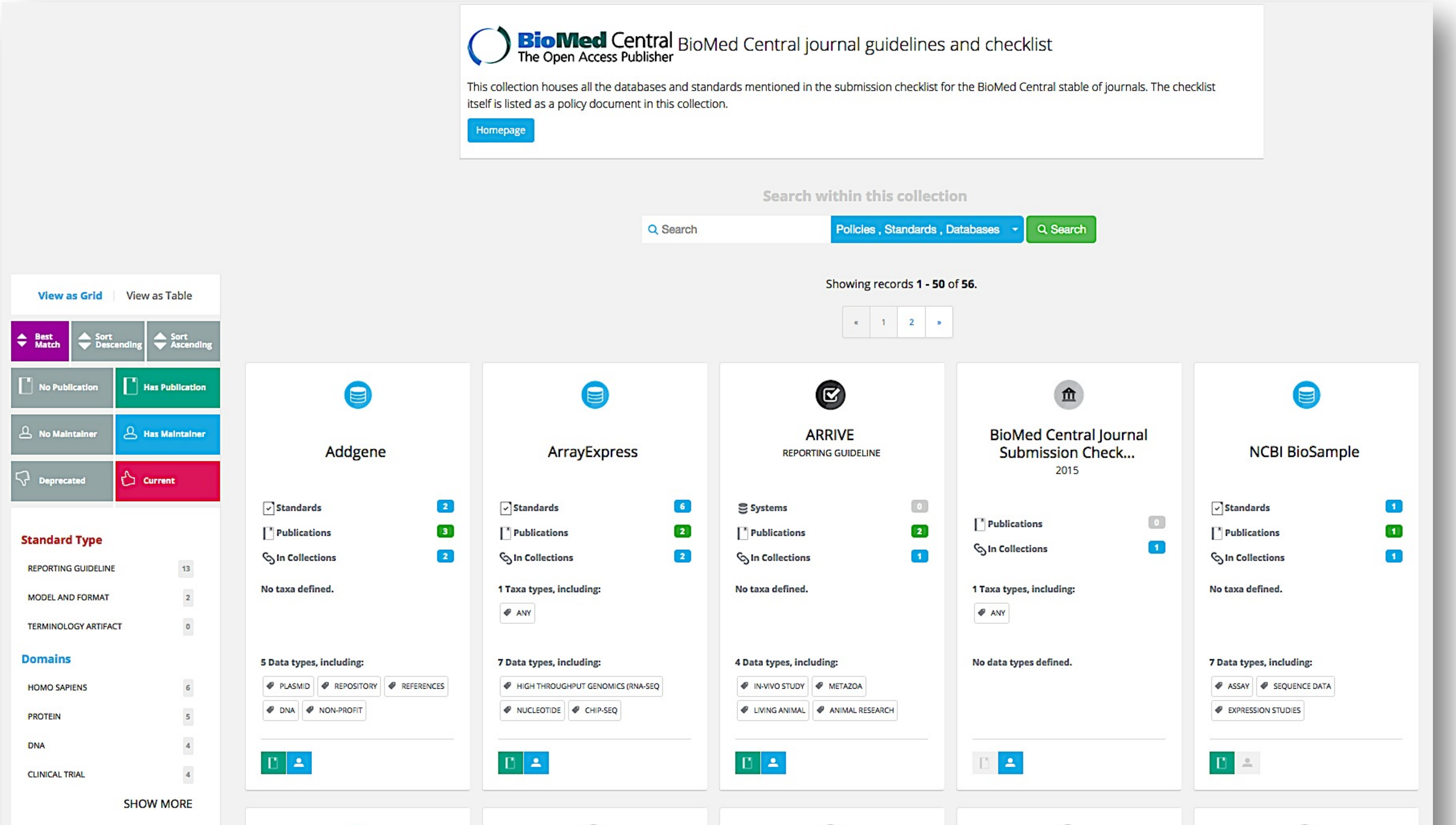
BMC Neuroscience

EDITORIAL — Open Access

CrossMark

Better reporting for better research: a checklist for reproducibility

Amye Kenall[1*], Scott Edmunds[2], Laurie Goodman[2], Liz Bal[1], Louisa Flintoft[3], Daniel R Shanahan[1] and Tim Shipley[4]



BioMed Central The Open Access Publisher BioMed Central journal guidelines and checklist

This collection houses all the databases and standards mentioned in the submission checklist for the BioMed Central stable of journals. The checklist itself is listed as a policy document in this collection.

Homepage

Search within this collection

Search | Policies , Standards , Databases | Search

Showing records **1 - 50** of **56.**

« 1 2 »

View as Grid | View as Table

Best Match | Sort Descending | Sort Ascending
No Publication | Has Publication
No Maintainer | Has Maintainer
Deprecated | Current

Standard Type

REPORTING GUIDELINE	13
MODEL AND FORMAT	2
TERMINOLOGY ARTIFACT	0

Domains

HOMO SAPIENS	6
PROTEIN	5
DNA	4
CLINICAL TRIAL	4

SHOW MORE

Addgene

Standards	2
Publications	3
In Collections	2

No taxa defined.

5 Data types, including: PLASMID, REPOSITORY, REFERENCES, DNA, NON-PROFIT

ArrayExpress

Standards	6
Publications	2
In Collections	2

1 Taxa types, including: ANY

7 Data types, including: HIGH THROUGHPUT GENOMICS (RNA-SEQ, NUCLEOTIDE, CHIP-SEQ

ARRIVE
REPORTING GUIDELINE

Systems	0
Publications	2
In Collections	1

No taxa defined.

4 Data types, including: IN-VIVO STUDY, METAZOA, LIVING ANIMAL, ANIMAL RESEARCH

BioMed Central Journal Submission Check...
2015

Publications	0
In Collections	1

1 Taxa types, including: ANY

No data types defined.

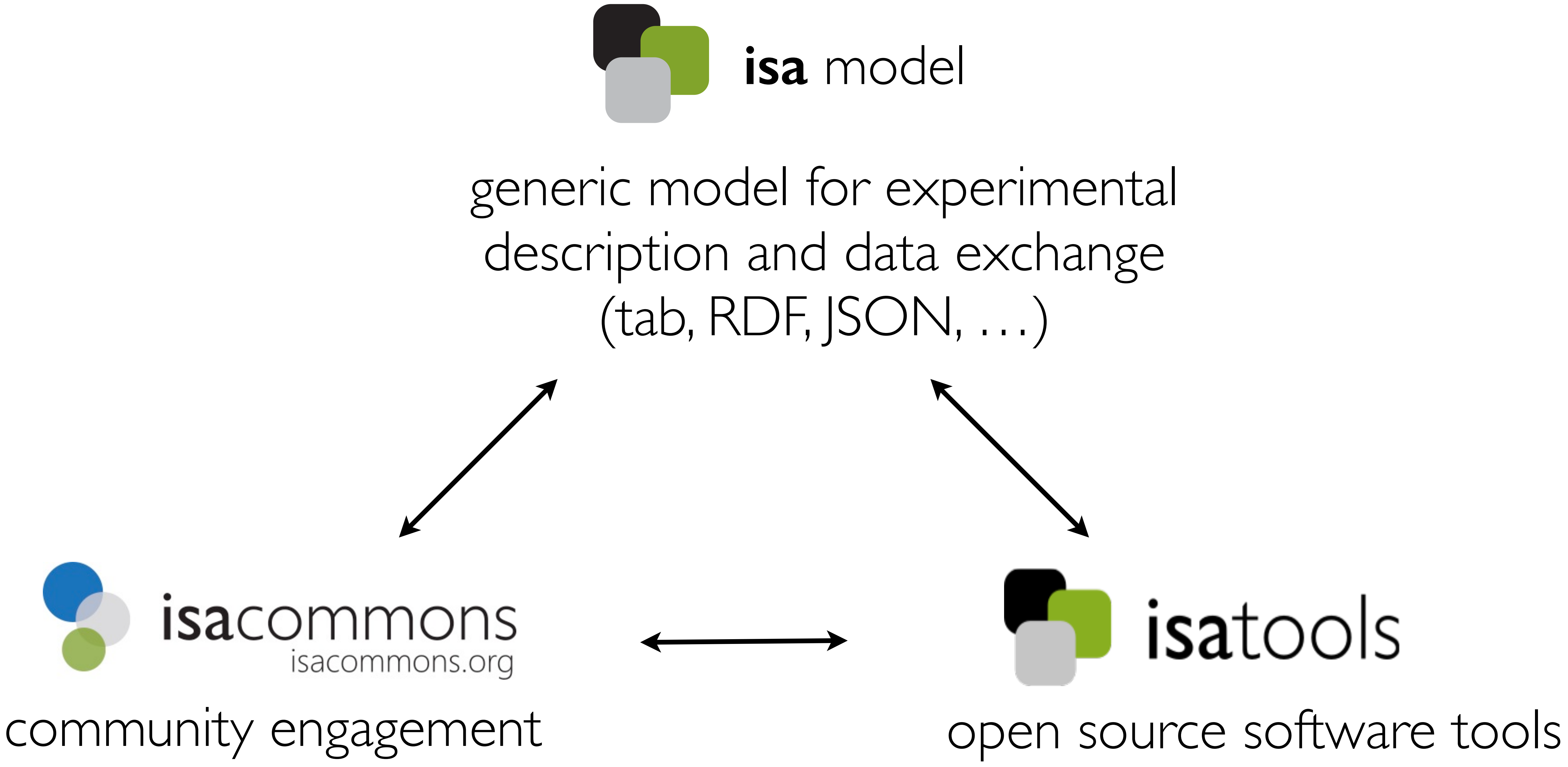
NCBI BioSample

Standards	1
Publications	1
In Collections	1

No taxa defined.

7 Data types, including: ASSAY, SEQUENCE DATA, EXPRESSION STUDIES

The isa infrastructure



isa model

generic model for experimental description and data exchange (tab, RDF, JSON, …)

isacommons
isacommons.org

community engagement

isatools

open source software tools

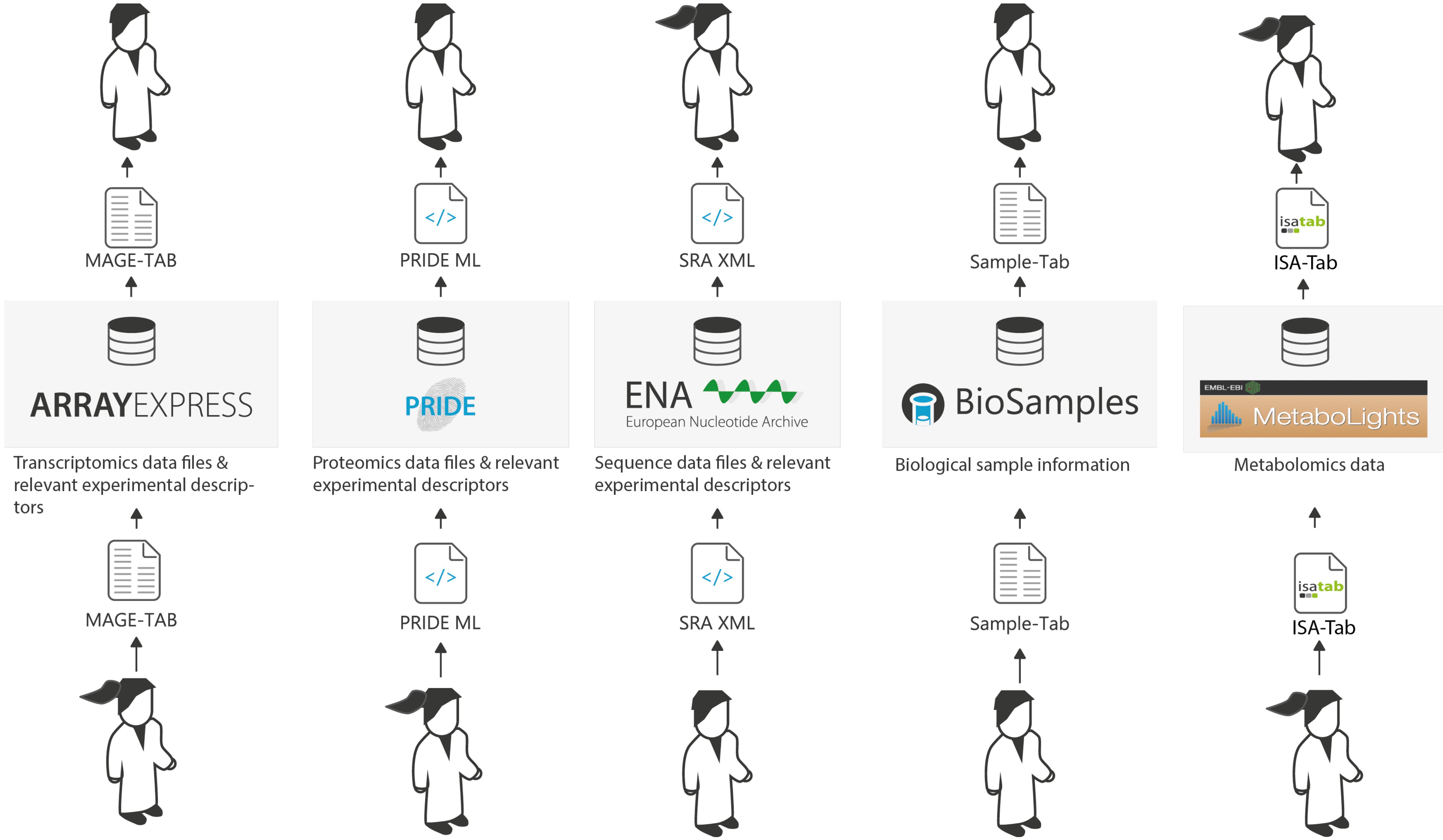
Towards interoperable bioscience data: Presenting the ISA Commons, authored by more than 50 collaborators at over 30 scientific organizations around the globe.

Sansone et al, 2012
Nature Genetics

ISA software suite:
Overview of ISA-Tab and first set of tools

Rocca-Serra et al, 2010
Bioinformatics

Formats & Database Fragmentation



MAGE-TAB

ARRAYEXPRESS

Transcriptomics data files & relevant experimental descriptors

MAGE-TAB

PRIDE ML

PRIDE

Proteomics data files & relevant experimental descriptors

PRIDE ML

SRA XML

ENA European Nucleotide Archive

Sequence data files & relevant experimental descriptors

SRA XML

Sample-Tab

BioSamples

Biological sample information

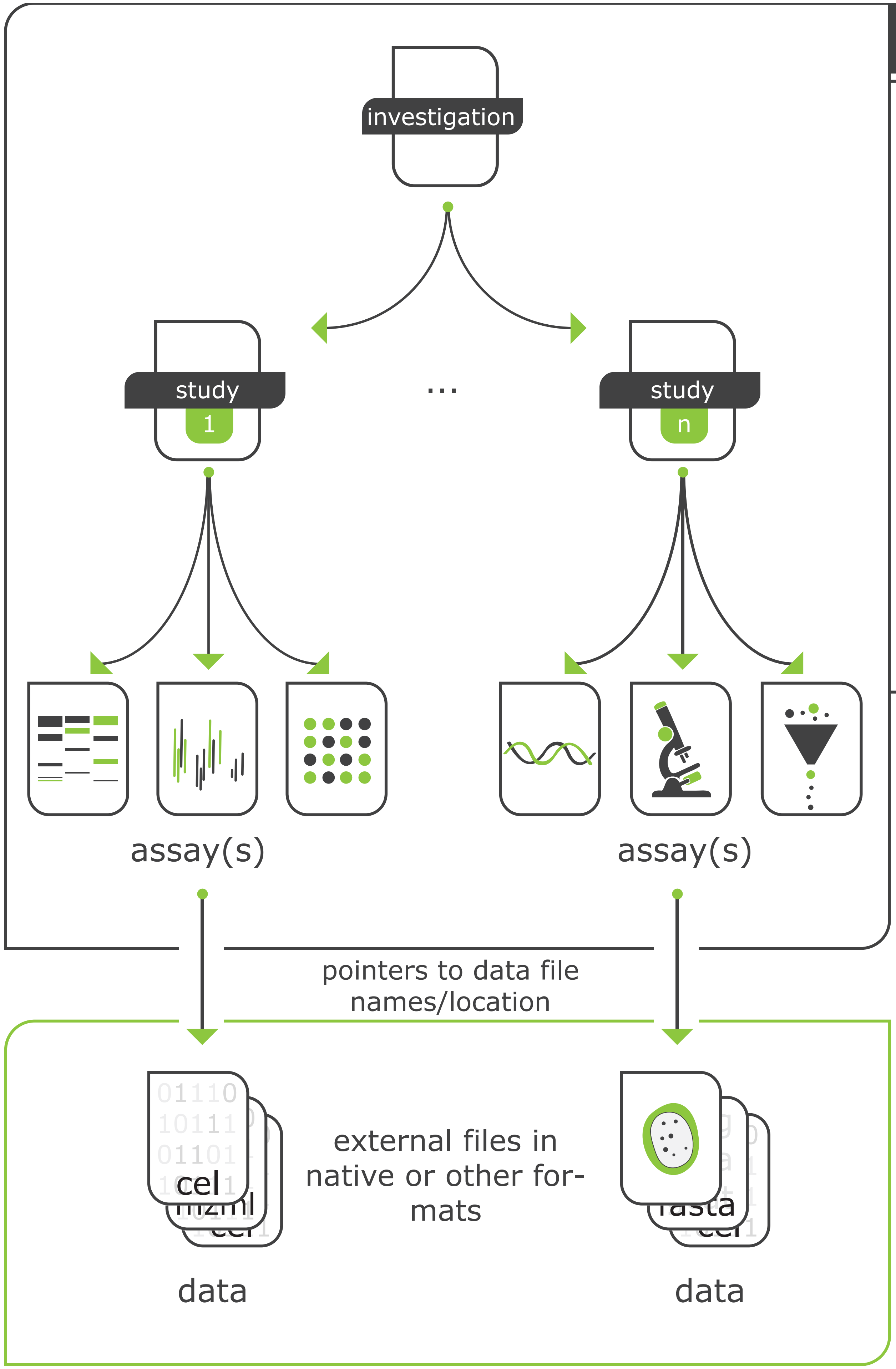
Sample-Tab

ISA-Tab

EMBL-EBI MetaboLights

Metabolomics data

ISA-Tab



isatab

investigation

high level concept to link related studies

study

the central unit, containing information on the subject under study, its characteristics and any treatments applied.

a study has associated ***assays***

assay

test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)

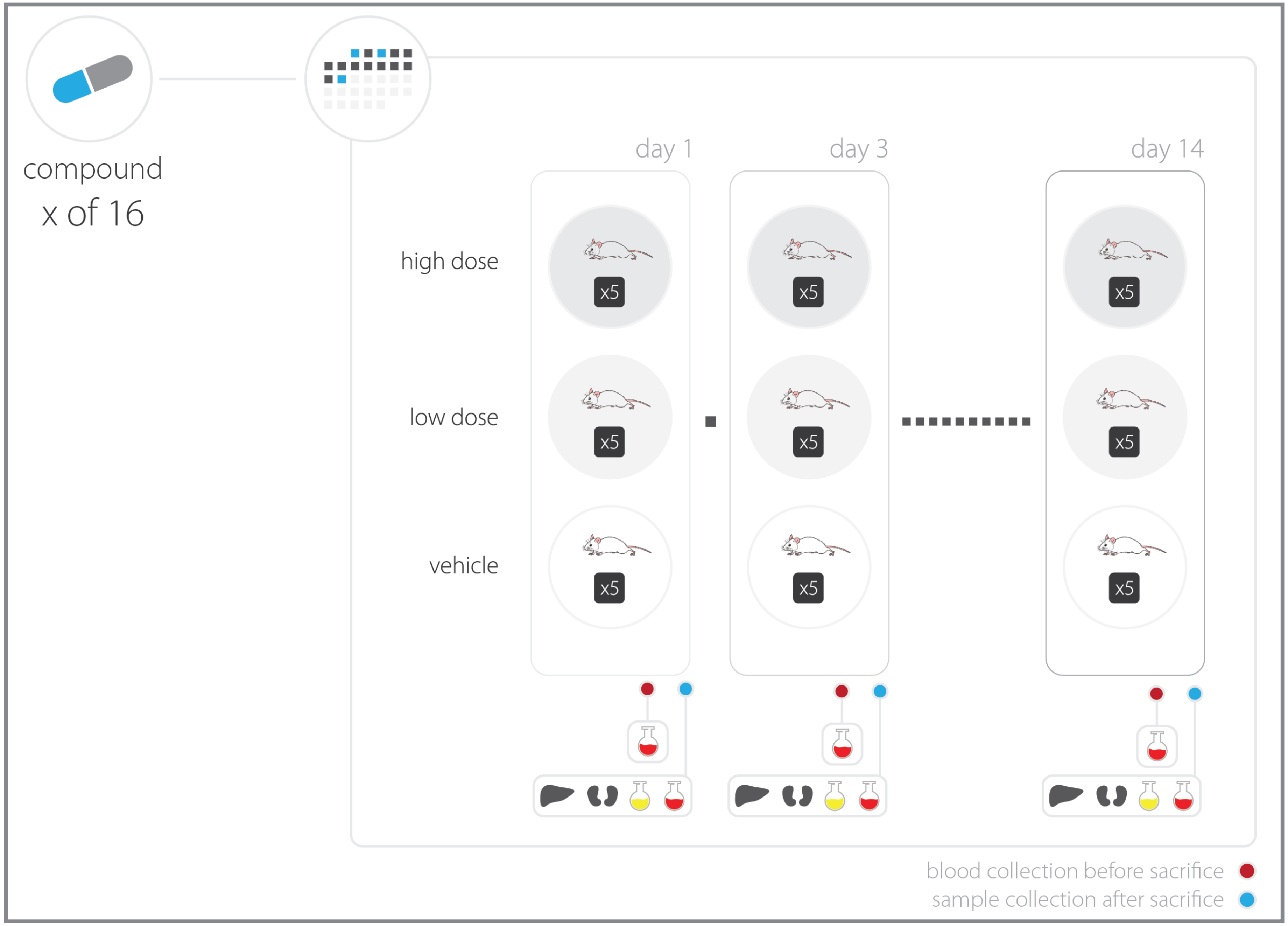
isatab

isacommons

isacommons.org

- environmental health
- environmental genomics
- metabolomics
- metagenomics
- nanotechnology
- proteomics
- stem cell discovery
- system biology
- transcriptomics
- toxicogenomics
- communities working to build a library of cellular signatures

The experimental plan


compound
x of 16
day 1
day 3
day 14
high dose
low dose
vehicle
x5
blood collection before sacrifice ●
sample collection after sacrifice ●

experimental design

sample characteristic(s)

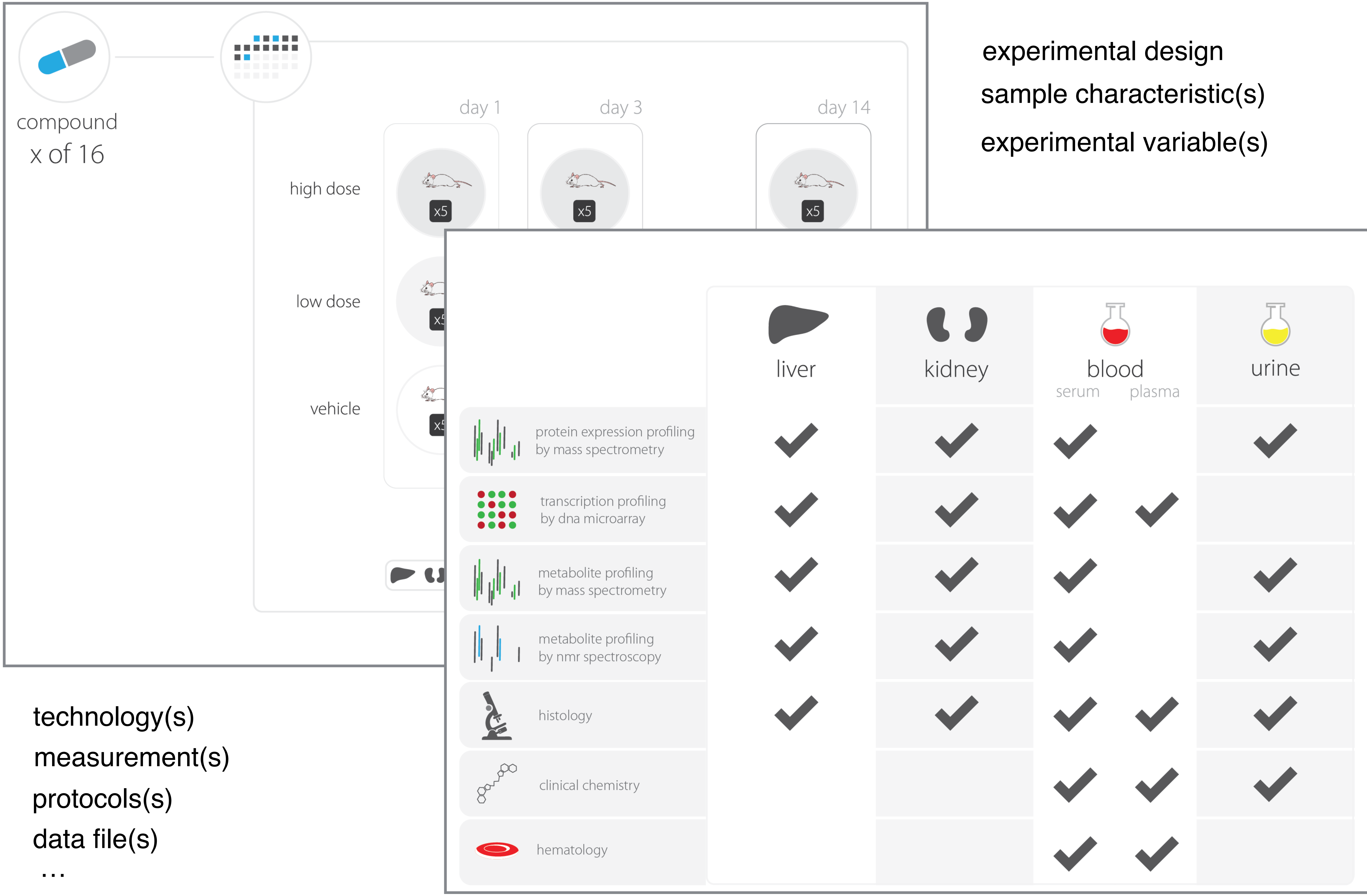
experimental variable(s)

InnoMed PredTox Project

2-week systemic rat study using male Wistar rats (N=15 per dose group)

14 proprietary drug candidates from participating companies and
2 reference toxic compounds

The experimental plan



	liver	kidney	blood		urine
			serum	plasma	
protein expression profiling by mass spectrometry	✓	✓	✓		✓
transcription profiling by dna microarray	✓	✓	✓	✓	
metabolite profiling by mass spectrometry	✓	✓	✓		✓
metabolite profiling by nmr spectroscopy	✓	✓	✓		✓
histology	✓	✓	✓	✓	✓
clinical chemistry			✓	✓	✓
hematology			✓	✓	

experimental design

sample characteristic(s)

experimental variable(s)

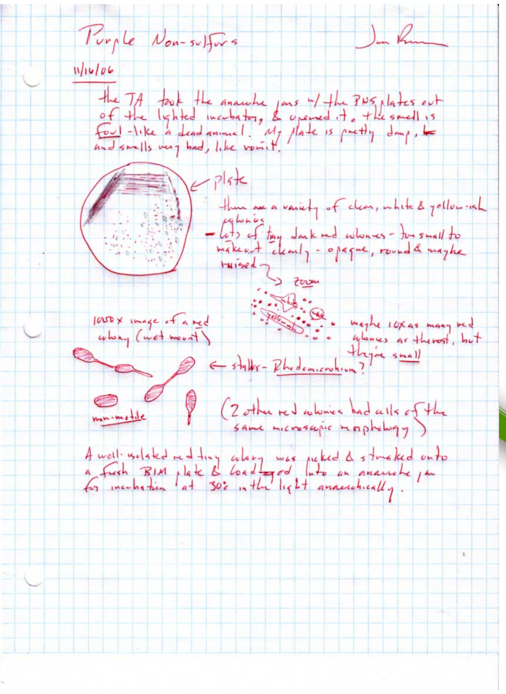
technology(s)

measurement(s)

protocols(s)

data file(s)

…



Purple Non-sulfurs

Jim R

11/16/06

the TA took the anaerobe jars w/ the PNS plates out of the lighted incubator, & opened it. the smell is foul - like a dead animal. My plate is pretty damp, and smells very bad, like vomit.

← plate

there are a variety of clear, white & yellow-ish colonies

- lots of tiny dark red colonies - too small to make out clearly - opaque, round & maybe raised

zoom

white

yellow-ish

white

maybe 10X as many red colonies as the rest, but they're small

1050 x image of a red colony (wet mount)

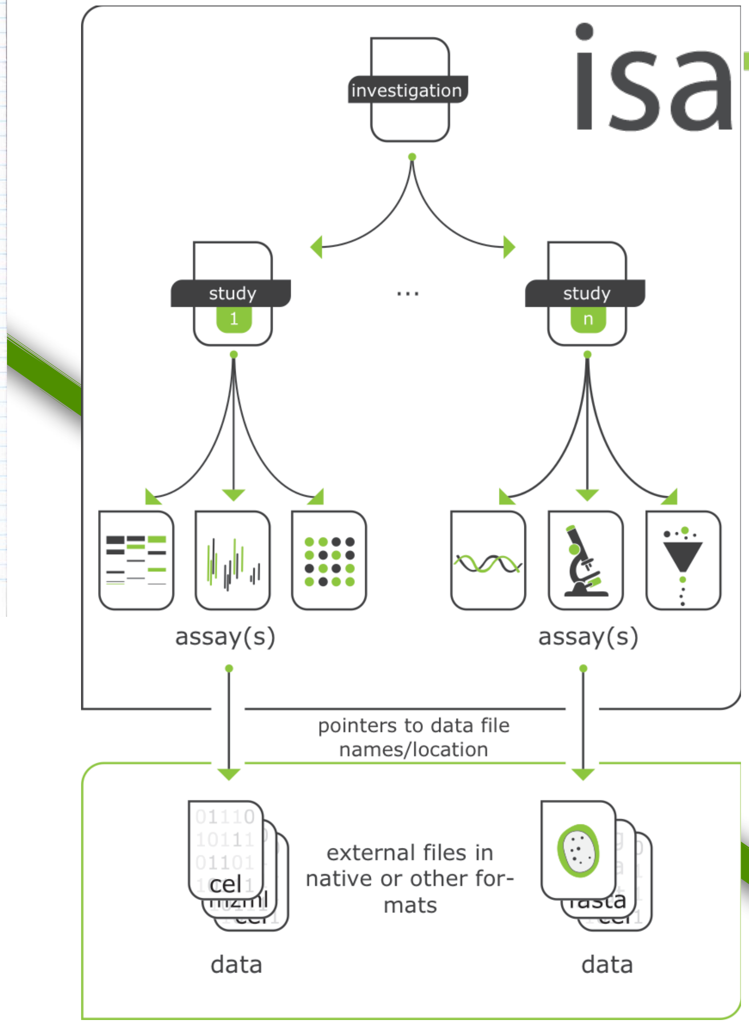
← stalks - Rhodomicrobium?

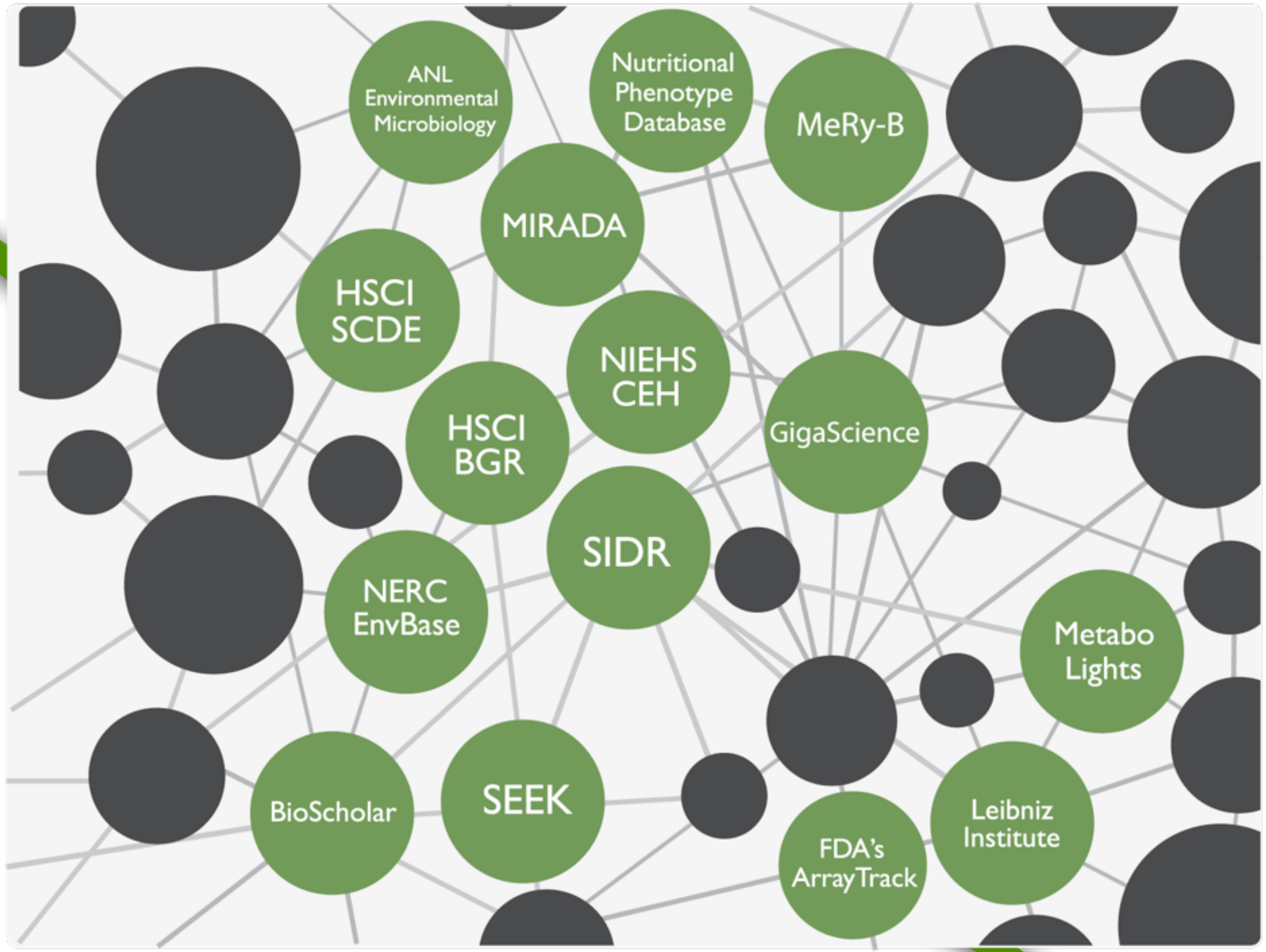
non-motile

(2 other red colonies had cells of the same microscopic morphology)

A well-isolated red tiny colony was picked & streaked onto a fresh BIM plate & loaded into an anaerobe jar for incubation at 30° in the light anaerobically.

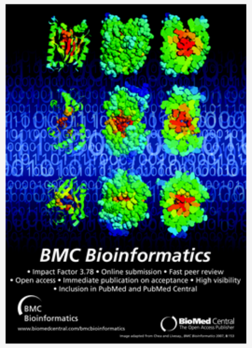
isatab


investigation
study 1
...
study n
assay(s)
assay(s)
pointers to data file names/location
cel
mzml
external files in native or other formats
fasta
data
data



linkedISA: semantic representation of ISA-Tab experimental metadata

Alejandra González-Beltrán*, Eamonn Maguire, Susanna-Assunta Sansone and Philippe Rocca-Serra

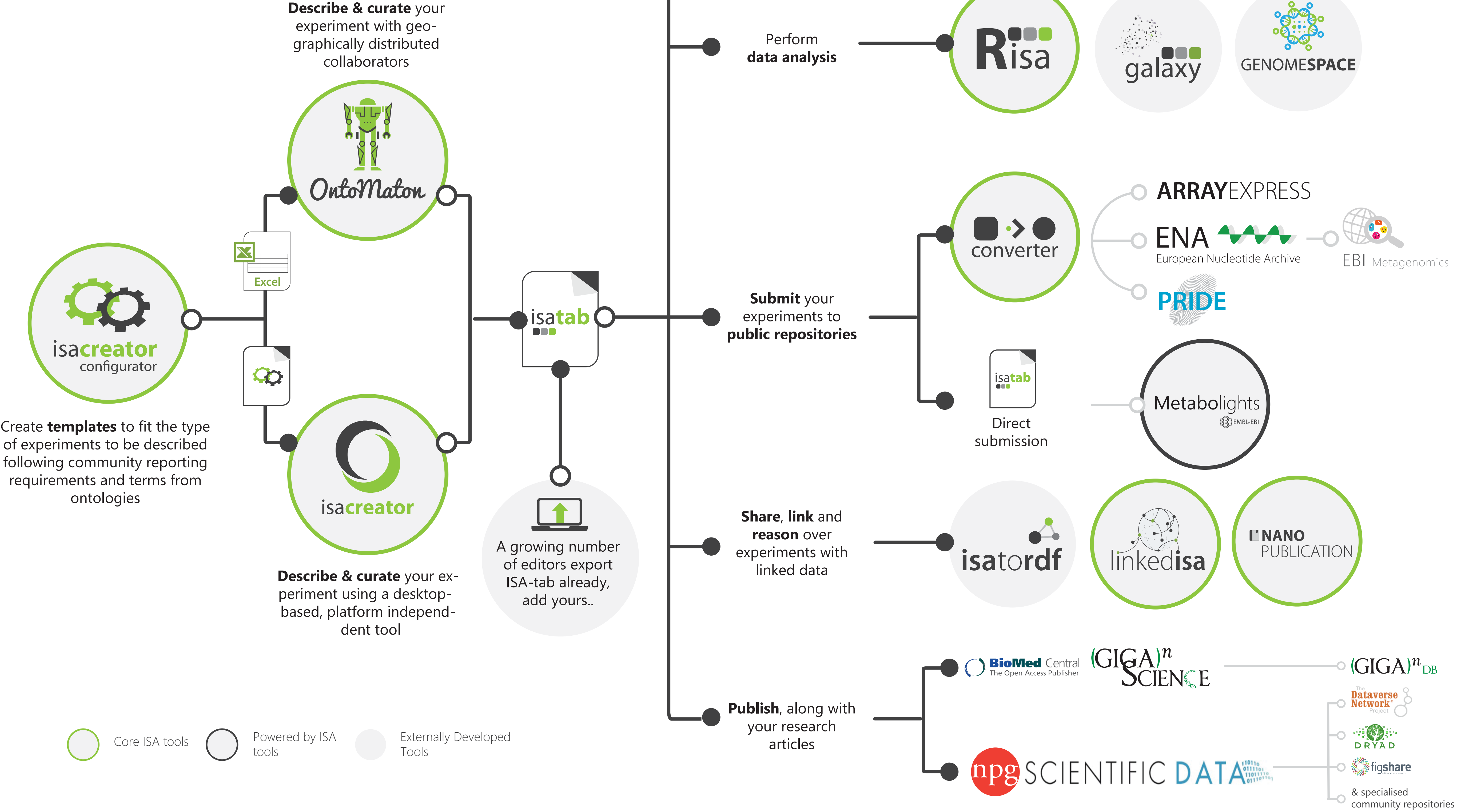


BMC Bioinformatics, Issue 15(Suppl 14):S4 , 2014. 10.1186/1471-2105-15-S14-S4.

isa-tools.org

github SOCIAL CODING

GITHUB.COM/ISA-TOOLS



Describe & curate your experiment with geographically distributed collaborators

Create **templates** to fit the type of experiments to be described following community reporting requirements and terms from ontologies

Describe & curate your experiment using a desktop-based, platform independent tool

A growing number of editors export ISA-tab already, add yours..

Create your **own repository**

Perform **data analysis**

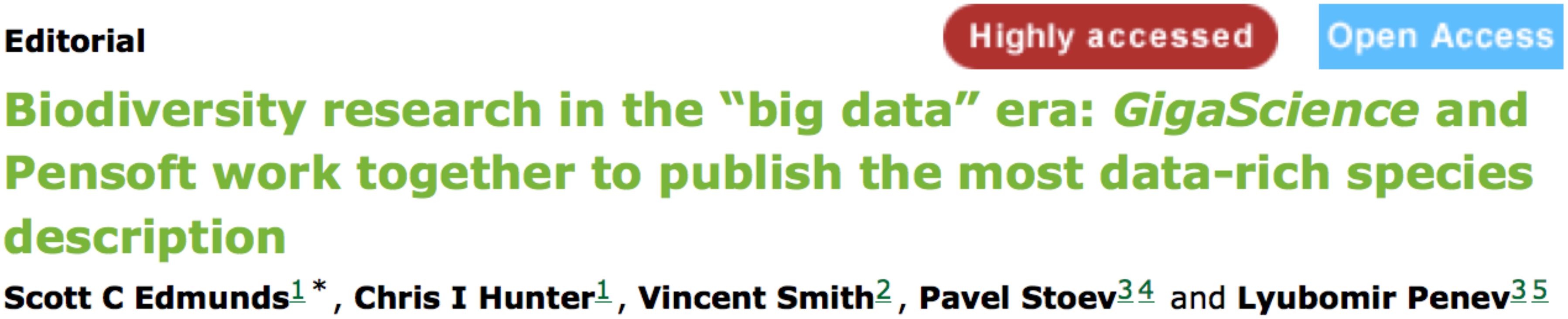
Submit your experiments to **public repositories**

Direct submission

Share, **link** and **reason** over experiments with linked data

Publish, along with your research articles

& specialised community repositories



Editorial

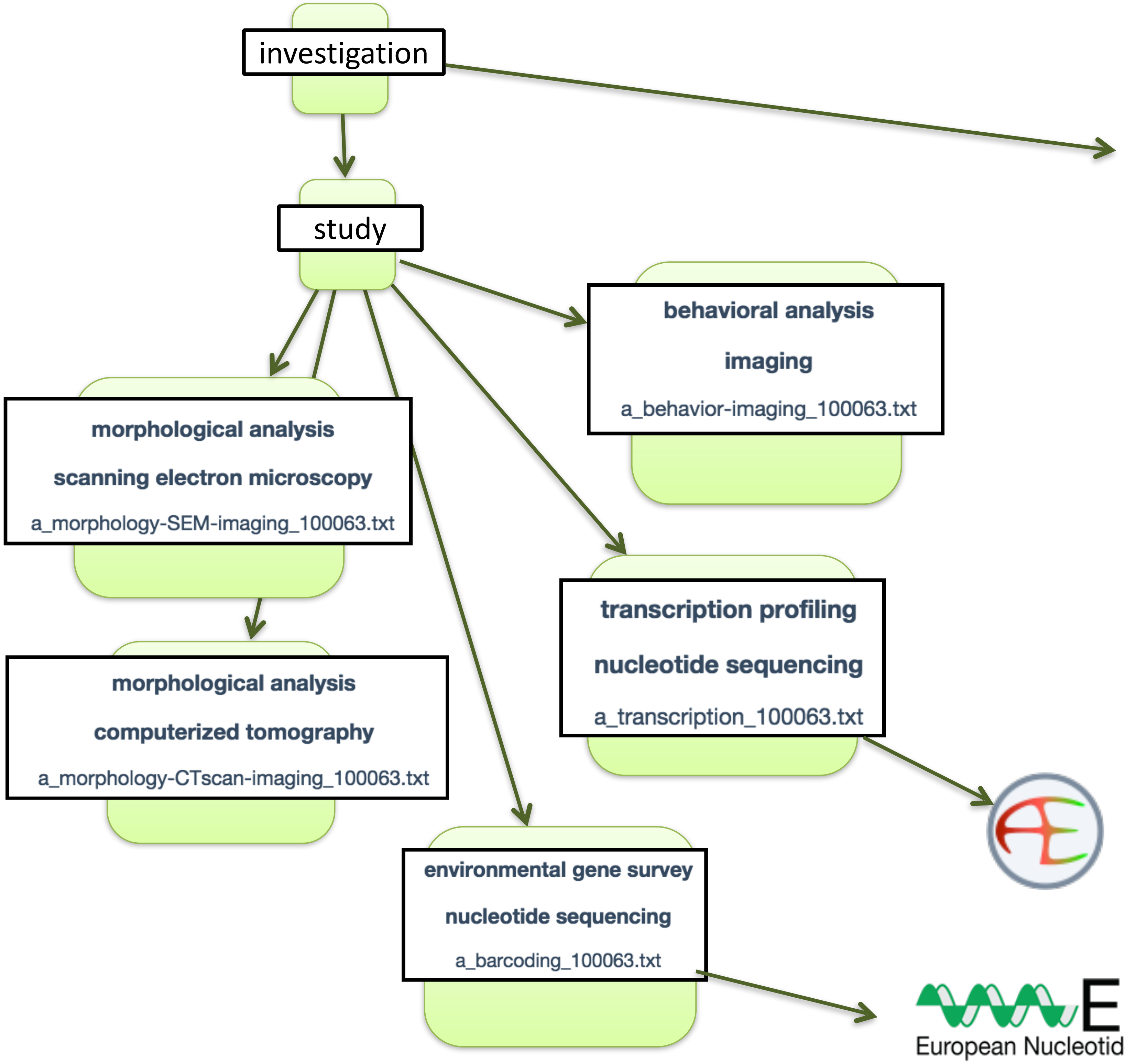
Highly accessed

Open Access

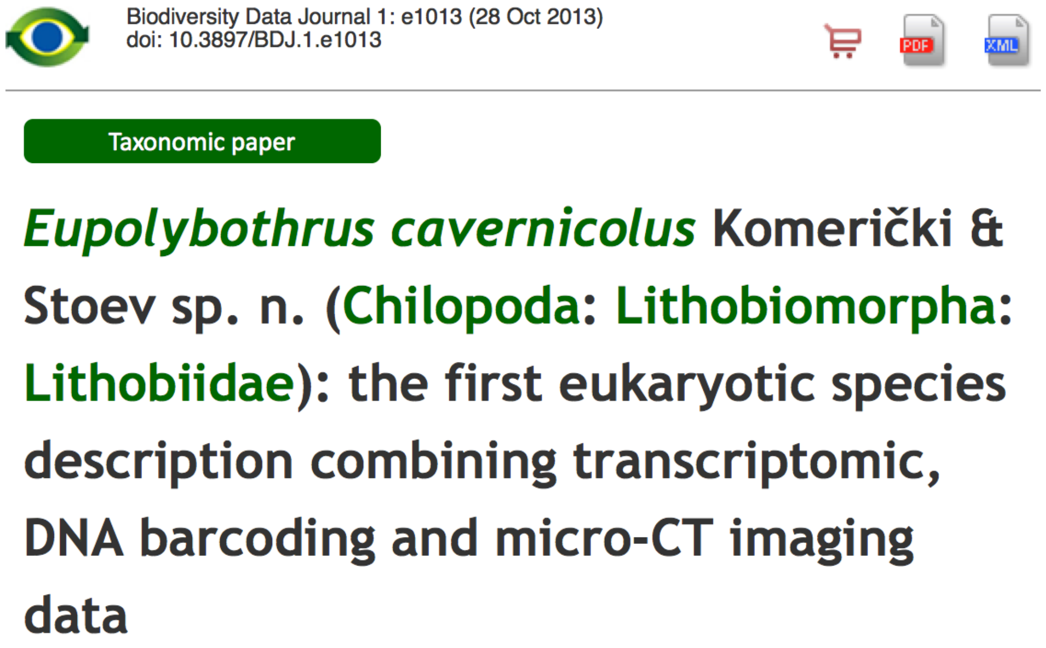
Biodiversity research in the "big data" era: *GigaScience* and Pensoft work together to publish the most data-rich species description

Scott C Edmunds[1]*, Chris I Hunter[1], Vincent Smith[2], Pavel Stoev[3 4] and Lyubomir Penev[3 5]

http://dx.doi.org/10.5524/100063



ENA European Nucleotide Archive

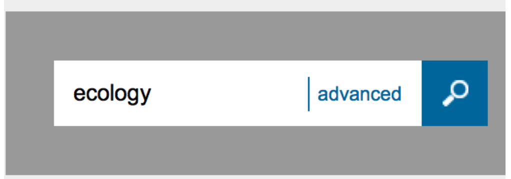


Biodiversity Data Journal 1: e1013 (28 Oct 2013)
doi: 10.3897/BDJ.1.e1013

Taxonomic paper

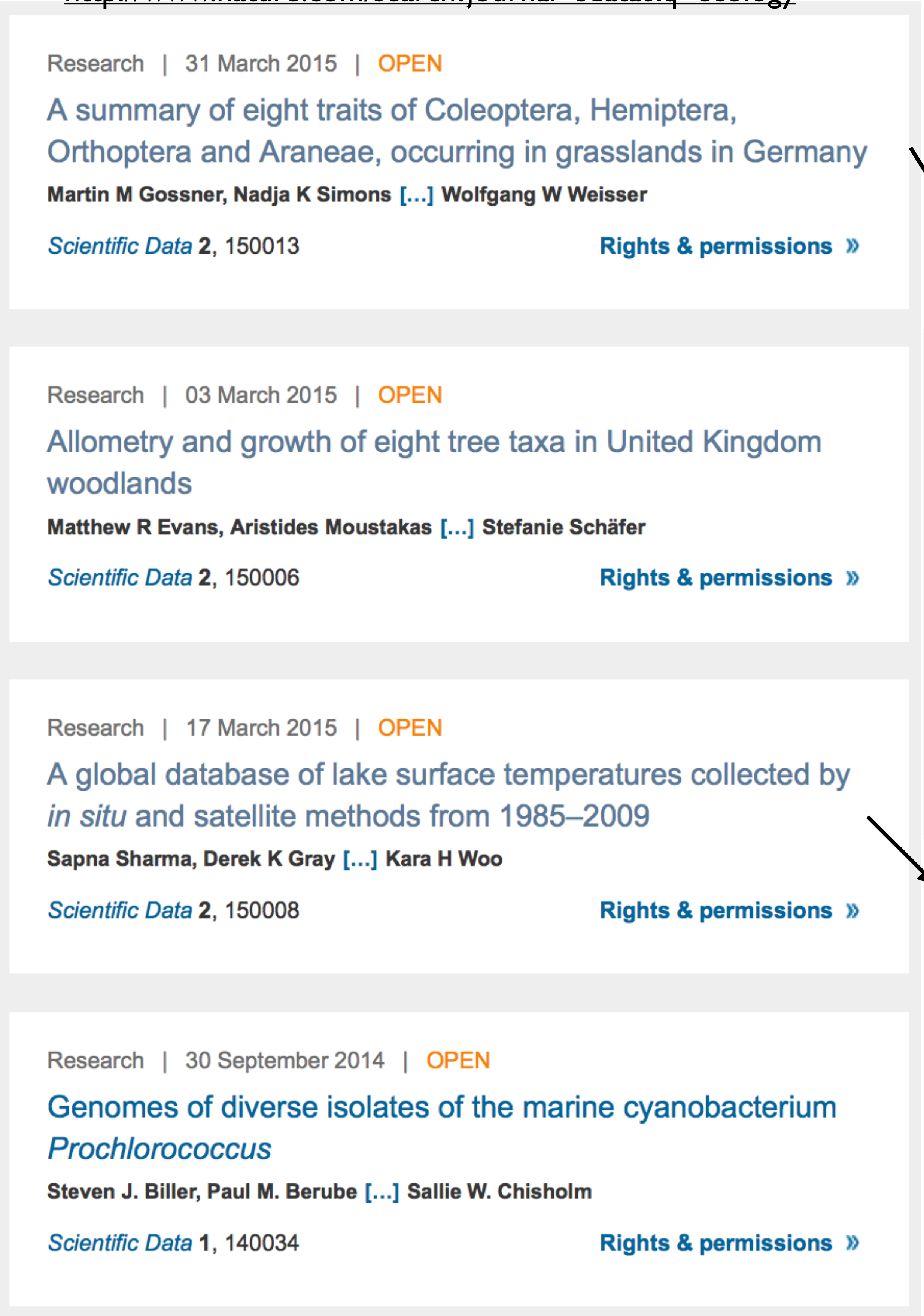
Eupolybothrus cavernicolus Komerički & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae): the first eukaryotic species description combining transcriptomic, DNA barcoding and micro-CT imaging data

SCIENTIFIC DATA



ecology | advanced

http://www.nature.com/search?journal=sdata&q=ecology



Research | 31 March 2015 | OPEN

A summary of eight traits of Coleoptera, Hemiptera, Orthoptera and Araneae, occurring in grasslands in Germany

Martin M Gossner, Nadja K Simons [...] Wolfgang W Weisser

Scientific Data **2**, 150013 **Rights & permissions »**

Research | 03 March 2015 | OPEN

Allometry and growth of eight tree taxa in United Kingdom woodlands

Matthew R Evans, Aristides Moustakas [...] Stefanie Schäfer

Scientific Data **2**, 150006 **Rights & permissions »**

Research | 17 March 2015 | OPEN

A global database of lake surface temperatures collected by *in situ* and satellite methods from 1985–2009

Sapna Sharma, Derek K Gray [...] Kara H Woo

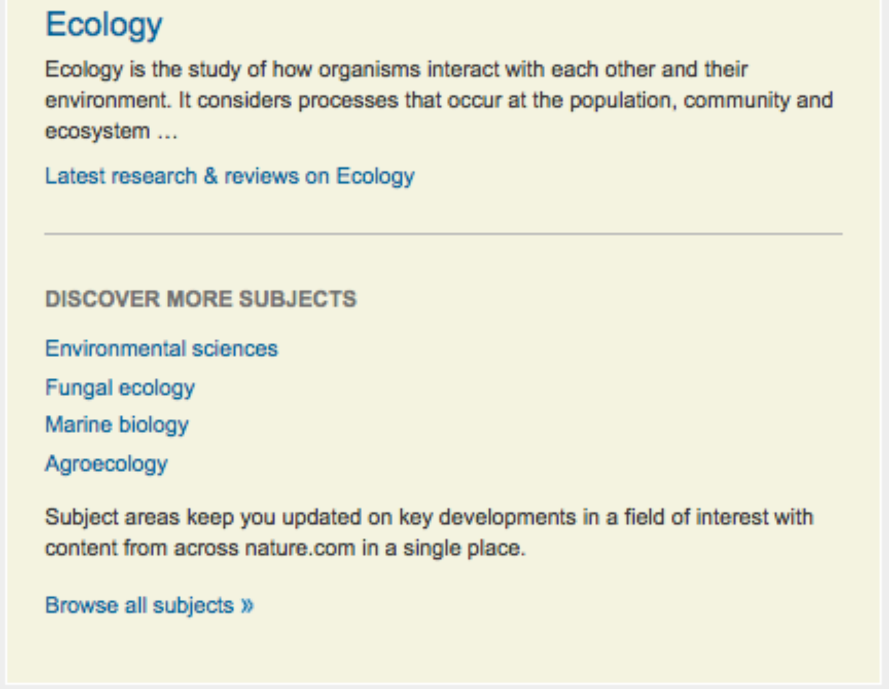
Scientific Data **2**, 150008 **Rights & permissions »**

Research | 30 September 2014 | OPEN

Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*

Steven J. Biller, Paul M. Berube [...] Sallie W. Chisholm

Scientific Data **1**, 140034 **Rights & permissions »**



Ecology

Ecology is the study of how organisms interact with each other and their environment. It considers processes that occur at the population, community and ecosystem ...

Latest research & reviews on Ecology

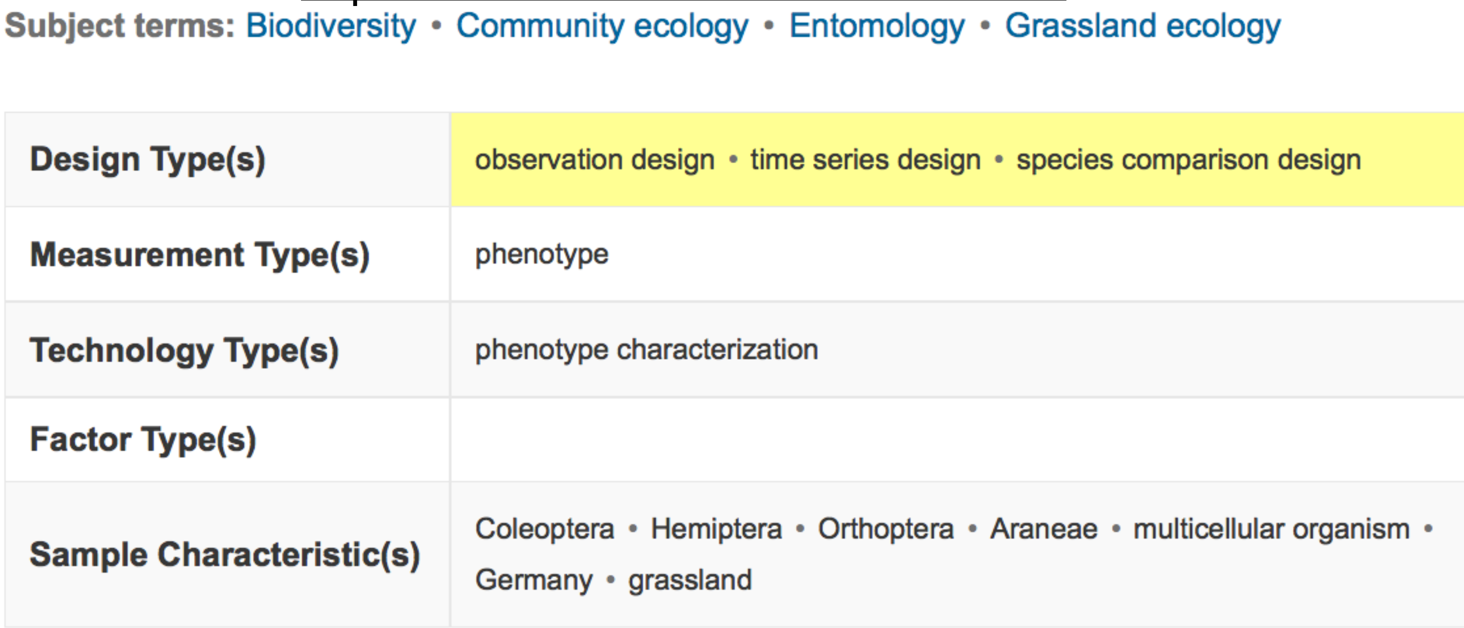
DISCOVER MORE SUBJECTS

Environmental sciences
Fungal ecology
Marine biology
Agroecology

Subject areas keep you updated on key developments in a field of interest with content from across nature.com in a single place.

Browse all subjects »

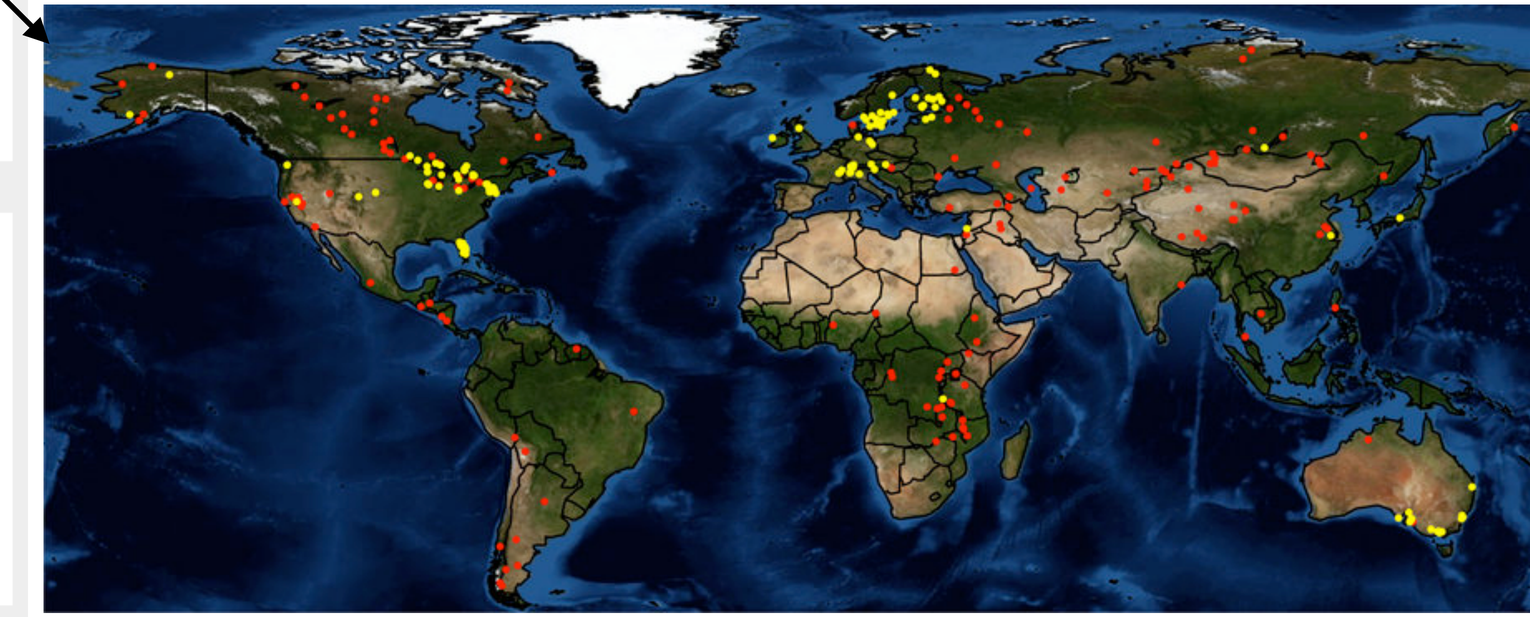
http://www.nature.com/articles/sdata201513



Subject terms: Biodiversity • Community ecology • Entomology • Grassland ecology

Design Type(s)	observation design • time series design • species comparison design
Measurement Type(s)	phenotype
Technology Type(s)	phenotype characterization
Factor Type(s)	
Sample Characteristic(s)	Coleoptera • Hemiptera • Orthoptera • Araneae • multicellular organism • Germany • grassland

http://www.nature.com/articles/sdata20158



1 **Map** the landscape of content standards via biosharing.org

2 Structure

Develop methods for creating templates.

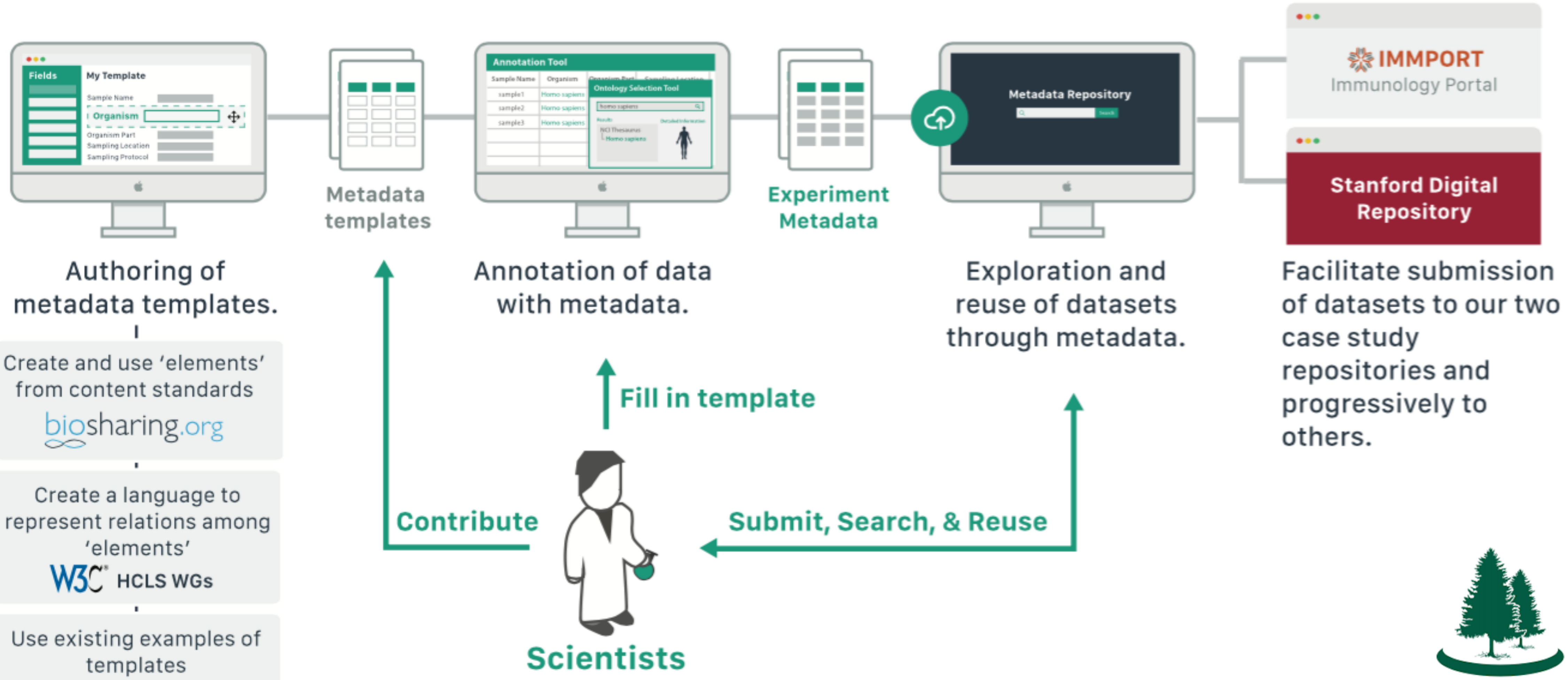
3 Annotate

Develop methods to ease the use of templates.

4 Explore

Create a repository of populated templates.

5 Case Studies



Authoring of metadata templates.

Annotation of data with metadata.

Exploration and reuse of datasets through metadata.

Facilitate submission of datasets to our two case study repositories and progressively to others.

Create and use 'elements' from content standards

biosharing.org

Create a language to represent relations among 'elements'

W3C® HCLS WGs

Use existing examples of templates

isatools IMMPORT HIPC

CEDAR

CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL

6

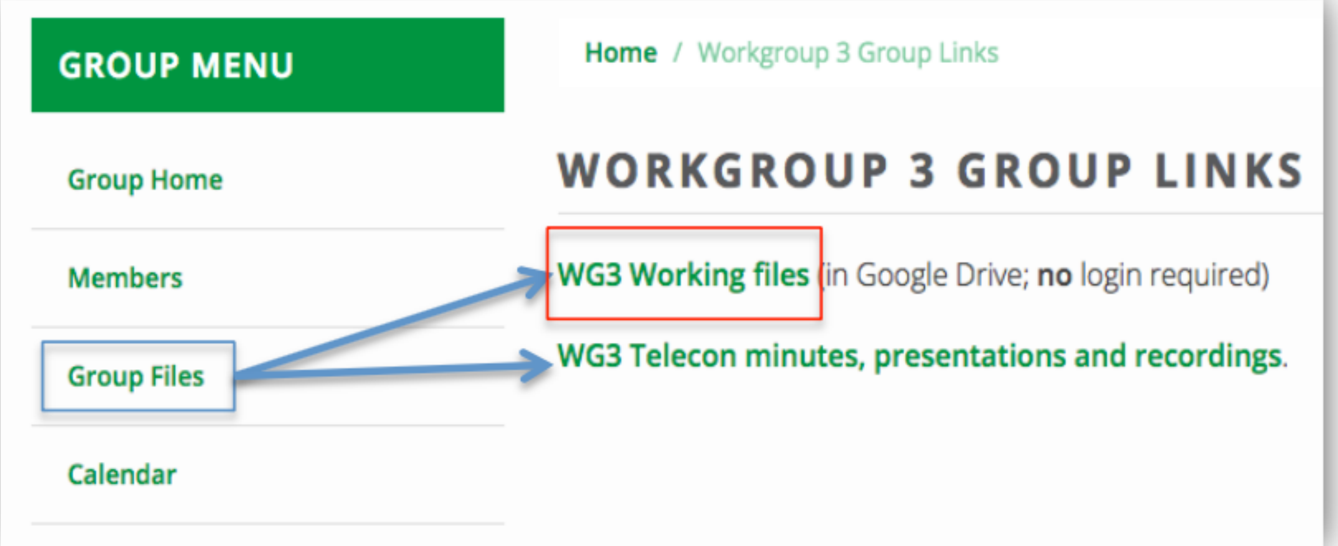
Analyze the CEDAR repository of populated templates to reveal patterns in the metadata that will enable the metadata tools to use predictive data entry to ease the task of filling out the templates.

Augment those metadata with links to the published literature (including secondary analyses and retractions).

Augment those metadata with links to follow-up experiments (in online databases and in the literature).

Allow the scientific community to comment on the experiment through structured metadata.

WG3 Metadata – work to date



Home / Workgroup 3 Group Links

WORKGROUP 3 GROUP LINKS

WG3 Working files (in Google Drive; **no** login required)

WG3 Telecon minutes, presentations and recordings.

① Standard Operating Procedure (SOP)

② Selected Competency Questions

③ Metadata Mapping

④ Core Metadata Elements

Done by:

Alejandra Gonzalez-Beltran - Oxford e-Research Centre, University of Oxford
Philippe Rocca-Serra - Oxford e-Research Centre, University of Oxford
Mary Vardigan - ICPSR, University of Michigan
Susanna-Assunta Sansone - University of Oxford

with contributions, comments from several WG 3 members and colleagues, in particular: Joan Starr, George Alter, Ian Fore, Kevin Read, Stian Soyland-Reyes, Muhammad Amith, Michel Dumontier…

Supported by the NIH grant 1U24 AI117966-01 to the University of California, San Diego

WG3 Metadata – SOP

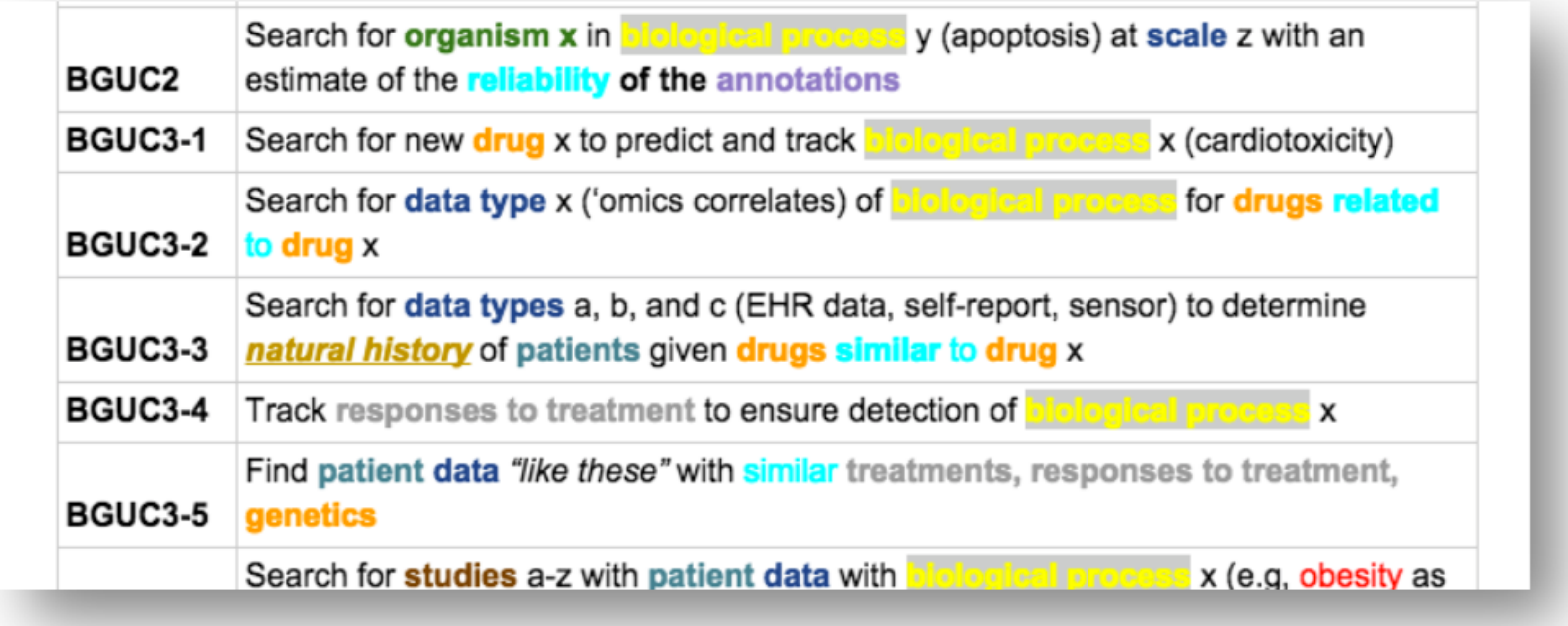
① Standard Operating Procedure (SOP) document:

- contains lists of material reviewed
  - data discovery initiatives and metadata initiatives
  - existing meta-models for representing metadata elements
- outlines the approach used to identify metadata descriptors
  - Via ***use cases*** and ***competency questions*** (top-down approach)
  - ***Mapping*** generic and life science-specific ***metadata schemas*** (bottom-up approach)
    - Listed in the BioSharing collection for bioCADDIE
- The results of both approaches has been compared and converged on the core set of metadata

WG3 Metadata – Use Cases

② Selected Competency Questions:

- representative set from use cases workshop, white paper, submitted by the community and from Phil Bourne
- questions have been abstracted and key metadata elements have been highlighted and color-coded and categorized
- as the set of core and extended metadata elements are defined, it will become clearer which questions the Data Discovery Index will not be able to answers if full and which only in part.



BGUC2	Search for **organism x** in **biological process** y (apoptosis) at **scale** z with an estimate of the **reliability of the annotations**
BGUC3-1	Search for new **drug** x to predict and track **biological process** x (cardiotoxicity)
BGUC3-2	Search for **data type** x ('omics correlates) of **biological process** for **drugs related to drug** x
BGUC3-3	Search for **data types** a, b, and c (EHR data, self-report, sensor) to determine ***natural history*** of **patients** given **drugs similar to drug** x
BGUC3-4	Track **responses to treatment** to ensure detection of **biological process** x
BGUC3-5	Find **patient data** *"like these"* with **similar treatments, responses to treatment, genetics**
	Search for **studies** a-z with **patient data** with **biological process** x (e.g, **obesity** as

Supported by the NIH grant 1U24 AI117966-01 to the University of California, San Diego

WG3 Metadata – Mapping

③ Metadata Mapping:

- both ***generic*** and (progressively more) ***life science-specific*** metadata schemas are being mapped to identify common metadata elements
- if available, formal representations such XML schema document (XSD) and semantic model (RDF/OWL representations) will be used as input material to the mapping process
  - provenance information each metadata schema/model is described and made available in the BioSharing collection, including: version; source of metadata elements (e.g. XSD), the URL where the model or schema has been sourced; documentation, including URL where the documentation is located

WG3 Metadata – Core list

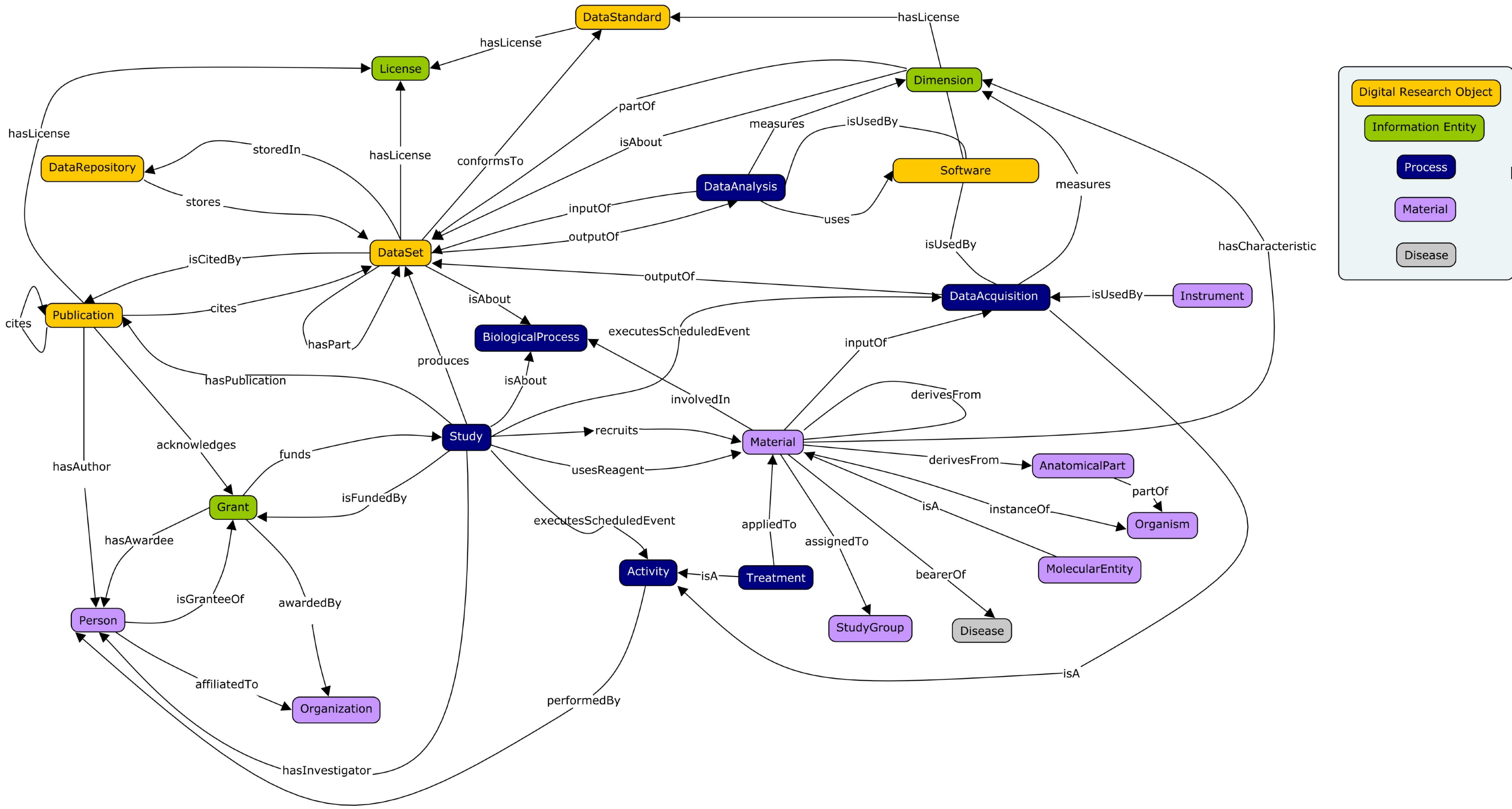
④ Core Metadata Elements:

- the result of the combined approaches, as outlined documents 1-3, is delivering a set of core metadata elements and progressively these will be extended to domain specific ones
- we aim to have maximum coverage of use cases with minimal number of data elements, but we do foreseen that not all questions can be answered in full

bio
CADDIE

http://dx.doi.org/10.5281/zenodo.28019

JSON-schemas
JSON examples



Open Biological and biomedical Ontologies (OBO) Foundry

- origins related to Gene Ontology (GO)
- collection of orthogonal reference ontologies in the biological and biomedical domain
  - e.g GO, chemical entities (ChEBI), investigations (OBI), phenotypes (PATO, MP), …
- agreed set of principles; best practices on ontology development
    - open
    - well-defined format, e.g. **obo** or **owl**
    - uses identifiers according to **obo** id policy
    - ontology life-cycle/versioning
    - clearly specified and delineated content
    - unambiguous definitions
    - uses or extends relations in the obo relations ontology
    - well-documented
    - plurality of users (mailing list, issue tracker)
    - developed collaboratively
    - orthogonal, modular

stato

statistics ontology

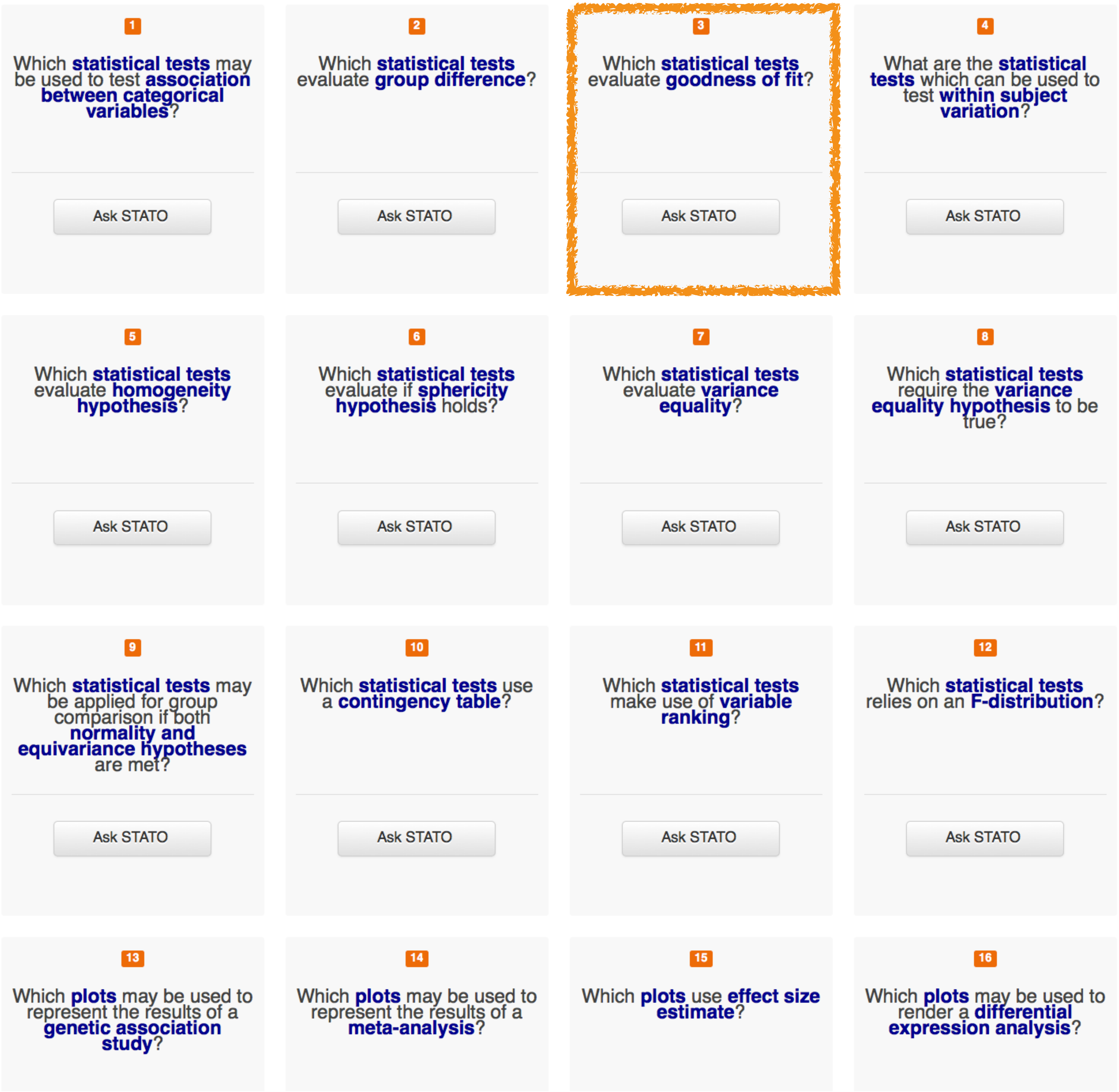
Developed in collaboration with Dr Burke, Senior Statistician, Nuffield Department of Population Health, University of Oxford

http://isa-tools.github.io/stato/

- General-purpose statistics ontology (formal logic-based representation)

- Coverage for processes (e.g. statistical tests and their condition of application) and information needed or resulting from statistical methods (e.g. probability distributions, variable, spread and variation metrics)

- STATO also benefits from: (i) **extensive documentation** with the provision of textual and formal definitions; (ii) an **associated R code snippets** using the dedicated R-command metadata tag, aiming at facilitating teaching and learning while relying of the popular R language; (iii) **query examples** documentation, highlighting how the ontology can be harnessed for reviewers/ tutors/student alike.



1

Which **statistical tests** may be used to test **association between categorical variables**?

Ask STATO

2

Which **statistical tests** evaluate **group difference**?

Ask STATO

3

Which **statistical tests** evaluate **goodness of fit**?

Ask STATO

4

What are the **statistical tests** which can be used to test **within subject variation**?

Ask STATO

5

Which **statistical tests** evaluate **homogeneity hypothesis**?

Ask STATO

6

Which **statistical tests** evaluate if **sphericity hypothesis** holds?

Ask STATO

7

Which **statistical tests** evaluate **variance equality**?

Ask STATO

8

Which **statistical tests** require the **variance equality hypothesis** to be true?

Ask STATO

9

Which **statistical tests** may be applied for group comparison if both **normality and equivariance hypotheses** are met?

Ask STATO

10

Which **statistical tests** use a **contingency table**?

Ask STATO

11

Which **statistical tests** make use of **variable ranking**?

Ask STATO

12

Which **statistical tests** relies on an **F-distribution**?

Ask STATO

13

Which **plots** may be used to represent the results of a **genetic association study**?

14

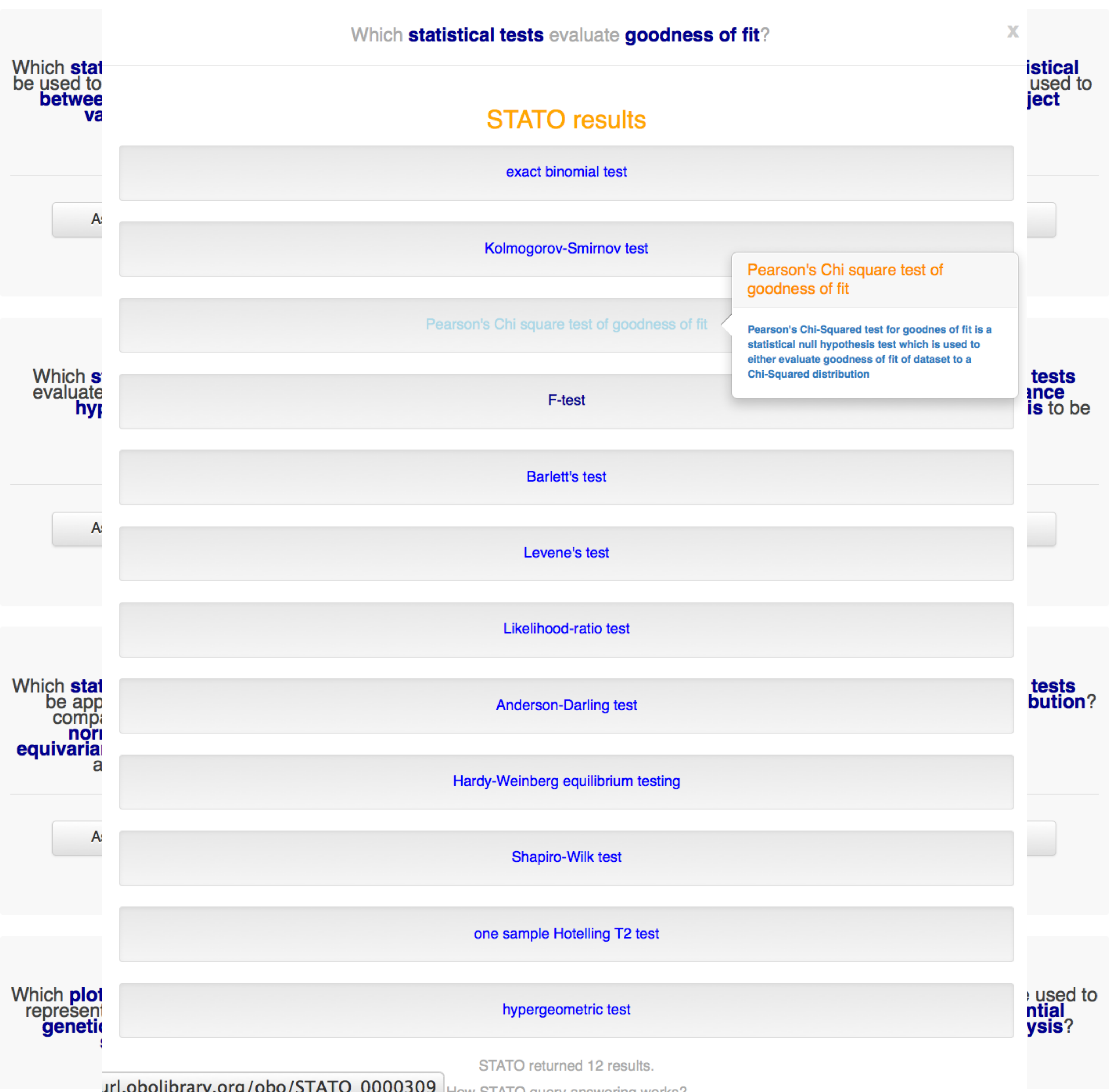
Which **plots** may be used to represent the results of a **meta-analysis**?

15

Which **plots** use **effect size estimate**?

16

Which **plots** may be used to render a **differential expression analysis**?



Which **statistical tests** evaluate **goodness of fit**?

STATO results

- exact binomial test
- Kolmogorov-Smirnov test
- Pearson's Chi square test of goodness of fit
- F-test
- Barlett's test
- Levene's test
- Likelihood-ratio test
- Anderson-Darling test
- Hardy-Weinberg equilibrium testing
- Shapiro-Wilk test
- one sample Hotelling T2 test
- hypergeometric test

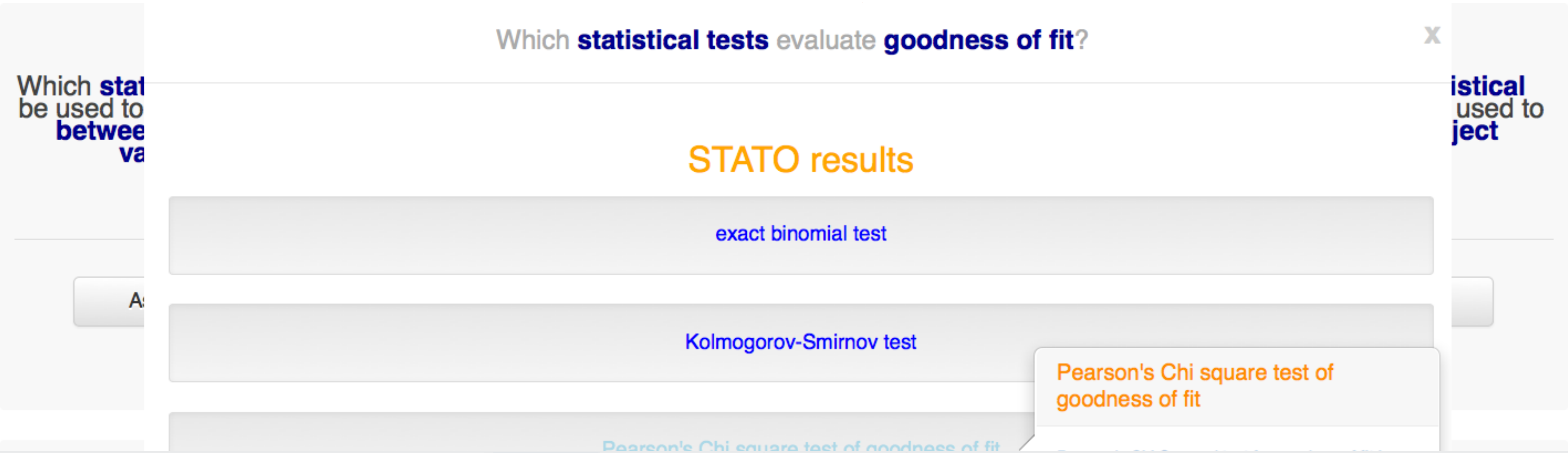
Pearson's Chi square test of goodness of fit

Pearson's Chi-Squared test for goodnes of fit is a statistical null hypothesis test which is used to either evaluate goodness of fit of dataset to a Chi-Squared distribution

STATO returned 12 results.

url.obolibrary.org/obo/STATO_0000309 How STATO query answering works?

Which stat be used to betwee va
A
Which s evaluate hyp
A
Which stat be app compa nor equivaria a
A
Which plot represent geneti
istical used to ject
tests ance is to be
tests bution?
used to ntial ysis?



Which **statistical tests** evaluate **goodness of fit**?

x

Which **stat**
be used to
betwee
va

istical
used to
ject

A

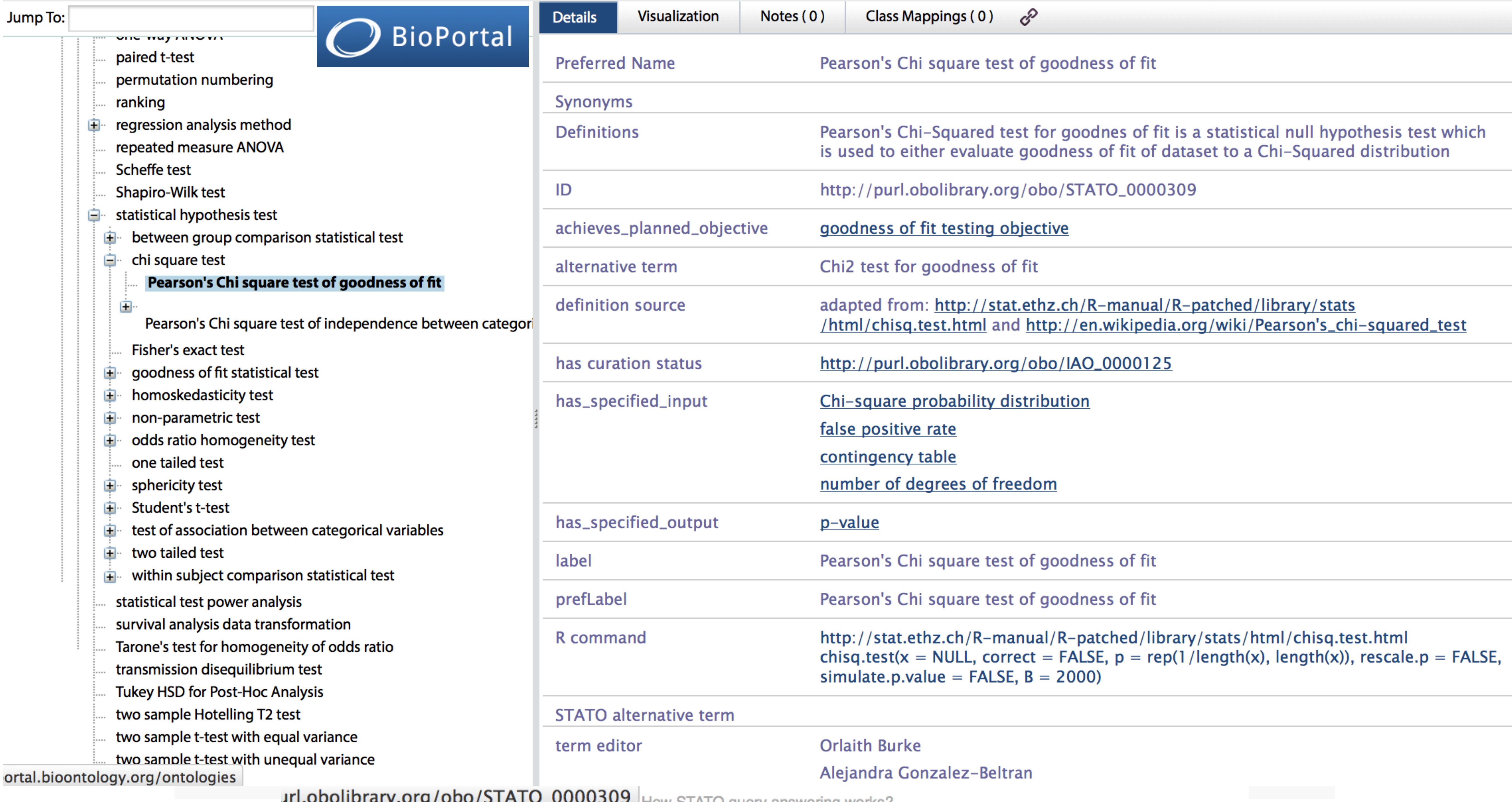
STATO results

exact binomial test

Kolmogorov-Smirnov test

Pearson's Chi square test of goodness of fit

Pearson's Chi square test of goodness of fit



Jump To:

- one-way ANOVA
- paired t-test
- permutation numbering
- ranking
- ⊞ regression analysis method
- repeated measure ANOVA
- Scheffe test
- Shapiro-Wilk test
- ⊟ statistical hypothesis test
  - ⊞ between group comparison statistical test
  - ⊟ chi square test
    - **Pearson's Chi square test of goodness of fit**
    - ⊞ Pearson's Chi square test of independence between categori
  - Fisher's exact test
  - ⊞ goodness of fit statistical test
  - ⊞ homoskedasticity test
  - ⊞ non-parametric test
  - ⊞ odds ratio homogeneity test
  - one tailed test
  - ⊞ sphericity test
  - ⊞ Student's t-test
  - ⊞ test of association between categorical variables
  - ⊞ two tailed test
  - ⊞ within subject comparison statistical test
- statistical test power analysis
- survival analysis data transformation
- Tarone's test for homogeneity of odds ratio
- transmission disequilibrium test
- Tukey HSD for Post-Hoc Analysis
- two sample Hotelling T2 test
- two sample t-test with equal variance
- two sample t-test with unequal variance

ortal.bioontology.org/ontologies

url.obolibrary.org/obo/STATO_0000309

How STATO query answering works?

Preferred Name	Pearson's Chi square test of goodness of fit
Synonyms	
Definitions	Pearson's Chi-Squared test for goodnes of fit is a statistical null hypothesis test which is used to either evaluate goodness of fit of dataset to a Chi-Squared distribution
ID	http://purl.obolibrary.org/obo/STATO_0000309
achieves_planned_objective	goodness of fit testing objective
alternative term	Chi2 test for goodness of fit
definition source	adapted from: http://stat.ethz.ch/R-manual/R-patched/library/stats/html/chisq.test.html and http://en.wikipedia.org/wiki/Pearson's_chi-squared_test
has curation status	http://purl.obolibrary.org/obo/IAO_0000125
has_specified_input	Chi-square probability distribution false positive rate contingency table number of degrees of freedom
has_specified_output	p-value
label	Pearson's Chi square test of goodness of fit
prefLabel	Pearson's Chi square test of goodness of fit
R command	http://stat.ethz.ch/R-manual/R-patched/library/stats/html/chisq.test.html chisq.test(x = NULL, correct = FALSE, p = rep(1/length(x), length(x)), rescale.p = FALSE, simulate.p.value = FALSE, B = 2000)
STATO alternative term	
term editor	Orlaith Burke Alejandra Gonzalez-Beltran

Thanks for your attention!

Questions?

bbsrc biotechnology and biological sciences research council

NATURAL ENVIRONMENT RESEARCH COUNCIL

NEBC

SEVENTH FRAMEWORK PROGRAMME

elixir UNITED KINGDOM

NIH Big Data to Knowledge (BD2K)

UNIVERSITY OF OXFORD

OXFORD e-Research CENTRE

You can email us...
isatools@googlegroups.com

View our websites

isa-tools.org

isacreator configurator · isacreator · bii · converter · Risa · linkedisa

stato-ontology.org

biosharing.org

View our Git repo & contribute
http://github.com/ISA-tools

View our blog
http://isatools.wordpress.com

Follow us on Twitter
@isatools