**Driving digital research and providing services to the Life, Natural and Medical Sciences**

# Create and publish reusable data in the life sciences

Susanna-Assunta Sansone, PhD

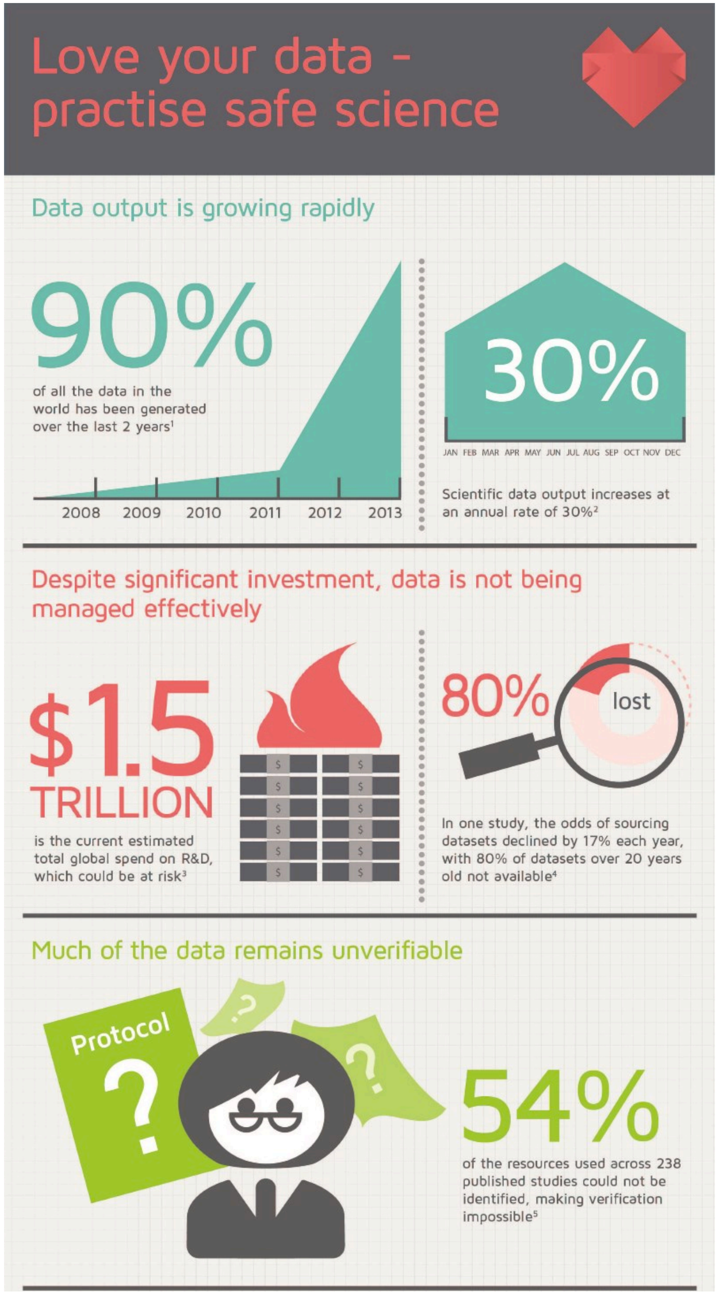
uk.linkedin.com/in/sasansone
@biosharing
@isatools
@scientificdata

Associate Director,
Principal Investigator

Oxford e-Research Centre
UNIVERSITY OF OXFORD

Data Consultant,
Honorary Academic Editor

nature publishing group
SCIENTIFIC DATA



# Love your data - practise safe science

## Data output is growing rapidly

**90%** of all the data in the world has been generated over the last 2 years[1]

Scientific data output increases at an annual rate of 30%[2]

## Despite significant investment, data is not being managed effectively

**$1.5 TRILLION** is the current estimated total global spend on R&D, which could be at risk[3]

**80%**

In one study, the odds of sourcing datasets declined by 17% each year, with 80% of datasets over 20 years old not available[4]

## Much of the data remains unverifiable

**54%** of the resources used across 238 published studies could not be identified, making verification impossible[5]

# Five Top Reasons to Protect Your Data and Practise Safe Science

Author: Julia Giddings
Date posted: January 13, 2014 in News

Part of the Digital Science family
projects.ac | figshare.com

©2014 Projects www.projects.ac

1. SINTEF (2013, May 22). Big Data, for better or worse: 90% of world's data generated over last two years. http://bit.ly/1IDXQ3Z. 2. Pryor, G. (2012). Why manage research data? In G. Pryor (Ed.), Managing research data (pp. 1-16). 3. 2013 Global R & D Funding Forecast, Advantage Business Media, http://bit.ly/UyETjN. 4. Vine T.H. et al. (2013), The availability of research data declines rapidly with article age. Current Biology (24)1: 94-97. 5. Vasilevsky N.A. et al. (2013), On the reproducibility of science: unique identification of research resources in the biomedical literature. PeerJ 1:e148. 6. Problems with scientific research: How science goes wrong (Oct 19th 2013). The Economist. 7. Steen R.G. et al. (2013), Why has the number of scientific retractions increased? PLoS ONE 8(7): e68397. 8. OECD members and partners: http://bit.ly/SQEyuD. 9. The Digital Curation Centre (DCC), Overview of funders' data policies: http://bit.ly/tLvdR7. 10. NIH Data Sharing Policy and Implementation Guidance: http://1.usa.gov/1cNNbNO

https://projects.ac/blog/five-top-reasons-to-protect-your-data-and-practise-safe-science/

## Time and money is wasted, impacting on science and society

**80,000** Since 2000, over 80,000 patients have taken part in clinical trials based on research that was later retracted because of error or fraud[6]

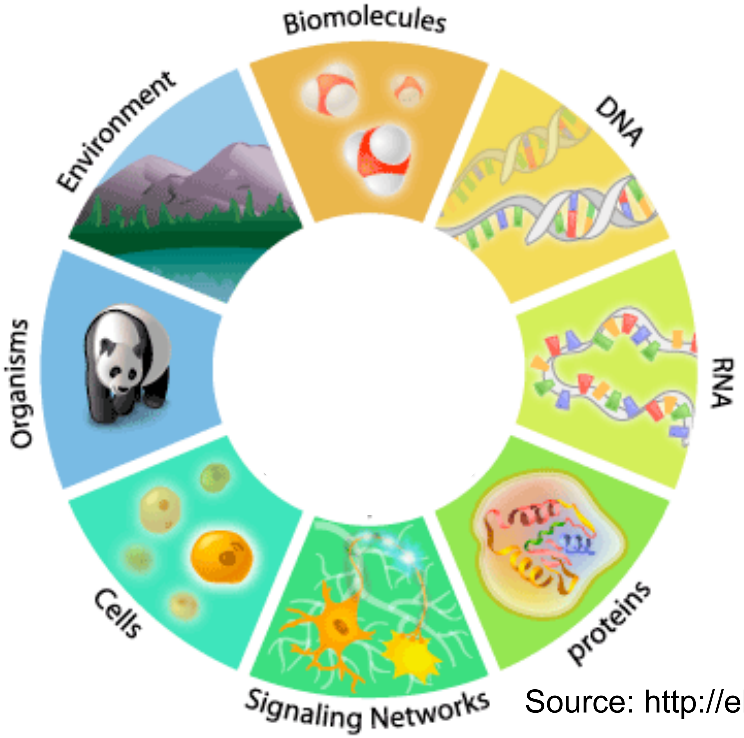
The number of retractions due to error has grown over fivefold since 1990[7]

## Funders now require data management and sharing policies

**34** countries have signed up to the "Declaration on Access to Research Data from Public Funding"[8]

Key funding bodies such as the NIH, MRC and Wellcome Trust now request data management plans be part of applications[9,10]

# Worldwide movement for FAIR data

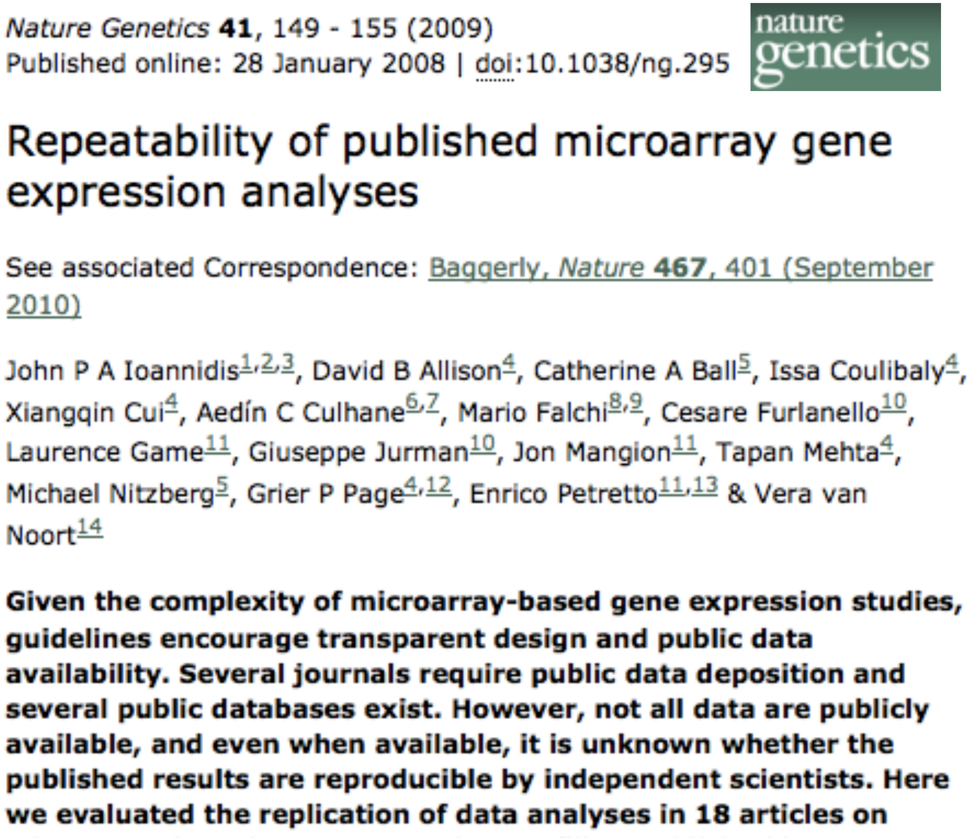


Source: http://ebbailey.wordpress.com

- **Researchers** and **bioinformaticians** in both *academic* and *commercial science*, along with **funding agencies** and **publishers**, embrace the concept that both
  - DATA: *entities of interest e.g.*, genes, metabolites, phenotypes and
  - METADATA: *experimental steps e.g.*, provenance of study materials, technology and measurement types

  should be **Findable**, **Accessible**, **Interoperable** and **Reusable**

# In all fairness, no much data is FAIR!



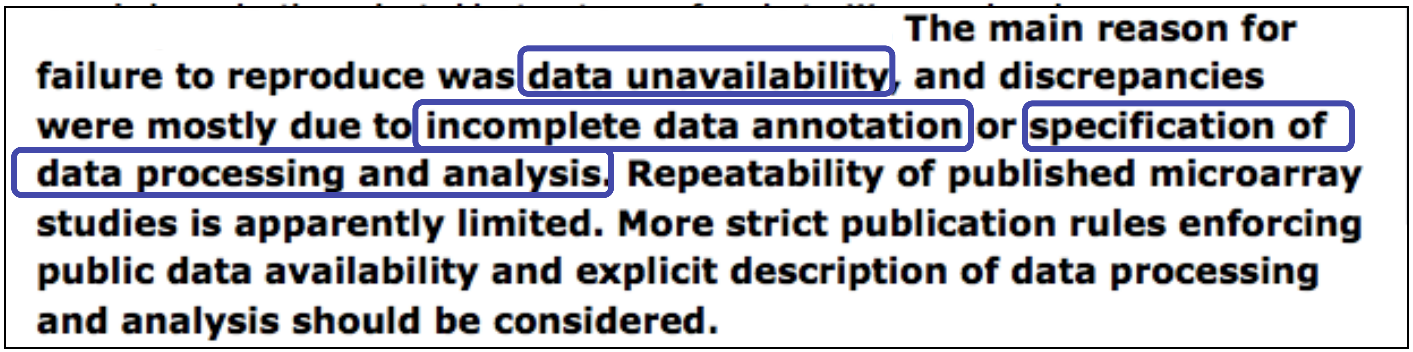
*Nature Genetics* **41**, 149 - 155 (2009)
Published online: 28 January 2008 | doi:10.1038/ng.295

Repeatability of published microarray gene expression analyses

See associated Correspondence: Baggerly, *Nature* **467**, 401 (September 2010)

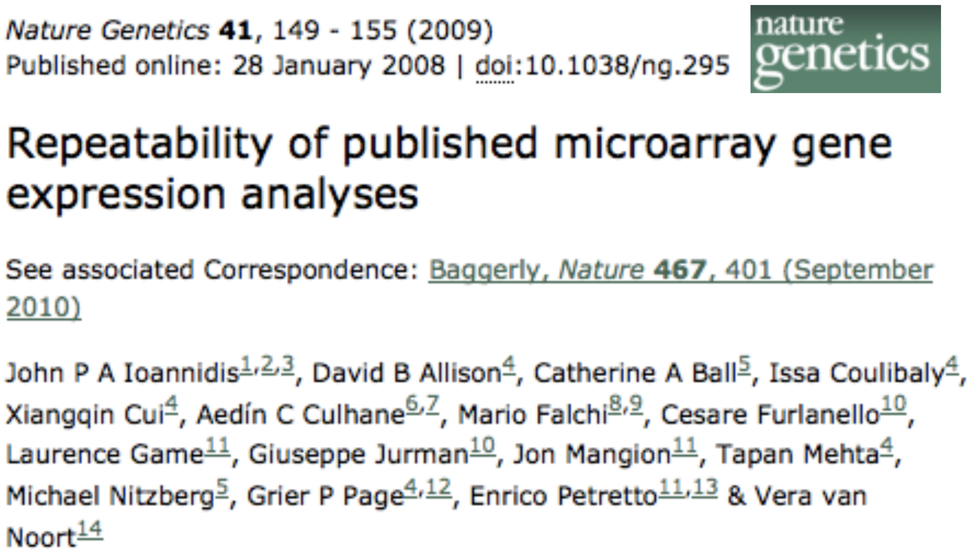
John P A Ioannidis[1,2,3], David B Allison[4], Catherine A Ball[5], Issa Coulibaly[4], Xiangqin Cui[4], Aedín C Culhane[6,7], Mario Falchi[8,9], Cesare Furlanello[10], Laurence Game[11], Giuseppe Jurman[10], Jon Mangion[11], Tapan Mehta[4], Michael Nitzberg[5], Grier P Page[4,12], Enrico Petretto[11,13] & Vera van Noort[14]

**Given the complexity of microarray-based gene expression studies, guidelines encourage transparent design and public data availability. Several journals require public data deposition and several public databases exist. However, not all data are publicly available, and even when available, it is unknown whether the published results are reproducible by independent scientists. Here we evaluated the replication of data analyses in 18 articles on microarray-based gene expression profiling published in *Nature Genetics* in 2005–2006. One table or figure from each article was**



**The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis. Repeatability of published microarray studies is apparently limited. More strict publication rules enforcing public data availability and explicit description of data processing and analysis should be considered.**

# In all fairness, no much data is FAIR!



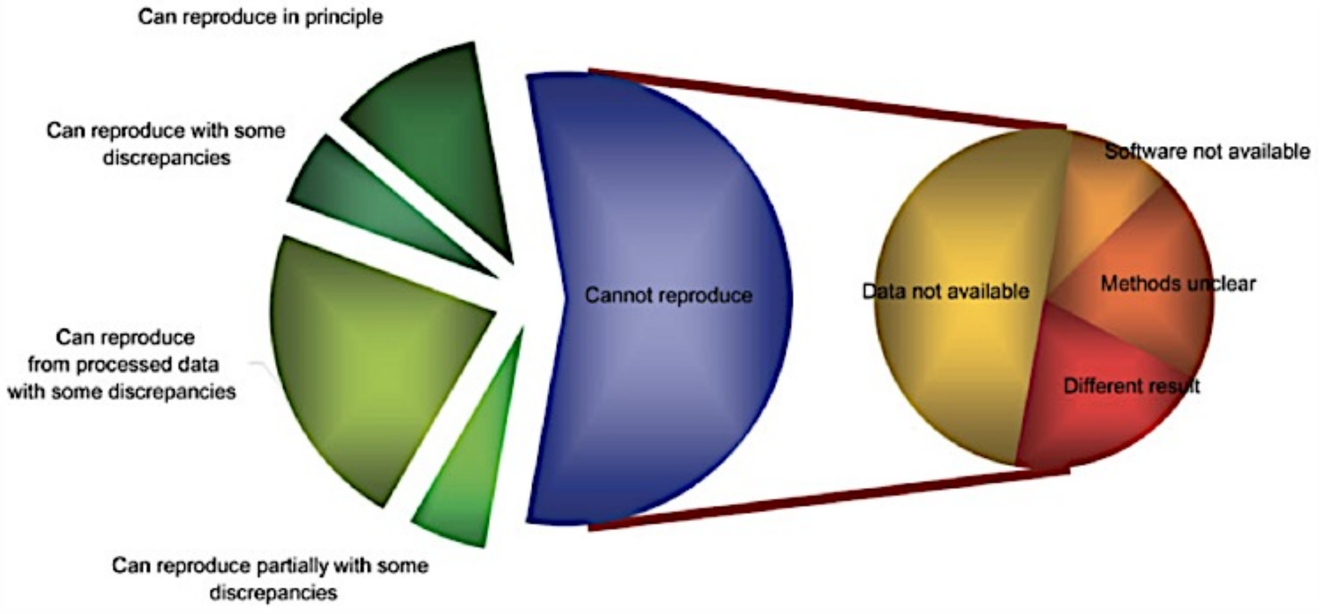
*Nature Genetics* **41**, 149 - 155 (2009)
Published online: 28 January 2008 | doi:10.1038/ng.295

nature genetics

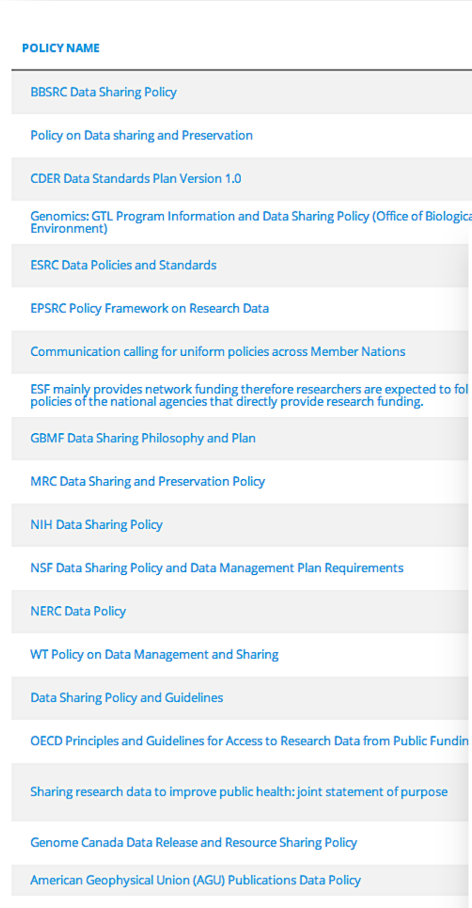
## Repeatability of published microarray gene expression analyses

See associated Correspondence: Baggerly, *Nature* **467**, 401 (September 2010)

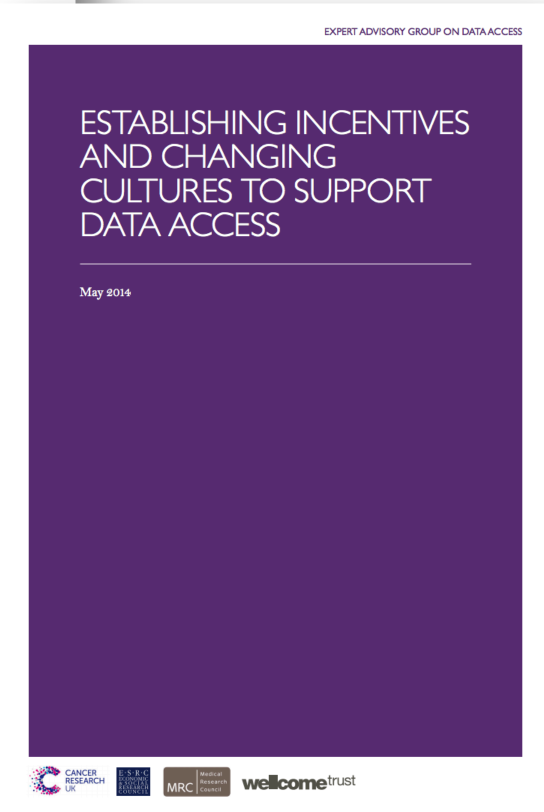
John P A Ioannidis[1,2,3], David B Allison[4], Catherine A Ball[5], Issa Coulibaly[4], Xiangqin Cui[4], Aedín C Culhane[6,7], Mario Falchi[8,9], Cesare Furlanello[10], Laurence Game[11], Giuseppe Jurman[10], Jon Mangion[11], Tapan Mehta[4], Michael Nitzberg[5], Grier P Page[4,12], Enrico Petretto[11,13] & Vera van Noort[14]



# Roles and responsibilities



| POLICY NAME |
|---|
| BBSRC Data Sharing Policy |
| Policy on Data sharing and Preservation |
| CDER Data Standards Plan Version 1.0 |
| Genomics: GTL Program Information and Data Sharing Policy (Office of Biological and Environment) |
| ESRC Data Policies and Standards |
| EPSRC Policy Framework on Research Data |
| Communication calling for uniform policies across Member Nations |
| ESF mainly provides network funding therefore researchers are expected to fol policies of the national agencies that directly provide research funding. |
| GBMF Data Sharing Philosophy and Plan |
| MRC Data Sharing and Preservation Policy |
| NIH Data Sharing Policy |
| NSF Data Sharing Policy and Data Management Plan Requirements |
| NERC Data Policy |
| WT Policy on Data Management and Sharing |
| Data Sharing Policy and Guidelines |
| OECD Principles and Guidelines for Access to Research Data from Public Fundin |
| Sharing research data to improve public health: joint statement of purpose |
| Genome Canada Data Release and Resource Sharing Policy |
| American Geophysical Union (AGU) Publications Data Policy |

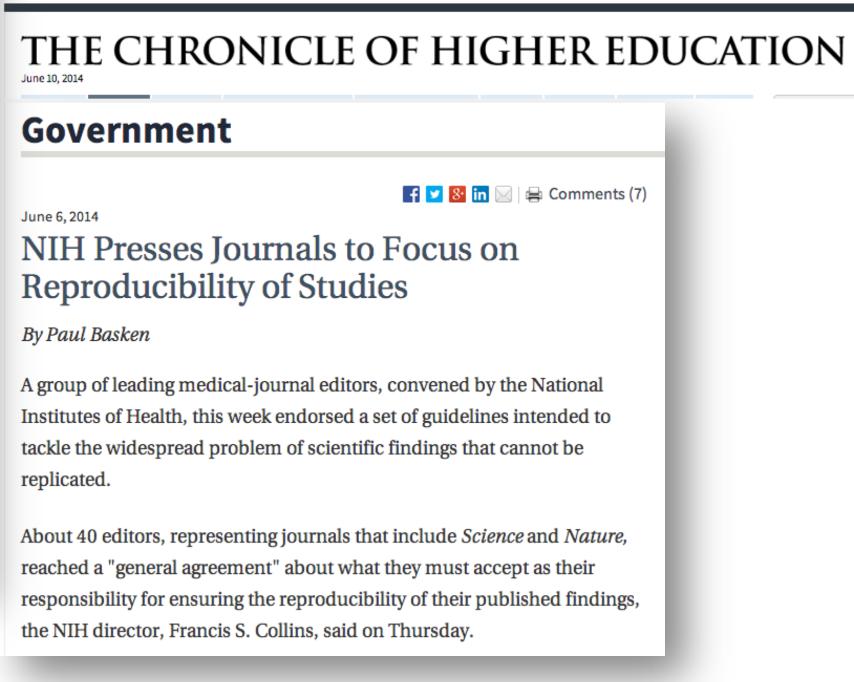


EXPERT ADVISORY GROUP ON DATA ACCESS

ESTABLISHING INCENTIVES AND CHANGING CULTURES TO SUPPORT DATA ACCESS

May 2014

http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtp056495.pdf



THE CHRONICLE OF HIGHER EDUCATION

June 10, 2014

Government

June 6, 2014

NIH Presses Journals to Focus on Reproducibility of Studies

*By Paul Basken*

A group of leading medical-journal editors, convened by the National Institutes of Health, this week endorsed a set of guidelines intended to tackle the widespread problem of scientific findings that cannot be replicated.

About 40 editors, representing journals that include *Science* and *Nature*, reached a "general agreement" about what they must accept as their responsibility for ensuring the reproducibility of their published findings, the NIH director, Francis S. Collins, said on Thursday.

http://chronicle.com/article/article-content/146951/

# LIFE, NATURAL & BIOMEDICAL SCIENCES

**Driving digital research and providing services to the Life, Natural and Medical Sciences**

Roche

elixir UNITED KINGDOM

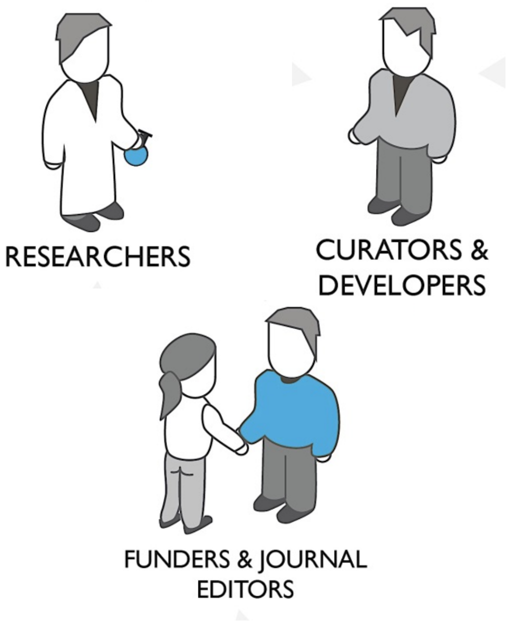
bbsrc biotechnology and biological sciences research council

MRC Medical Research Council

NATURAL ENVIRONMENT RESEARCH COUNCIL

SEVENTH FRAMEWORK PROGRAMME

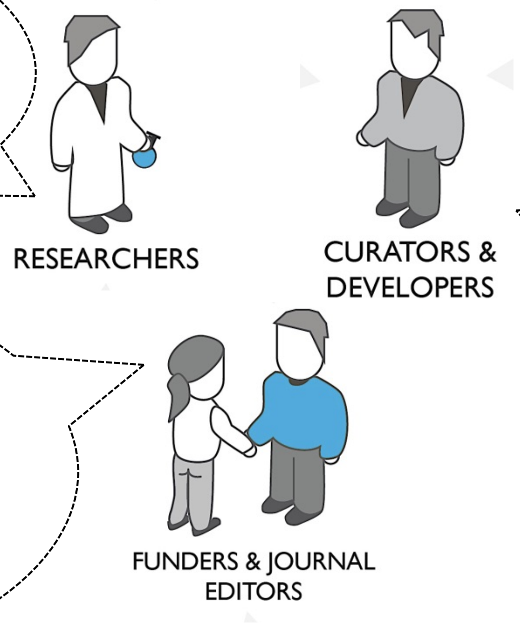
*Communities we work with/for:*



# LIFE, NATURAL & BIOMEDICAL SCIENCES

**Driving digital research and providing services to the Life, Natural and Medical Sciences**

*Communities we work with/for:*



- Describe my experiments
- Share info with my group
- Store for query & access
- Upload to analysis tool
- Compare with similar data
- etc……

- Support the needs of our researchers
- Enhancing our existing tools; develop new components, or connect with other tools
- Comply to community standards
- etc….

- Drive data reusability agenda
- Recommend tools, data resources and standards in data policies
- Criteria for endorsement
- Mapping the ecosystem of efforts
- etc……

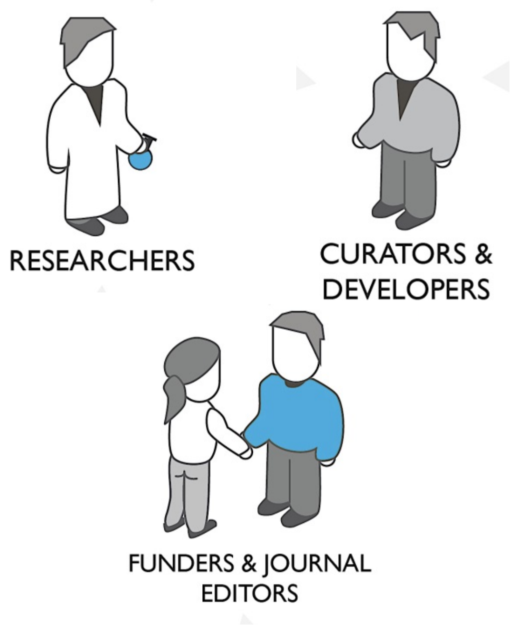
# LIFE, NATURAL & BIOMEDICAL SCIENCES

**Driving digital research and providing services to the Life, Natural and Medical Sciences**

*Key areas of activity:*

- Data capture and curation
- Database development
- Data (nano)publication
- Data provenance
- Open, community ontologies and standards
- Semantic web
- Social engineering
- Software development
- Training
- Visualization (collaboration/ jointly with Prof Min Chen)

*Communities we work with/for:*



# LIFE, NATURAL & BIOMEDICAL SCIENCES

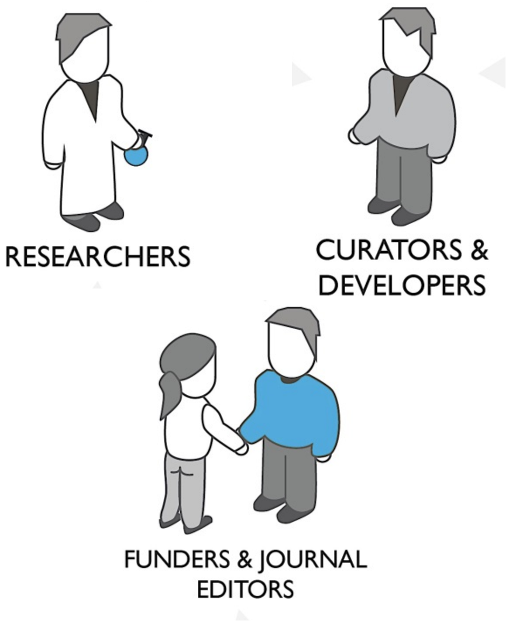
**Driving digital research and providing services to the Life, Natural and Medical Sciences**

bbsrc biotechnology and biological sciences research council
Roche
MRC Medical Research Council
elixir UNITED KINGDOM
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME

*Key areas of activity:*

- Data capture and curation
- Database development
- Data (nano)publication
- Data provenance
- Open, community ontologies and standards
- Semantic web
- Social engineering
- Software development
- Training
- Visualization (collaboration/ jointly with Prof Min Chen)

*Communities we work with/for:*



RESEARCHERS

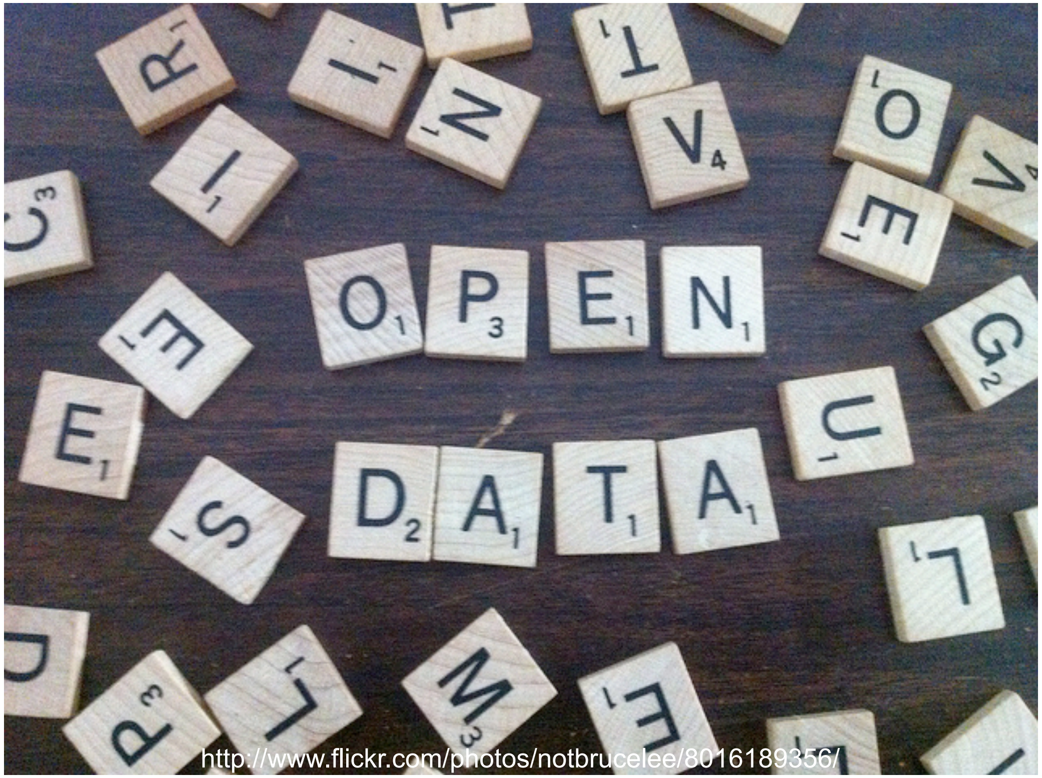
CURATORS & DEVELOPERS

FUNDERS & JOURNAL EDITORS

*As part of:*

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

EMBL-EBI
nature publishing group npg
RDA RESEARCH DATA ALLIANCE
DRYAD
NATIONAL CANCER INSTITUTE Nanotechnology Informatics Working Group
UNIVERSITY OF OXFORD
Oxford Interdisciplinary Bioscience DTP
Nuffield Department of Population Health
Imperial College London
MANCHESTER 1824 The University of Manchester
STANFORD SCHOOL OF MEDICINE
HARVARD SCHOOL OF PUBLIC HEALTH



http://www.flickr.com/photos/notbrucelee/8016189356/

Our activities are *around* and *in* support of
data **curation**, **management** and **publication**
and their pivotal roles in
enabling FAIR data and research, driving science and discoveries

# LIFE, NATURAL & BIOMEDICAL SCIENCES

Roche
bbsrc biotechnology and biological sciences research council
MRC Medical Research Council
elixir UNITED KINGDOM
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME

## CURRENT PROJECTS

isatools

isacommons

### ISA – standards-compliant data management tools

Providing a tool kit to facilitate standards compliant collection, curation and local management of experiments in the life-sciences

find out more ❯

### Digital platforms for scholarly publishing

Collaborations with scientific, technical and medical publishers to develop novel data platforms to track and publish scholarly outputs

find out more ❯

- **Nature Publishing Group's Scientific Data**
- **BioMedCentral and BGI's GigaScience**
- **F1000 Research**
- **Oxford University Press**

elixir

### ELIXIR's UK Node

The UK Node contributes the country's substantial expertise in bioinformatics training for researchers, computer scientists and data managers in the Life, Natural and Medical Sciences.

find out more ❯

### Metagenomics Data Infrastructure

Coordinated by EMBL-European Bioinformatics Institute, the Metagenomics service is being developed to be an automated pipeline for the curation, archiving and analysis of metagenomic data.

find out more ❯

cosmos

### COSMOS - COordination Of Standards In MetabOlomicS

Coordinated by EMBL-European Bioinformatics Institute, COSMOS (Coordination of Standards in Metabolomics) has brought together European metabolomics data providers to set and promote community standards.

find out more ❯

### OBI - Ontology for BioMedical Investigation

stato
statistics ontology

biosharing

### BioSharing – standards, policies and databases

a web-based catalogue to centralize bioscience data policies, reporting standards and links to other related portals; and a communication forum to maintain linkages between funders, journals & standardization leaders

find out more ❯

Two more new funded projects!

# LIFE, NATURAL & BIOMEDICAL SCIENCES

## CURRENT PROJECTS

### ISA – standards-compliant data management tools

Providing a tool kit to facilitate standards compliant collection, curation and local management of experiments in the life-sciences

find out more ❯

### Digital platforms for scholarly publishing

Collaborations with scientific, technical and medical publishers to develop novel data platforms to track and publish scholarly outputs

find out more ❯

- **Nature Publishing Group's Scientific Data**
- **BioMedCentral and BGI's GigaScience**
- **F1000 Research**
- **Oxford University Press**

### ELIXIR's UK Node

The UK Node contributes the country's substantial expertise in bioinformatics training for researchers, computer scientists and data managers in the Life, Natural and Medical Sciences.

find out more ❯

### Metagenomics Data Infrastructure

Coordinated by EMBL-European Bioinformatics Institute, the Metagenomics service is being developed to be an automated pipeline for the curation, archiving and analysis of metagenomic data.

find out more ❯

### COSMOS - COordination Of Standards In MetabOlomicS

Coordinated by EMBL-European Bioinformatics Institute, COSMOS (Coordination of Standards in Metabolomics) has brought together European metabolomics data providers to set and promote community standards.

find out more ❯

### OBI - Ontology for BioMedical Investigation

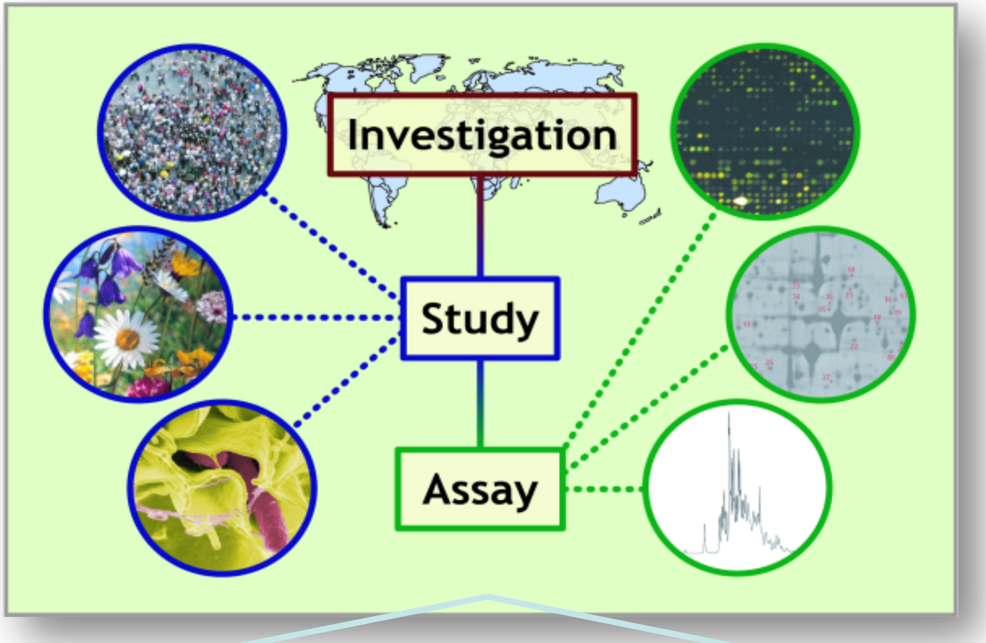
stato
statistics ontology

### BioSharing – standards, policies and databases

a web-based catalogue to centralize bioscience data policies, reporting standards and links to other related portals; and a communication forum to maintain linkages between funders, journals & standardization leaders

find out more ❯

Two more new funded projects!



isatools
isa-tools.org

isatab format

collect and curate, following standards

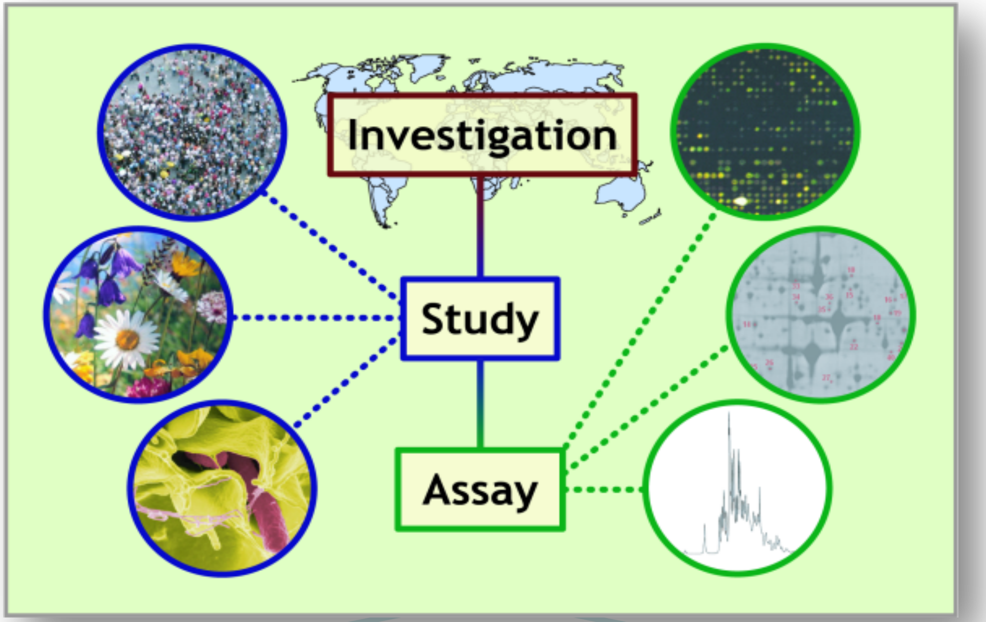
store and browse, locally or publicly

submit to public repositories

analyse with existing tools

release, reason and nanopublish

publish data along your article



isatools
isa-tools.org

isatab format

user community

isacommons
isacommons.org

**ISA powers data collection, curation resources and repositories, e.g.:**

EMBL-EBI

HARVARD School of Public Health

HSCI HARVARD STEM CELL INSTITUTE

NOVARTIS INSTITUTES FOR BIOMEDICAL RESEARCH

diXa

SEURAT-1

npg nature publishing group

SCIENTIFIC DATA

experimental design

sample characteristic(s)

experimental variable(s)

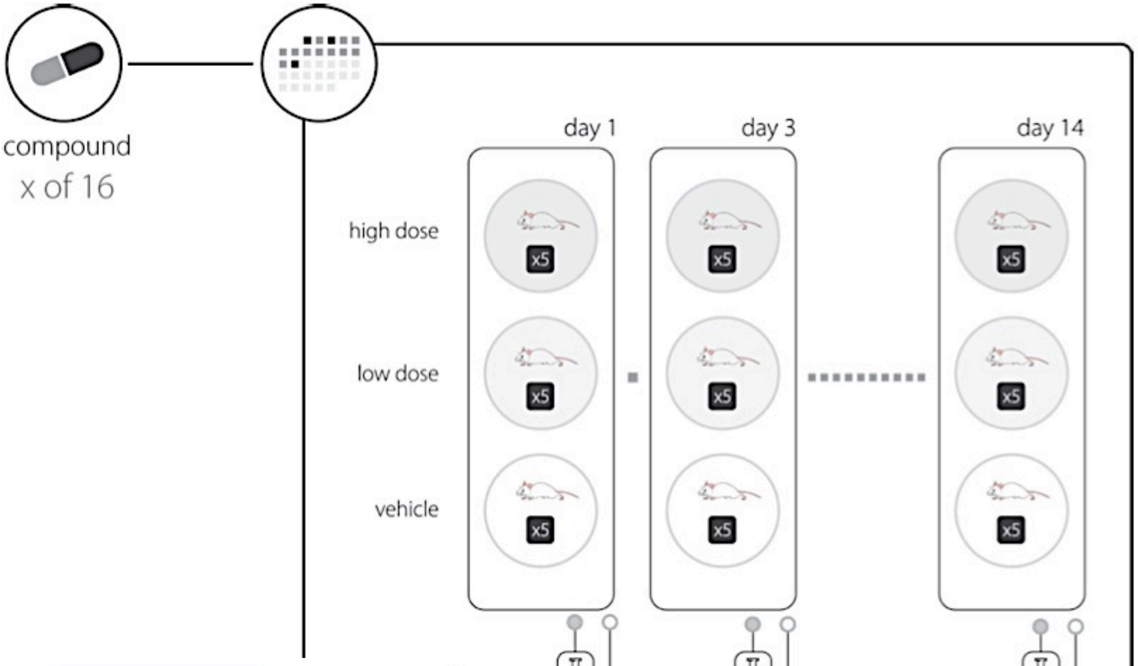
technology(s)

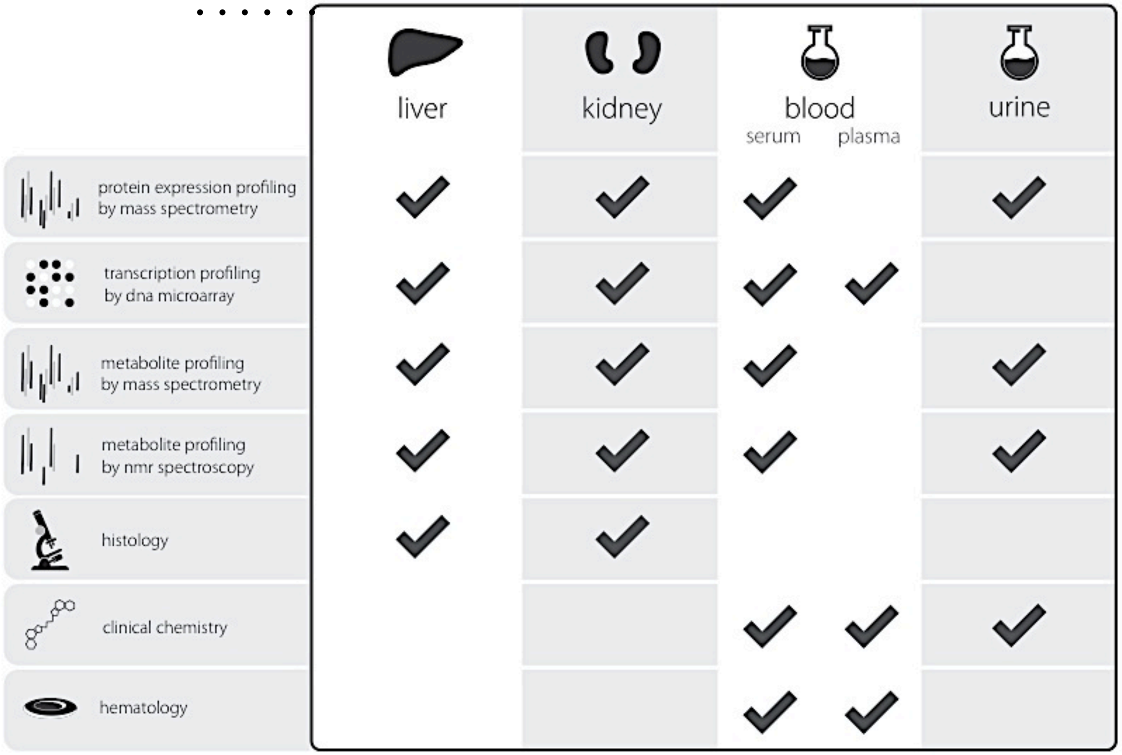
measurement(s)

protocols(s)

data file(s)

.....





| | 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 | | | ✓ | ✓ | |

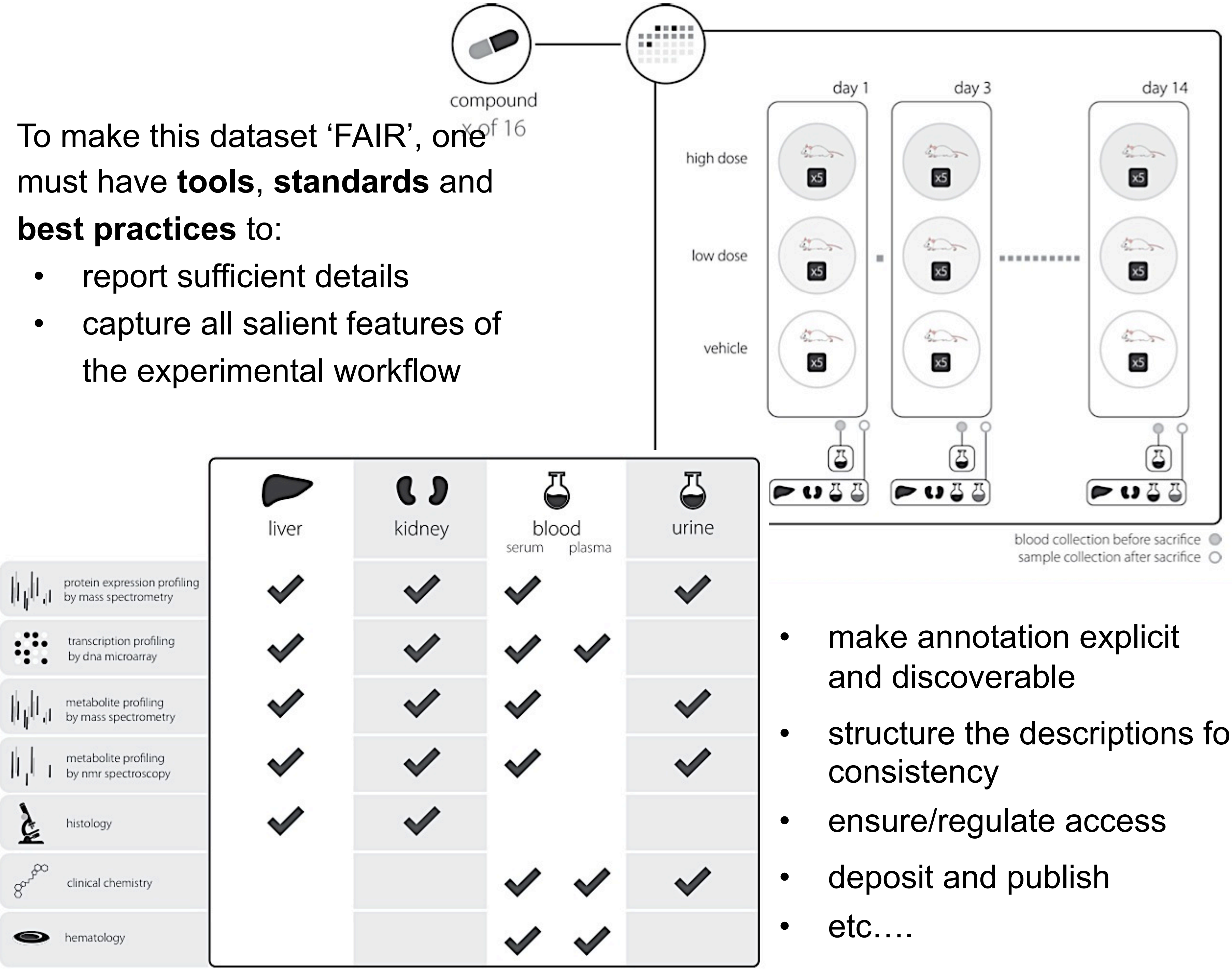
720 animals

16 compounds

3 doses

~20,000 assays

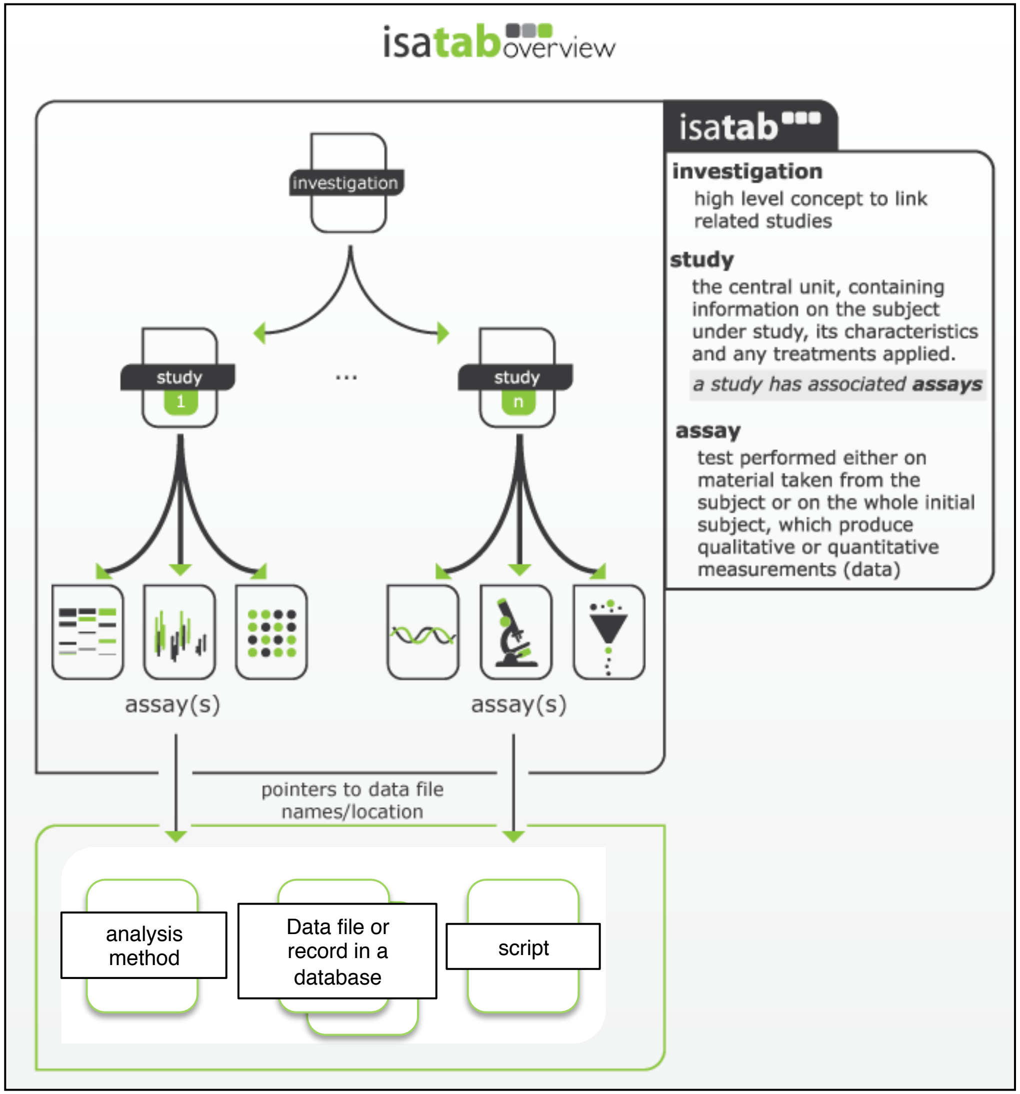
- To make this dataset ‘FAIR’, one must have **tools**, **standards** and **best practices** to:
    - report sufficient details
    - capture all salient features of the experimental workflow



| | 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 | | | ✓ | ✓ | |

- make annotation explicit and discoverable
- structure the descriptions for consistency
- ensure/regulate access
- deposit and publish
- etc….

# isatab



**investigation**
high level concept to link related studies

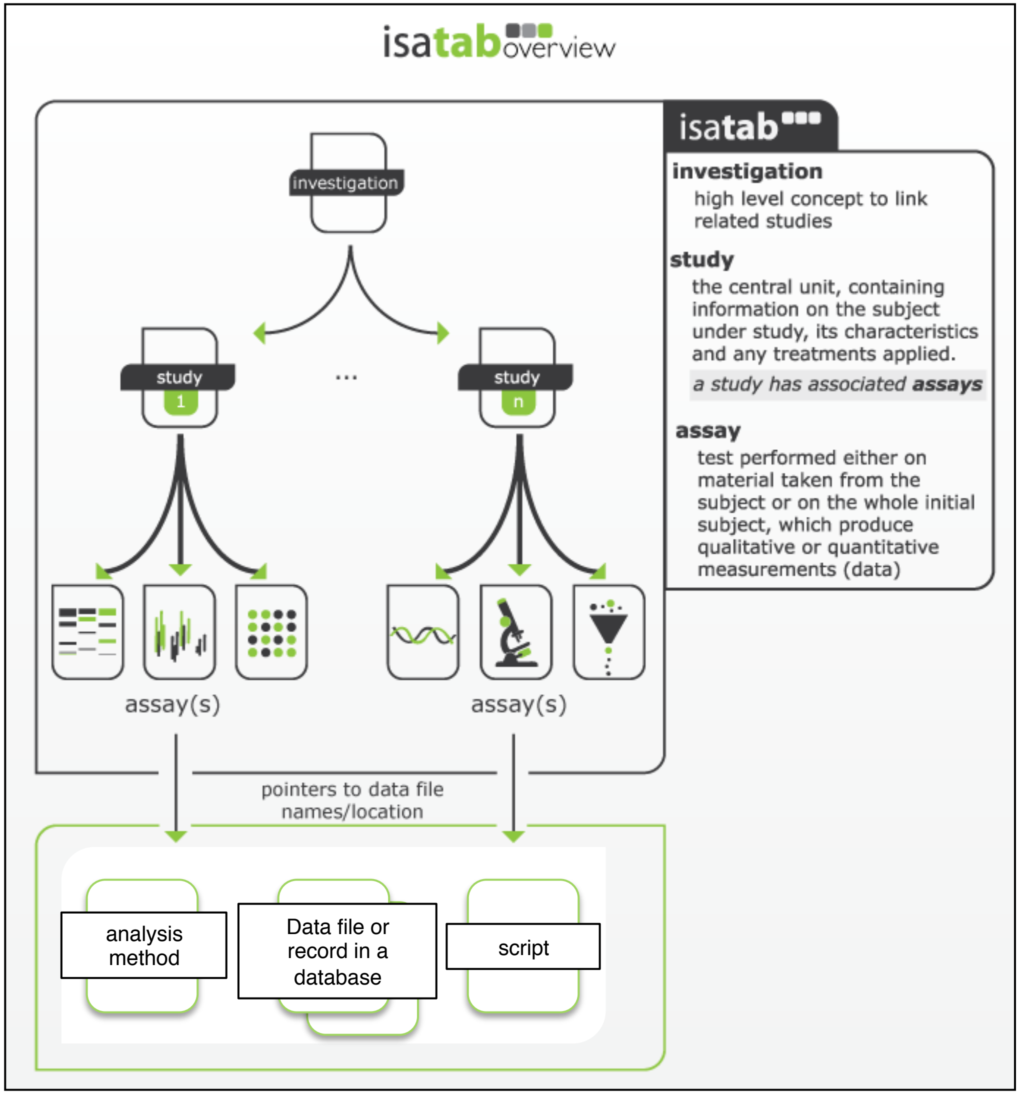
**study**
the central unit, containing information on the subject under study, its characteristics and any treatments applied.

*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)

isatools
COLLECT | CURATE | ANALYSE | SHARE | PUBLISH
*funded by:*
bbsrc
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME
UNIVERSITY OF OXFORD

# isatab



**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)

General-purpose, configurable format, designed to support:

- description of the experimental workflow, making the annotation explicit and discoverable
- provenance tracking
- use community standards, such as minimal reporting guidelines and terminologies
  - over 300 ‘ontologies’ and over 60 guidelines
- conversions to - a growing number of - other metadata formats
  - e.g. used by EBI repositories
  - and as linked data

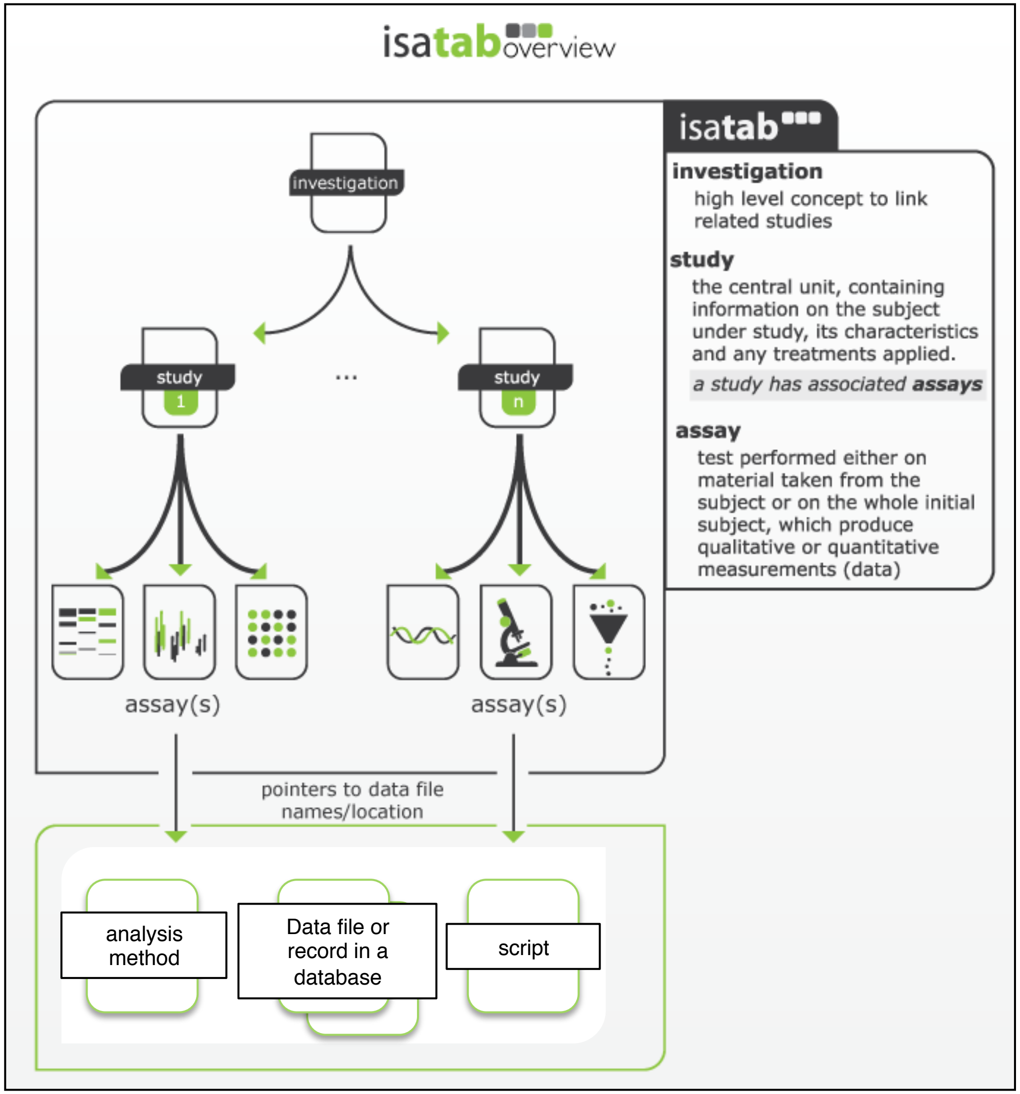
biosharing

converter
mage-tab | pride-ml | sra-xml |
ArrayExpress PRIDE ENA

isatoowl

isatools *funded by:*

# isatab



**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)

General-purpose, configurable format, designed to support:

- description of the experimental workflow, making the annotation explicit and discoverable
- provenance tracking
- use community standards, such as minimal reporting guidelines and terminologies
  - over 300 'ontologies' and over 60 guidelines
- conversions to - a growing number of - other metadata formats
  - e.g. used by EBI repositories
  - and as linked data

biosharing
Run By OXFORD e-Research CENTRE

converter
mage-tab | pride-ml | sra-xml |
ArrayExpress PRIDE ENA

isatoowl

ISA is implemented by several service systems that are

- local, institute-based
- project, consortium-based
- global, international repositories
- and another 'data journal, *GigaScience* in GigaDB

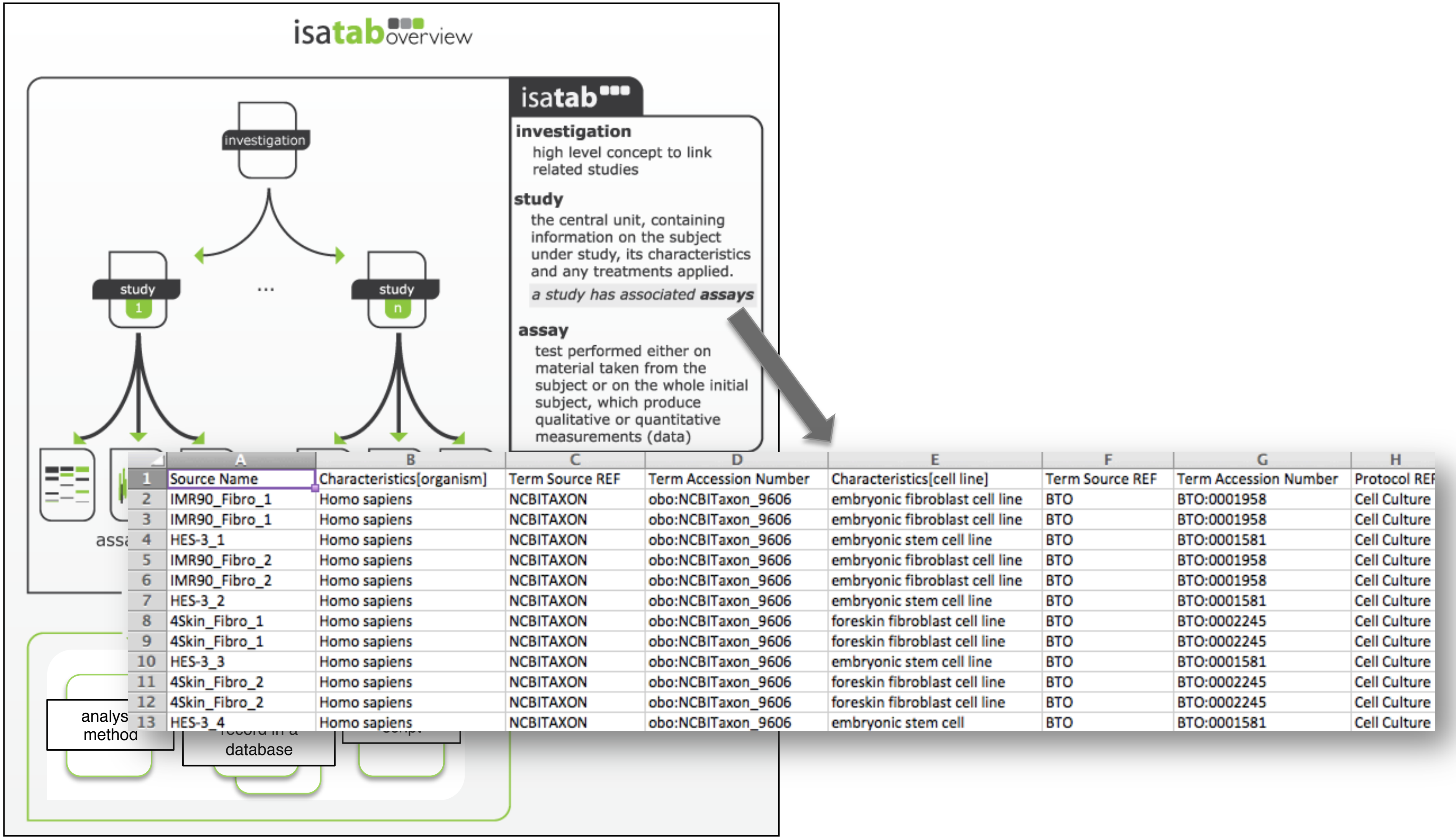
isacommons
isacommons.org

isatools
COLLECT | CURATE | ANALYSE | SHARE | PUBLISH

*funded by:*

bbsrc
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME
UNIVERSITY OF OXFORD

# isatab



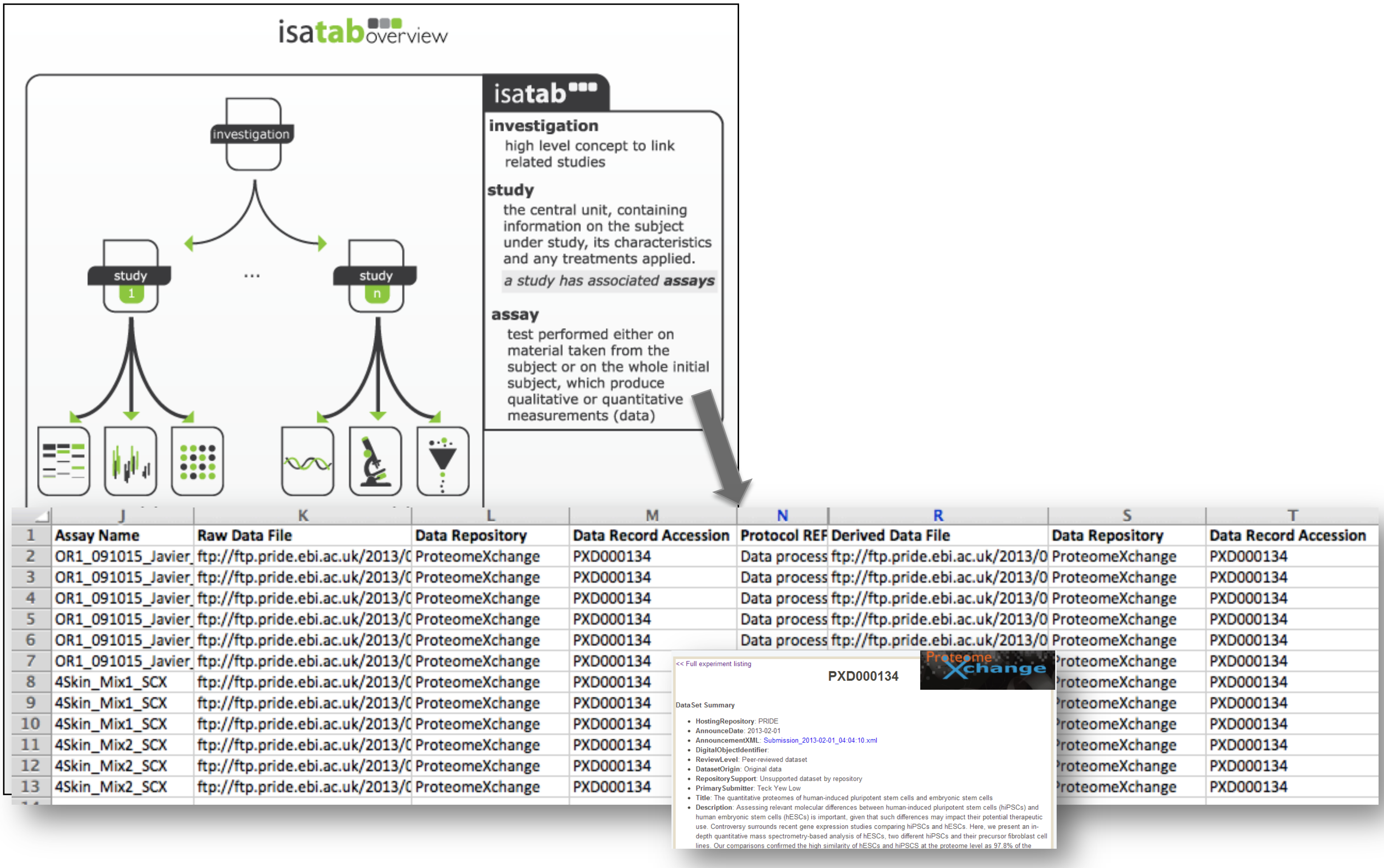
**investigation**
high level concept to link related studies

**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)

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Source Name | Characteristics[organism] | Term Source REF | Term Accession Number | Characteristics[cell line] | Term Source REF | Term Accession Number | Protocol REF |
| 2 | IMR90_Fibro_1 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic fibroblast cell line | BTO | BTO:0001958 | Cell Culture |
| 3 | IMR90_Fibro_1 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic fibroblast cell line | BTO | BTO:0001958 | Cell Culture |
| 4 | HES-3_1 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic stem cell line | BTO | BTO:0001581 | Cell Culture |
| 5 | IMR90_Fibro_2 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic fibroblast cell line | BTO | BTO:0001958 | Cell Culture |
| 6 | IMR90_Fibro_2 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic fibroblast cell line | BTO | BTO:0001958 | Cell Culture |
| 7 | HES-3_2 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic stem cell line | BTO | BTO:0001581 | Cell Culture |
| 8 | 4Skin_Fibro_1 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | foreskin fibroblast cell line | BTO | BTO:0002245 | Cell Culture |
| 9 | 4Skin_Fibro_1 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | foreskin fibroblast cell line | BTO | BTO:0002245 | Cell Culture |
| 10 | HES-3_3 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic stem cell line | BTO | BTO:0001581 | Cell Culture |
| 11 | 4Skin_Fibro_2 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | foreskin fibroblast cell line | BTO | BTO:0002245 | Cell Culture |
| 12 | 4Skin_Fibro_2 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | foreskin fibroblast cell line | BTO | BTO:0002245 | Cell Culture |
| 13 | HES-3_4 | Homo sapiens | NCBITAXON | obo:NCBITaxon_9606 | embryonic stem cell | BTO | BTO:0001581 | Cell Culture |

# isatab



**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***

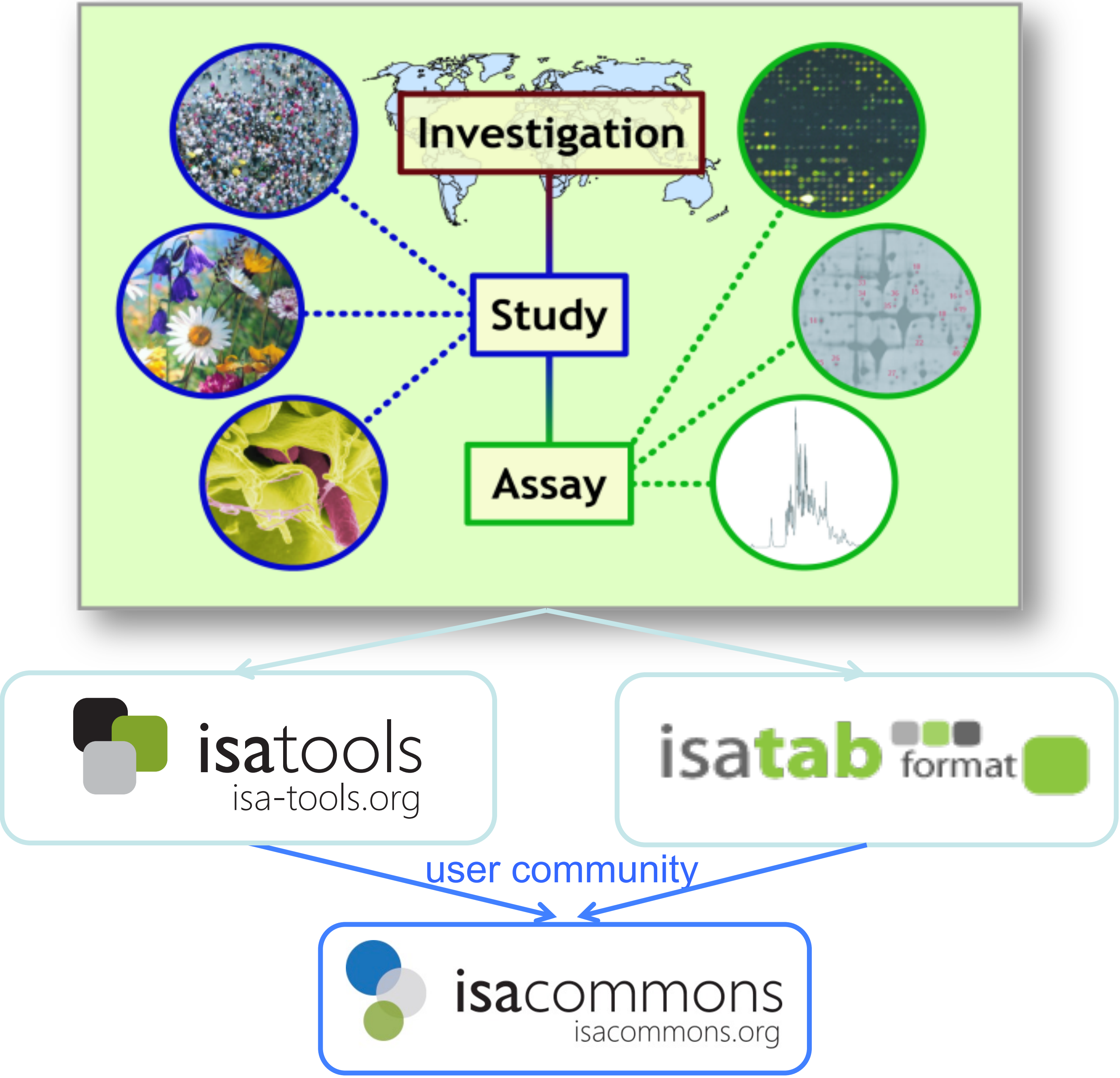
**assay**
test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)

| | J | K | L | M | N | R | S | T |
|---|---|---|---|---|---|---|---|---|
| 1 | **Assay Name** | **Raw Data File** | **Data Repository** | **Data Record Accession** | **Protocol REF** | **Derived Data File** | **Data Repository** | **Data Record Accession** |
| 2 | OR1_091015_Javier_ | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | Data process | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 |
| 3 | OR1_091015_Javier_ | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | Data process | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 |
| 4 | OR1_091015_Javier_ | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | Data process | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 |
| 5 | OR1_091015_Javier_ | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | Data process | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 |
| 6 | OR1_091015_Javier_ | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | Data process | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 |
| 7 | OR1_091015_Javier_ | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |
| 8 | 4Skin_Mix1_SCX | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |
| 9 | 4Skin_Mix1_SCX | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |
| 10 | 4Skin_Mix1_SCX | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |
| 11 | 4Skin_Mix2_SCX | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |
| 12 | 4Skin_Mix2_SCX | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |
| 13 | 4Skin_Mix2_SCX | ftp://ftp.pride.ebi.ac.uk/2013/0 | ProteomeXchange | PXD000134 | | | roteomeXchange | PXD000134 |

**PXD000134**

DataSet Summary

- **HostingRepository**: PRIDE
- **AnnounceDate**: 2013-02-01
- **AnnouncementXML**: Submission_2013-02-01_04:04:10.xml
- **DigitalObjectIdentifier**:
- **ReviewLevel**: Peer-reviewed dataset
- **DatasetOrigin**: Original data
- **RepositorySupport**: Unsupported dataset by repository
- **PrimarySubmitter**: Teck Yew Low
- **Title**: The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells
- **Description**: Assessing relevant molecular differences between human-induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) is important, given that such differences may impact their potential therapeutic use. Controversy surrounds recent gene expression studies comparing hiPSCs and hESCs. Here, we present an in-depth quantitative mass spectrometry-based analysis of hESCs, two different hiPSCs and their precursor fibroblast cell lines. Our comparisons confirmed the high similarity of hESCs and hiPSCS at the proteome level as 97.8% of the

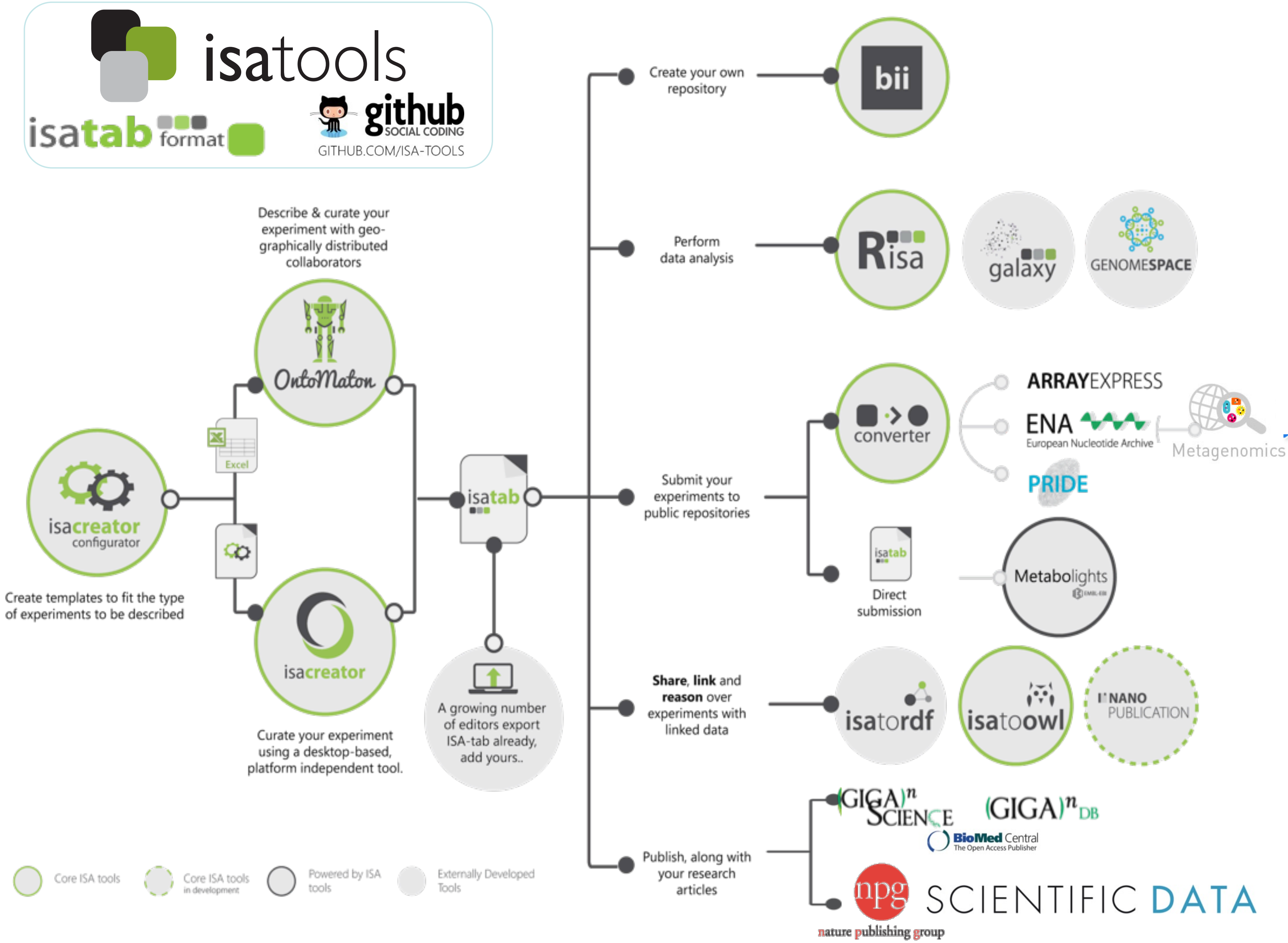


isatools
isa-tools.org

isatab format

user community

isacommons
isacommons.org



Create your own repository

Perform data analysis

Describe & curate your experiment with geographically distributed collaborators

Create templates to fit the type of experiments to be described

Submit your experiments to public repositories

Direct submission

Curate your experiment using a desktop-based, platform independent tool.

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

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

Publish, along with your research articles

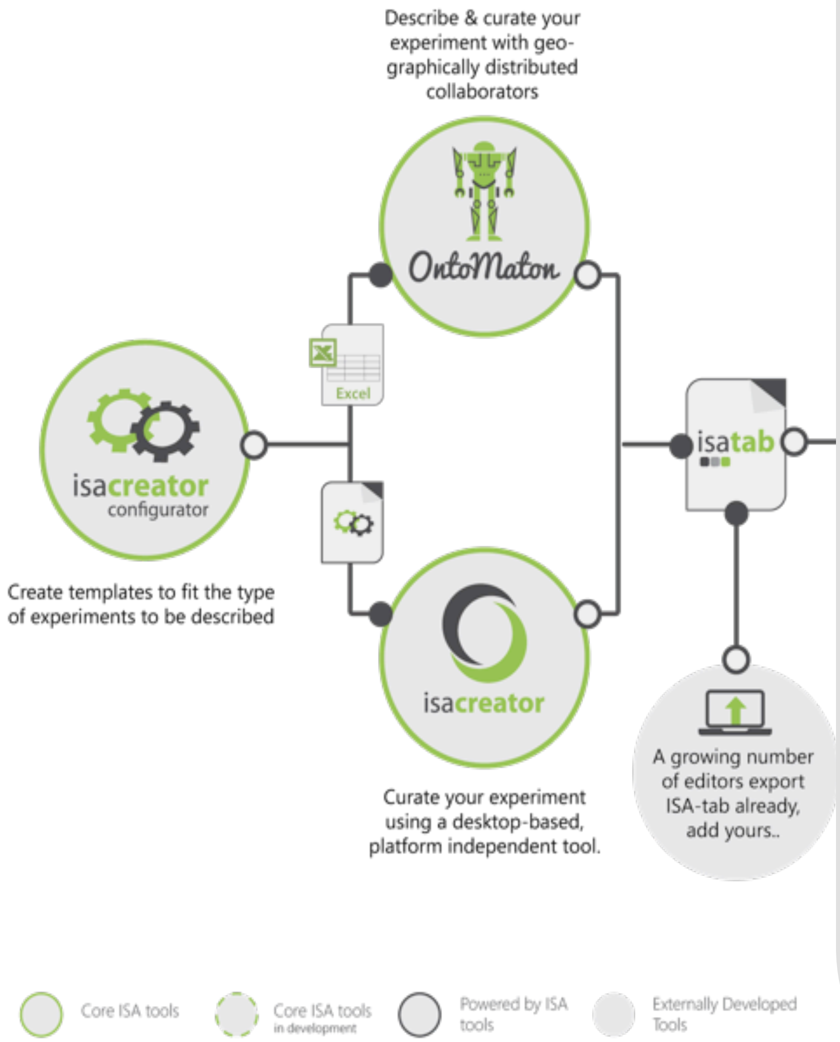
- Core ISA tools
- Core ISA tools in development
- Powered by ISA tools
- Externally Developed Tools

isatools

isatab format

github SOCIAL CODING

GITHUB.COM/ISA-TOOLS

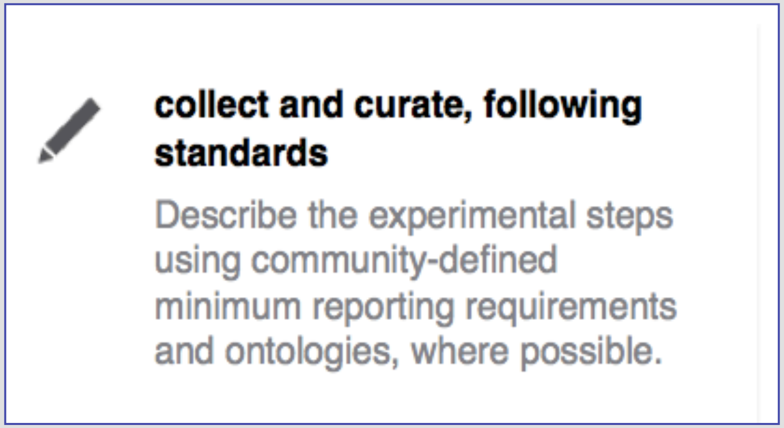

Describe & curate your experiment with geo-graphically distributed collaborators

Create templates to fit the type of experiments to be described

Curate your experiment using a desktop-based, platform independent tool.

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

1


**collect and curate, following standards**

Describe the experimental steps using community-defined minimum reporting requirements and ontologies, where possible.

isatools

isatab format

github SOCIAL CODING

GITHUB.COM/ISA-TOOLS

2

**store and browse, locally or publicly**

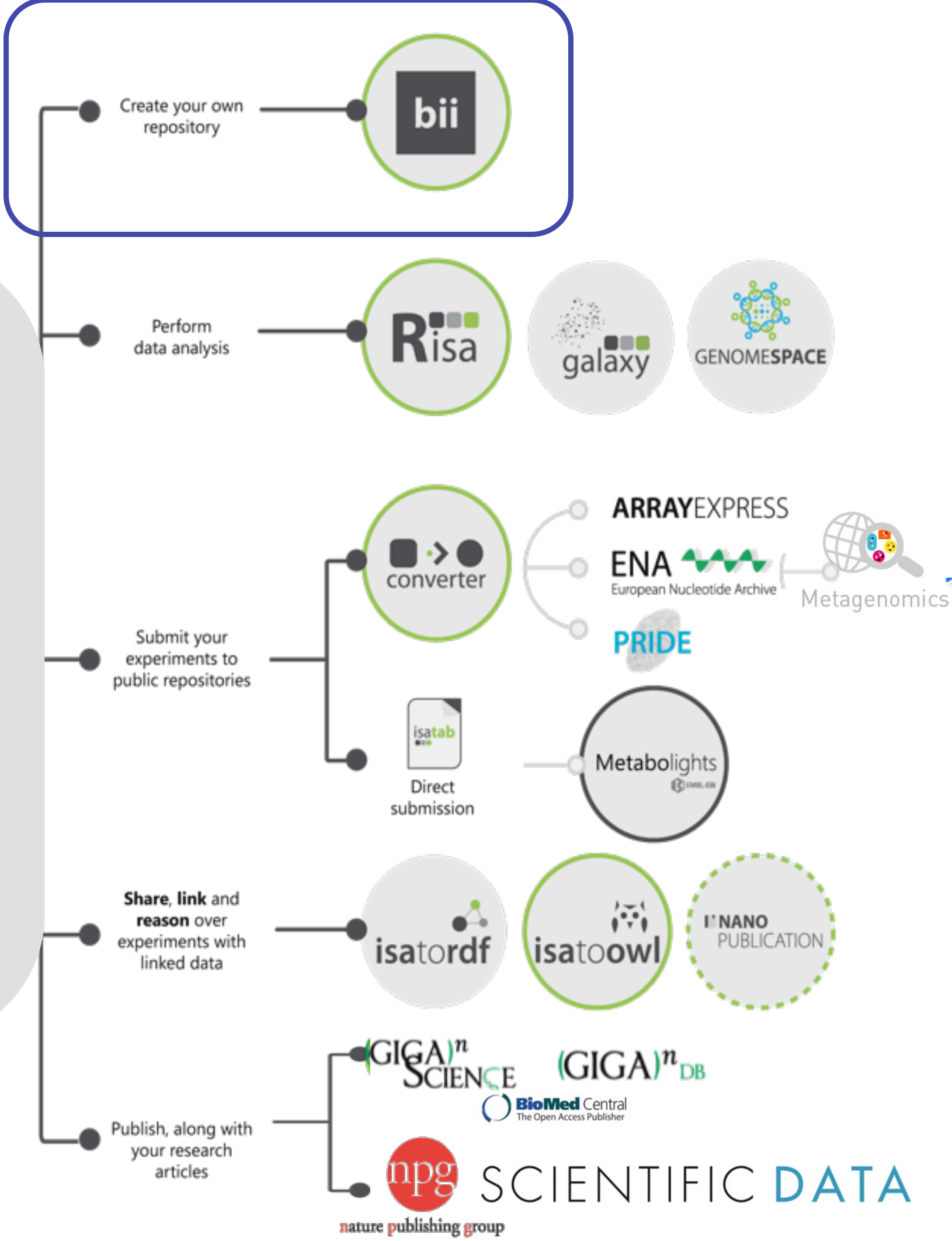
create your own repository to search and browse the experimental description and associated data making it close or open.

○ Core ISA tools

○ Core ISA tools in development

○ Powered by ISA tools

○ Externally Developed Tools



Create your own repository

bii

Perform data analysis

Risa

galaxy

GENOMESPACE

Submit your experiments to public repositories

converter

ARRAYEXPRESS

ENA European Nucleotide Archive

Metagenomics

PRIDE

isatab

Direct submission

Metabolights

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

isatordf

isatoowl

NANO PUBLICATION

Publish, along with your research articles

(GIGA)$^n$ SCIENCE

(GIGA)$^n$ DB

BioMed Central The Open Access Publisher

npg nature publishing group

SCIENTIFIC DATA

isatools

isatab format

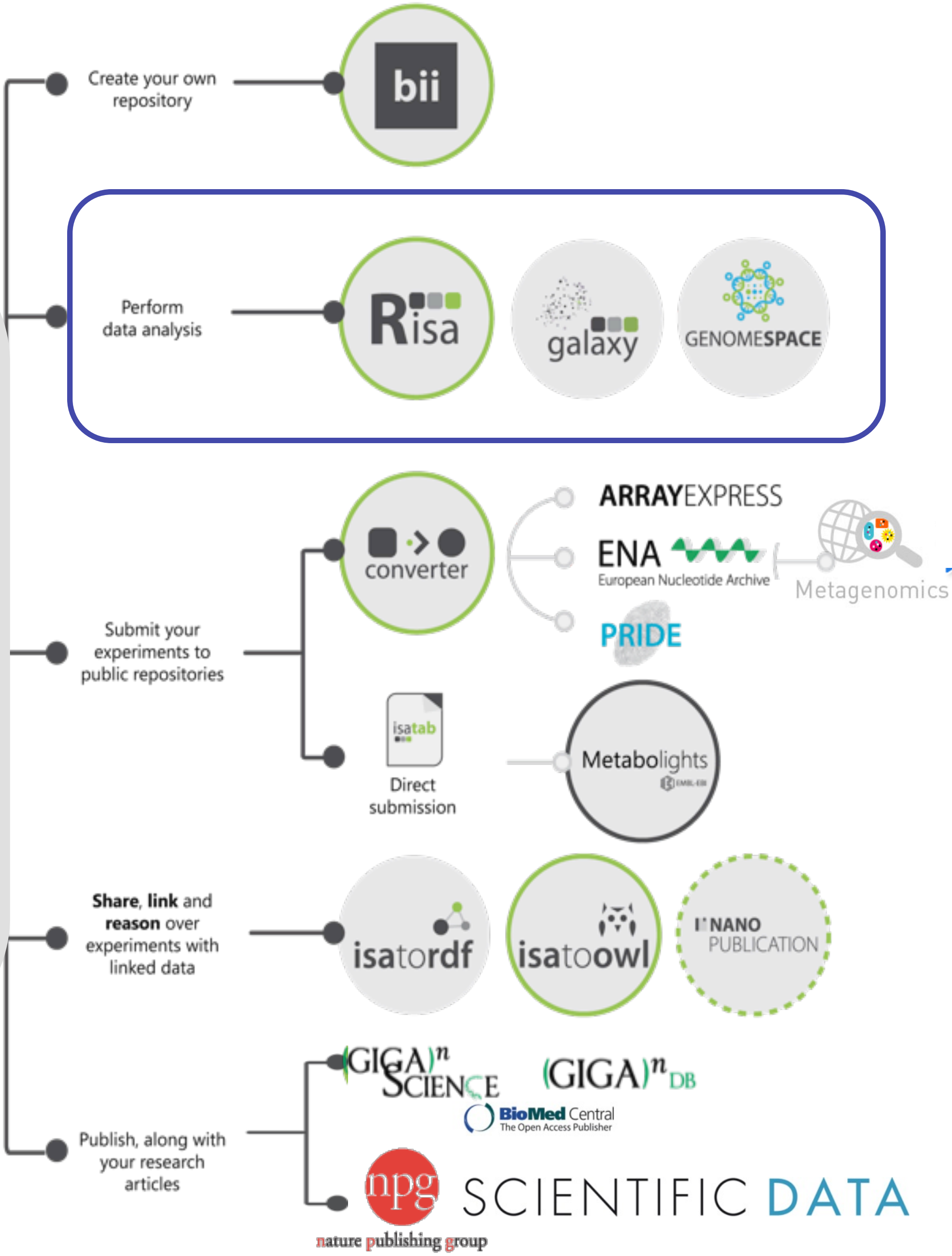
github SOCIAL CODING

GITHUB.COM/ISA-TOOLS

3

**analyse with existing tools**

upload experimental description and associated data to a growing number of well-known analysis systems, ISA connects with.


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

○ Core ISA tools ○ Core ISA tools in development ○ Powered by ISA tools ○ Externally Developed Tools

isatools

isatab format

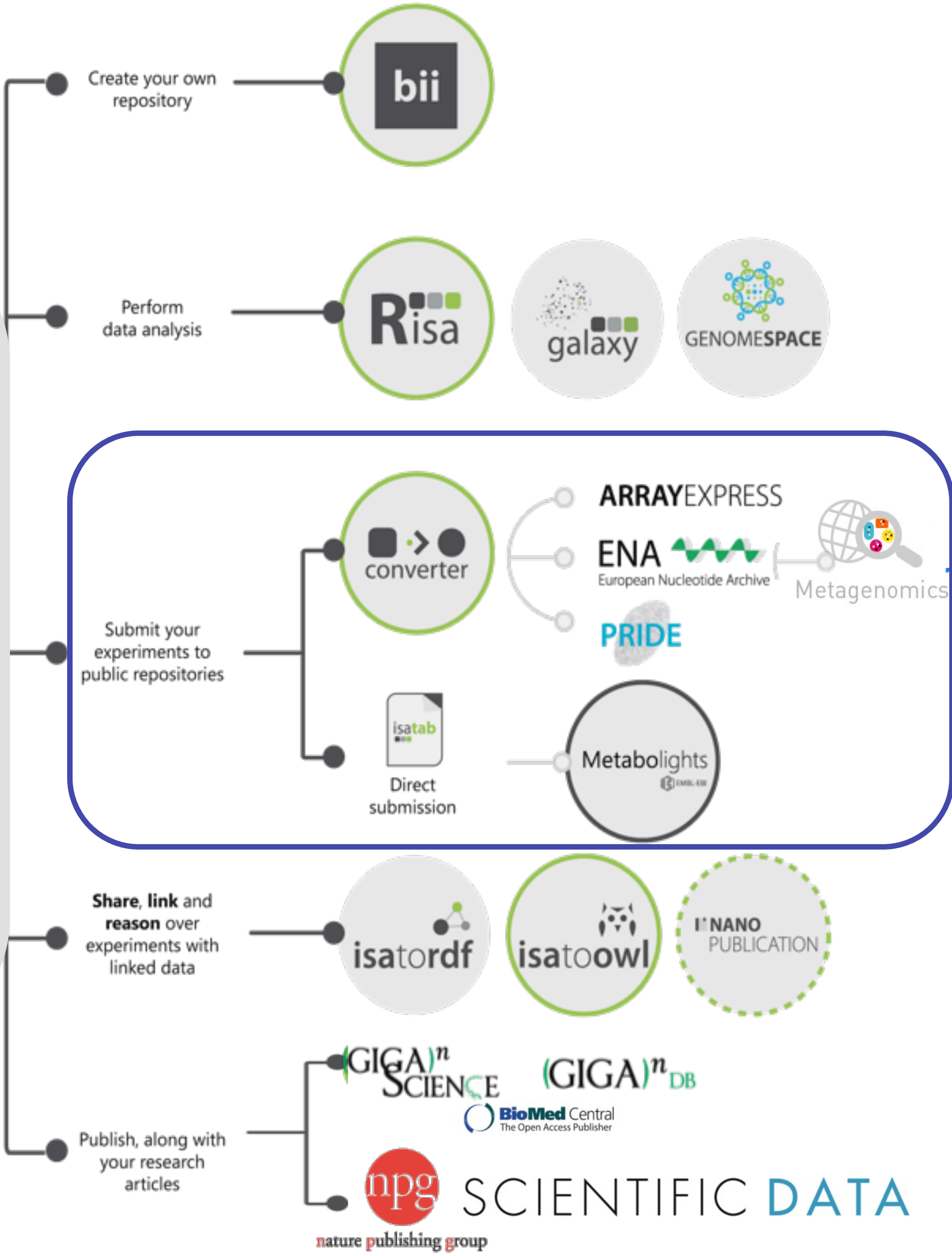
github SOCIAL CODING

GITHUB.COM/ISA-TOOLS

4

## submit to public repositories

when required, reformat the experiments for submission to supported public repositories or directly export to those already using ISA-Tab.



○ Core ISA tools ○ Core ISA tools in development ○ Powered by ISA tools ○ Externally Developed Tools

isatools

isatab format

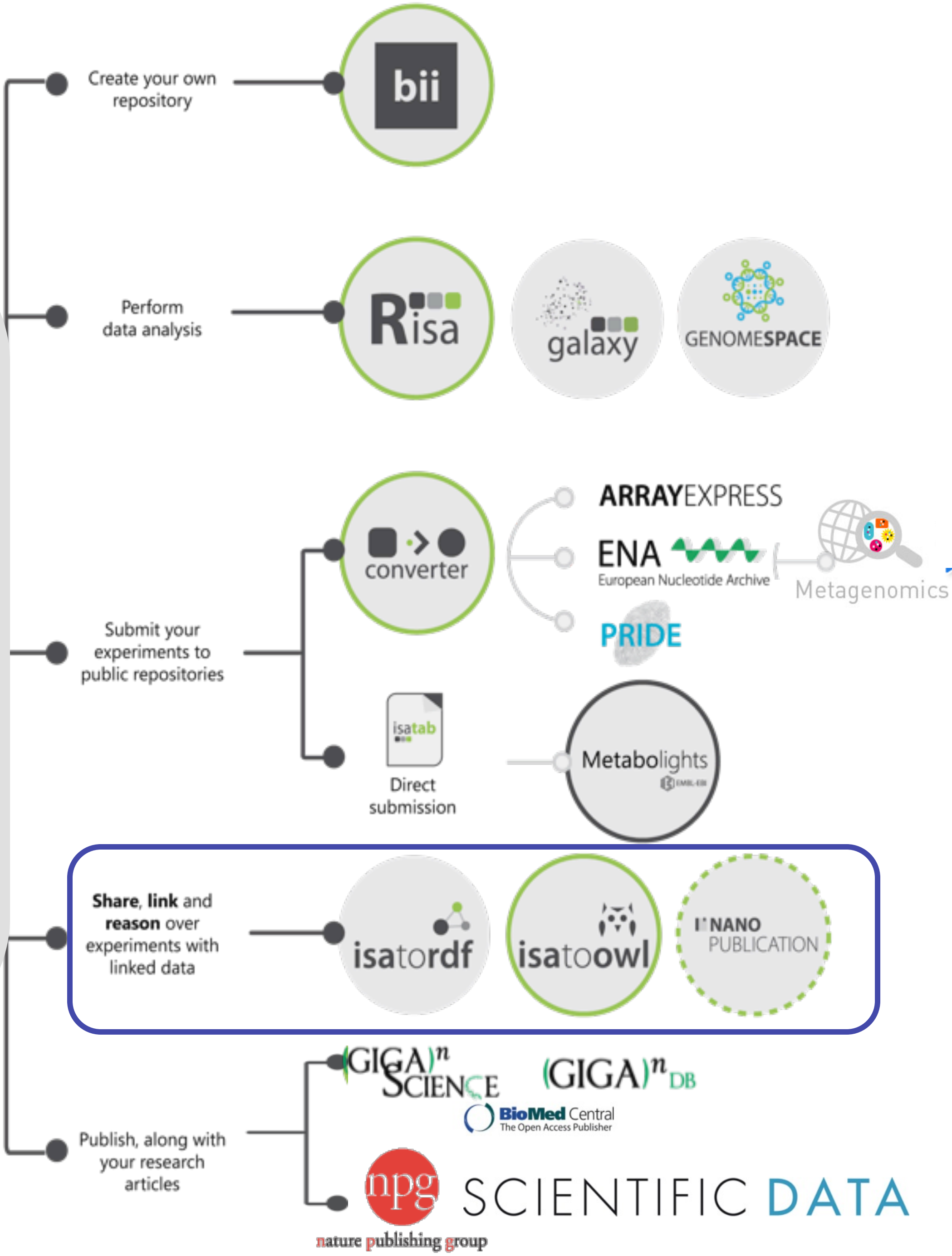
github SOCIAL CODING

GITHUB.COM/ISA-TOOLS

5

**release, reason and nanopublish**

explore how to reason over your experiments, open them to the linked data universe, or publish nano-statements of your discoveries.



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

○ Core ISA tools

○ Core ISA tools in development

○ Powered by ISA tools

● Externally Developed Tools

## *Community involvement and uptake*

1st ISA-Tab workshop

2nd ISA-Tab workshop

3rd ISA-Tab workshop

User workshops/visits - start

Other tools implement ISA-Tab

1st public instance: Harvard Stem Cell Discovery Engine - - - - - - - - - - →

Growing number of systems starts to adopt ISA framework - - - - - - - - →

## *Core developments*

Straw man ISA-Tab spec

Final ISA-Tab spec

ISA software v1

Database instance at EBI

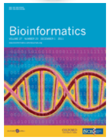
Conversions to Pride-XML/SRA-XML/MAGE-Tab - - - - - - -

Links to analysis tools starts - - - - - - - - →

RDF/OWL format starts - - - - - - - - →

2007 2008 2009 2010 2011 2012 2013

## *Publications*



**The ISA software suite: supporting standards-compliant curation at the community level**

Bioinformatics

**The ISA metadata tracking framework *in action* in the Harvard's Stem Cell Discovery Engine**

Nucleic Acids Research

**ISA chapter in : Open Source Software in Life Science Research**

Woodhead Publishing

**Cite the ISA Commons! Towards interoperable bioscience data.**

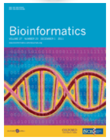
Nature Genetics

**MetaboLights: an open-access repository for metabolomics at EBI powered by ISA.**

Nucleic Acids Research

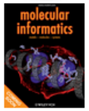
**Standardizing data, introducing ISA-Tab-Nano**

Nature Nanotechnology



**OntoMaton: a Bioportal powered ontology widget for Google Spreadsheets.**

Bioinformatics



**The ToxBank Data Warehouse: a research cluster of 7 EU FP7 Health systems toxicology and toxicogenomics projects.**

Molecular Informatics

# LIFE, NATURAL & BIOMEDICAL SCIENCES

## CURRENT PROJECTS

isatools

isacommons

### ISA – standards-compliant data management tools

Providing a tool kit to facilitate standards compliant collection, curation and local management of experiments in the life-sciences

find out more ❯

### Digital platforms for scholarly publishing

Collaborations with scientific, technical and medical publishers to develop novel data platforms to track and publish scholarly outputs

find out more ❯

- **Nature Publishing Group's Scientific Data**
- **BioMedCentral and BGI's GigaScience**
- **F1000 Research**
- **Oxford University Press**

### ELIXIR's UK Node

The UK Node contributes the country's substantial expertise in bioinformatics training for researchers, computer scientists and data managers in the Life, Natural and Medical Sciences.

find out more ❯

### Metagenomics Data Infrastructure

Coordinated by EMBL-European Bioinformatics Institute, the Metagenomics service is being developed to be an automated pipeline for the curation, archiving and analysis of metagenomic data.

find out more ❯

cosmos

### COSMOS - COordination Of Standards In MetabOlomicS

Coordinated by EMBL-European Bioinformatics Institute, COSMOS (Coordination of Standards in Metabolomics) has brought together European metabolomics data providers to set and promote community standards.

find out more ❯

### OBI - Ontology for BioMedical Investigation

stato
statistics ontology

biosharing

### BioSharing – standards, policies and databases

a web-based catalogue to centralize bioscience data policies, reporting standards and links to other related portals; and a communication forum to maintain linkages between funders, journals & standardization leaders

find out more ❯

Two more new funded projects!

ABOUT

- why we need ELIXIR
- how it works
- ELIXIR's leadership
- data integration
- what people are saying
- feedback
- join us

## About ELIXIR

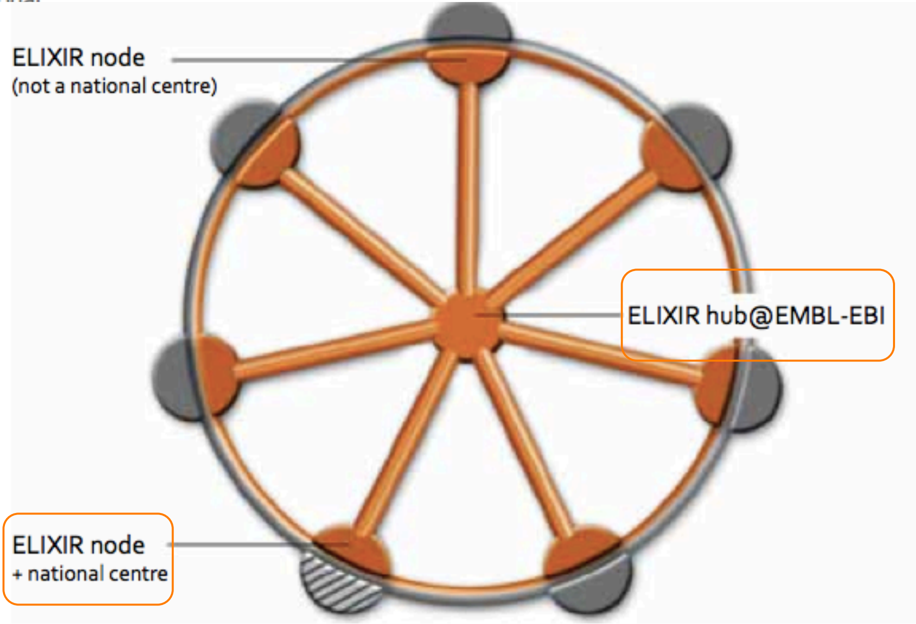
The purpose of ELIXIR is to construct and operate a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine and the environment, the bio-industries and society.

Because of new technologies such as next-generation DNA sequencing, data produced in biological experiments is doubling every few months, and this rate is increasing. In addition, new types of data are constantly emerging that need to be integrated meaningfully.

The collection, curation, storage, archiving, integration and deployment of biomolecular data is an immense challenge that cannot be handled by a single organisation or by one country alone, but requires international coordination.

elixir UNITED KINGDOM

Oxford e-Research Centre (Sansone SA):
Data Standards and Curation training



# LIFE, NATURAL & BIOMEDICAL SCIENCES

Roche
bbsrc biotechnology and biological sciences research council
MRC Medical Research Council
elixir UNITED KINGDOM
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME

## CURRENT PROJECTS

isatools
isacommons

### ISA – standards-compliant data management tools

Providing a tool kit to facilitate standards compliant collection, curation and local management of experiments in the life-sciences

find out more ❯

### Digital platforms for scholarly publishing

Collaborations with scientific, technical and medical publishers to develop novel data platforms to track and publish scholarly outputs

find out more ❯

- **Nature Publishing Group's Scientific Data**
- **BioMedCentral and BGI's GigaScience**
- **F1000 Research**
- **Oxford University Press**

elixir

### ELIXIR's UK Node

The UK Node contributes the country's substantial expertise in bioinformatics training for researchers, computer scientists and data managers in the Life, Natural and Medical Sciences.

find out more ❯

### Metagenomics Data Infrastructure

Coordinated by EMBL-European Bioinformatics Institute, the Metagenomics service is being developed to be an automated pipeline for the curation, archiving and analysis of metagenomic data.

find out more ❯

cosmos

### COSMOS - COordination Of Standards In MetabOlomicS

Coordinated by EMBL-European Bioinformatics Institute, COSMOS (Coordination of Standards in Metabolomics) has brought together European metabolomics data providers to set and promote community standards.

find out more ❯

### OBI - Ontology for BioMedical Investigation

stato
statistics ontology

biosharing

### BioSharing – standards, policies and databases

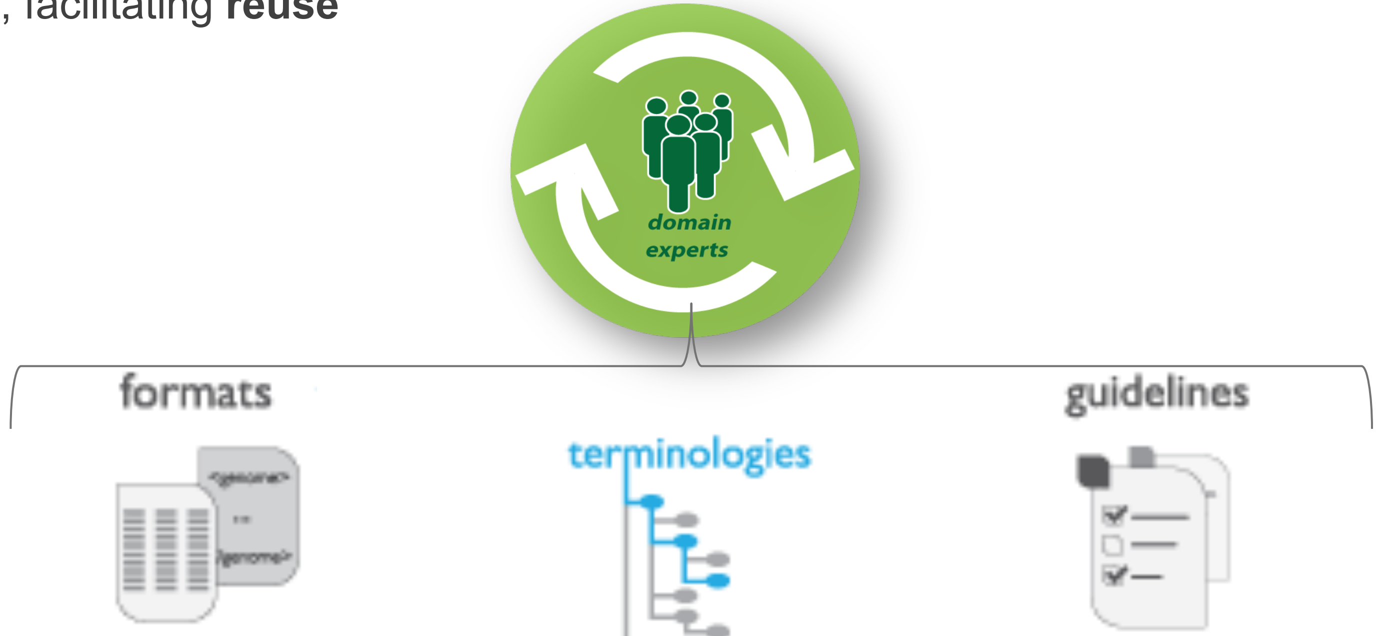
a web-based catalogue to centralize bioscience data policies, reporting standards and links to other related portals; and a communication forum to maintain linkages between funders, journals & standardization leaders

find out more ❯

Two more new funded projects!

# Types of reporting standards

**Researchers** and **bioinformaticians** in both *academic* and *commercial* arenas, along with **funding agencies** and **publishers**, embrace the concept that **community-developed**, **standards** are pivotal to ***structure***, ***enrich*** the description and ***share*** dataset, facilitating **reuse**

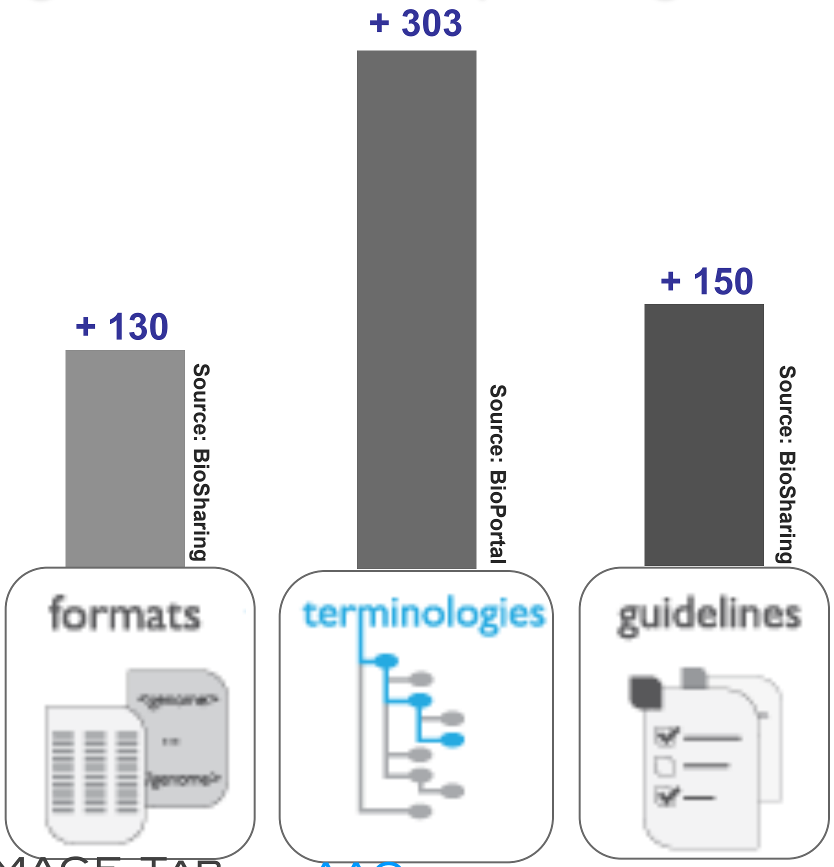


Including ***conceptual model***, ***conceptual schema*** from which an exchange format is derived to allow data to flow from one system to another

Including ***controlled vocabularies***, ***taxonomies***, ***thesauri***, ***ontologies*** etc. to use the same word and refer to the same 'thing'

Including ***minimum information reporting requirements***, or ***checklists*** to report the same core, essential information

# Growing number of reporting standards



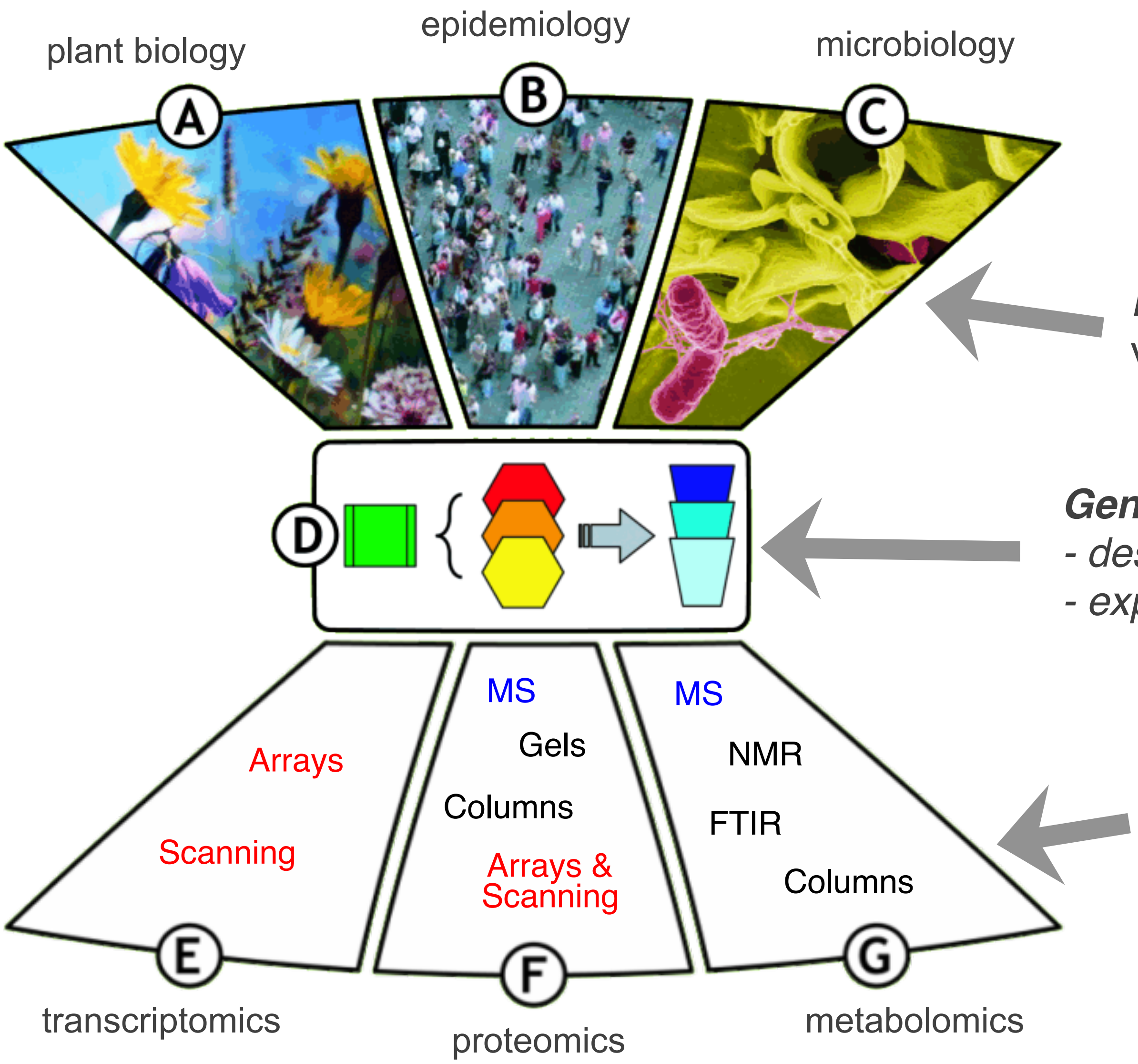
MAGE-Tab
GCDML
SRAxml
SOFT
FASTA
CML
DICOM
GELML
SBRML
MITAB
MzML
ISA-Tab SEDML…

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

MIAME
MIAPA
MIRIAM
MIX MIQAS
REMARK
MIGEN
MIQE
MIAPE
CIMR CONSORT
MIASE MISFISHIE….

Databases, annotation, curation tools implementing standards

# Fragmentation, duplications and gaps



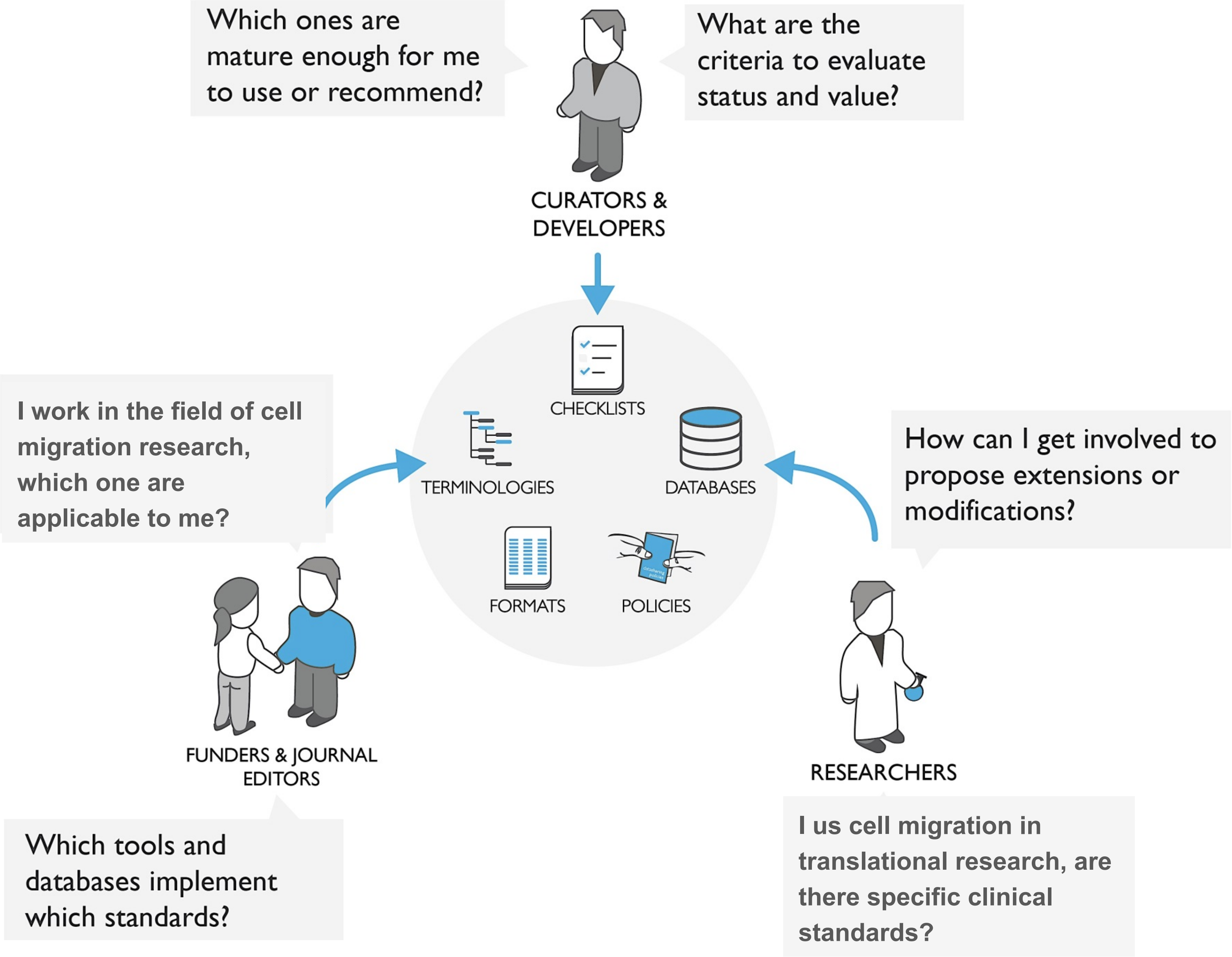
***Biologically***-delineated views of the world

***Generic*** features ( ‘common core’ )
- *description of source biomaterial*
- *experimental design components*

***Technologically***-delineated views of the world

To compare and integrate data we need interoperable standards

# How much do we know and which standards can we use?



Which ones are mature enough for me to use or recommend?

What are the criteria to evaluate status and value?

**I work in the field of cell migration research, which one are applicable to me?**

How can I get involved to propose extensions or modifications?

Which tools and databases implement which standards?

**I us cell migration in translational research, are there specific clinical standards?**

biosharing.org

## POLICIES

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

## STANDARDS

FORMATS | TERMINOLOGIES | CHECKLISTS

BioPortal
OBO

mibbi
equator network

A catalogue of reporting standards (minimum reporting guidelines, exchange formats and terminologies) and organizations that develop these.

BioPortal

## DATABASES

BioDBcore
ISB International Society for Biocuration

A catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them; compiled in collaboration with 2012 NAR Database.

re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES
OXFORD UNIVERSITY PRESS

### Journals

BioMed Central The Open Access Publisher
F1000 FACULTY of 1000
GIGA SCIENCE
nature genetics
PeerJ
npg SCIENTIFIC DATA
OXFORD UNIVERSITY PRESS

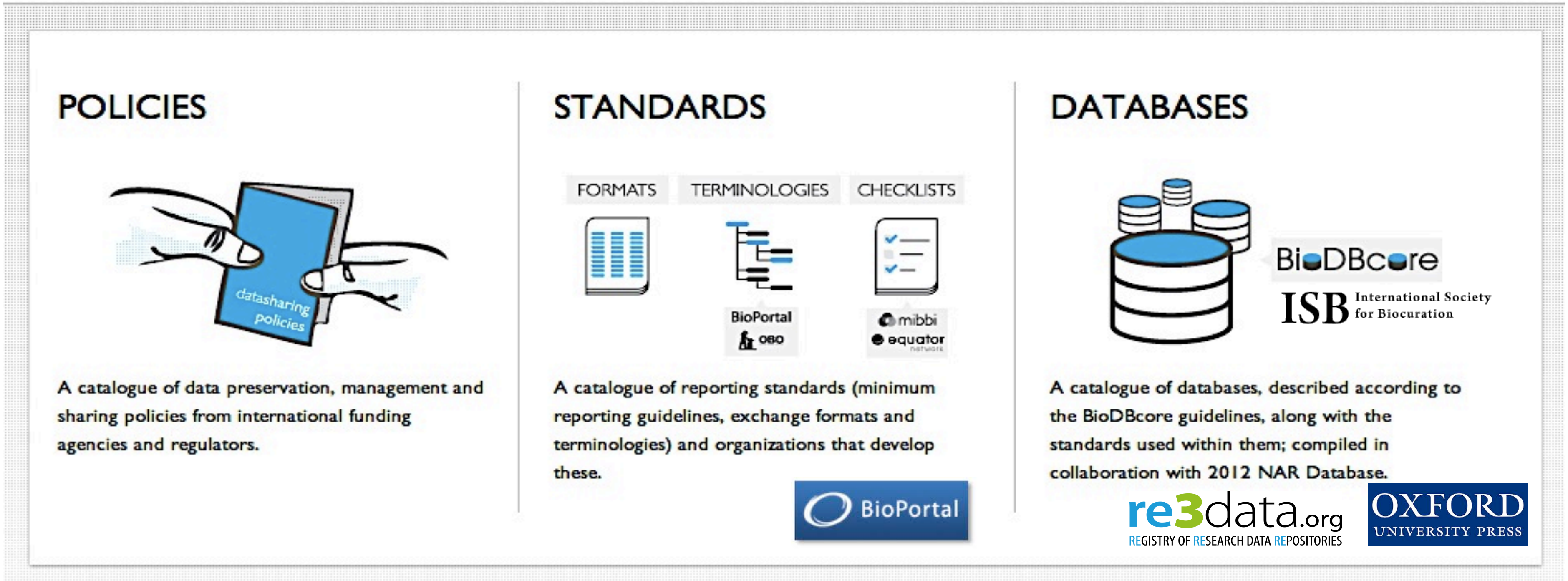
### Systems

isatools
DRYAD
RightField
SEEK for Science
OLS
DataCite
re3data.org
BioPortal
BioCatalogue
NITRC
Science EXCHANGE
NIF
BioGPS

### Groups

THE HUMAN VARIOME PROJECT
Pistoia Alliance
OBO
Sage
mibbi
combine
genomic observatories
UK PUBMED CENTRAL
BRITISH LIBRARY
DCC
ISB International Society for Biocuration
THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY
DIN
FuGE
science commons
Proteomics Standards Initiative
genomic STANDARDS consortium
equator network

biosharing.org



**POLICIES**

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

**STANDARDS**

A catalogue of reporting standards (minimum reporting guidelines, exchange formats and terminologies) and organizations that develop these.

**DATABASES**

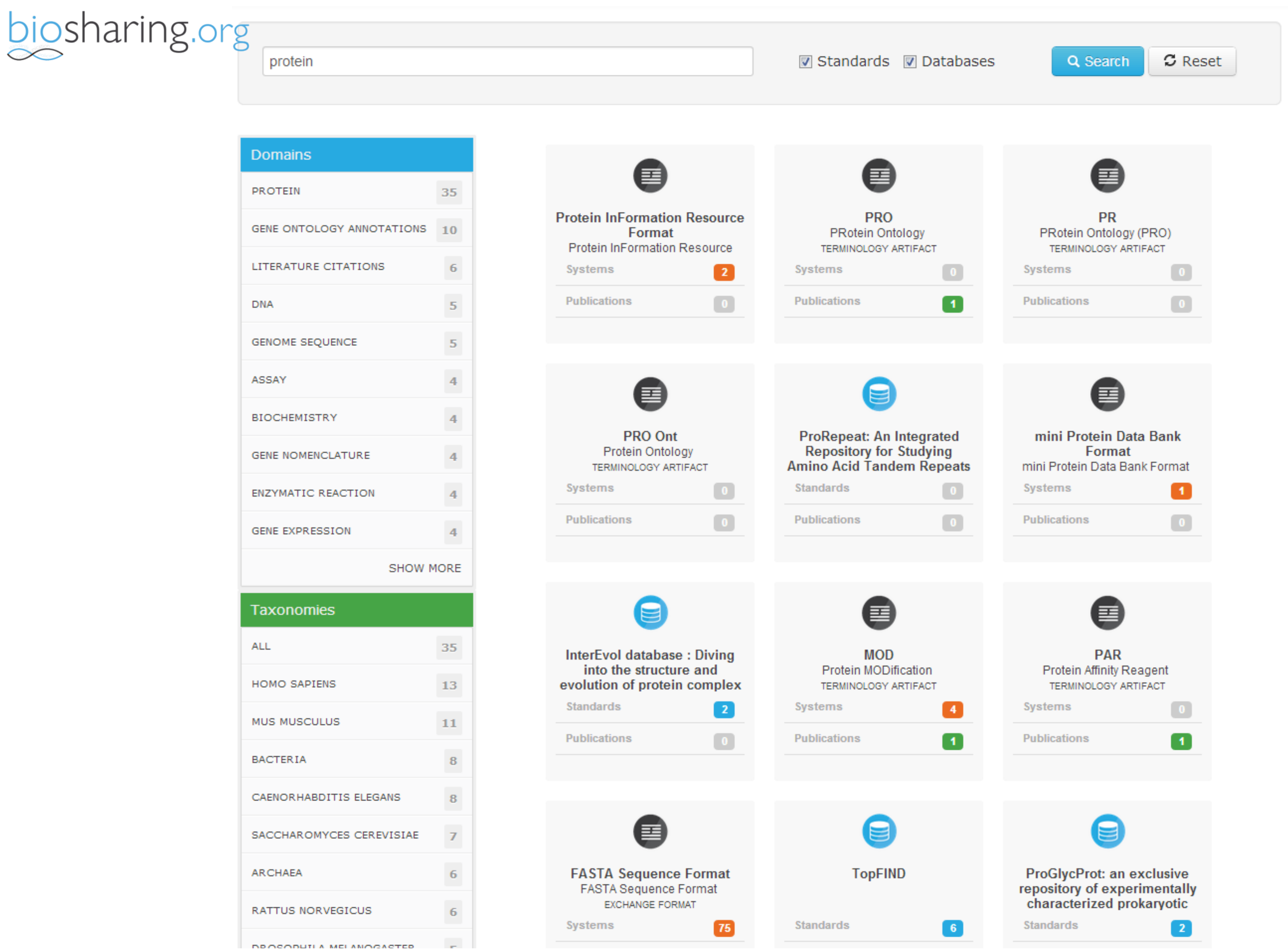
A catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them; compiled in collaboration with 2012 NAR Database.

Journals

BioMed Central The Open Access Publisher · F1000 FACULTY of 1000 · GIGA SCIENCE · nature genetics · PeerJ · npg SCIENTIFIC DATA · OXFORD UNIVERSITY PRESS

**Registering and cataloging is just step one; the next one are:**

- Develop assessment criteria for usability and popularity of standards
- Associate standards to data policies and databases
- Assemble journal and funder policies re data storage
- Make fully cross-searchable
- *Intended goal: help stakeholders make informed decisions*



biosharing.org

protein

☑ Standards ☑ Databases

Search

Reset

**Domains**

| | |
|---|---|
| PROTEIN | 35 |
| GENE ONTOLOGY ANNOTATIONS | 10 |
| LITERATURE CITATIONS | 6 |
| DNA | 5 |
| GENOME SEQUENCE | 5 |
| ASSAY | 4 |
| BIOCHEMISTRY | 4 |
| GENE NOMENCLATURE | 4 |
| ENZYMATIC REACTION | 4 |
| GENE EXPRESSION | 4 |

SHOW MORE

**Taxonomies**

| | |
|---|---|
| ALL | 35 |
| HOMO SAPIENS | 13 |
| MUS MUSCULUS | 11 |
| BACTERIA | 8 |
| CAENORHABDITIS ELEGANS | 8 |
| SACCHAROMYCES CEREVISIAE | 7 |
| ARCHAEA | 6 |
| RATTUS NORVEGICUS | 6 |
| DROSOPHILA MELANOGASTER | |

**Protein InFormation Resource Format**
Protein InFormation Resource
Systems 2
Publications 0

**PRO**
PRotein Ontology
TERMINOLOGY ARTIFACT
Systems 0
Publications 1

**PR**
PRotein Ontology (PRO)
TERMINOLOGY ARTIFACT
Systems 0
Publications 0

**PRO Ont**
Protein Ontology
TERMINOLOGY ARTIFACT
Systems 0
Publications 0

**ProRepeat: An Integrated Repository for Studying Amino Acid Tandem Repeats**
Standards 0
Publications 0

**mini Protein Data Bank Format**
mini Protein Data Bank Format
Systems 1
Publications 0

**InterEvol database : Diving into the structure and evolution of protein complex**
Standards 2
Publications 0

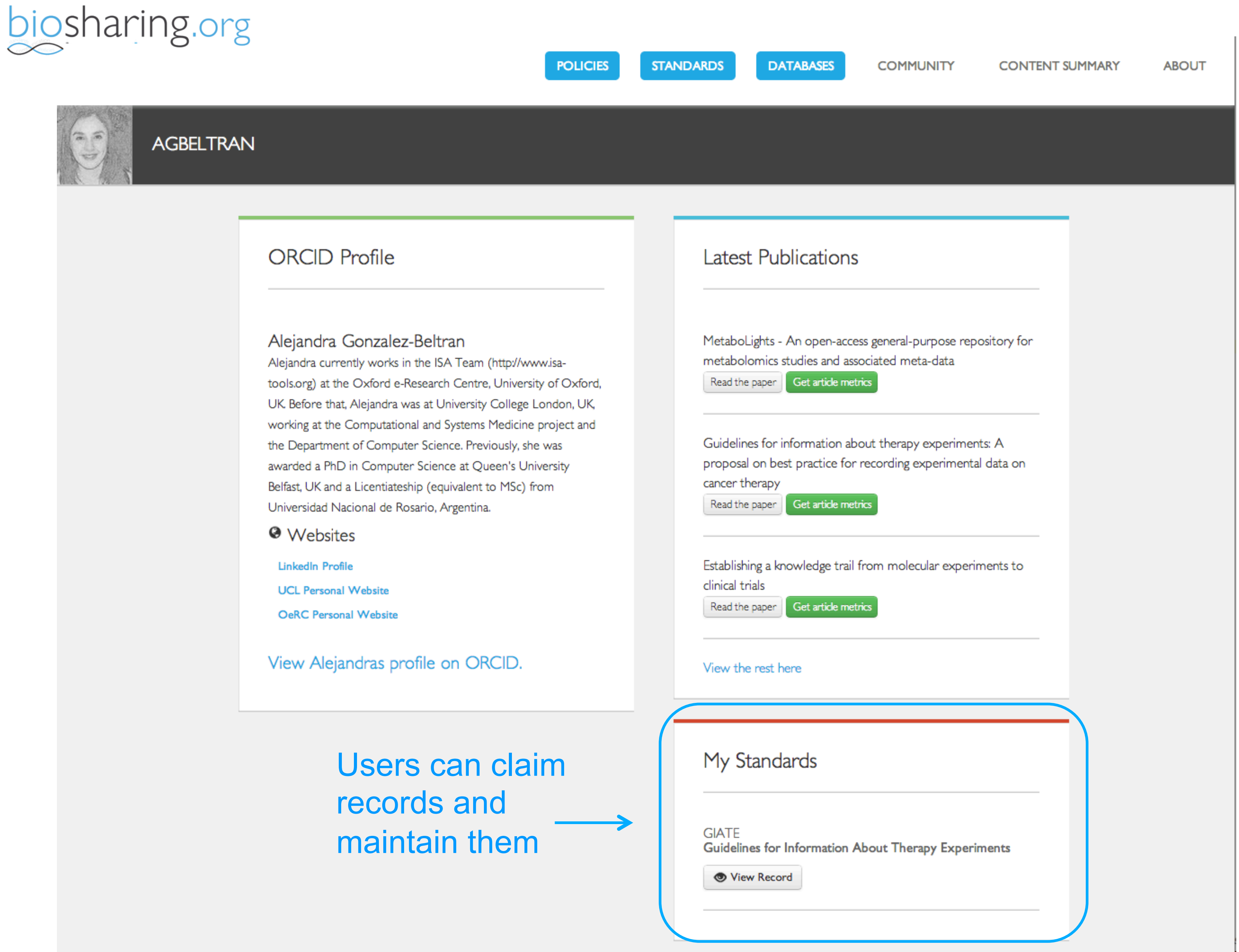
**MOD**
Protein MODification
TERMINOLOGY ARTIFACT
Systems 4
Publications 1

**PAR**
Protein Affinity Reagent
TERMINOLOGY ARTIFACT
Systems 0
Publications 1

**FASTA Sequence Format**
FASTA Sequence Format
EXCHANGE FORMAT
Systems 75

**TopFIND**
Standards 6

**ProGlycProt: an exclusive repository of experimentally characterized prokaryotic**
Standards 2



AGBELTRAN

## ORCID Profile

### Alejandra Gonzalez-Beltran

Alejandra currently works in the ISA Team (http://www.isa-tools.org) at the Oxford e-Research Centre, University of Oxford, UK. Before that, Alejandra was at University College London, UK, working at the Computational and Systems Medicine project and the Department of Computer Science. Previously, she was awarded a PhD in Computer Science at Queen's University Belfast, UK and a Licentiateship (equivalent to MSc) from Universidad Nacional de Rosario, Argentina.

### Websites

- LinkedIn Profile
- UCL Personal Website
- OeRC Personal Website

View Alejandras profile on ORCID.

## Latest Publications

MetaboLights - An open-access general-purpose repository for metabolomics studies and associated meta-data

Guidelines for information about therapy experiments: A proposal on best practice for recording experimental data on cancer therapy

Establishing a knowledge trail from molecular experiments to clinical trials

View the rest here

## My Standards

GIATE
**Guidelines for Information About Therapy Experiments**

Users can claim records and maintain them →

# LIFE, NATURAL & BIOMEDICAL SCIENCES

## CURRENT PROJECTS

isatools

isacommons

### ISA – standards-compliant data management tools

Providing a tool kit to facilitate standards compliant collection, curation and local management of experiments in the life-sciences

find out more

### Digital platforms for scholarly publishing

Collaborations with scientific, technical and medical publishers to develop novel data platforms to track and publish scholarly outputs

find out more

- **Nature Publishing Group's Scientific Data**
- **BioMedCentral and BGI's GigaScience**
- **F1000 Research**
- **Oxford University Press**

### ELIXIR's UK Node

The UK Node contributes the country's substantial expertise in bioinformatics training for researchers, computer scientists and data managers in the Life, Natural and Medical Sciences.

find out more

### Metagenomics Data Infrastructure

Coordinated by EMBL-European Bioinformatics Institute, the Metagenomics service is being developed to be an automated pipeline for the curation, archiving and analysis of metagenomic data.

find out more

### COSMOS - COordination Of Standards In MetabOlomicS

Coordinated by EMBL-European Bioinformatics Institute, COSMOS (Coordination of Standards in Metabolomics) has brought together European metabolomics data providers to set and promote community standards.

find out more

### OBI - Ontology for BioMedical Investigation

stato

statistics ontology

### BioSharing – standards, policies and databases

a web-based catalogue to centralize bioscience data policies, reporting standards and links to other related portals; and a communication forum to maintain linkages between funders, journals & standardization leaders

find out more

Two more new funded projects!

# Working with data publication platforms:

SCIENTIFIC DATA

Search Go

Home | About | For Authors | Advisory and Editorial Board | Open Access | Contacts | FAQ

Helping you publish, discover, and reuse research data

**Credit**
Credit, through a citable publication, for depositing & sharing your data

**Reuse**
Complete, curated & standardized descriptions enable the reuse of your data

**Quality**
Rigorous community based peer review

**Discovery**
Find datasets relevant to your research

**Open**
Promotes & endorses open science principles & available to all through a Creative Commons license

**Service**
In-house curation, rapid peer review & publication of your data descriptions

Sign up for Scientific Data e-alert

Scientific Data Updates

Introducing Scientific Data Updates
April 4, 2013

If not now then when – my view from within
April 3, 2013

[Press Release] NPG to launch Scientific Data to help scientists publish and reuse research data
April 3, 2013

Follow Us

npg nature publishing group

DRYAD

figshare credit for all your research

isatools COLLECT | CURATE | ANALYSE | SHARE | PUBLISH

F1000Research

Articles · For Authors · For Referees

(GIGA)$^n$DB

Promoting reproducible research

*GigaScience's* integrated "big-data" repository containing citable data available for public download and use

(GIGA)$^n$ SCIENCE

华大基因 BGI

BioMed Central
The Open Access Publisher

# LIFE, NATURAL & BIOMEDICAL SCIENCES

Roche
bbsrc
MRC Medical Research Council
elixir UNITED KINGDOM
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME

## CURRENT PROJECTS

isatools
isacommons

### ISA – standards-compliant data management tools

Providing a tool kit to facilitate standards compliant collection, curation and local management of experiments in the life-sciences

find out more ❯

### Digital platforms for scholarly publishing

Collaborations with scientific, technical and medical publishers to develop novel data platforms to track and publish scholarly outputs

find out more ❯

- **Nature Publishing Group's Scientific Data**
- **BioMedCentral and BGI's GigaScience**
- **F1000 Research**
- **Oxford University Press**

elixir

### ELIXIR's UK Node

The UK Node contributes the country's substantial expertise in bioinformatics training for researchers, computer scientists and data managers in the Life, Natural and Medical Sciences.

find out more ❯

### Metagenomics Data Infrastructure

Coordinated by EMBL-European Bioinformatics Institute, the Metagenomics service is being developed to be an automated pipeline for the curation, archiving and analysis of metagenomic data.

find out more ❯

cosmos

### COSMOS - COordination Of Standards In MetabOlomicS

Coordinated by EMBL-European Bioinformatics Institute, COSMOS (Coordination of Standards in Metabolomics) has brought together European metabolomics data providers to set and promote community standards.

find out more ❯

### OBI - Ontology for BioMedical Investigation

stato
statistics ontology

biosharing

### BioSharing – standards, policies and databases

a web-based catalogue to centralize bioscience data policies, reporting standards and links to other related portals; and a communication forum to maintain linkages between funders, journals & standardization leaders

find out more ❯

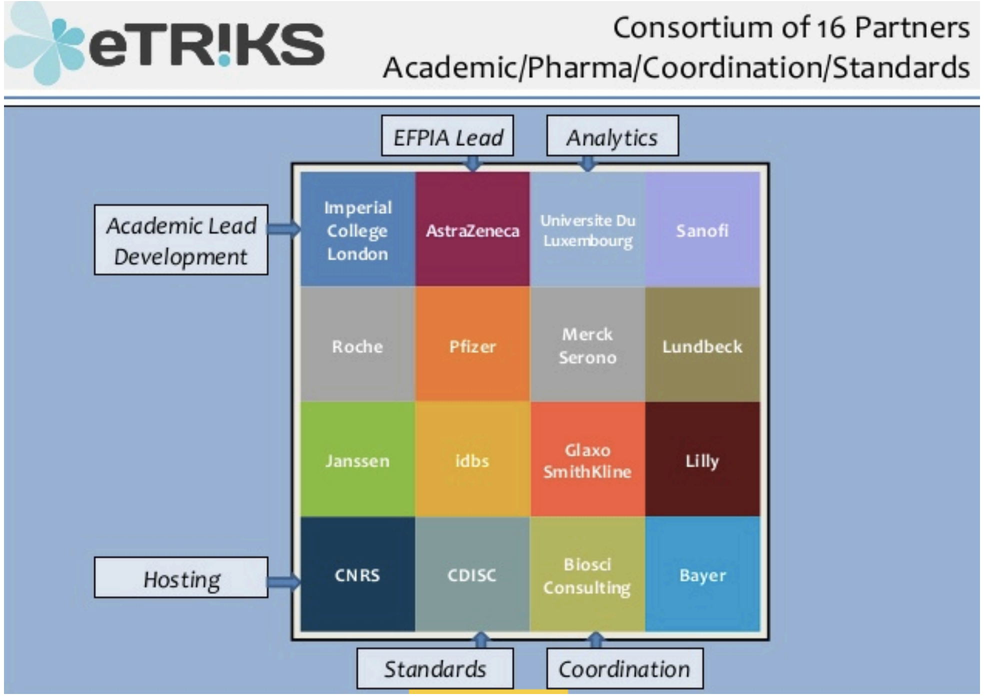
Two more new funded projects!

We are in the process of joining

eTRIKS
European Translational Information and Knowledge Management Services

efpia imi

Funded by Roche



eTRIKS

Consortium of 16 Partners
Academic/Pharma/Coordination/Standards

EFPIA Lead
Analytics
Academic Lead Development
Hosting
Standards
Coordination

| | | | |
|---|---|---|---|
| Imperial College London | AstraZeneca | Universite Du Luxembourg | Sanofi |
| Roche | Pfizer | Merck Serono | Lundbeck |
| Janssen | idbs | Glaxo SmithKline | Lilly |
| CNRS | CDISC | Biosci Consulting | Bayer |

We just received addition funds from

bbsrc
biotechnology and biological sciences research council

To co-develop an open
**Plant Science Bioinformatics platform**,
with
TGAC, Warwick, GARNet and EBI

and engage with

UNIVERSITY OF OXFORD
DEPARTMENT OF PLANT SCIENCES

# LIFE, NATURAL & BIOMEDICAL SCIENCES

**Driving digital research and providing services to the Life, Natural and Medical Sciences**

Roche

bbsrc
biotechnology and biological sciences research council

MRC Medical Research Council

elixir UNITED KINGDOM

NATURAL ENVIRONMENT RESEARCH COUNCIL

SEVENTH FRAMEWORK PROGRAMME

**Susanna-Assunta Sansone,** PhD (*PI*)

**Philippe Rocca-Serra,** PhD (*technical coordinator*)

**Alejandra Gonzalez-Beltran,** PhD (*senior developer*)

**Eamonn Maguire,** DPhil candidate (*senior developer*)

*and new team member - developer - to be recruited*

***Alumni***

**Annapaola Santarsiero**, MSc *(developer)*

**Pavlos Georgiou,** MSc student (*developer*)

# Journal publishing: the changing landscape

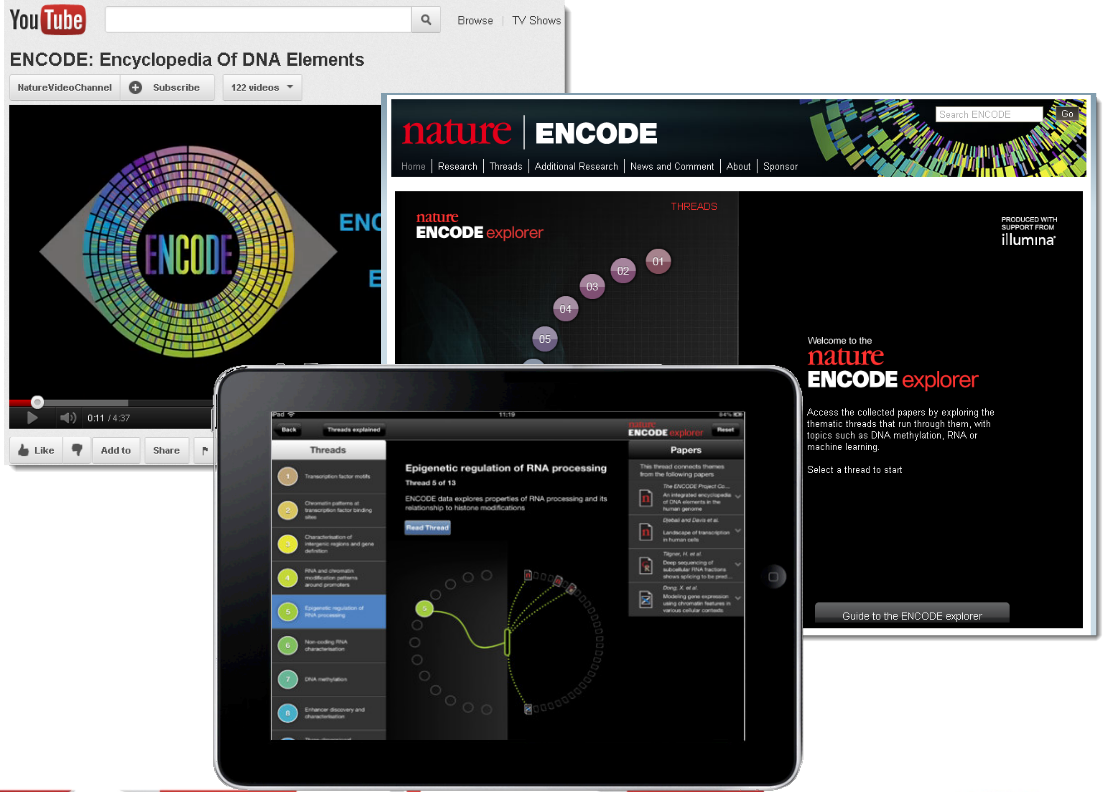
Human Genome 2001

62 Pages, 150 Authors,
49 Figure, 27 tables

# Journal publishing: the changing landscape

Encode Project 2012

30 papers, 3 Journals



ENCODE: Encyclopedia Of DNA Elements

Welcome to the nature ENCODE explorer

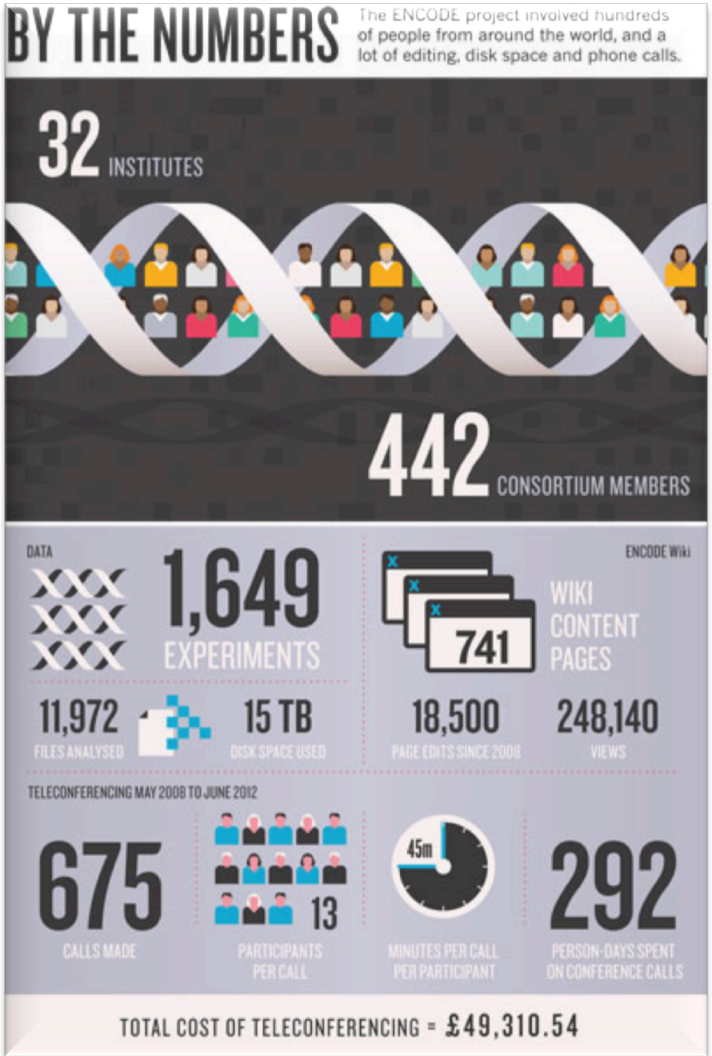
Access the collected papers by exploring the thematic threads that run through them, with topics such as DNA methylation, RNA or machine learning.

Select a thread to start

Epigenetic regulation of RNA processing

Thread 5 of 13

ENCODE data explores properties of RNA processing and its relationship to histone modifications



## BY THE NUMBERS

The ENCODE project involved hundreds of people from around the world, and a lot of editing, disk space and phone calls.

32 INSTITUTES

442 CONSORTIUM MEMBERS

DATA

1,649 EXPERIMENTS

11,972 FILES ANALYSED

15 TB DISK SPACE USED

ENCODE Wiki

741 WIKI CONTENT PAGES

18,500 PAGE EDITS SINCE 2008

248,140 VIEWS

TELECONFERENCING MAY 2008 TO JUNE 2012

675 CALLS MADE

13 PARTICIPANTS PER CALL

45m MINUTES PER CALL PER PARTICIPANT

292 PERSON-DAYS SPENT ON CONFERENCE CALLS

TOTAL COST OF TELECONFERENCING = £49,310.54

# SCIENTIFIC DATA

## Helping you publish, discover and reuse research data

**Honorary Academic Editor**
Susanna-Assunta Sansone, PhD
(University of Oxford)

**Managing Editor**
Andrew L Hufton, PhD

**Victoria Newman**
Editorial Curator

**Publisher**
Ruth Wilson

**Visit**
nature.com/scientificdata

**Email**
scientificdata@nature.com

**Tweet**
@ScientificData

Supported by: DRYAD figshare biosharing isatools

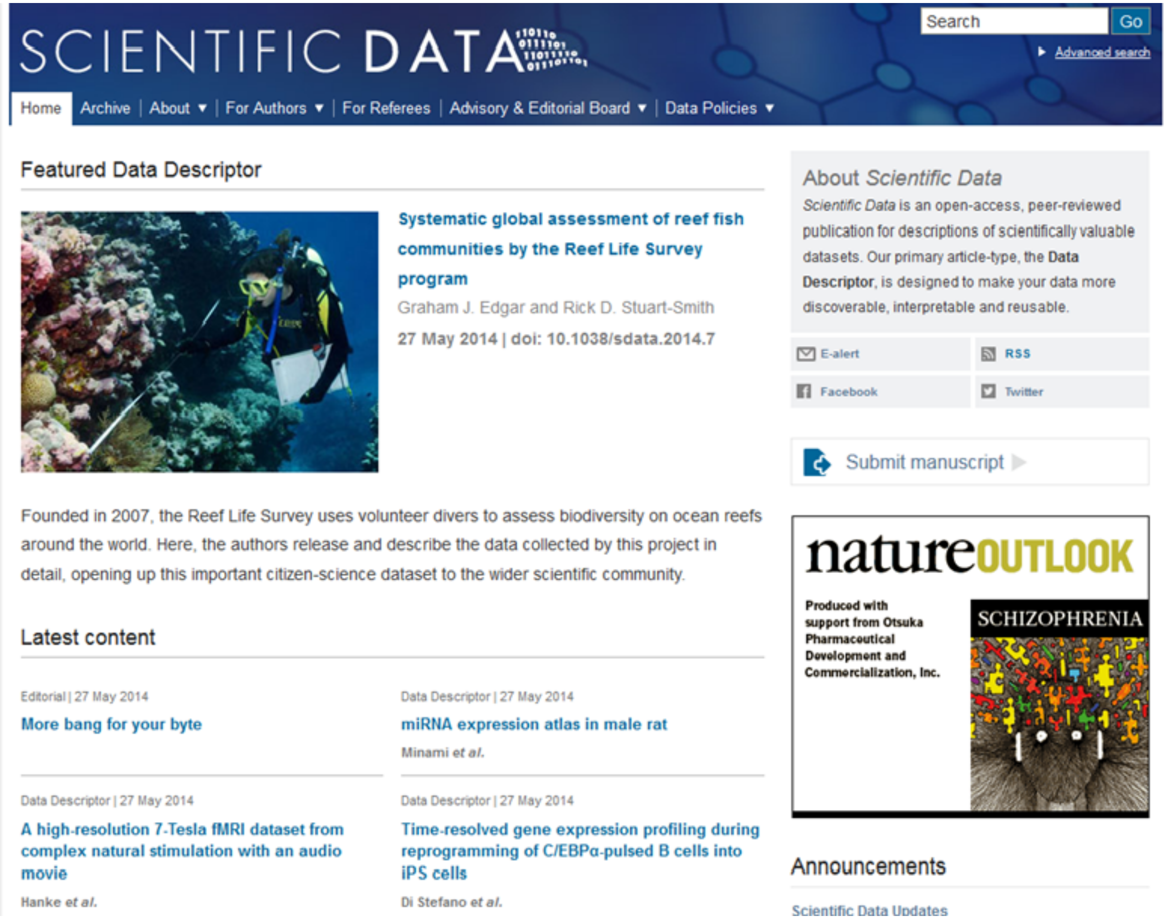
nature publishing group npg

- Introduction
  - concepts and principles
  - working with repositories
- Data Descriptor
  - structured experimental meta data
- Launch content
  - examples

# SCIENTIFIC DATA

## Helping you publish, discover and reuse research data

## NOW LIVE!!



SCIENTIFIC DATA

Search Go

Advanced search

Home | Archive | About ▾ | For Authors ▾ | For Referees | Advisory & Editorial Board ▾ | Data Policies ▾

**Featured Data Descriptor**

**Systematic global assessment of reef fish communities by the Reef Life Survey program**

Graham J. Edgar and Rick D. Stuart-Smith

27 May 2014 | doi: 10.1038/sdata.2014.7

Founded in 2007, the Reef Life Survey uses volunteer divers to assess biodiversity on ocean reefs around the world. Here, the authors release and describe the data collected by this project in detail, opening up this important citizen-science dataset to the wider scientific community.

**Latest content**

Editorial | 27 May 2014
**More bang for your byte**

Data Descriptor | 27 May 2014
**miRNA expression atlas in male rat**
Minami *et al.*

Data Descriptor | 27 May 2014
**A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie**
Hanke *et al.*

Data Descriptor | 27 May 2014
**Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells**
Di Stefano *et al.*

**About *Scientific Data***

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

E-alert | RSS | Facebook | Twitter

Submit manuscript ▶

natureOUTLOOK

Produced with support from Otsuka Pharmaceutical Development and Commercialization, Inc.

SCHIZOPHRENIA

**Announcements**

Scientific Data Updates

- Credit for sharing your data
- Focused on reuse and reproducibility
- Peer reviewed, curated
- Promoting Community Data Repositories
- Open Access

nature publishing group npg

# What does *Scientific Data* offer?

- Get Credit for Sharing Your Data
- Publications will be listed in the major indexes and will be citeable

- Focused on Data Reuse
- All the information others need to reuse the data; no interpretative analysis or hypothesis testing

- Open-access
- Authors select from three Creative Commons licences for the main
- Data Descriptor. Each publication supported by curated CC0 metadata

- Peer-reviewed
- Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards

- Promoting Community Data Repositories
- Data stored in community data repositories

# What does *Scientific Data* offer?

- Get Credit for Sharing Your Data
- Publications will be listed in the major indexes and will be citeable

- Focused on Data Reuse
- All the information others need to reuse the data; no interpretative analysis or hypothesis testing

- Open-access
- Authors select from three Creative Commons licences for the main
- Data Descriptor. Each publication supported by curated CC0 metadata

- Peer-reviewed
- Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards

- Promoting Community Data Repositories
- Data stored in community data repositories

# What does *Scientific Data* offer?

- Get Credit for Sharing Your Data
- Publications will be listed in the major indexes and will be citeable

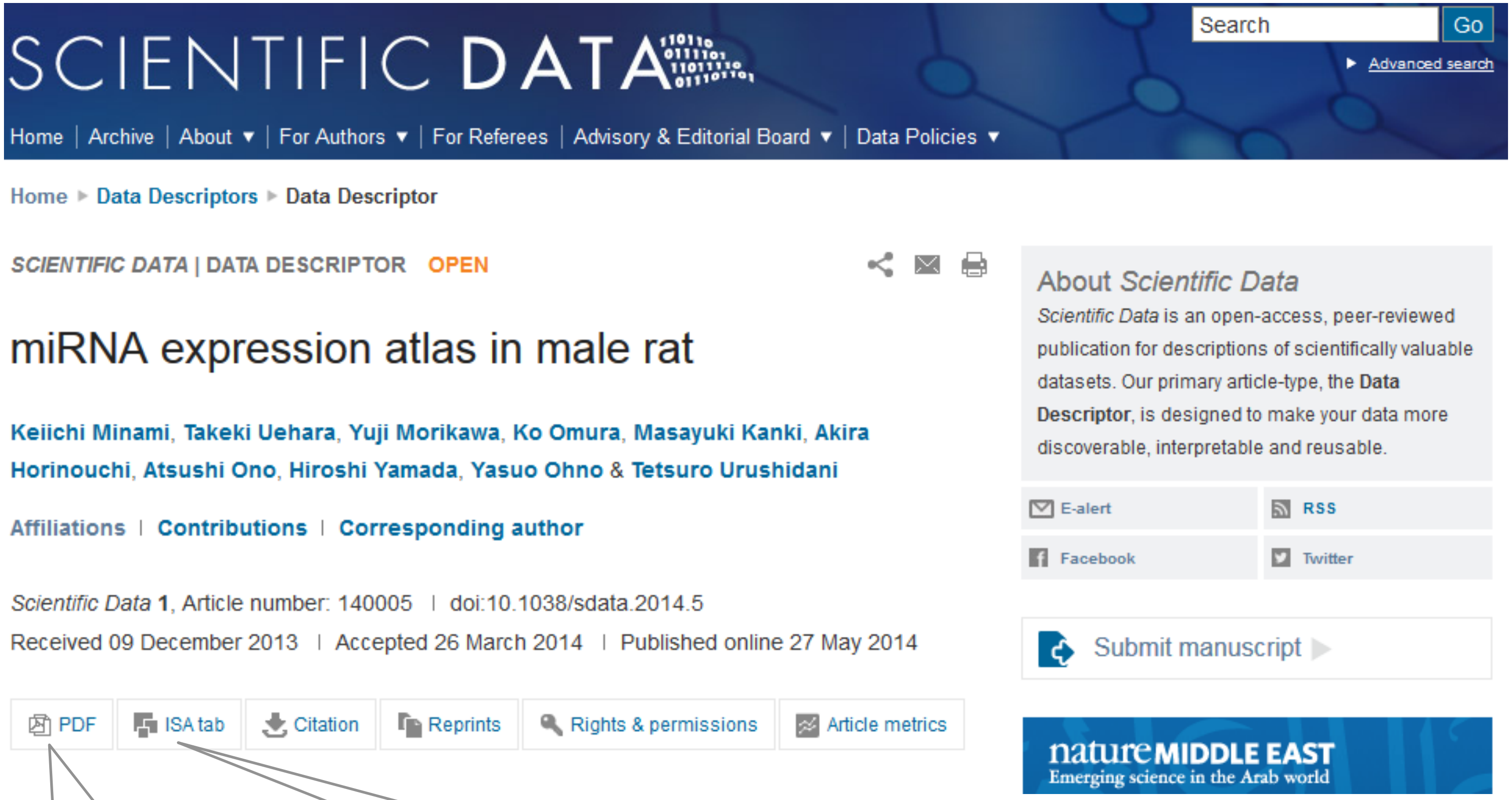
- Focused on Data Reuse
- All the information others need to reuse the data; no interpretative analysis or hypothesis testing

- Open-access
- Authors select from three Creative Commons licences for the main
- Data Descriptor. Each publication supported by curated CC0 metadata

- Peer-reviewed
- Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards

- Promoting Community Data Repositories
- Data stored in community data repositories

# Data Descriptor: narrative and structure



*SCIENTIFIC DATA* | DATA DESCRIPTOR OPEN

## miRNA expression atlas in male rat

**Keiichi Minami, Takeki Uehara, Yuji Morikawa, Ko Omura, Masayuki Kanki, Akira Horinouchi, Atsushi Ono, Hiroshi Yamada, Yasuo Ohno & Tetsuro Urushidani**

**Affiliations | Contributions | Corresponding author**

*Scientific Data* **1**, Article number: 140005 | doi:10.1038/sdata.2014.5

Received 09 December 2013 | Accepted 26 March 2014 | Published online 27 May 2014

About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

**Article or *narrative* component**

(PDF and HTML)

**Experimental metadata or *structured* component**

(in-house curated, machine-readable formats)

# Data Descriptor: narrative

**Focus on data reuse**

Detailed descriptions of the methods and technical analyses supporting the quality of the measurements.

Does not contain tests of new scientific hypotheses

Sections:

- **Title**
- **Abstract**
- **Background & Summary**
- **Methods**
- **Technical Validation**
- **Data Records**
- **Usage Notes**
- **Figures & Tables**
- **References**
- Data Citations

In traditional publications this information is not provided in a sufficiently detailed manner

However this information is essential for understanding, reusing, and reproducing datasets

# Data Descriptor: narrative

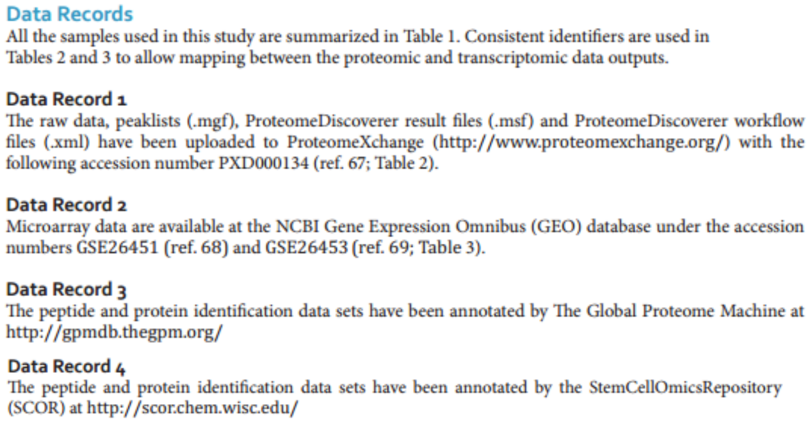
**Focus on data reuse**

Detailed descriptions of the methods and technical analyses supporting the quality of the measurements.

Does not contain tests of new scientific hypotheses

Sections:

- **Title**
- **Abstract**
- **Background & Summary**
- **Methods**
- **Technical Validation**
- **Data Records**
- **Usage Notes**
- **Figures & Tables**
- **References**
- **Data Citations**



**Data Records**

All the samples used in this study are summarized in Table 1. Consistent identifiers are used in Tables 2 and 3 to allow mapping between the proteomic and transcriptomic data outputs.

**Data Record 1**

The raw data, peaklists (.mgf), ProteomeDiscoverer result files (.msf) and ProteomeDiscoverer workflow files (.xml) have been uploaded to ProteomeXchange (http://www.proteomexchange.org/) with the following accession number PXD000134 (ref. 67; Table 2).

**Data Record 2**

Microarray data are available at the NCBI Gene Expression Omnibus (GEO) database under the accession numbers GSE26451 (ref. 68) and GSE26453 (ref. 69; Table 3).

**Data Record 3**

The peptide and protein identification data sets have been annotated by The Global Proteome Machine at http://gpmdb.thegpm.org/

**Data Record 4**

The peptide and protein identification data sets have been annotated by the StemCellOmicsRepository (SCOR) at http://scor.chem.wisc.edu/

# Data Descriptor: narrative

**Focus on data reuse**

Detailed descriptions of the methods and technical analyses supporting the quality of the measurements.

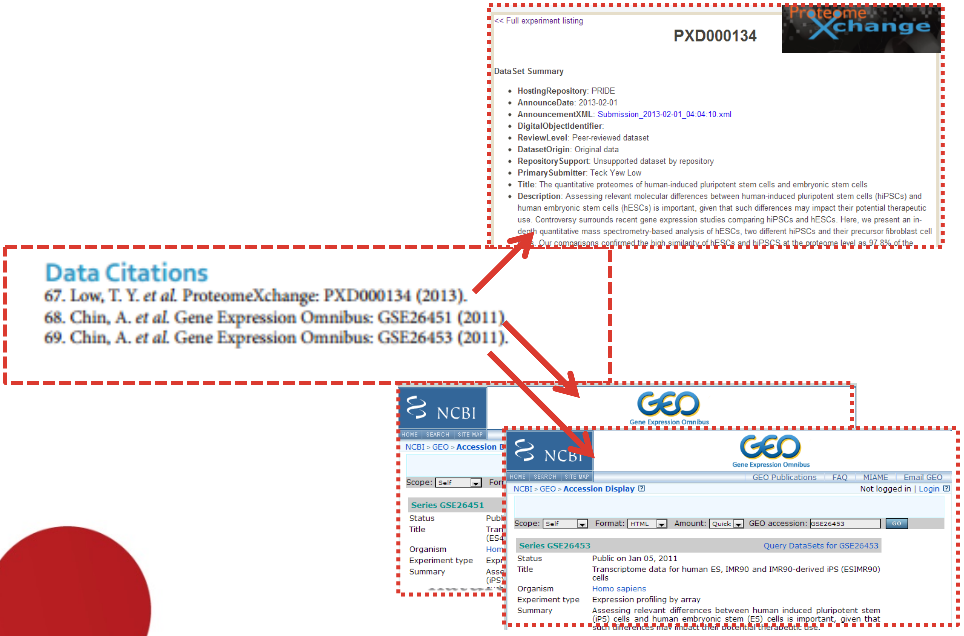
Does not contain tests of new scientific hypotheses

Sections:

- **Title**
- **Abstract**
- **Background & Summary**
- **Methods**
- **Technical Validation**
- **Data Records**
- **Usage Notes**
- **Figures & Tables**
- **References**
- **Data Citations**

***Joint Declaration of Data Citation Principles*** by the Data Citation Synthesis Group, incl.:

- CODATA
- Research Data Alliance (RDA),
- Force11



PXD000134

DataSet Summary

- **HostingRepository**: PRIDE
- **AnnounceDate**: 2013-02-01
- **AnnouncementXML**: Submission_2013-02-01_04:04:10.xml
- **DigitalObjectIdentifier**:
- **ReviewLevel**: Peer-reviewed dataset
- **DatasetOrigin**: Original data
- **RepositorySupport**: Unsupported dataset by repository
- **PrimarySubmitter**: Teck Yew Low
- **Title**: The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells
- **Description**: Assessing relevant molecular differences between human-induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) is important, given that such differences may impact their potential therapeutic use. Controversy surrounds recent gene expression studies comparing hiPSCs and hESCs. Here, we present an in-depth quantitative mass spectrometry-based analysis of hESCs, two different hiPSCs and their precursor fibroblast cell lines. Our comparisons confirmed the high similarity of hESCs and hiPSCS at the proteome level as 97.8% of the

**Data Citations**

67. Low, T. Y. *et al.* ProteomeXchange: PXD000134 (2013).
68. Chin, A. *et al.* Gene Expression Omnibus: GSE26451 (2011).
69. Chin, A. *et al.* Gene Expression Omnibus: GSE26453 (2011).

NCBI > GEO > Accession D

Scope: Self For

Series GSE26451

Status Publ

Title Tran (ES4

Organism Hom

Experiment type Expr

Summary Asse (iPS)

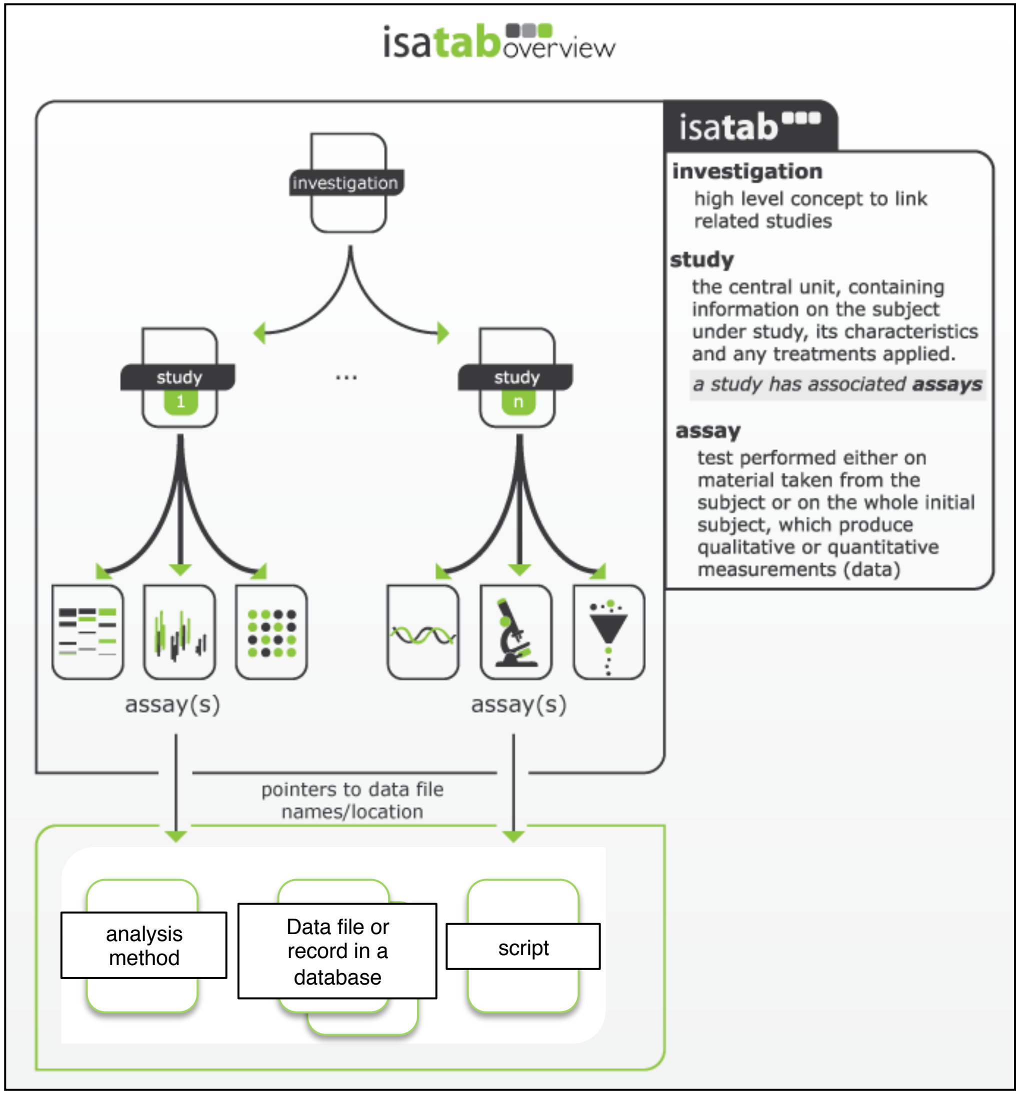
NCBI > GEO > Accession Display

Scope: Self Format: HTML Amount: Quick GEO accession: GSE26453

Series GSE26453 Query DataSets for GSE26453

| | |
|---|---|
| Status | Public on Jan 05, 2011 |
| Title | Transcriptome data for human ES, IMR90 and IMR90-derived iPS (ESIMR90) cells |
| Organism | Homo sapiens |
| Experiment type | Expression profiling by array |
| Summary | Assessing relevant differences between human induced pluripotent stem (iPS) cells and human embryonic stem (ES) cells is important, given that such differences may impact their potential therapeutic use. |

# Data Descriptor: structure (CC0)



**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)

**In-house curation team:**

- assists users to submit the structured content via simple templates and an internal authoring tool
- performs value-added semantic annotation of the experimental metadata

For advanced users/service providers willing to export ISA-Tab for direct submission, we will release a technical specification:


nature publishing group npg

SCIENTIFIC DATA

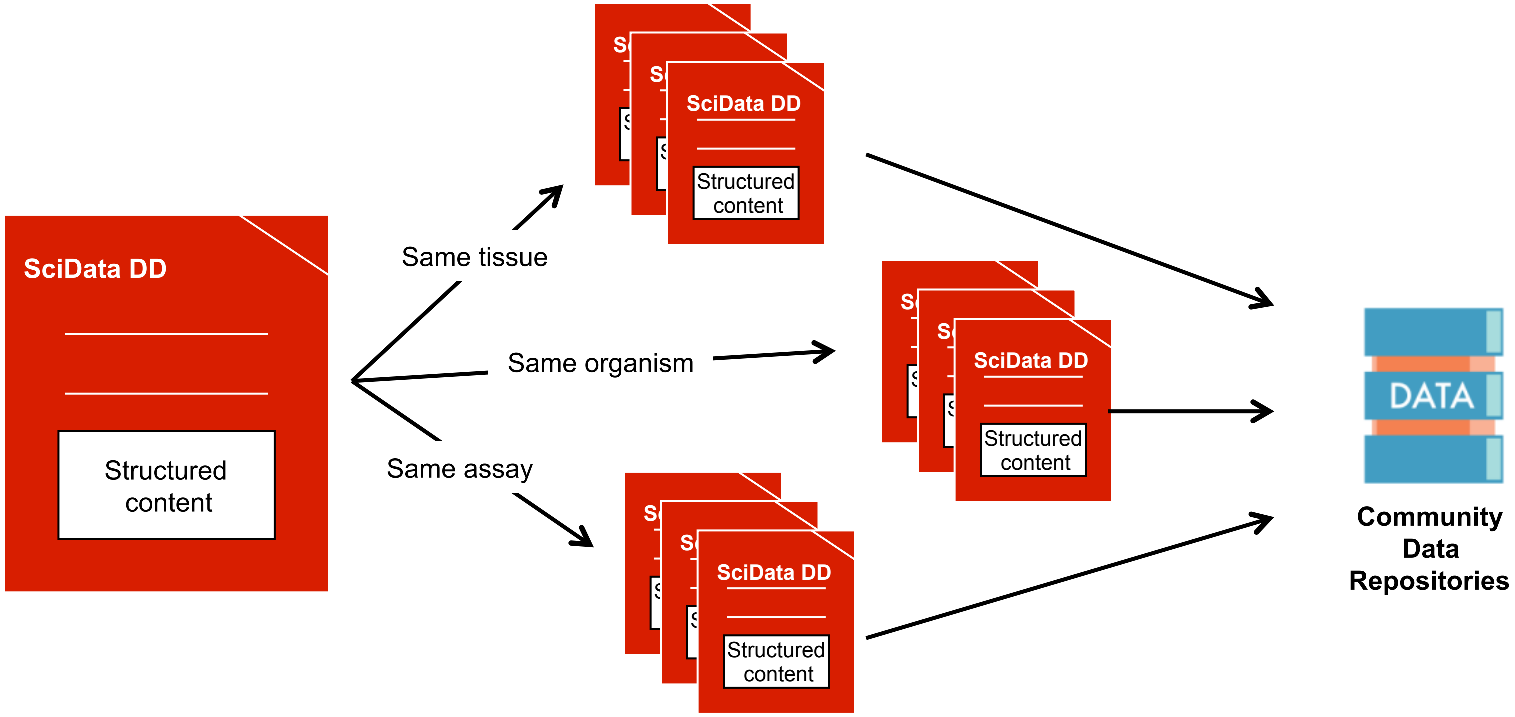
**ISA-Tab configuration specification for the experimental metadata component of a Data Descriptor**

**Status**

Production version 1$^{12}$.

# Discover similar datasets

Structured content allows users to link, with one click, to other datasets studying the same tissue, disease, organism, or using the same experimental platform



# What does *Scientific Data* offer?

- Get Credit for Sharing Your Data
- Publications will be listed in the major indexes and will be citeable

- Focused on Data Reuse
- All the information others need to reuse the data; no interpretative analysis or hypothesis testing

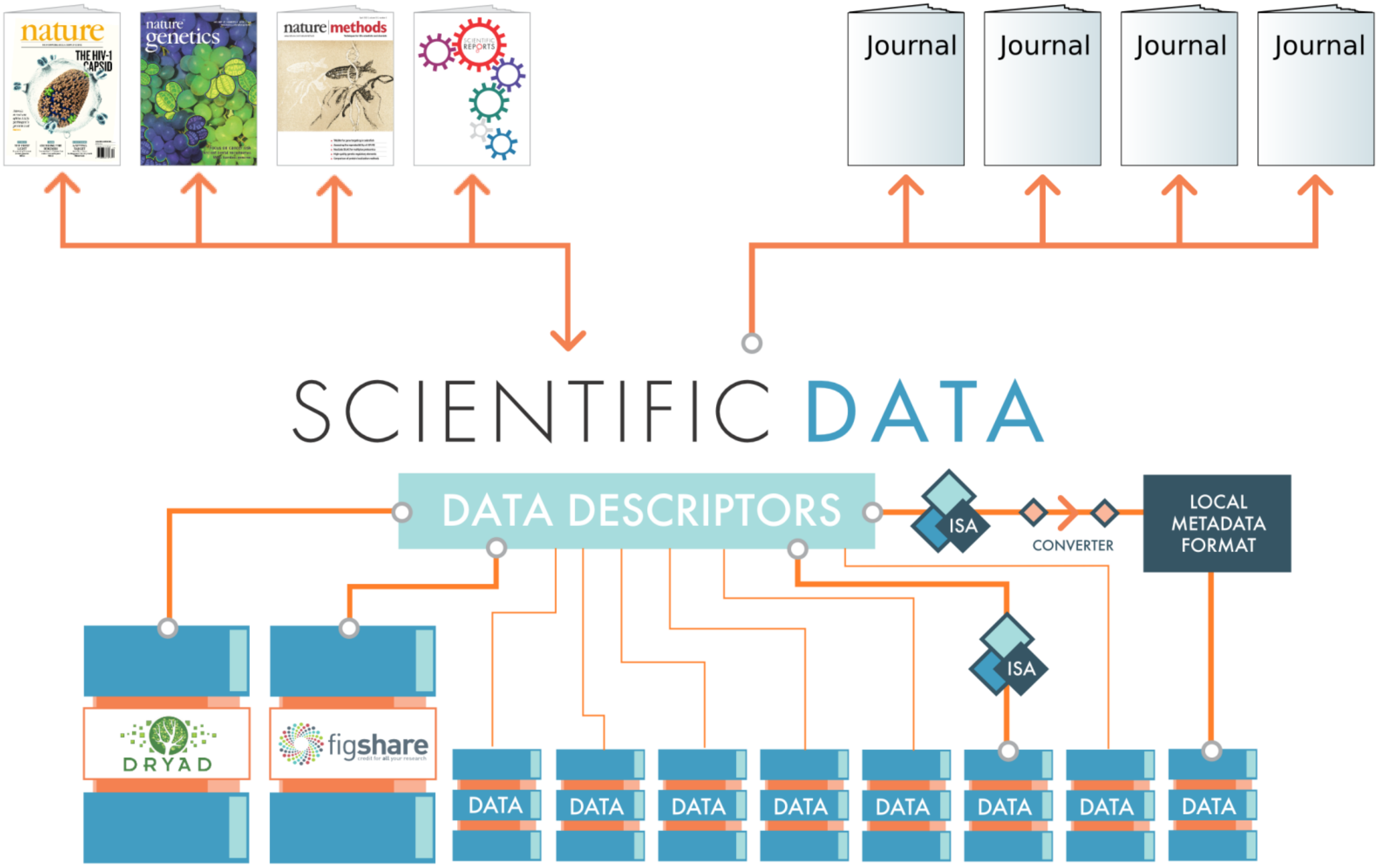
- Open-access
- Authors select from three Creative Commons licences for the main
- Data Descriptor. Each publication supported by curated CC0 metadata

- Peer-reviewed
- Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards

- Promoting Community Data Repositories
- Data stored in community data repositories

# Making data discoverable

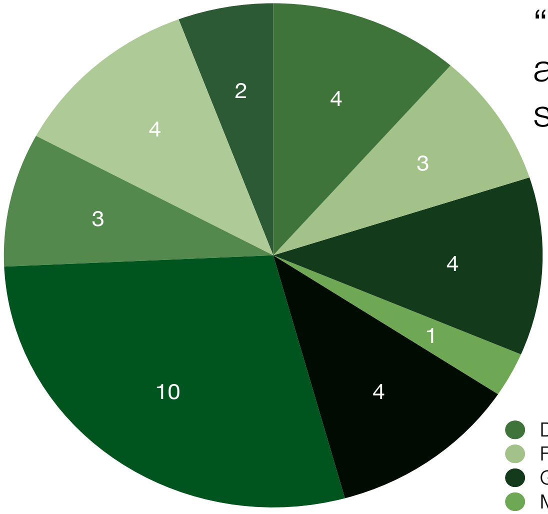
Linking between **research papers**, **Data Descriptors**, and **data records**



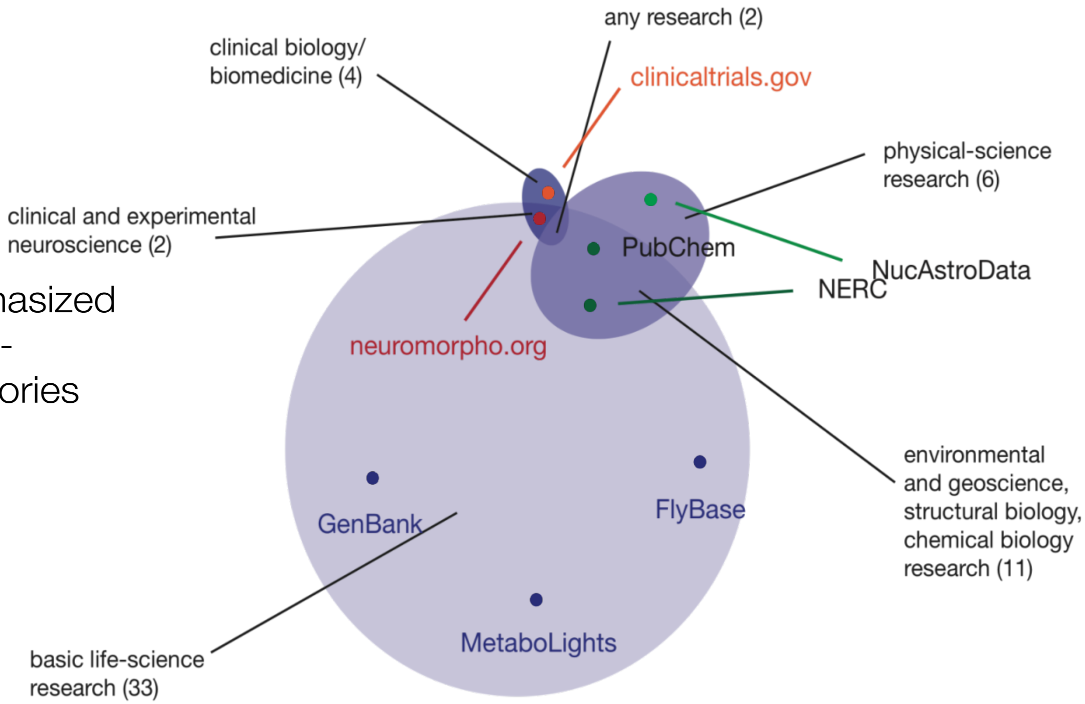
# Helping authors find the right place for the data

- We currently recognize over 50 public data repositories, and provide advice on the best place for authors to archive their data
- We have integrated systems with both:

DRYAD figshare credit for all your research



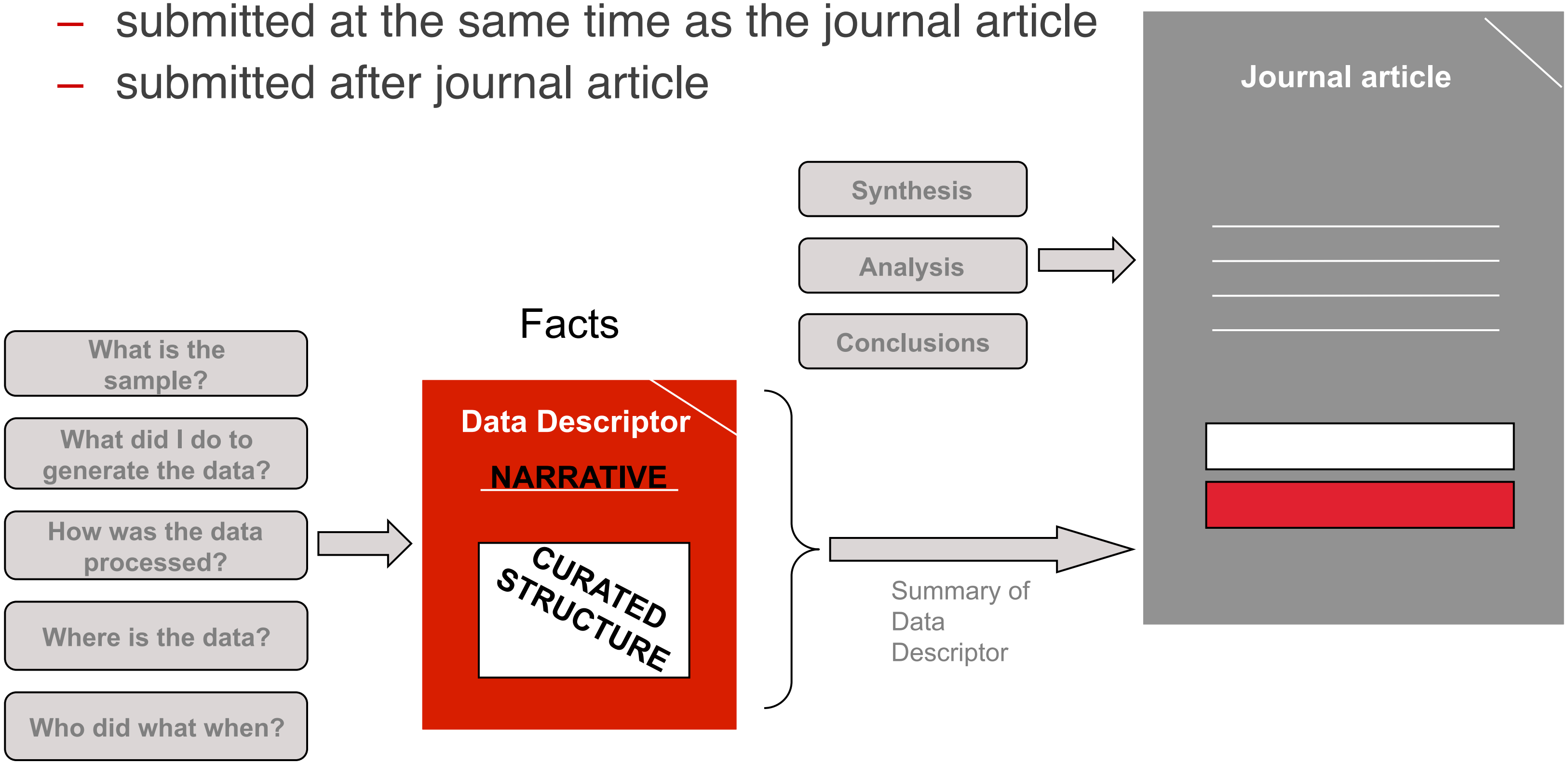
"Omics" is emphasized among basic life-sciences repositories



# Data Descriptor: relation with traditional articles

- The data descriptor is only concerned with the facts behind the methodology of data generation/collection and processing
- A data descriptor can be:
  - submitted prior to journal article
  - submitted at the same time as the journal article
  - submitted after journal article

Interpretation



Journal article

Synthesis

Analysis

Conclusions

Facts

What is the sample?

What did I do to generate the data?

How was the data processed?

Where is the data?

Who did what when?

**Data Descriptor**

**NARRATIVE**

**CURATED STRUCTURE**

Summary of Data Descriptor

# Data Descriptor: relation with traditional articles

**Complementary**

Methods and technical analyses supporting the quality of the measurements.

Do not contain tests of new scientific hypotheses

SCIENTIFIC DATA

Search Go

Advanced search

Home | Archive | About ▾ | For Authors ▾ | For Referees | Advisory & Editorial Board ▾ | Data Policies ▾

Home ▸ Data Descriptors ▸ Data Descriptor

*SCIENTIFIC DATA* | DATA DESCRIPTOR OPEN

## Systematic global assessment of reef fish communities by the Reef Life Survey program

Graham J Edgar & Rick D Stuart-Smith

Affiliations | Contributions | Corresponding author

*Scientific Data* **1**, Article number: 140007 | doi:10.1038/sdata.2014.7

Received 14 February 2014 | Accepted 15 April 2014 | Published online 27 May 2014

PDF | ISA tab | Citation | Reprints | Rights & permissions | Article metrics

About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

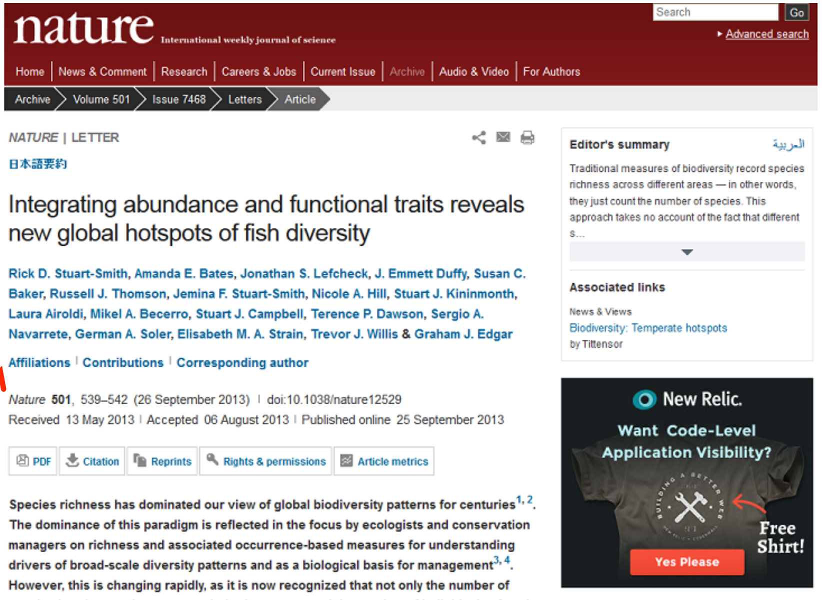
E-alert | RSS | Facebook | Twitter

**Associated Links**

*Nature* | Article

Integrating abundance and functional traits reveals new global hotspots of fish diversity

by Rick D. Stuart-Smith *et al*



nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive › Volume 501 › Issue 7468 › Letters › Article

*NATURE* | LETTER

日本語要約

## Integrating abundance and functional traits reveals new global hotspots of fish diversity

Rick D. Stuart-Smith, Amanda E. Bates, Jonathan S. Lefcheck, J. Emmett Duffy, Susan C. Baker, Russell J. Thomson, Jemina F. Stuart-Smith, Nicole A. Hill, Stuart J. Kininmonth, Laura Airoldi, Mikel A. Becerro, Stuart J. Campbell, Terence P. Dawson, Sergio A. Navarrete, German A. Soler, Elisabeth M. A. Strain, Trevor J. Willis & Graham J. Edgar

Affiliations | Contributions | Corresponding author

*Nature* **501**, 539–542 (26 September 2013) | doi:10.1038/nature12529

Received 13 May 2013 | Accepted 06 August 2013 | Published online 25 September 2013

PDF | Citation | Reprints | Rights & permissions | Article metrics

**Species richness has dominated our view of global biodiversity patterns for centuries[1, 2]. The dominance of this paradigm is reflected in the focus by ecologists and conservation managers on richness and associated occurrence-based measures for understanding drivers of broad-scale diversity patterns and as a biological basis for management[3, 4]. However, this is changing rapidly, as it is now recognized that not only the number of species but the species present, their phenotypes and the number of individuals of each species are critical in determining the nature and strength of the relationships between**

**Editor's summary** العربية

Traditional measures of biodiversity record species richness across different areas — in other words, they just count the number of species. This approach takes no account of the fact that different s...

**Associated links**

News & Views

Biodiversity: Temperate hotspots

by Tittensor

New Relic.
Want Code-Level Application Visibility?
Free Shirt!
Yes Please

**Editors' pick**

What did I do to generate the data?
How was the data processed?
Where is the data?
Who did what when

Synthesis
Analysis
Conclusions

# When submitting – your options

- *Early!*
  - *Interested in getting your data to the community as soon as possible?*
    - See our prior publication policy with the *Nature*-titled journals

- Complementary
  - Publish Data Descriptors alongside your research publications

- After
  - Publish Data Descriptor that expand on earlier research publications when they would help scientists reuse important datasets

# What does *Scientific Data* offer?

- Get Credit for Sharing Your Data
- Publications will be listed in the major indexes and will be citeable

- Focused on Data Reuse
- All the information others need to reuse the data; no interpretative analysis or hypothesis testing

- Open-access
- Authors select from three Creative Commons licences for the main
- Data Descriptor. Each publication supported by curated CC0 metadata

- Peer-reviewed
- Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards

- Promoting Community Data Repositories
- Data stored in community data repositories

# Open Access – APC supported

**Data:** the primary datasets will reside in public repositories.  Partnering with FigShare and Dryad, which are both **CC0**

**Data Descriptor - structured component (ISA-Tab):** as NPG has already done with its existing Linked Data Portal, the metadata about data descriptors in *Scientific Data* will be **CC0**

**Data Descriptor - narrative component:** describing the methodology of data generation/collection and processing will be licensed under either of the following, by author choice:

- cc BY
- cc BY NC
- cc BY NC SA

# What does *Scientific Data* offer?

- Get Credit for Sharing Your Data
- Publications will be listed in the major indexes and will be citeable

- Focused on Data Reuse
- All the information others need to reuse the data; no interpretative analysis or hypothesis testing

- Open-access
- Authors select from three Creative Commons licences for the main
- Data Descriptor. Each publication supported by curated CC0 metadata

- **Peer-reviewed**
- **Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards**

- Promoting Community Data Repositories
- Data stored in community data repositories

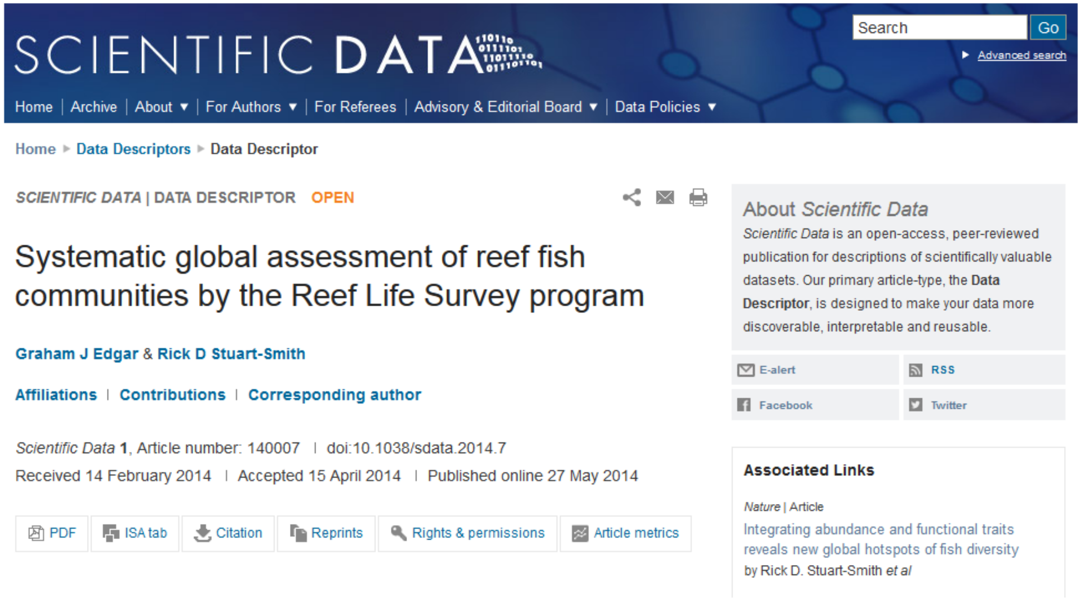
# Peer review process focused on quality and reuse

Evaluation is ***not*** be based on the perceived impact or novelty of the findings

- **Experimental Rigour and Technical Data Quality**
  - Were the data produced in a rigorous and methodologically sound manner?
  - Was the technical quality of the data supported convincingly with technical validation experiments and statistical analyses of data quality or error, as needed?
  - Are the depth, coverage, size, and/or completeness of these data sufficient for the types of applications or research questions outlined by the authors?
- **Completeness of the Description**
  - Are the methods and any data-processing steps described in sufficient detail to allow others to reproduce these steps?
  - Did the authors provide all the information needed for others to reuse this dataset or integrate it with other data?
  - Is this Data Descriptor, in combination with any repository metadata, consistent with relevant minimum information or reporting standards?
- **Integrity of the Data Files and Repository Record**
  - Have you confirmed that the data files deposited by the authors are complete and match the descriptions in the Data Descriptor?
  - Have these data files been deposited in the most appropriate available data repository?

- Introduction
  - concepts and principles
  - working with repositories
- Data Descriptor
  - structured experimental meta data
- Launch content
  - examples

# Edgar: Ecology



*SCIENTIFIC DATA* | DATA DESCRIPTOR OPEN

Systematic global assessment of reef fish communities by the Reef Life Survey program

Graham J Edgar & Rick D Stuart-Smith

Affiliations | Contributions | Corresponding author

*Scientific Data* **1**, Article number: 140007 | doi:10.1038/sdata.2014.7

Received 14 February 2014 | Accepted 15 April 2014 | Published online 27 May 2014

About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

**Associated Links**

*Nature* | Article

Integrating abundance and functional traits reveals new global hotspots of fish diversity

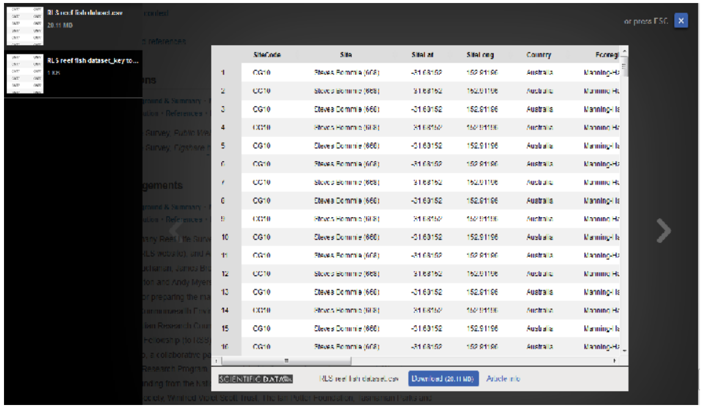
by Rick D. Stuart-Smith *et al*

**Associated *Nature* Articles**

Data in **figshare**

Integrated **figshare** data viewer

*Citizen Science project*

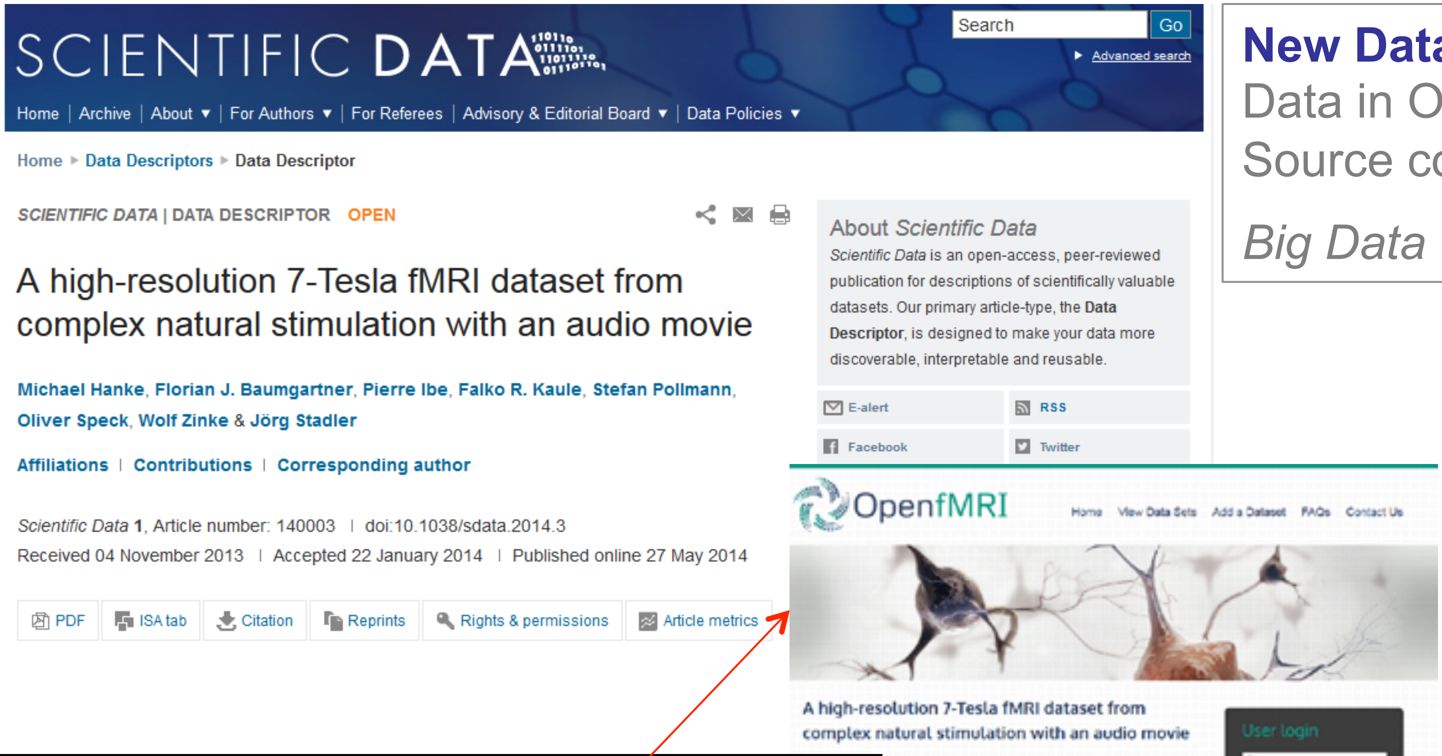


**Data Citations**

Abstract • Background & Summary • Methods • Data Records • Technical Validation • Usage Notes • Additional information • References • Data Citations • Acknowledgements • Author information

1. Reef Life Survey, *Public Website* www.reeflifesurvey.com (2014).
2. Reef Life Survey, *Figshare* http://dx.doi.org/10.6084/m9.figshare.934319 (2014).

# Hanke: Neuroscience



*SCIENTIFIC DATA* | DATA DESCRIPTOR OPEN

## A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie

**Michael Hanke**, **Florian J. Baumgartner**, **Pierre Ibe**, **Falko R. Kaule**, **Stefan Pollmann**, **Oliver Speck**, **Wolf Zinke** & **Jörg Stadler**

Affiliations | Contributions | Corresponding author

*Scientific Data* **1**, Article number: 140003 | doi:10.1038/sdata.2014.3
Received 04 November 2013 | Accepted 22 January 2014 | Published online 27 May 2014

About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie

**New Dataset**
Data in OpenfMRI
Source code in GitHub

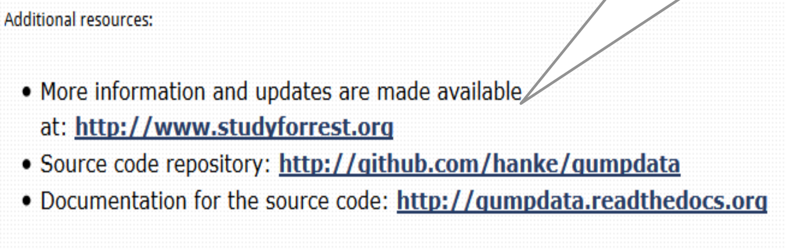
*Big Data*

### Data Citations

Abstract • Background & Summary • Methods • Data Records • Technical Validation • Usage Notes • Additional information • References • Data Citations • Acknowledgements • Author information • Supplementary information

1. Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., Zinke, W., & Stadler, J. *OpenfMRI* ds000113 (2014).

**Code in GitHub**

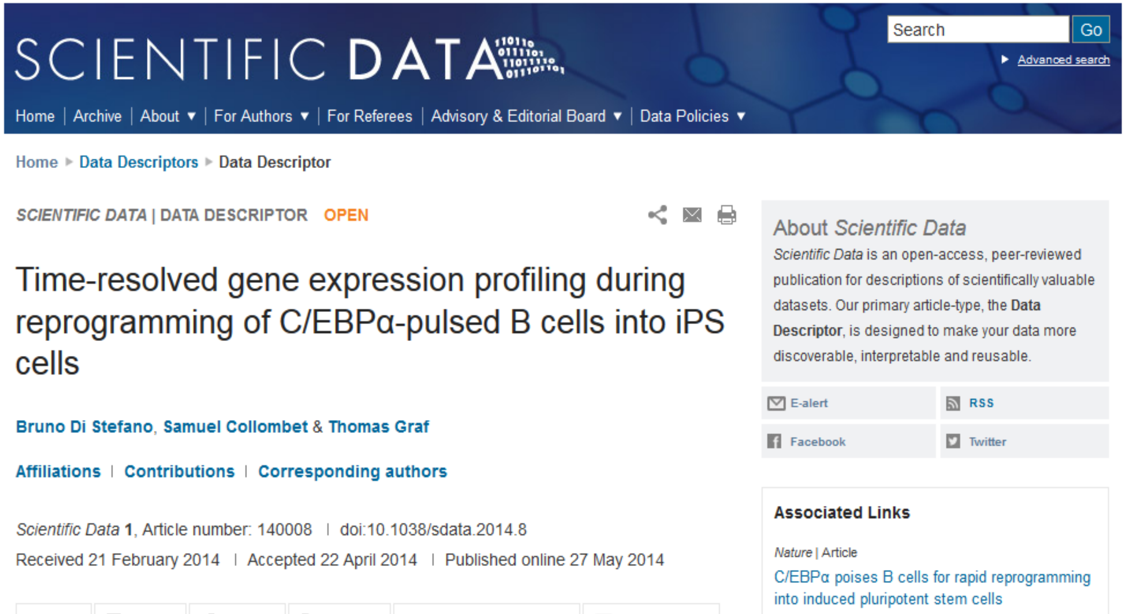


Additional resources:

- More information and updates are made available at: http://www.studyforrest.org
- Source code repository: http://github.com/hanke/gumpdata
- Documentation for the source code: http://gumpdata.readthedocs.org

nature publishing group npg

# Stefano: Stem Cells



*SCIENTIFIC DATA* | DATA DESCRIPTOR OPEN

Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells

**Bruno Di Stefano**, **Samuel Collombet** & **Thomas Graf**

**Affiliations** | **Contributions** | **Corresponding authors**

*Scientific Data* **1**, Article number: 140008 | doi:10.1038/sdata.2014.8
Received 21 February 2014 | Accepted 22 April 2014 | Published online 27 May 2014

About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

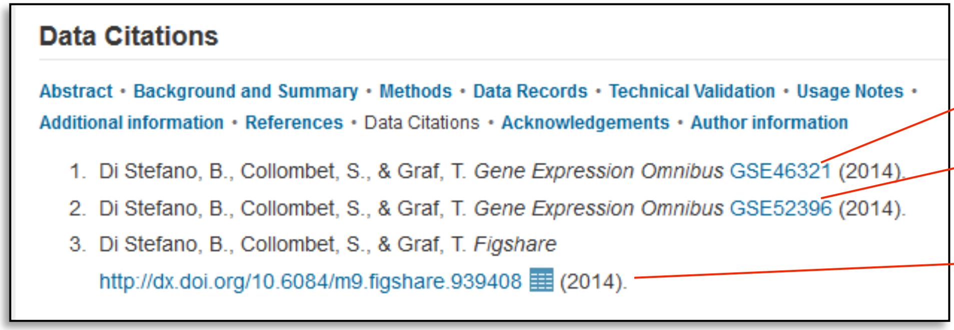
**Associated Links**

*Nature* | Article
C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells
by Bruno Di Stefano *et al*

## **Associated *Nature* Article**

Data
- **fig**share
- NCBI GEO

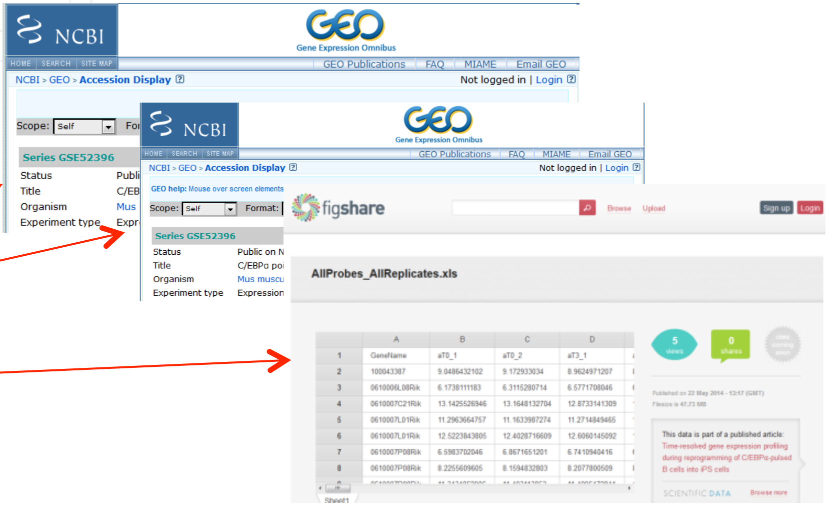
Integrated **fig**share data viewer



**Data Citations**

Abstract • Background and Summary • Methods • Data Records • Technical Validation • Usage Notes • Additional information • References • Data Citations • Acknowledgements • Author information

1. Di Stefano, B., Collombet, S., & Graf, T. *Gene Expression Omnibus* GSE46321 (2014).
2. Di Stefano, B., Collombet, S., & Graf, T. *Gene Expression Omnibus* GSE52396 (2014).
3. Di Stefano, B., Collombet, S., & Graf, T. *Figshare* http://dx.doi.org/10.6084/m9.figshare.939408 (2014).



Series GSE52396

Status Publi
Title C/EB
Organism Mus
Experiment type Expr

Series GSE52396

Status Public on N
Title C/EBPα poi
Organism Mus muscu
Experiment type Expression

AllProbes_AllReplicates.xls

| | A | B | C | D |
|---|---|---|---|---|
| 1 | GeneName | aT0_1 | aT0_2 | aT3_1 |
| 2 | 100043387 | 9.0486432102 | 9.172933034 | 8.9624971207 |
| 3 | 0610006L08Rik | 6.1738111183 | 6.3115280714 | 6.5771708046 |
| 4 | 0610007C21Rik | 13.1425526946 | 13.1648132704 | 12.8733141309 |
| 5 | 0610007L01Rik | 11.2963664757 | 11.1633987274 | 11.2714849465 |
| 6 | 0610007L01Rik | 12.5223843805 | 12.4028716609 | 12.6060145092 |
| 7 | 0610007P08Rik | 6.5983702046 | 6.8671651201 | 6.7410940416 |
| 8 | 0610007P08Rik | 8.2255609605 | 8.1594832803 | 8.2077800509 |

Published on 22 May 2014 - 13:17 (GMT)
Filesize is 47.73 MB

This data is part of a published article:
Time-resolved gene expression profiling during reprogramming of C/EBPα-pulsed B cells into iPS cells

# Hao: Environmental

Home ▸ Data Descriptors ▸ Data Descriptor

*SCIENTIFIC DATA* | DATA DESCRIPTOR OPEN

Global integrated drought monitoring and prediction system

Zengchao Hao, Amir AghaKouchak, Navid Nakhjiri & Alireza Farahmand

Affiliations | Contributions | Corresponding author

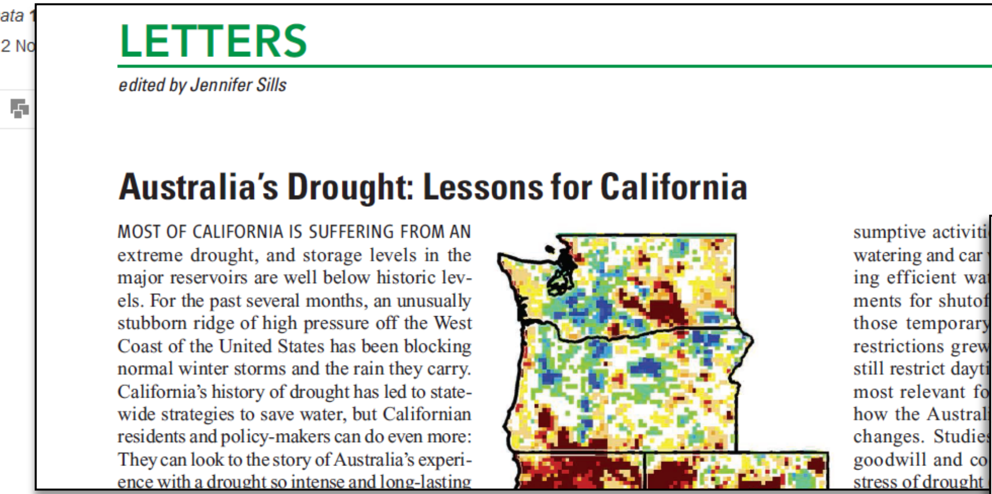
About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

*Scientific Data* 1

Received 12 No

PDF



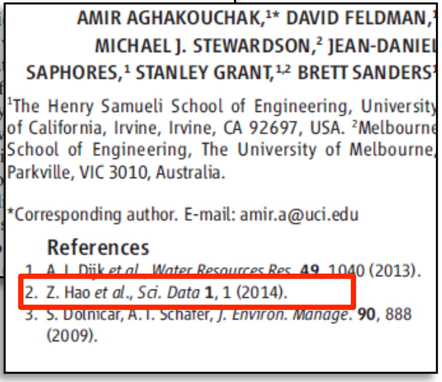
LETTERS

*edited by Jennifer Sills*

Australia's Drought: Lessons for California

MOST OF CALIFORNIA IS SUFFERING FROM AN extreme drought, and storage levels in the major reservoirs are well below historic levels. For the past several months, an unusually stubborn ridge of high pressure off the West Coast of the United States has been blocking normal winter storms and the rain they carry. California's history of drought has led to statewide strategies to save water, but Californian residents and policy-makers can do even more: They can look to the story of Australia's experience with a drought so intense and long-lasting

sumptive activiti
watering and car
ing efficient wa
ments for shutof
those temporary
restrictions grew
still restrict dayt
most relevant fo
how the Austral
changes. Studie
goodwill and co
stress of drought



AMIR AGHAKOUCHAK,[1*] DAVID FELDMAN,
MICHAEL J. STEWARDSON,[2] JEAN-DANIEL
SAPHORES,[1] STANLEY GRANT,[1,2] BRETT SANDERS[1]

[1]The Henry Samueli School of Engineering, University of California, Irvine, Irvine, CA 92697, USA. [2]Melbourne School of Engineering, The University of Melbourne, Parkville, VIC 3010, Australia.

*Corresponding author. E-mail: amir.a@uci.edu

References

1. A. I. Dijk *et al.*, *Water Resources Res.* **49**, 1040 (2013).
2. Z. Hao *et al.*, *Sci. Data* **1**, 1 (2014).
3. S. Dolnicar, A. I. Schafer, *J. Environ. Manage.* **90**, 888 (2009).

**New Dataset**

Data in **fig**share

Code in figshare

Integrated figshare data viewer

Cited in Science

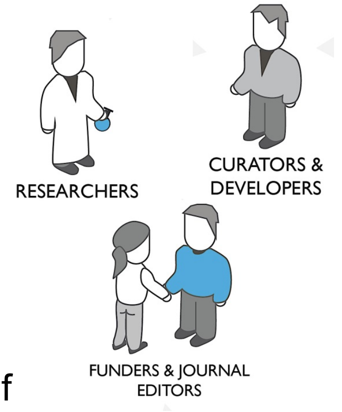
# LIFE, NATURAL & BIOMEDICAL SCIENCES

**Driving digital research and providing services to the Life, Natural and Medical Sciences**

bbsrc biotechnology and biological sciences research council
Roche
MRC Medical Research Council
elixir UNITED KINGDOM
NATURAL ENVIRONMENT RESEARCH COUNCIL
SEVENTH FRAMEWORK PROGRAMME

*Do you need help or collaborators in the following areas?*

- Data capture and curation
- Database development
- Data (nano)publication
- Data provenance
- Open, community ontologies and standards
- Semantic web
- Social engineering
- Software development
- Training
- Visualization (collaboration/jointly with Prof Min Chen)



*Are you interested in exploring, using or extending our tools?*

isatools

isatab format

github SOCIAL CODING
GITHUB.COM/ISA-TOOLS

biosharing.org

*If yes, please, contact us!*