

Research Data Publication in Principle and Practice

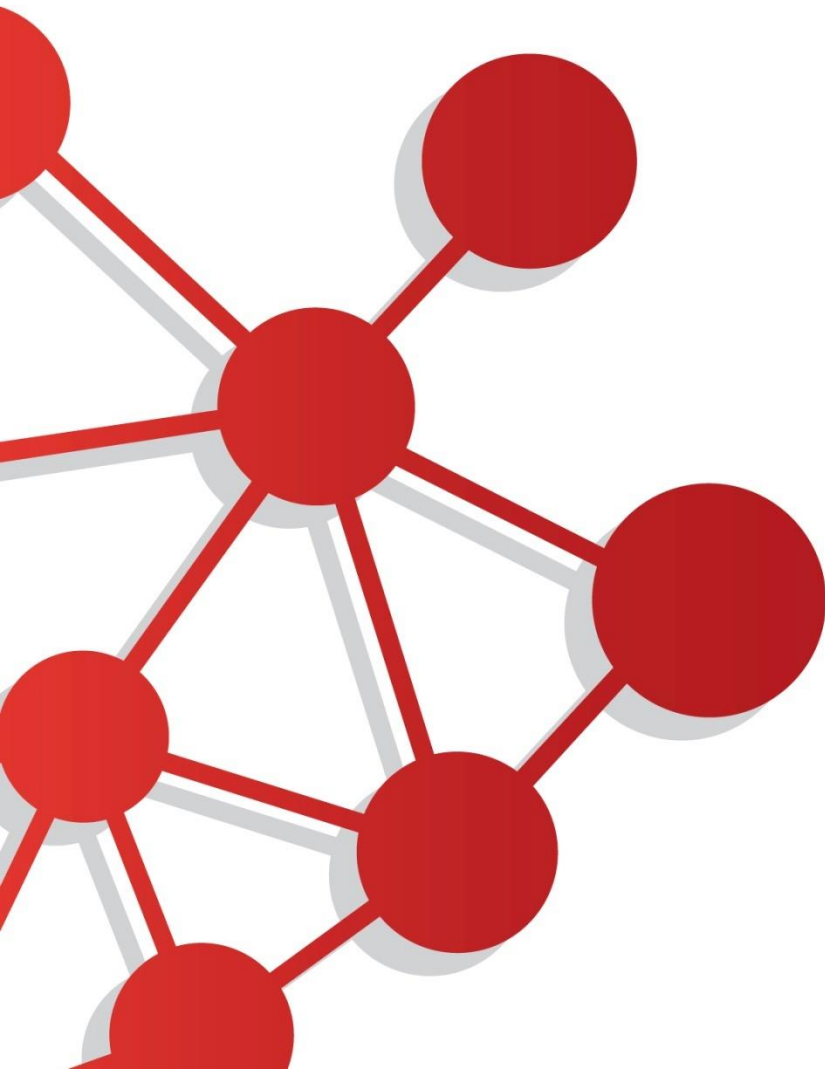
IDCC conference workshop 2014

Ruth Wilson – Publisher, Nature Publishing Group

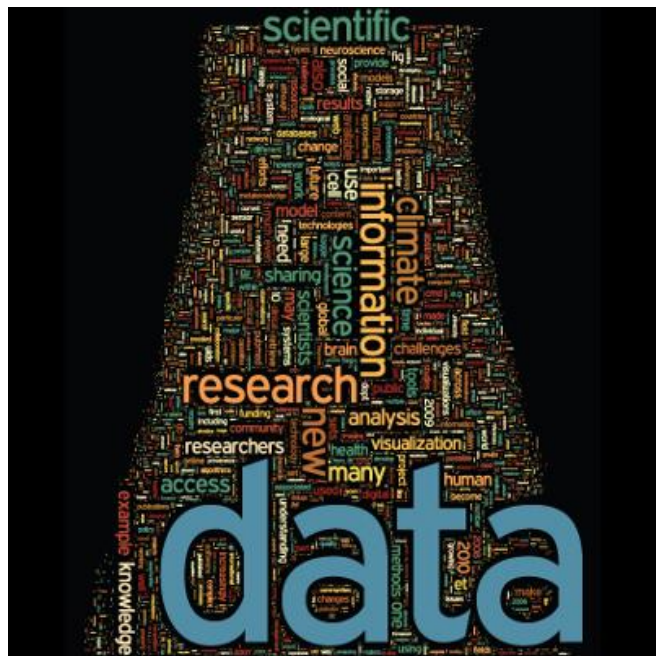
Principles

Practice

1. Scientific Data
2. PubChem deposition



Publishing Data



Two important factors are driving data publication:

- To ensure the scientific process is transparent and can be scrutinised and research results reproduced
- To speed the scientific process, lead to new insights and reduce duplicated and repeated work

Scientific Data



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**

The Data Descriptor

www.nature.com/scientificdata

Sample Notice
This is a sample Data Descriptor derived from a publication at *Molecular Systems Biology*. It should not be considered an independent publication. The original article (Munoz, J. et al. *Mol. Syst. Biol.* 7, 550; 2011) should be cited in all scholarly publications.

SCIENTIFIC DATA


SUBJECT CATEGORIES
» Induced pluripotent stem cells
» Proteomic analysis
» Microarray analysis

Proteomic profiles of human embryonic stem cells, induced-pluripotent stem cells and precursor fibroblasts

Javier Munoz¹ and Albert J.R. Heck^{2,3}

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 a dataset comprising quantitative mass spectrometry-based measurements of the proteomes of hESCs, two different hiPSCs and their precursor fibroblast cell lines, along with matching gene expression profiles for each sample. These data are suitable for in depth comparative analysis of the proteomes of both somatic and pluripotent cells, and have been deposited in three different public repositories to maximize ease of reuse by the community.

Design Type(s)	cell type comparison design • growth condition intervention design
Measurement Type(s)	protein expression profiling • transcription profiling assay
Technology Type(s)	mass spectrometry assay • DNA microarray
Factor Type(s)	cell line • growth condition
Sample Characteristic(s)	<i>Homo sapiens</i> • embryonic stem cell line • embryonic fibroblast cell line • foreskin fibroblast cell line

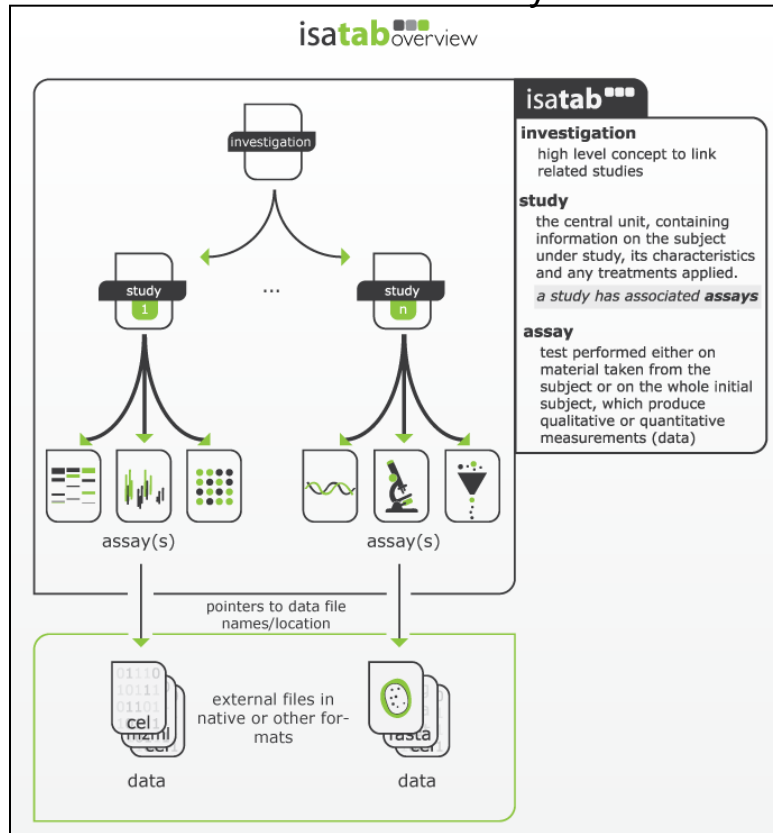
¹Proteomics Unit, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.
²Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
³Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.

SCIENTIFIC DATA | 1:1 | DOI: 10.1038/00000

1

Structured experimental meta-data

Submit ISA-tab files directly



OR

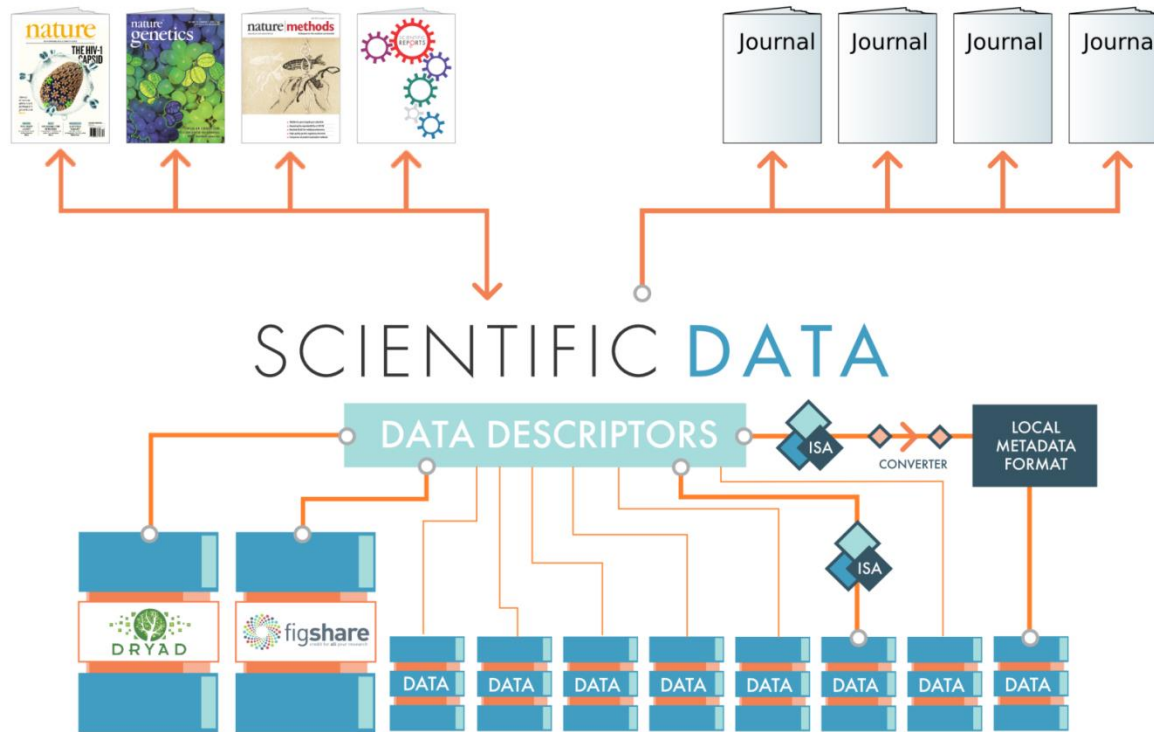
Submission tools and simple templates help authors provide the information without special tools

	A	B	C	D	E	F
1	Samples					
2	Key information on the samples; note that sample here refers inter alia to an organism, or tissue, or an					
3						
4	Sample ID	Description	Material type*	Organism*	Organism part*	Factor value
5	sample_1	description text with any additional information on sample	tissue	Drosophila melanogaster	brain	brain
6	sample_2	description text with any additional information on sample	tissue	Drosophila melanogaster	heart	heart
7	sample_3	description text with any additional information on sample	tissue	Drosophila melanogaster	muscle	muscle
8	...					
9						
10						
11						
12						
13						
14						
15						
16						
17						



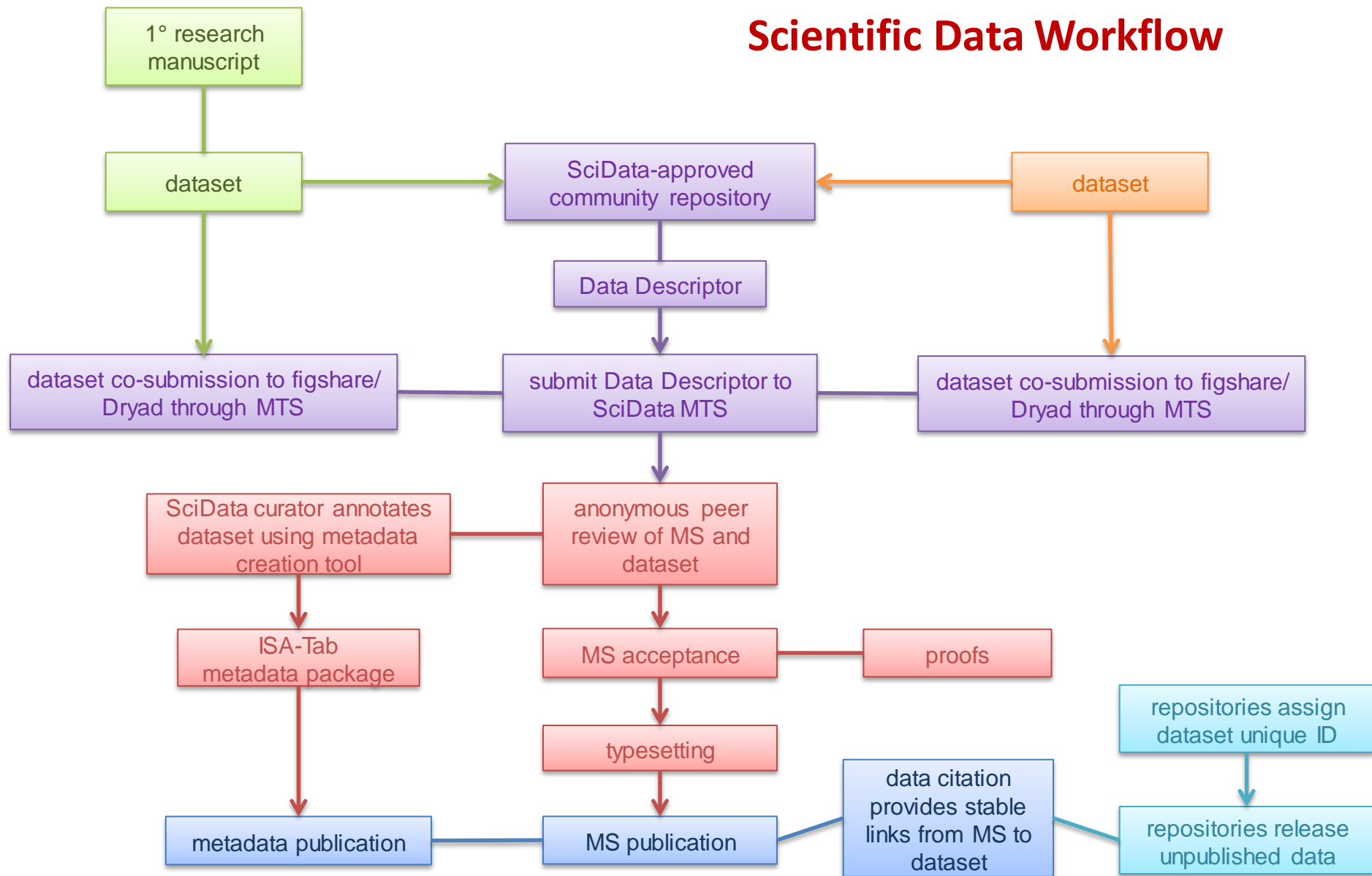
In-house curator
standardizes the
structured content

Repository Criteria



- Be supported and recognized within their scientific community
- Ensure long-term persistence and preservation of datasets
- Provide expert curation
- Implement relevant community endorsed reporting requirements
- Provide for confidential review of submitted datasets
- Provide stable identifiers for submitted datasets
- Allow public access to data without unnecessary restrictions

Scientific Data Workflow



Green: author submits dataset from previous publication; orange: author submits dataset without previous publication; purple: author data and MS submission process; red: SciData and production; teal: repository; blue: publication; MTS: manuscript tracking system

Who are we?

Susanna-Assunta Sansone - Honorary Academic Editor



Andrew L Hufton - Managing Editor

Advisory Panel

[Judith A. Blake](#)

The Jackson Laboratory, USA

[Chris Bowler](#)

IBENS, France

[Piero Carninci](#)

RIKEN Omics Science Center,
Japan

[David Carr](#)

Wellcome Trust, UK

[Stephen Chanock](#)

National Cancer Institute, USA

[Patricia Cruse](#)

CDL, USA

[Simon Hodson](#)

CODATA,, France

[Joseph R. Ecker](#)

Salk Institute, USA

[Mark Forster](#)

Syngenta, UK

[Stephen Friend](#)

Sage Bionetworks, USA

[Pascale Gaudet](#)

Swiss Institute of Bioinformatics,
Switzerland

[Anne-Claude Gavin](#)

EMBL, Germany

[Albert J. R. Heck](#)

Utrecht University, The Netherlands

[Wolfram Horstmann](#)

University of Oxford, UK

[Johanna McEntyre](#)

EMBL-EBI, European Bioinformatics Institute, UK

[Anthony Rowe](#)

Johnson & Johnson, USA

[Richard H. Scheuermann](#)

J. Craig Venter Institute, USA

[Caroline Shamu](#)

Harvard Medical School, USA

[Jessica Tenenbaum](#)

Duke Translational Medicine Institute, USA

[Weida Tong](#)

National Center for Toxicological Research, FDA,
USA

Supported by

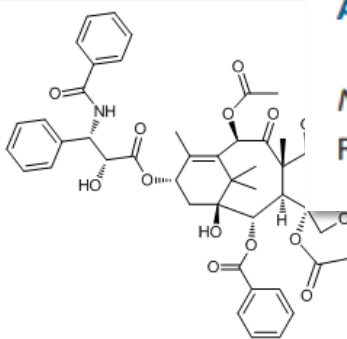


PubChem structure deposition

Compound 1

From **Scalable enantioselective total synthesis of taxanes**
Abraham Mendoza, Yoshihiro Ishihara & Phil S. Baran
Nature Chemistry 4, 21–25 (2012)
[Supplementary Information](#)

Compound 1
Taxol®



[View in PubChem](#) | [View in 3D \(10 KB\)](#) | [Download ChemDraw file of structure \(7 KB\)](#) | [Download CML file \(9 KB\)](#) | [Download Molfile \(6 KB\)](#)

NATURE CHEMISTRY | ARTICLE

Scalable enantioselective total synthesis of taxanes

Abraham Mendoza, Yoshihiro Ishihara & Phil S. Baran

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

Nature Chemistry 4, 21–25 (2012) | doi:10.1038/nchem.1196
Received 22 August 2011 | Accepted 10 October 2011 | Published online 06 November 2011

PubChem Structure Database

Journal creates Nature compound page