Reproducibility in the Article of the Future

Hylke Koers, Head of Content Innovation, Elsevier

What is needed for effective reproduction of research results?

A pragmatic* check-list:

1. <u>Access</u> to materials that underpin the results, such as primary data and methods (including computer code)

2. <u>Unambiguous reference</u> to resources and experimental conditions

3. <u>Provenance:</u> a clear and unambiguous description of the full research chain, all the way from primary data to results

4. All of this made available to peer researchers in a way that's <u>understandable and actionable</u>!

*) For a thorough introduction in the subject, see e.g. "<u>What is Reproducibility</u>" by Prof. Carole Goble

How does CI help reproducibility?
1. Access to research material

Content Innovation "data visualization" features give readers access to materials (notably data and code) that are not traditionally included in the article – even though they are an important part of the research carried out and are often needed to reproduce results.

- **Data**: CI data viewers; including 3D viewers, interactive plots, Virtual Microscope, etc. – provide readers access to research data underpinning a publication

- **Computer code**: The R-code viewer, Inline Supplementary Material – provide readers access to computer code that was used as a method to obtain results reported in a paper

- **Multimedia**: Videos on ScienceDirect (though not technically a CI feature) may contain details of a procedure or represent actual data

How does CI help reproducibility?
2. Unambiguous reference

A number of Content Innovation "context and reference" features are based on codes or accession numbers in the article that describe an entity or resource without any ambiguity – this in contrast to colloquial terms or product names that can mean different things. See also "*Minimal Data Standards for neuroscience articles: Resource Identification Initiative*"

(An additional benefit is that proper tagging of such codes makes it easier to search across articles).
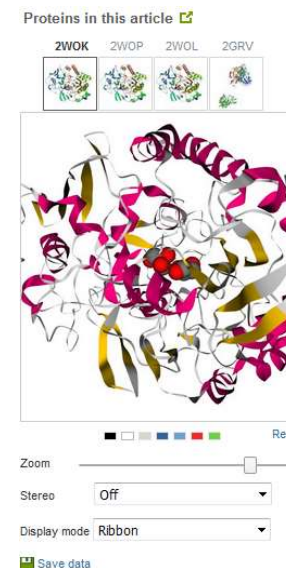
Content Innovation	Code / accession number scheme	What do they identify?
Antibody Data Viewer	AntibodyRegistry or RRID resource identifiers	Antibodies
Genome Viewer	Genbank accession numbers	Genes
Gene Expression Omnibus App	GEO accession numbers	Genes and more
Protein Viewer	Protein Data Bank accession numbers	Proteins
PubChem	PubChem codes for chemical compounds	Chemical compounds
Arabidopsis Genome Viewer	TAIR genome codes	Genes

How does CI help reproducibility?
3. Provenance

Or: how does all that code and data fit together?

- **R-code viewer**: This viewer does not only show the R-code, but also the underlying data – and describes how the code depends on the data.

- **The Executable Paper**: A pilot project (now stopped) to capture the entire computation chain, including code and data, in a fully re-executable way with the article.

- **Quod next?**
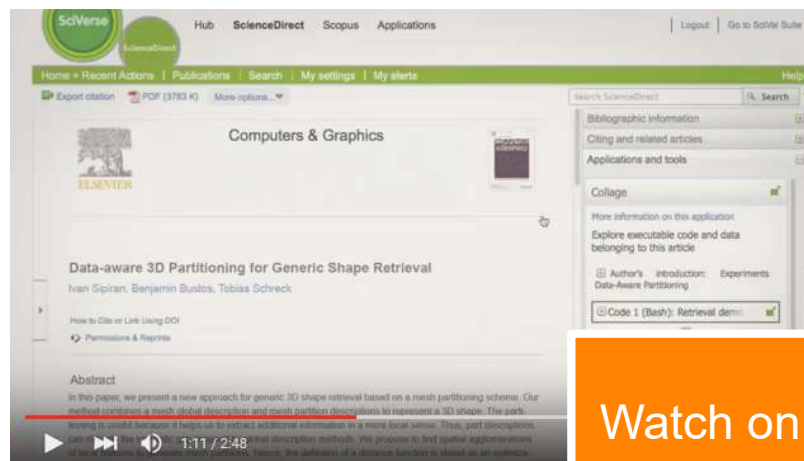
Watch on YouTube

How does CI help reproducibility?
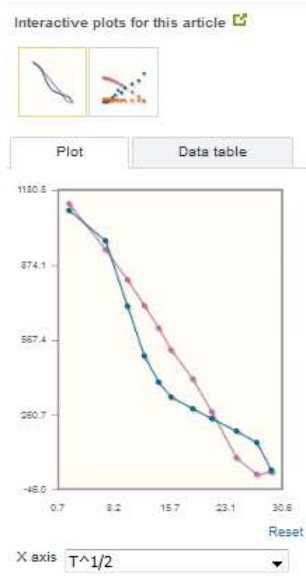# 4. Make it all <u>understandable & actionable</u>

Supplementary 3D models

You can zoom, rotate and pan this 3D reconstruction

Rotate

Zoom

Stereo

Save data

Interactive plots for this article

Plot | Data table

1180.8

874.1

567.4

260.7

-46.0

0.7   8.2   15.7   23.1   30.6

Reset

X axis T^1/2

Code files for this article

benefits syntax 3-27-14 (R, 7.5 kB)
Download Preview

benefitsforRstandardizedwithBig5 (CSV, 12.1 MB)
Download

benefitsforRstandardizedwithBig5Agr (CSV, 2.7 MB)
Download

benefitsforRstandardizedwithBig5Con (CSV, 2.7 MB)
Download

benefitsforRstandardizedwithBig5Ext (CSV, 2.2 MB)
Download

benefitsforRstandardizedwithBig5Neur (CSV, 4.5 MB)
Download

benefitsforRstandardizedwithBig5Open (CSV, 3.6 MB)
Download

Most data and code viewers give reader access to the material for download – so they can easily use it for **further analysis or re-do some of the work** reporting in the paper

Data viewers (such as the 3D viewers) enable readers to inspect data in detail and **build a better understanding**
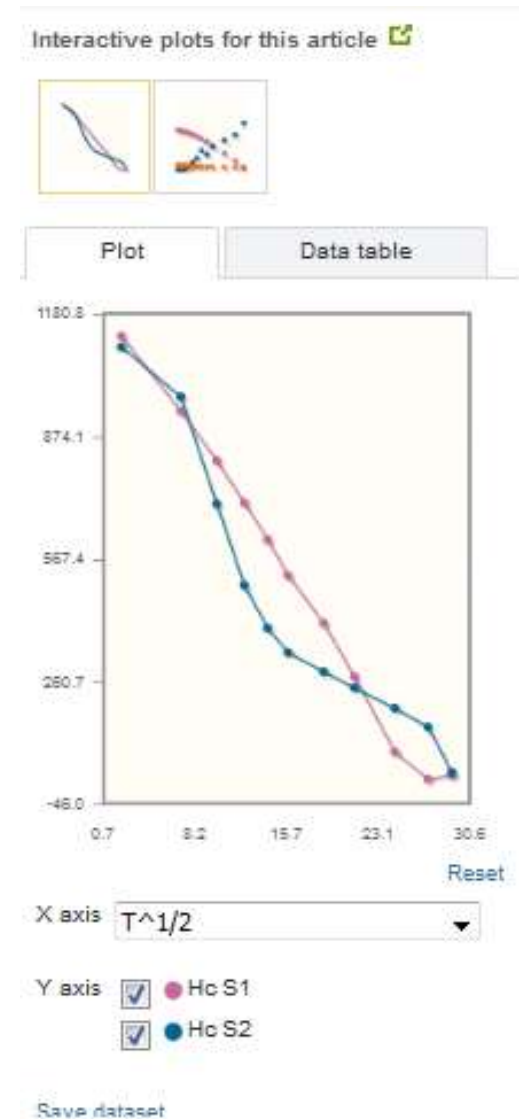
Annex

Slides for selected CI's

Interactive Plot Viewer

Sharing data, bringing plots to life

- Making available the valuable data that is now often "buried" in plot images.
- Readers can inspect and interact with data right from within the article, building additional insights into the work presented
- Facilitates re-use and reproducibility of research results

How?
- Authors are invited to submit data underlying plots as a supplementary CSV file
- The CSV file is displayed as an interactive plot next to the article

Interactive plots for this article

Plot | Data table

X axis   T^1/2

Y axis   ☑ ● Hc S1
         ☑ ● Hc S2

Save dataset

More info: http://www.elsevier.com/books-and-journals/content-innovation/iplots
Example article: http://www.sciencedirect.com/science/article/pii/S0375960115003102
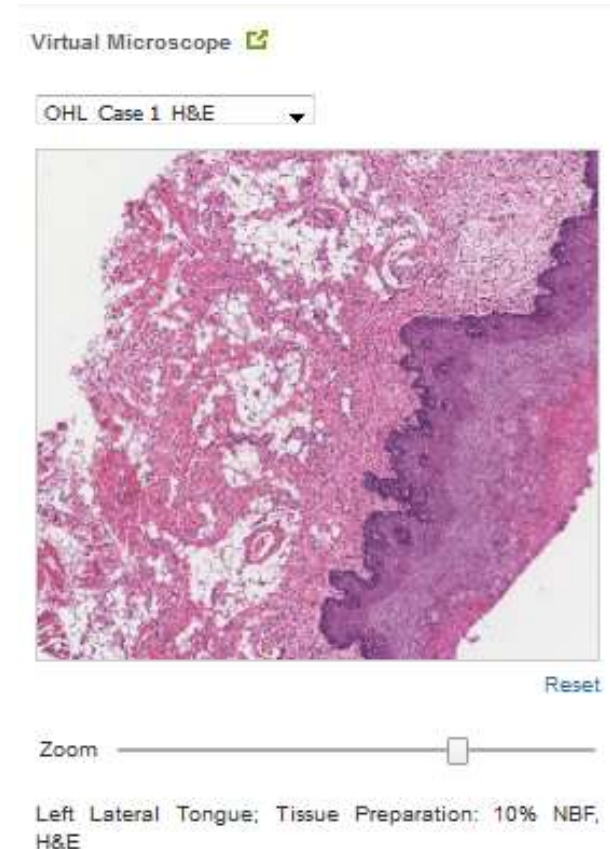
Virtual Microscope

High-resolution slide imagery at your fingertips

- Including whole slide samples right with the online article
- Readers can see the slide as a whole, or zoom in up to 40x magnification to inspect details in any region

How?

- Authors are invited to upload digital slides with their publication (scanning services for physical slides available)
- All slides are pre-processed for optimal viewing on both desktop and mobile devices

Virtual Microscope

OHL_Case 1 H&E

Reset

Zoom

Left Lateral Tongue; Tissue Preparation: 10% NBF, H&E

Proud recipient of the 2015 PROSE award for best eProduct/App

More info: http://www.elsevier.com/books-and-journals/content-innovation/virtual-microscope
Example article: http://www.sciencedirect.com/science/article/pii/S2212440314004635

3D Model Viewers

Letting 3D come to life

- Bringing 3D material to life in an intuitive yet powerful way
- Readers can use standard 3D controls like rotate and zoom, and change visualization styles

How?
- Authors are invited to upload 3D models using a variety of formats
- Depending on the format, the most appropriate visualization style is chosen (including "fly-through" virtual environments)
- Fully supported on mobile devices

Supplementary 3D models

You can zoom, rotate and pan this 3D reconstruction

Rotate
Zoom
Stereo   Off
Save data (83.7 MB)

More info: http://www.elsevier.com/books-and-journals/content-innovation/3d-modeling
Example article: http://www.sciencedirect.com/science/article/pii/S2212054813000027

R-code viewer

Making research truly reproducible

- Displaying computer code in the 'R' language, together with related data files
- Readers can see summary of all code and data used in the article, and inspect the code in details using syntax highlighting
- Supports format-preserving copy-and-paste of code, and download for all code and data files

How?
- Authors are invited to upload 'R' code files and related data (in any format) as a supplementary ZIP file

Code files for this article

benefits syntax 3-27-14 (R, 7.5 kB)
Download Preview

benefitsforRstandardizedwithBig5 (CSV, 12.1 MB)
Download

benefitsforRstandardizedwithBig5Agr (CSV, 2.7 MB)
Download

benefitsforRstandardizedwithBig5Con (CSV, 2.7 MB)
Download

benefitsforRstandardizedwithBig5Ext (CSV, 2.2 MB)
Download

benefitsforRstandardizedwithBig5Neur (CSV, 4.5 MB)
Download

benefitsforRstandardizedwithBig5Open (CSV, 3.6 MB)
Download

Download all (ZIP, 915 kB)

More info: http://www.elsevier.com/books-and-journals/content-innovation/rcodeviewer
Example article: http://www.sciencedirect.com/science/article/pii/S0092656614000488

Inline Supplementary Material

Presenting supplementary material at the right place in the right context

- Displaying supplementary material at the right place in the article
- Readers will see an expandable text box that gives access to the supplementary material in the appropriate context
- Supports supplementary images, tables, and computer code

How?
- Authors are invited to upload supplementary material with their article, and indicate where it should be placed in the manuscript

PROV data can be visualized as a directed labeled graph in whic activities and agents and edges represent influences between e events (plus annotations of nodes and edges, such as timestam for the example above is shown in Fig. 2, where entities are sho rectangles, and agents are pentagons. The following is the Sup related to this article.

⊞ Inline Supplementary Computer Code S1

⊟ Inline Supplementary Computer Code S2

```
document
prefix bnode <http://openprovenance.org/provtoolbox/bnode/>
prefix xsd <http://www.w3.org/2001/XMLSchema>
prefix ex <http://www.example.org#>
prefix rdf <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
activity(ex:publish,-,-)
```

More info: https://www.elsevier.com/books-and-journals/content-innovation/inline-supplementary-material
Example article: http://www.sciencedirect.com/science/article/pii/S1570826815000177

ELSEVIER

Antibody Data

Connecting and disambiguating antibodies

- Presenting relevant information about antibodies mentioned in the article
- Readers see a list of relevant antibodies with key information and links to the Neuroscience Information Framework (NIF) for further details.
- Helps drive proper resource identification standards

How?
- Authors are invited to tag antibodies in their article (e.g. "AntibodyRegistry: AB_177645")
- Relevant antibody data is displayed next to the article, using data from NIF Antibody Registry
- In collaboration with NIF and Force11 "Resource Identification Initiative"

Antibody data for this article

⊟ **Anti-Huntingtin Protein, clone mEM48 antibody**
Antibody ID   AB_177645
Antibody Target   Huntingtin Protein clone mEM48
Vendor   EMD Millipore
Catalog Num   MAB5374
Clonality   monoclonal antibody
Source Organism   mouse

⊞ Anti-NeuN, clone A60 antibody

Powered By NIF

More info: http://www.elsevier.com/books-and-journals/content-innovation/antibody-data
Example article: http://www.sciencedirect.com/science/article/pii/S0969996114002526

Genome Viewer

Connecting the article with a wealth of genomics data

- Showing sequence data for relevant genes mentioned in the article
- Readers can view, inspect and analyze gene sequence data – zoom, flip strands, find a specific location, and more
- Deep-links to NCBI for further data
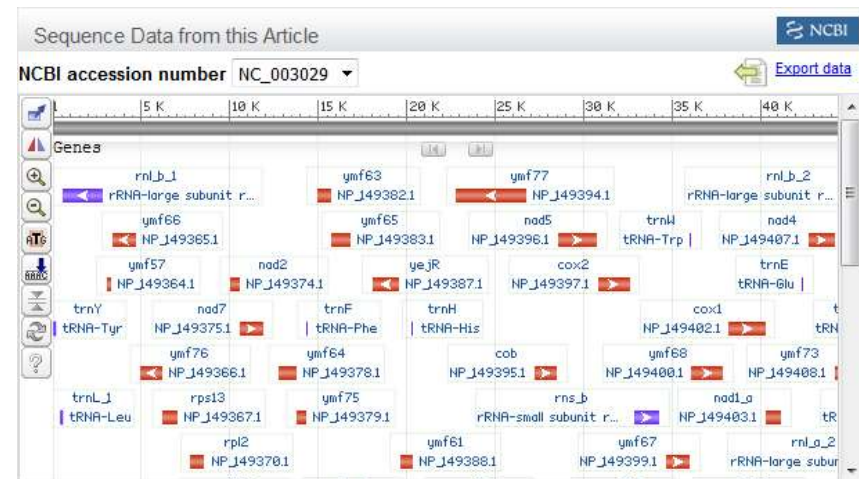
How?
- Authors are invited to tag Genbank accession numbers in their article (e.g. "Genbank: NC_003029")
- Genome Viewer is displayed within the article, using data from NCBI for author-tagged genes
- In collaboration with the NCBI Sequence Viewer team

More info: https://www.elsevier.com/books-and-journals/content-innovation/genome-viewer
Example article: http://www.sciencedirect.com/science/article/pii/S0378111911000941

Gene Expression Omnibus (GEO) App

Connecting the article with a wealth of functional genomics data

- Presenting information about genes and other GEO (Gene Expression Omnibus) entities mentioned in the article
- Readers see a summary overview for each GEO entity, with deep-links for further detail and data.

How?
- Authors are invited to use GEO identifiers for genes and other entities in their article (e.g. "GEO: GPL570")
- Developed in collaboration with NCBI

GEO data referred to in this article

Platform (1) Sample (0) Series (2)

GEO entity: GPL570

Platform GPL570

Status: public on Nov 07, 2003

Title: [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array

Technology type: in situ oligonucleotide

Organism: Homo sapiens

GEO
Gene Expression Omnibus

More info: http://www.elsevier.com/books-and-journals/content-innovation/geo-app
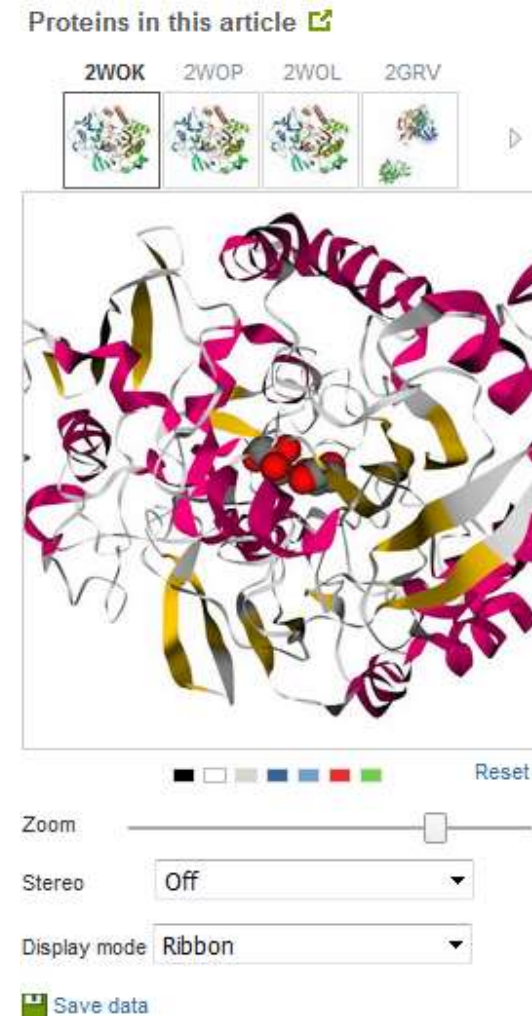Example article: http://www.sciencedirect.com/science/article/pii/S0888754311002217

Protein Viewer

Live 3D structures from the Protein Data Bank

- Displaying 3D structures for the articles most relevant proteins
- Readers can explore 3D structure properties using zoom & rotate, and change visualization settings as desired

How?
- Authors are invited to tag Protein Data Bank (PDB) accession numbers in their article (e.g. "PDB: 2WOK")
- Protein Viewer is displayed next to the article, using data from PDB for author-tagged proteins
- In collaboration with PDB

More info: https://www.elsevier.com/books-and-journals/content-innovation/protein-viewer
Example article: http://www.sciencedirect.com/science/article/pii/S0022283609014399

PubChem Compound Viewer

Connecting chemistry, biology, and pharma

- Connecting the article to <u>NCBI PubChem Compound database</u> that provides information on the biological activities of small molecules
- Readers see a a short summary overview for every compound, with links to PubChem for further information

How?
- Authors are invited to include relevant PubChem identifiers in a section "Chemical compounds studied in this article"

Studied compounds

Glutamic acid
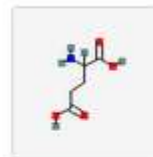
PubChem CID :611

Molecular Weight: 147.12926

Molecular Formula: $C_5H_9NO_4$

IUPAC name: 2-aminopentanedioic acid

InChIKey: WHUUTDBJXJRKMK-UHFFFAOYSA-N

Pub©hem

More info: http://www.elsevier.com/books-and-journals/content-innovation/pubchem
Example article: http://www.sciencedirect.com/science/article/pii/S0168365913002101

ELSEVIER

TAIR Arabidopsis Data Viewer

Connecting with a world of Arabidopsis genetic data

- Presenting key information about Arabidopis loci mentioned in the article
- Readers see a summary overview for each locus, with deep-links to The Arabidopsis Information Resource (TAIR) for further detail

How?
- Authors are invited to use TAIR identifiers for Arabidopsis loci in their article (e.g. "TAIR: At3g17690")
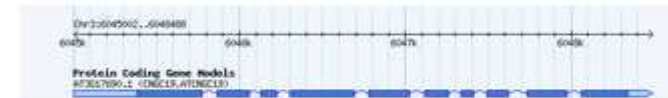- In collaboration with TAIR

TAIR Arabidopsis

Go to overview of loci in this article

Locus   At3g17690

● Details   ○ Occurrences (1)

Member of cyclic nucleotide gated channel family

click for full size application

Other names

ATCNGC19,CNGC19,CYCLIC NUCLEOTIDE GATED CHANNEL 19,CYCLIC NUCLEOTIDE-GATED CHANNEL 19

Powered by TAIR

More info: http://www.elsevier.com/books-and-journals/content-innovation/arabidopsis-gene-viewer
Example article: http://www.sciencedirect.com/science/article/pii/S0981942813000235