Data Citation Implementation Pilot

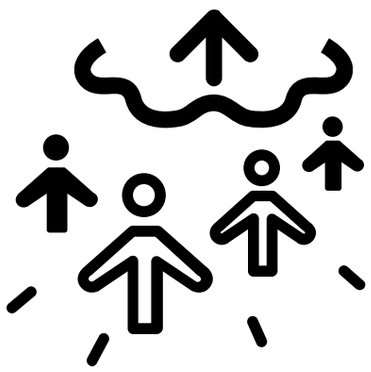
Focus on Repositories: What do they need to do?

Martin Fenner

DataCite Technical Director

http://orcid.org/0000-0003-1419-2405

Assumptions



Based on Joint Declaration of Data Citation Principles, e.g.

Importance

Unique Identification

Persistence

Principles

Realistic goals for pilot

Use existing workflows and tools for article citation as much as possible

Volume 16 Number 10 1988 Nucleic Acids Research

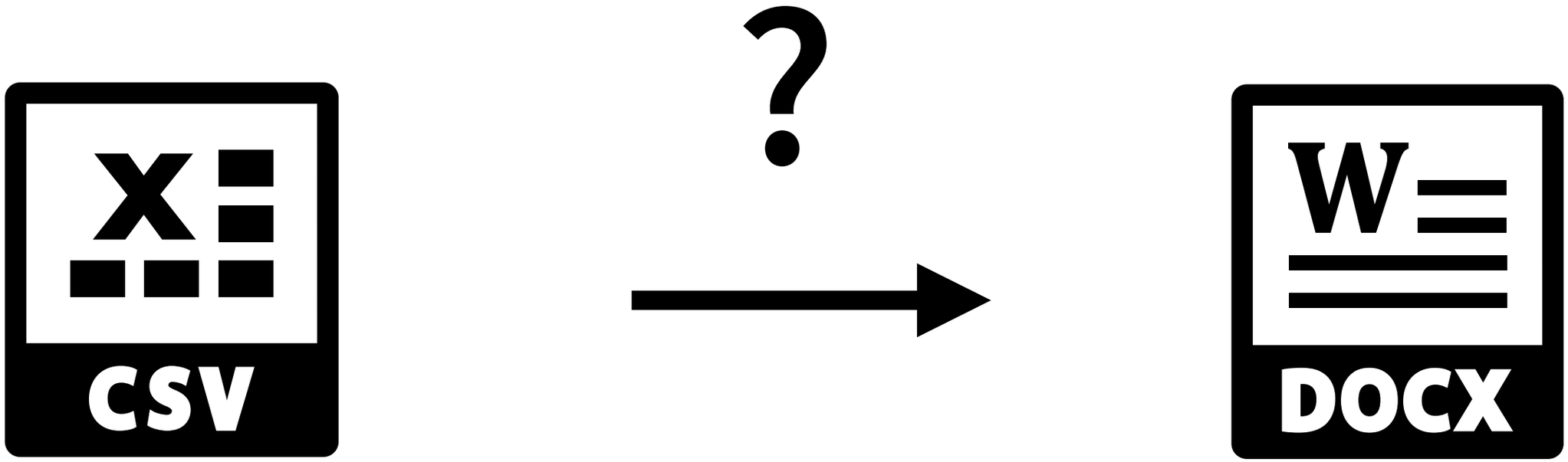
NAR's new requirement for data submission to the EMBL data library: information for authors

Patricia Kahn and David Hazledine

EMBL Data Library, European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

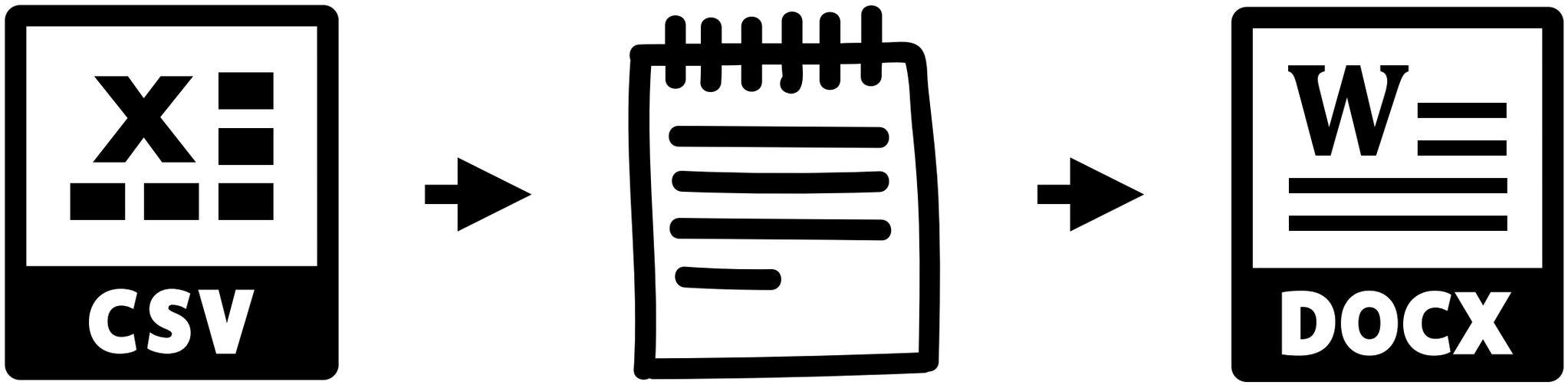
As of 1 January 1988, manuscripts submitted to Nucleic Acids Research (NAR) and containing or discussing sequence data must be accompanied by evidence that the data have been deposited with the EMBL Data Library. The background to this new policy and a general description of how it is being implemented were discussed in a recent NAR article (volume 15, number 18).

The following is a set of instructions describing how researchers can submit their data to the EMBL Data Library and obtain an accession number as quickly as possible.

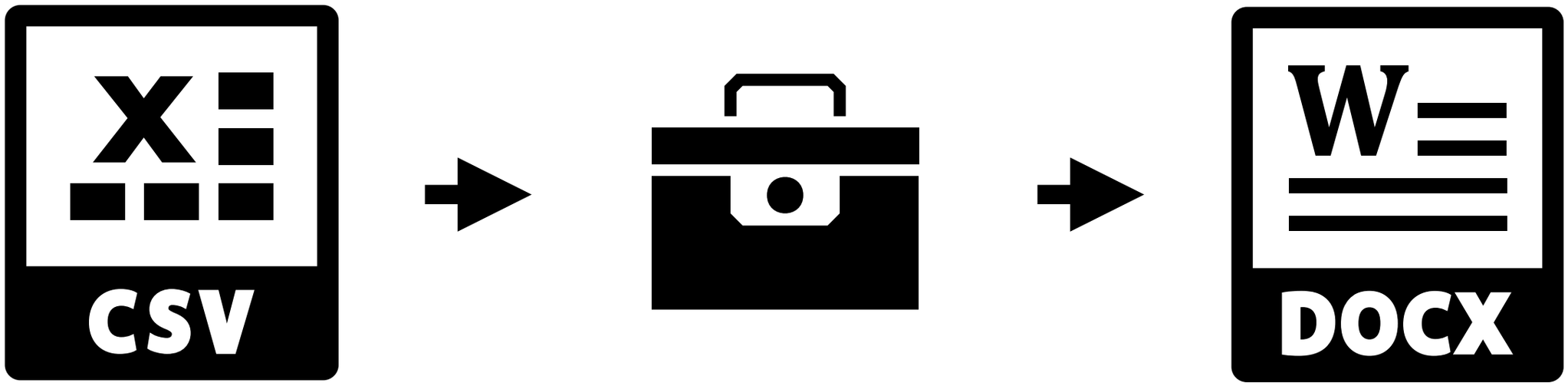


repository

publisher



repository author publisher



repository tools publisher

1. Find Data

repository
indexing service
reference manager

search functionality
expose metadata/data
for harvesting

2. Fetch Metadata for Citation

required citation metadata
on landing page
that are machine readable
using standard tags

Metadata on Landing Pages

</>

Place each dataset on a separate HTML page.
Export bibliographic data in HTML "<meta>" tags.
Use Highwire Press tags, PRISM tags, or Dublin Core tags.

https://scholar.google.no/intl/en/scholar/inclusion.html#indexing

<meta name="DC.creator" content="Vrana, Andrea" />
<meta name="DC.creator" content="Hotz-Boendermaker, Sabina" />
<meta name="DC.creator" content="Stämpfli, Philipp" />
<meta name="DC.creator" content="Hänggi, Jürgen" />
<meta name="DC.creator" content="Seifritz, Erich" />
<meta name="DC.creator" content="Humphreys, B. Kim" />
<meta name="DC.creator" content="Meier, Michael L." />
<meta name="DC.identifier" content="doi:10.5061/dryad.2h0q3" />
<meta name="DCTERMS.issued" content="2015-11-16" scheme="DCTERMS.W3CDTF" />
<meta name="DC.title" content="Data from: Differential neural processing during motor imagery of daily activities in chronic low back pain patients" />
<meta name="DC.type" content="Article" />

http://doi.org/10.5061/dryad.2h0q3

3. Cite Data in Manuscript

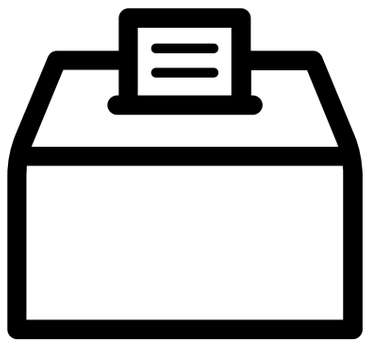
included in references not manually, but using reference manager adjusting to community practices

Figure 3. Multiple Alignment of Ten Conserved Motifs in the RAG1 Core Proteins and *Transib* TPases

The motifs are underlined and numbered from 1 to 10. Starting positions of the motifs immediately follow the corresponding protein names. Distances between the motifs are indicated in numbers of aa residues. Black circles denote conserved residues that form the RAG1/*Transib* catalytic DDE triad. The RAG1 proteins are as follows: RAG1_XL (GenBank GI no. 2501723, Xenopus laevis, frog), RAG1_HS (4557841, *Homo sapiens*,human), RAG1_GG (131826, *Gallus gallus*, chicken), RAG1_CL (1470117,Carcharhinus leucas, bull shark), RAG1_FR (4426834, Fugu rubripes, fugu fish).

http.//doi.org/10.1371/journal.pbio.0030181

4. Submit Manuscript



ideally with references as machine readable metadata in practice as formatted references in main text

Conclusions

Handoff of data citation metadata repository -> manuscript -> publisher is broken

Recommendations for Repositories

Machine-readable landing page

Guidance to authors to include data in references

Thank you!

Icons from the Noun Project
https://thenounproject.com/