Crossref

Measuring the impact of digital repositories

Jennifer Lin, PhD
@jenniferlin15
orcid.org/0000-0002-9680-2328

NITRD Big Data Workshop, 28 February 2017

FORCE11 - Joint Declaration of Data Citation Principles

1. Importance
2. Credit and Attribution
3. Evidence
4. Unique Identification
5. Access
6. Persistence
7. Specificity and Verifiability
8. Interoperability and flexibility

FORCE11 The Future of Research Communications and e-Scholarship + CODATA ICSU + DataCite FIND, ACCESS, AND REUSE DATA

+ contributions from 25 orgs (DCC, ORCID, Publishers, Repositories …)

↓

DC1 Data Citation Principles

Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [https://www.force11.org/group/joint-declaration-data-citation-principles-final].

Operationalizing the JDCCP

FORCE11 Data Citation Implementation group recommendations:

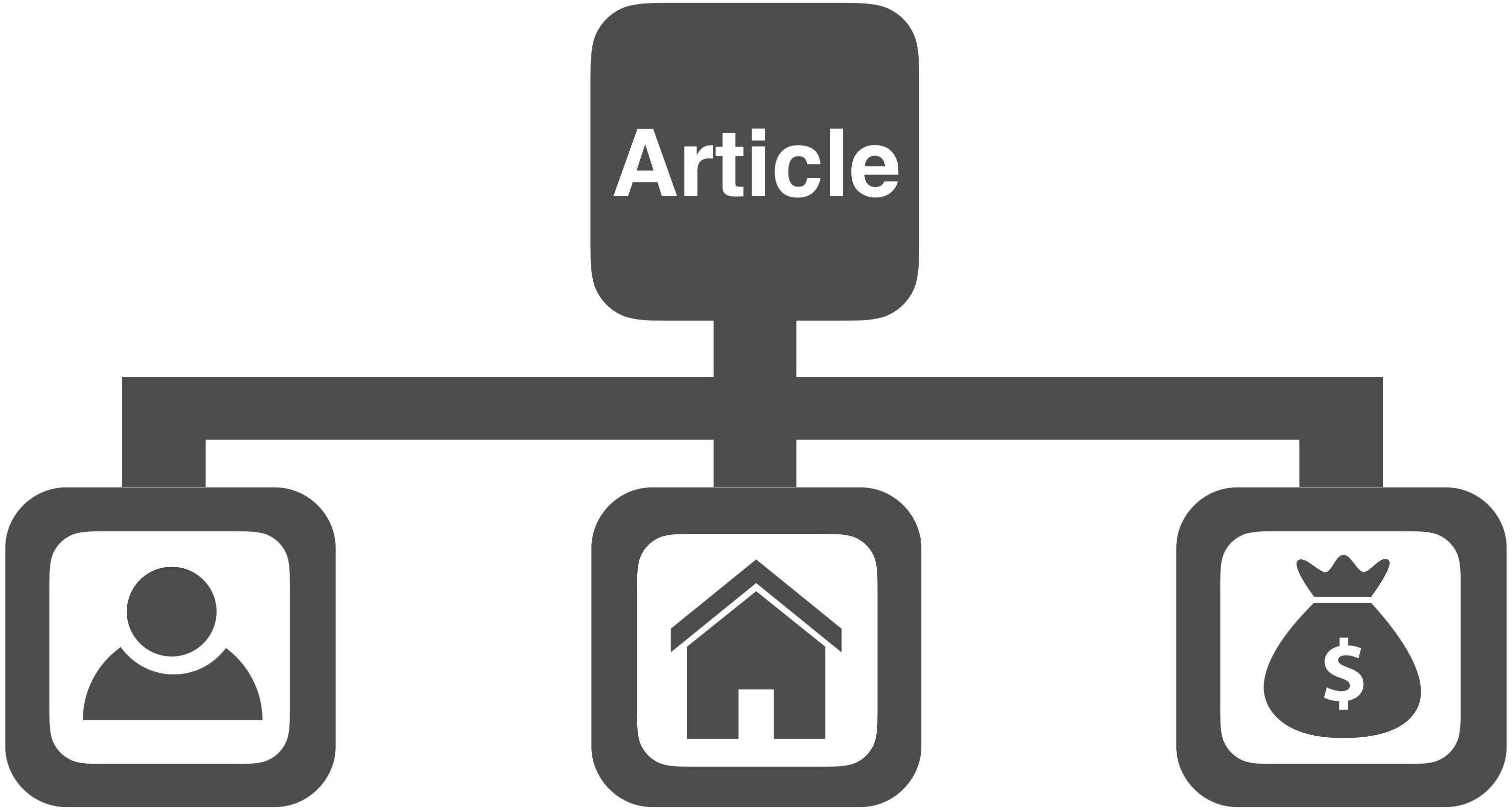
- identifier schemes
- identifier resolution behavior
- required metadata elements
- best practices for realizing programmatic machine actionability of cited data

Starr J, Castro E, Crosas M, Dumontier M, Downs RR, Duerr R, Haak L, Haendel M, Herman I, Hodson S, Hourclé J, Kratz JE, Lin J, Nielsen LH, Nurnberger A, Pröll S, Rauber A, Sacchi S, Smith AP, Taylor M, Clark T. (2015) Achieving human and machine accessibility of cited data in scholarly publications. PeerJ PrePrints 3:e697v4 https://doi.org/10.7287/peerj.preprints.697v4

Metadata enable connections

Crossref

Literature 🔗 Associated research entities

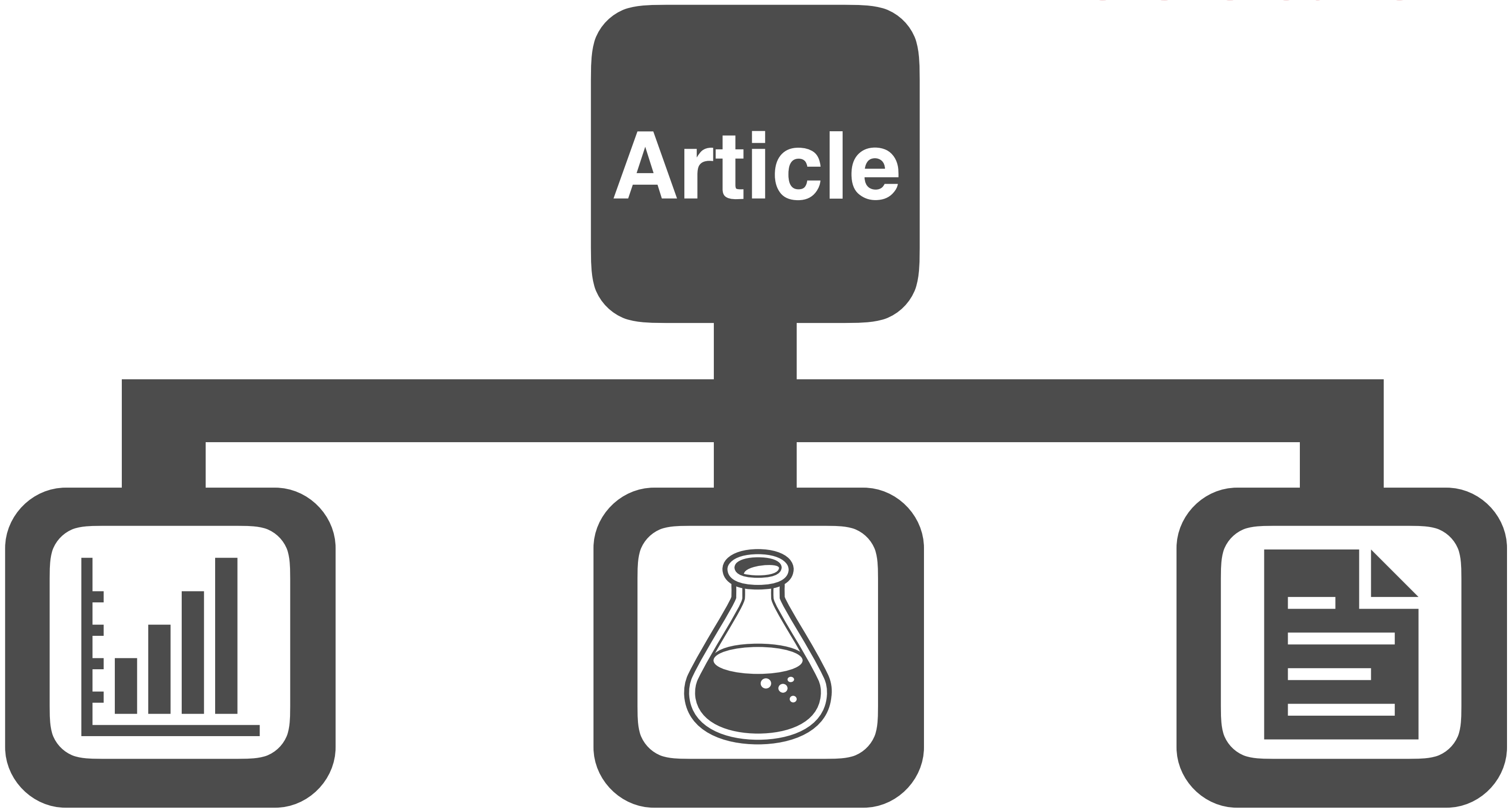


- authors
- collaborators

- reviewers
- editors

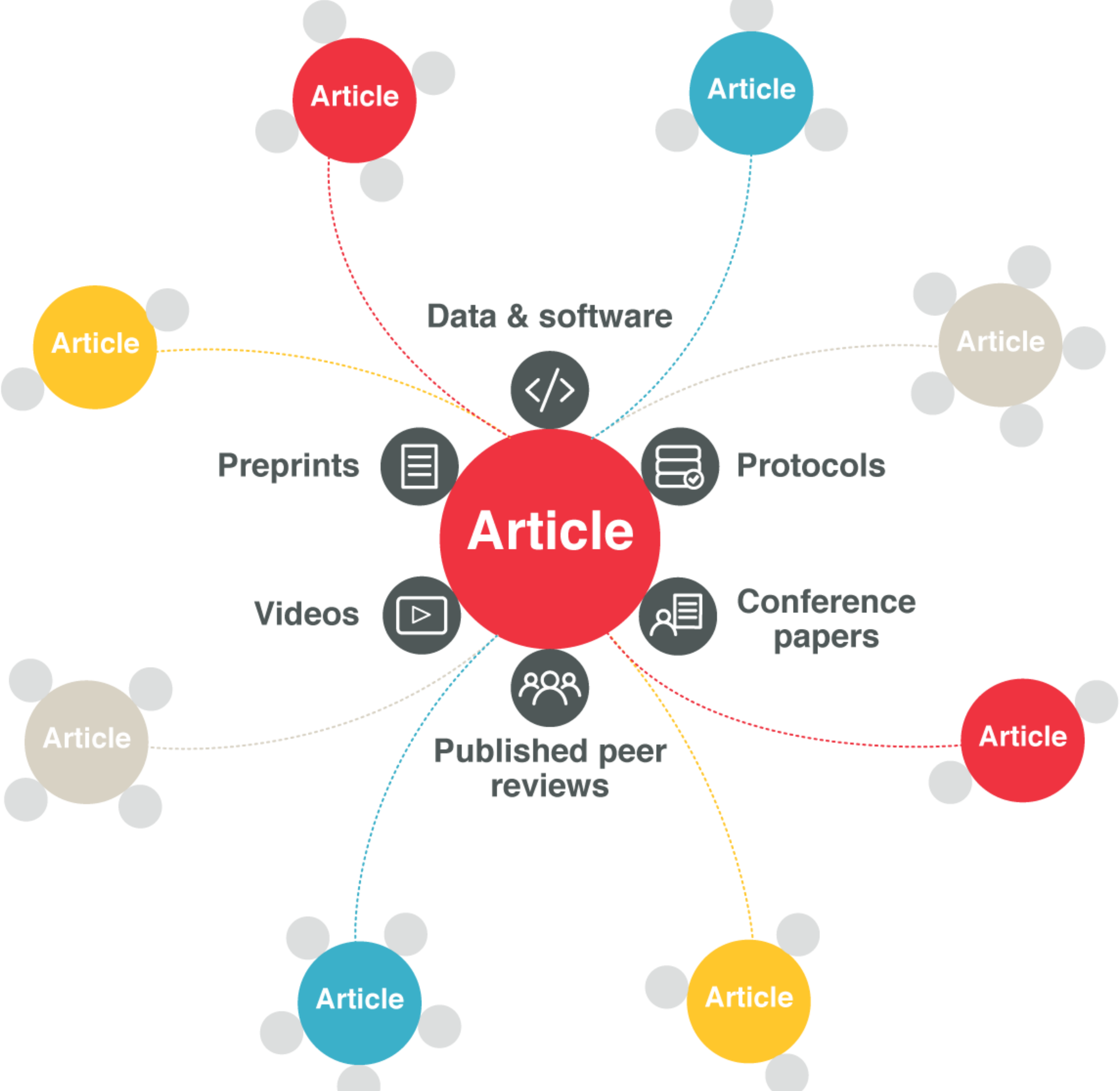
- funders
- affiliations

Literature 🔗 Associated research outputs

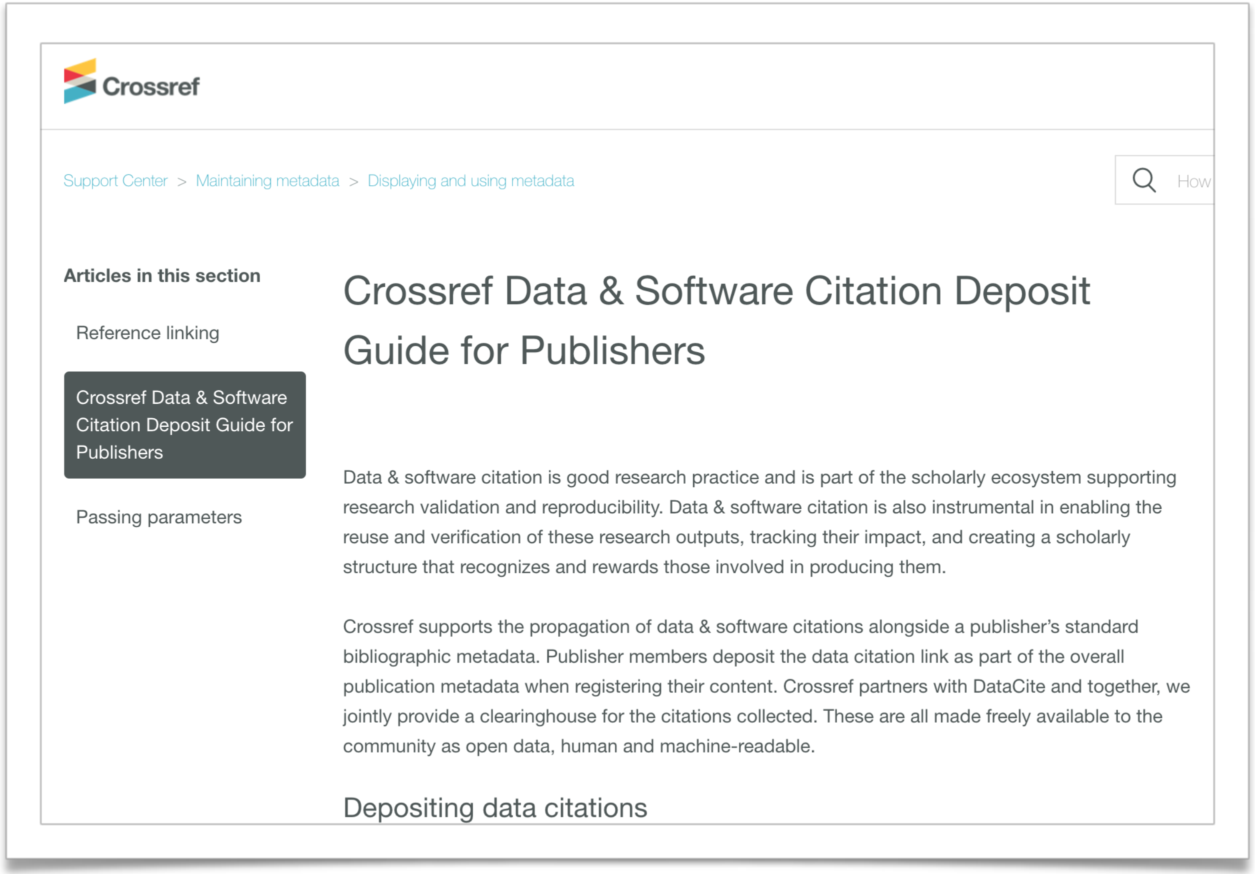


- datasets
- software
- protocols
- materials
- preprints
- conf papers
- peer reviews
- translations…

It’s all about relations: relationship types connect the article with its resources



Data - literature links: metadata needed



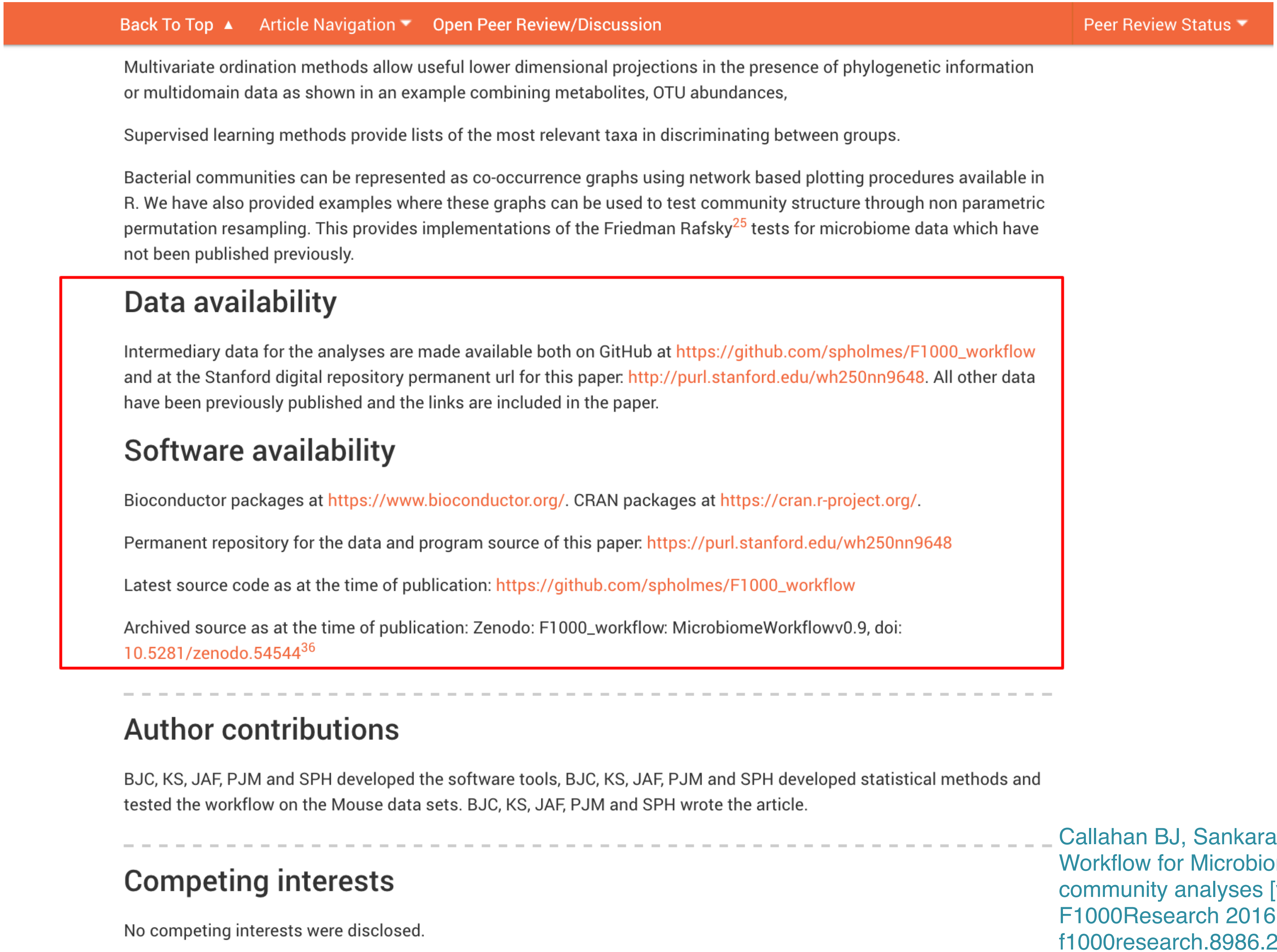
Crossref Data & Software Citation Deposit Guide for Publishers

Data & software citation is good research practice and is part of the scholarly ecosystem supporting research validation and reproducibility. Data & software citation is also instrumental in enabling the reuse and verification of these research outputs, tracking their impact, and creating a scholarly structure that recognizes and rewards those involved in producing them.

Crossref supports the propagation of data & software citations alongside a publisher's standard bibliographic metadata. Publisher members deposit the data citation link as part of the overall publication metadata when registering their content. Crossref partners with DataCite and together, we jointly provide a clearinghouse for the citations collected. These are all made freely available to the community as open data, human and machine-readable.

Depositing data citations

https://support.crossref.org



Multivariate ordination methods allow useful lower dimensional projections in the presence of phylogenetic information or multidomain data as shown in an example combining metabolites, OTU abundances,

Supervised learning methods provide lists of the most relevant taxa in discriminating between groups.

Bacterial communities can be represented as co-occurrence graphs using network based plotting procedures available in R. We have also provided examples where these graphs can be used to test community structure through non parametric permutation resampling. This provides implementations of the Friedman Rafsky[25] tests for microbiome data which have not been published previously.

Data availability

Intermediary data for the analyses are made available both on GitHub at https://github.com/spholmes/F1000_workflow and at the Stanford digital repository permanent url for this paper: http://purl.stanford.edu/wh250nn9648. All other data have been previously published and the links are included in the paper.

Software availability

Bioconductor packages at https://www.bioconductor.org/. CRAN packages at https://cran.r-project.org/.

Permanent repository for the data and program source of this paper: https://purl.stanford.edu/wh250nn9648

Latest source code as at the time of publication: https://github.com/spholmes/F1000_workflow

Archived source as at the time of publication: Zenodo: F1000_workflow: MicrobiomeWorkflowv0.9, doi: 10.5281/zenodo.54544[36]

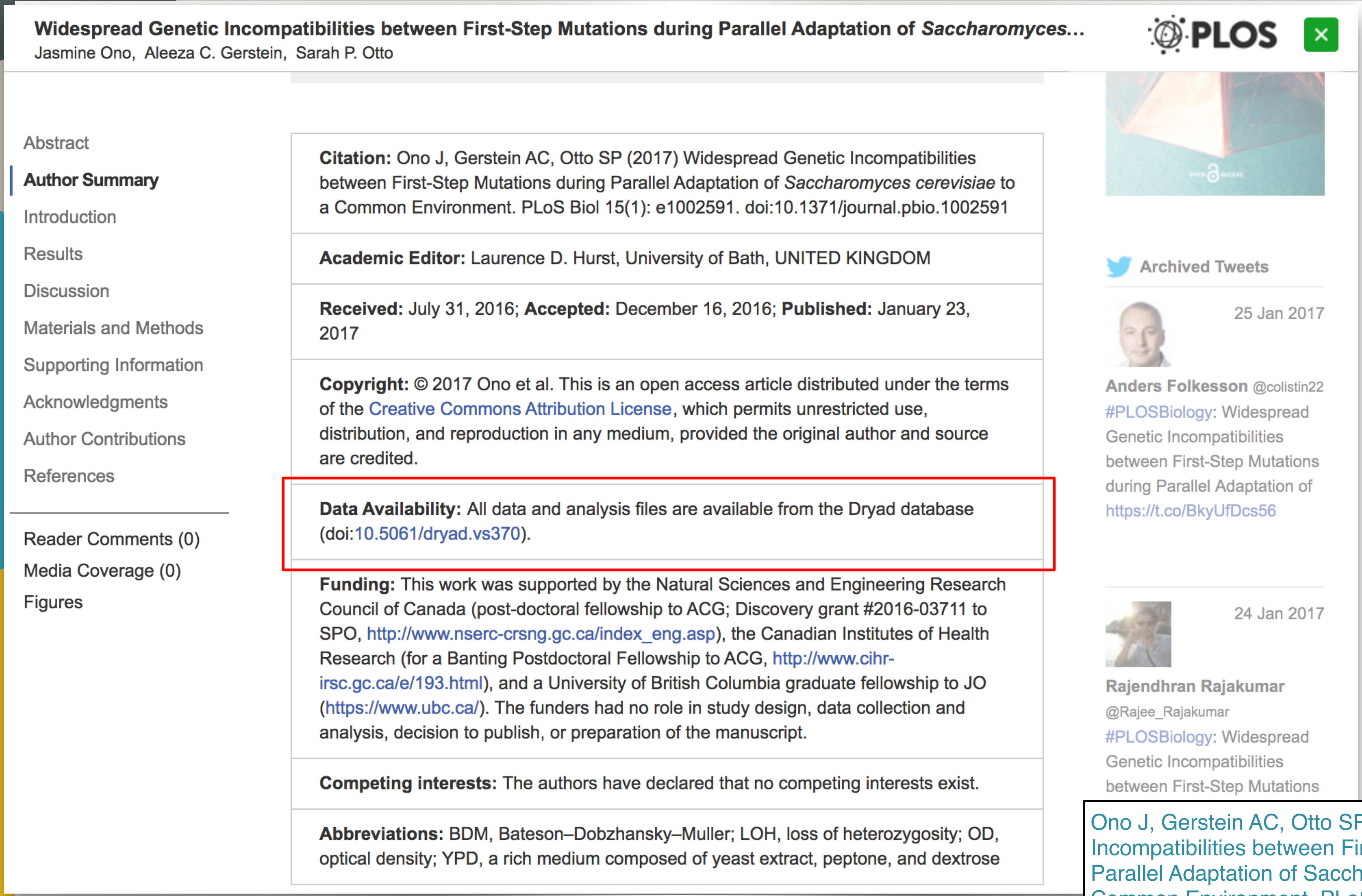
Author contributions

BJC, KS, JAF, PJM and SPH developed the software tools, BJC, KS, JAF, PJM and SPH developed statistical methods and tested the workflow on the Mouse data sets. BJC, KS, JAF, PJM and SPH wrote the article.

Competing interests

No competing interests were disclosed.

Callahan BJ, Sankaran K, Fukuyama JA et al. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses [version 2; referees: 3 approved]. F1000Research 2016, 5:1492. doi: http://doi.org/10.12688/f1000research.8986.2)



Widespread Genetic Incompatibilities between First-Step Mutations during Parallel Adaptation of *Saccharomyces*…
Jasmine Ono, Aleeza C. Gerstein, Sarah P. Otto

Citation: Ono J, Gerstein AC, Otto SP (2017) Widespread Genetic Incompatibilities between First-Step Mutations during Parallel Adaptation of *Saccharomyces cerevisiae* to a Common Environment. PLoS Biol 15(1): e1002591. doi:10.1371/journal.pbio.1002591

Academic Editor: Laurence D. Hurst, University of Bath, UNITED KINGDOM

Received: July 31, 2016; **Accepted:** December 16, 2016; **Published:** January 23, 2017

Copyright: © 2017 Ono et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: All data and analysis files are available from the Dryad database (doi:10.5061/dryad.vs370).

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada (post-doctoral fellowship to ACG; Discovery grant #2016-03711 to SPO, http://www.nserc-crsng.gc.ca/index_eng.asp), the Canadian Institutes of Health Research (for a Banting Postdoctoral Fellowship to ACG, http://www.cihr-irsc.gc.ca/e/193.html), and a University of British Columbia graduate fellowship to JO (https://www.ubc.ca/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

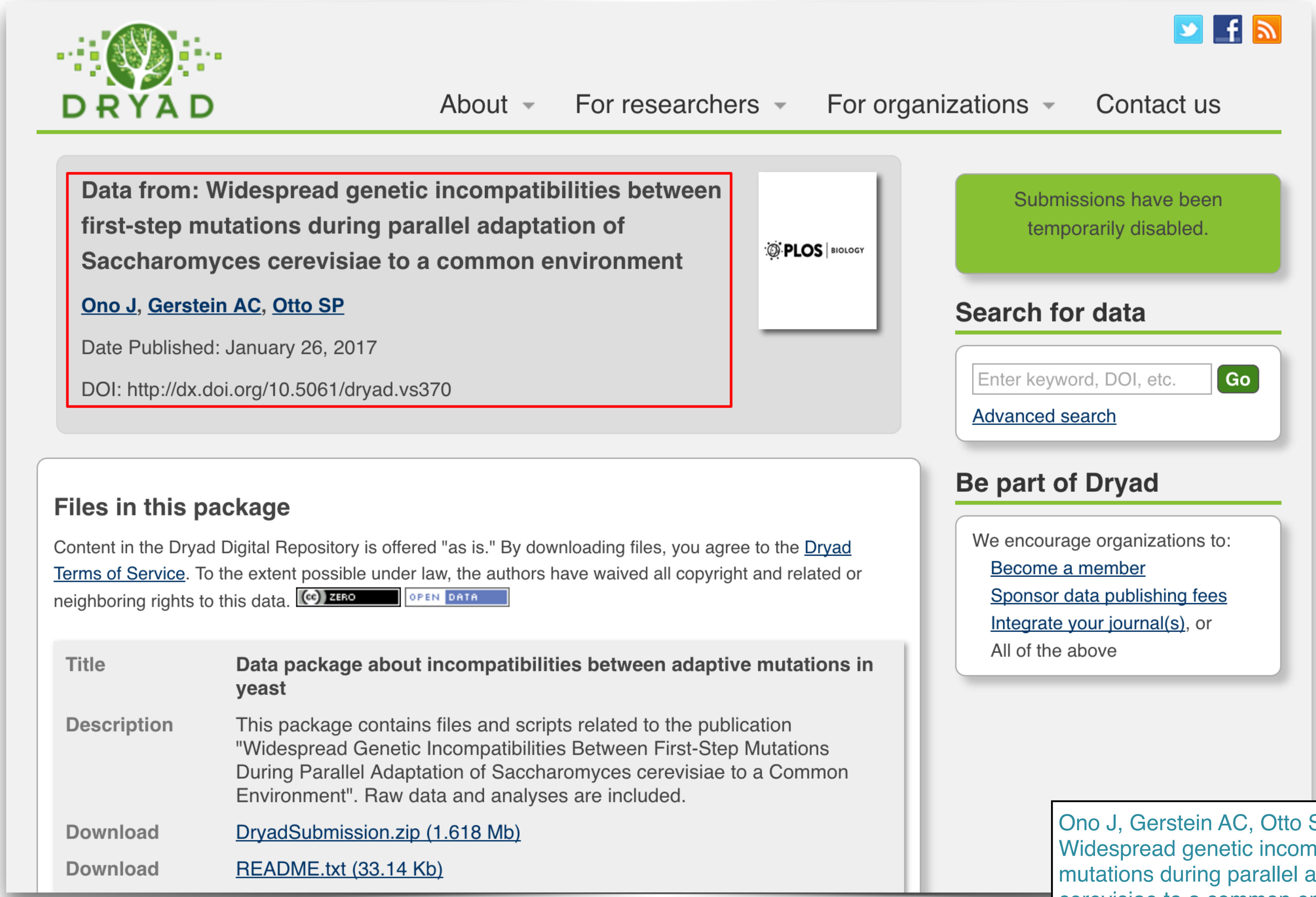
Competing interests: The authors have declared that no competing interests exist.

Abbreviations: BDM, Bateson–Dobzhansky–Muller; LOH, loss of heterozygosity; OD, optical density; YPD, a rich medium composed of yeast extract, peptone, and dextrose

25 Jan 2017
Anders Folkesson @colistin22
#PLOSBiology: Widespread Genetic Incompatibilities between First-Step Mutations during Parallel Adaptation of https://t.co/BkyUfDcs56

24 Jan 2017
Rajendhran Rajakumar @Rajee_Rajakumar
#PLOSBiology: Widespread Genetic Incompatibilities between First-Step Mutations

Ono J, Gerstein AC, Otto SP (2017) Widespread Genetic Incompatibilities between First-Step Mutations during Parallel Adaptation of Saccharomyces cerevisiae to a Common Environment. PLoS Biol 15(1): e1002591. doi: 10.1371/journal.pbio.1002591



Data from: Widespread genetic incompatibilities between first-step mutations during parallel adaptation of Saccharomyces cerevisiae to a common environment

Ono J, Gerstein AC, Otto SP

Date Published: January 26, 2017

DOI: http://dx.doi.org/10.5061/dryad.vs370

Search for data

Advanced search

Be part of Dryad

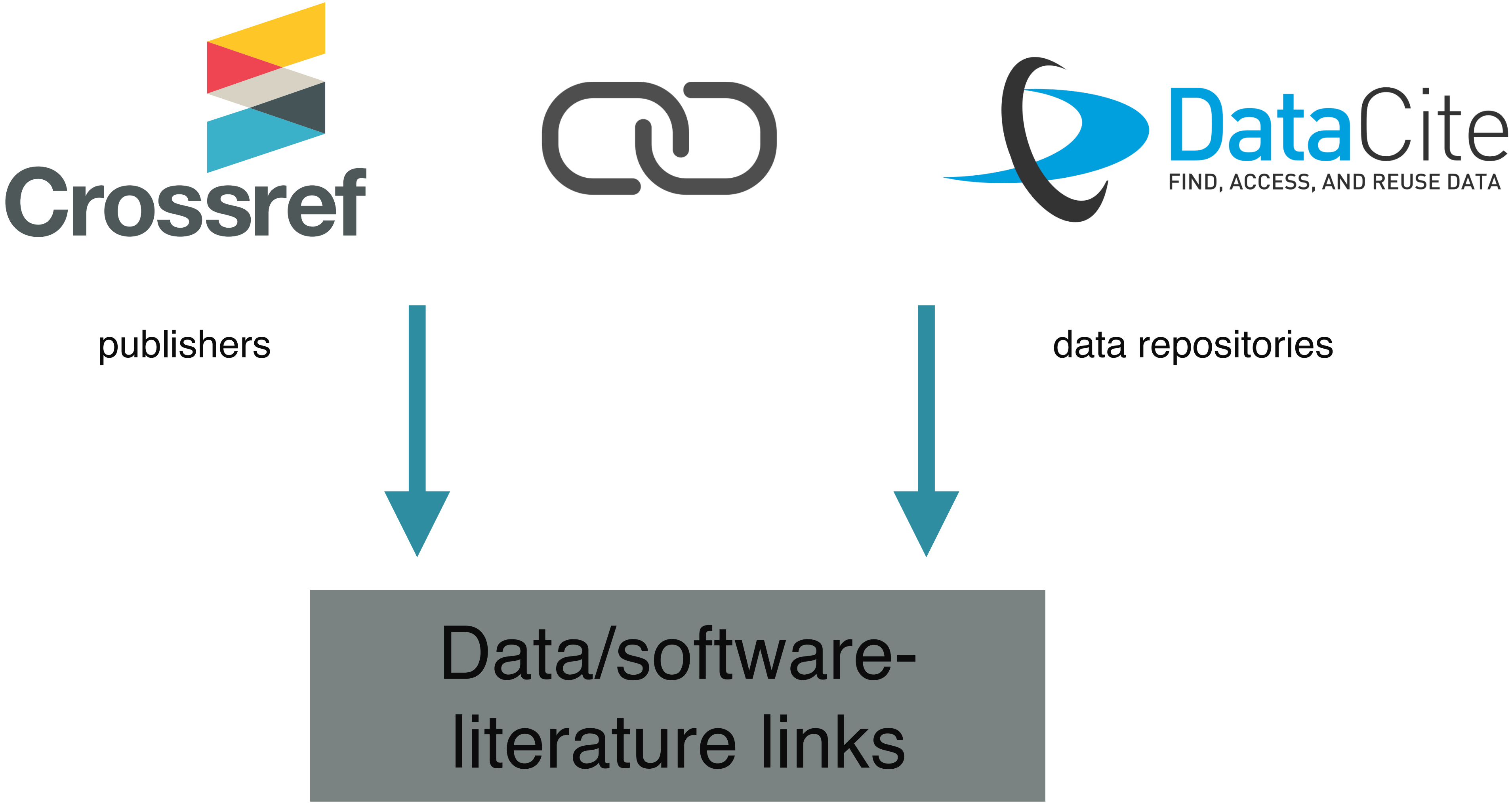
We encourage organizations to:
- Become a member
- Sponsor data publishing fees
- Integrate your journal(s), or
- All of the above

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the Dryad Terms of Service. To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

Title	**Data package about incompatibilities between adaptive mutations in yeast**
Description	This package contains files and scripts related to the publication "Widespread Genetic Incompatibilities Between First-Step Mutations During Parallel Adaptation of Saccharomyces cerevisiae to a Common Environment". Raw data and analyses are included.
Download	DryadSubmission.zip (1.618 Mb)
Download	README.txt (33.14 Kb)

Ono J, Gerstein AC, Otto SP (2017) Data from: Widespread genetic incompatibilities between first-step mutations during parallel adaptation of Saccharomyces cerevisiae to a common environment. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.vs370



Crossref

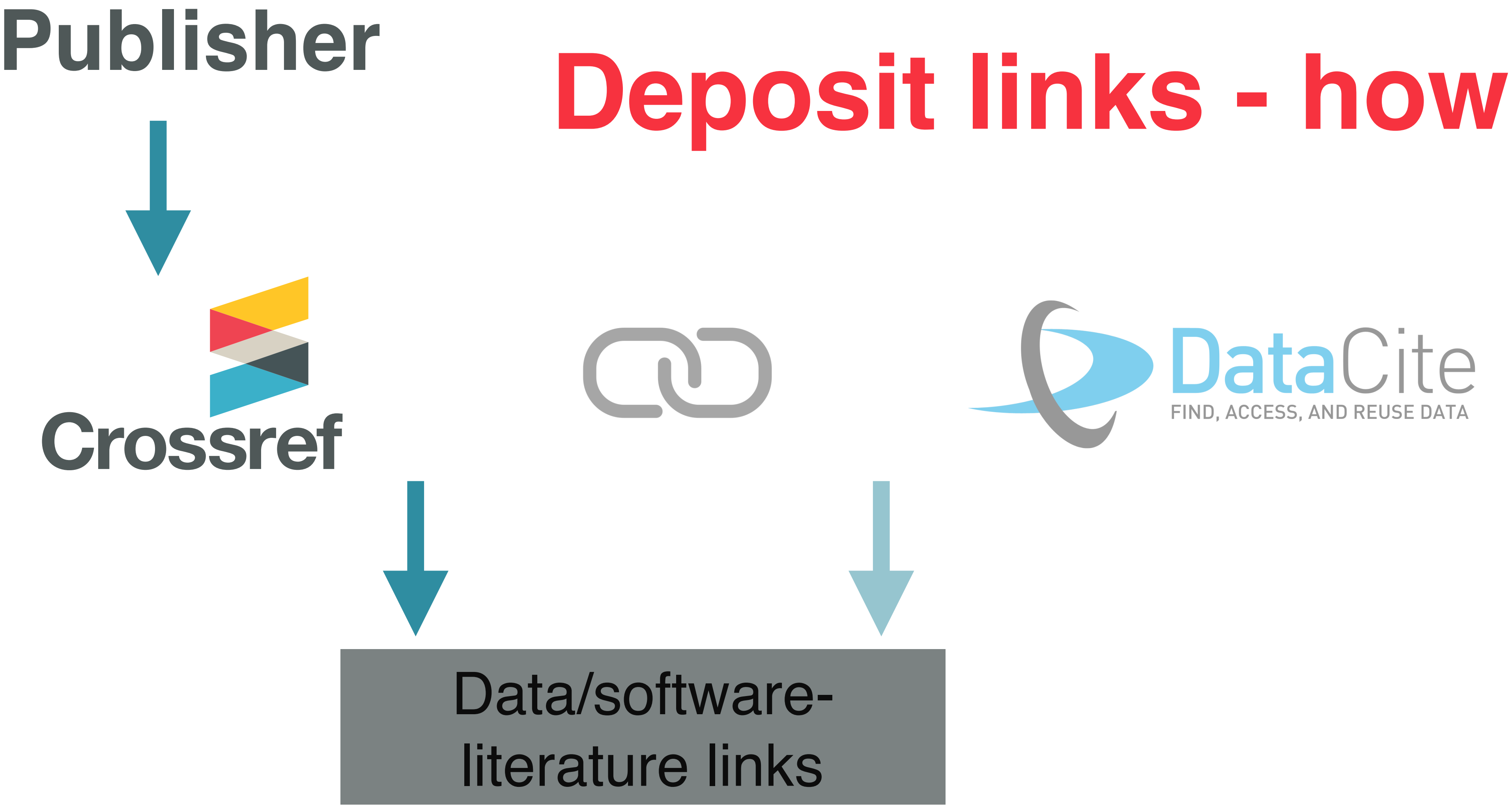
DataCite
FIND, ACCESS, AND REUSE DATA

publishers

data repositories

Data/software-
literature links

Deposit links - how?



Publisher

Crossref

DataCite

FIND, ACCESS, AND REUSE DATA

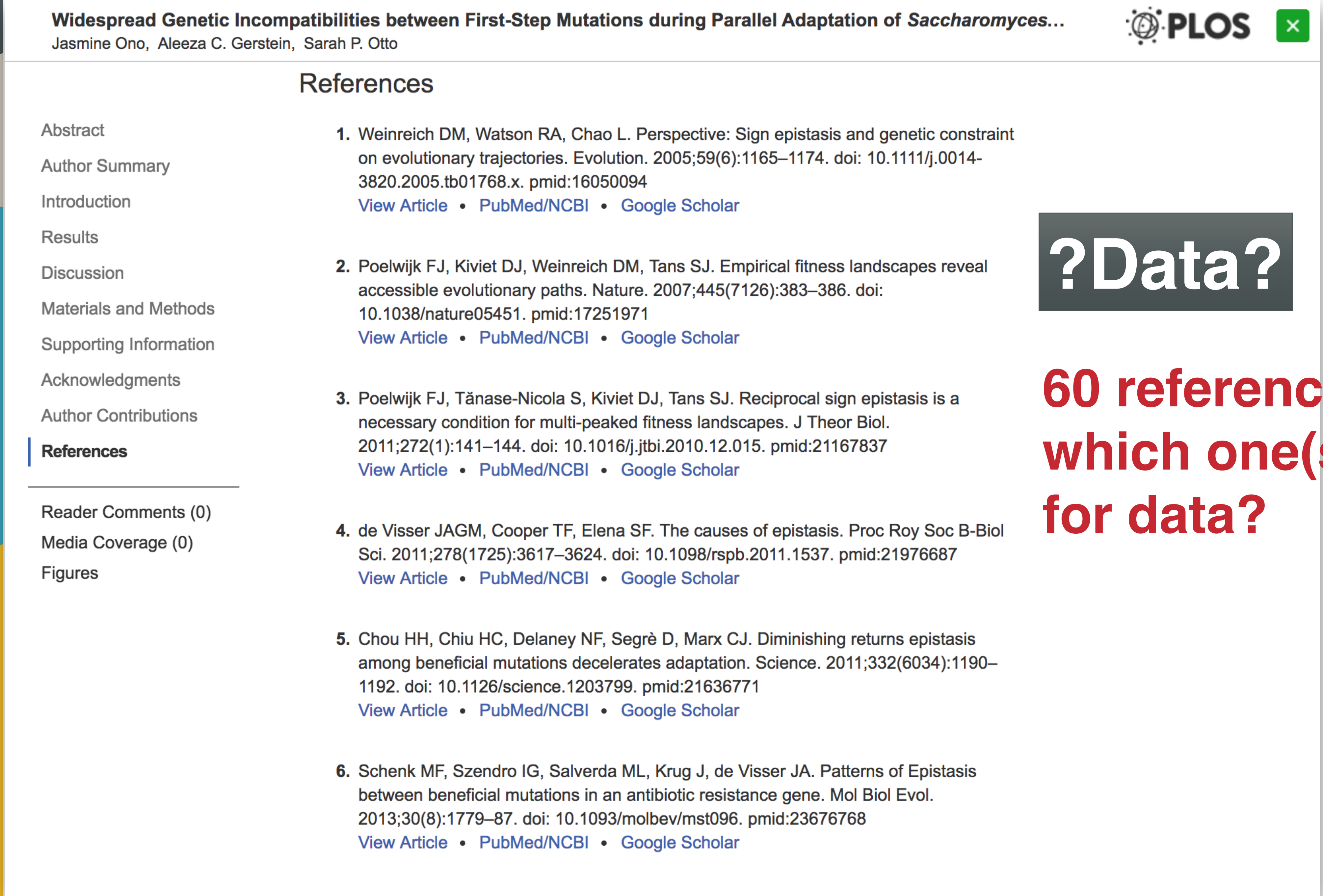
Data/software-
literature links

standard content registration process:
no workflow changes

Method 1: Bibliographic references

Dataset citation structure: [Author(s), Year, Dataset Title, Data repository or archive, Version, Global Persistent Identifier]

- Authors cite dataset or software in the reference section.
- Publishers deposit references to Crossref when registering content.
- Crossref checks every reference deposited for a DOI. If DOI is DataCite's, we make the link.



Widespread Genetic Incompatibilities between First-Step Mutations during Parallel Adaptation of *Saccharomyces*...

Jasmine Ono, Aleeza C. Gerstein, Sarah P. Otto

References

1. Weinreich DM, Watson RA, Chao L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 2005;59(6):1165–1174. doi: 10.1111/j.0014-3820.2005.tb01768.x. pmid:16050094
   View Article • PubMed/NCBI • Google Scholar
2. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;445(7126):383–386. doi: 10.1038/nature05451. pmid:17251971
   View Article • PubMed/NCBI • Google Scholar
3. Poelwijk FJ, Tănase-Nicola S, Kiviet DJ, Tans SJ. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J Theor Biol. 2011;272(1):141–144. doi: 10.1016/j.jtbi.2010.12.015. pmid:21167837
   View Article • PubMed/NCBI • Google Scholar
4. de Visser JAGM, Cooper TF, Elena SF. The causes of epistasis. Proc Roy Soc B-Biol Sci. 2011;278(1725):3617–3624. doi: 10.1098/rspb.2011.1537. pmid:21976687
   View Article • PubMed/NCBI • Google Scholar
5. Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science. 2011;332(6034):1190–1192. doi: 10.1126/science.1203799. pmid:21636771
   View Article • PubMed/NCBI • Google Scholar
6. Schenk MF, Szendro IG, Salverda ML, Krug J, de Visser JA. Patterns of Epistasis between beneficial mutations in an antibiotic resistance gene. Mol Biol Evol. 2013;30(8):1779–87. doi: 10.1093/molbev/mst096. pmid:23676768
   View Article • PubMed/NCBI • Google Scholar

?Data?

60 references total - which one(s) is/are for data?

Method 2: Relation type

Insert link into article metadata as part of standard content registration process. This includes:

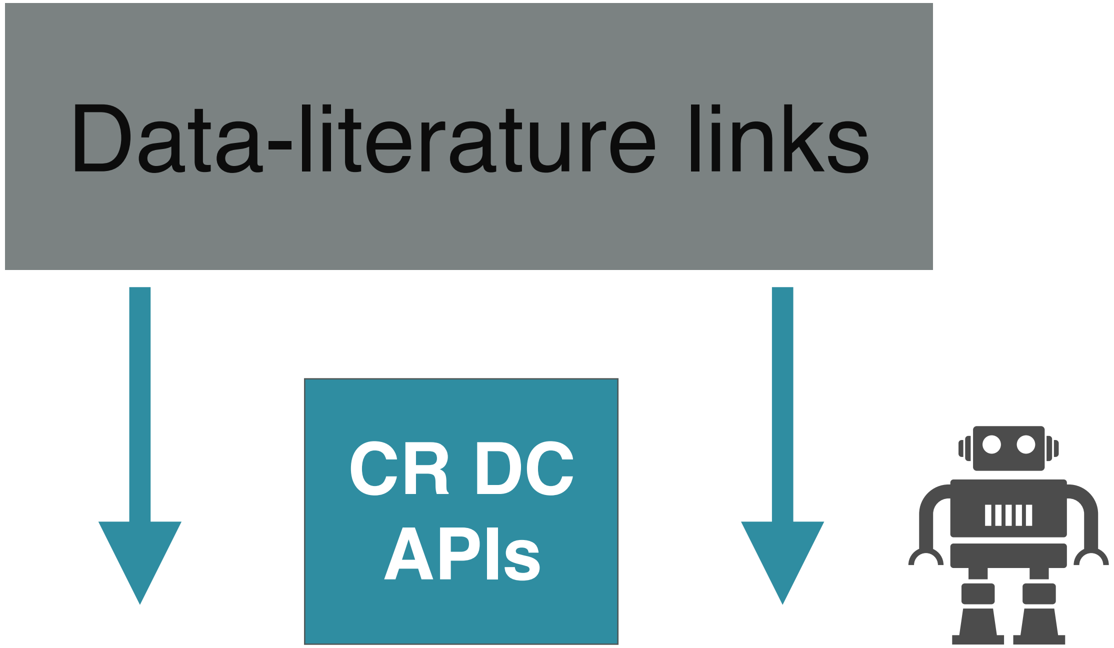
- description of dataset or software (optional)
- dataset or software identifier
- identifier type
- relationship type: <references>* or < isSupplementedBy>**

* Dataset produced by a different set of researchers or previously published

** Dataset generated as part of research results

Crossref

DataCite
FIND, ACCESS, AND REUSE DATA



Data-literature links

CR DC
APIs

research ecosystem robots
publishers data repositories
&&&…&&& (Scholix, etc.)

Data-literature links propagate to:

- Scholix
- Funders
- Institutions
- Archives & repositories
- Research councils
- Data centres
- Professional networks
- Patent offices
- Indexing services
- Publishing vendors
- Peer review systems
- Reference manager systems
- Lab & diagnostics suppliers
- Info mgmt systems
- Educational tools
- Data analytics systems
- Literature discovery services
- PID providers, registration agencies

Thank you

Jennifer Lin, PhD
jlin@crossref.org
@jenniferlin15
orcid.org/0000-0002-9680-2328

Crossref