Measuring the impact of digital repositories
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

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

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

Metadata enable connections
Literature

Associated research entities

• authors
• collaborators
• reviewers
• editors
• funders
• affiliations
Literature

Associated research outputs

Article

- datasets
- protocols
- preprints
- peer reviews
- software
- materials
- conf papers
- translations...
It's all about relations: relationship types connect the article with its resources.
Data - literature links: metadata needed

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 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**


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.545445

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

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)
Data/software-literature links

publishers  data repositories

Crossref  DataCite
Deposit links - how?

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.
References

   View Article • PubMed/NCBI • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

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
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