

# Data Sharing as Publication: Developing a Data Journal for Archaeology

Eric C. Kansa<sup>1</sup> and Sarah Witcher Kansa<sup>2</sup>

1. Open Context (<http://opencontext.org>) & University of California, Berkeley, [ekansa@ischool.berkeley.edu](mailto:ekansa@ischool.berkeley.edu)
2. The Alexandria Archive Institute (<http://alexandriaarchive.org>), [skansa@alexandriaarchive.org](mailto:skansa@alexandriaarchive.org)

More about  
this project



## Introduction

### Publishing models to promote data quality and sharing

Improved data dissemination can promote analytic rigor and transparency, reduce inefficiencies, and open new research opportunities for larger scale, multidisciplinary inquiry. To make this happen, researchers need to see greater incentives to participate in data dissemination.

This project proposes that a publication model for data sharing will increase scholarly participation in data dissemination and use. We are testing this theory by piloting a “data journal” for archaeology that will help set, communicate, and maintain expectations for quality of datasets in this discipline.

The data journal will combine the scholarly communications expertise and infrastructure of the California Digital Library (CDL), a unit that runs many of the University of California’s leading scholarly communications and data preservation efforts, with datasets from Open Context, a recognized open access data publication venue referenced by NSF and NEH for archaeology data management plans.

## Aims

### Making data a “first class citizen” in research

This study aims to increase participation in data dissemination while improving the quality and usability of published data. Data journal editorial review processes will help improve data quality and align data with disciplinary standards. We hope this will better set and communicate expectations of quality and improve professional recognition for data sharing. Through integration with citation infrastructures and services, a data journal will better integrate with the mainstream of scholarly communications.

The primary goals are:

1. Encourage the practice of data publication among researchers by providing a channel that builds upon existing scholarly communications reward and incentive structures.
2. Increase the impact of published data because datasets will be easier to find, have better documentation, and will earn more trust (having undergone review before inclusion in the data journal).

## Further Information



More about Open Context:  
<http://opencontext.org>



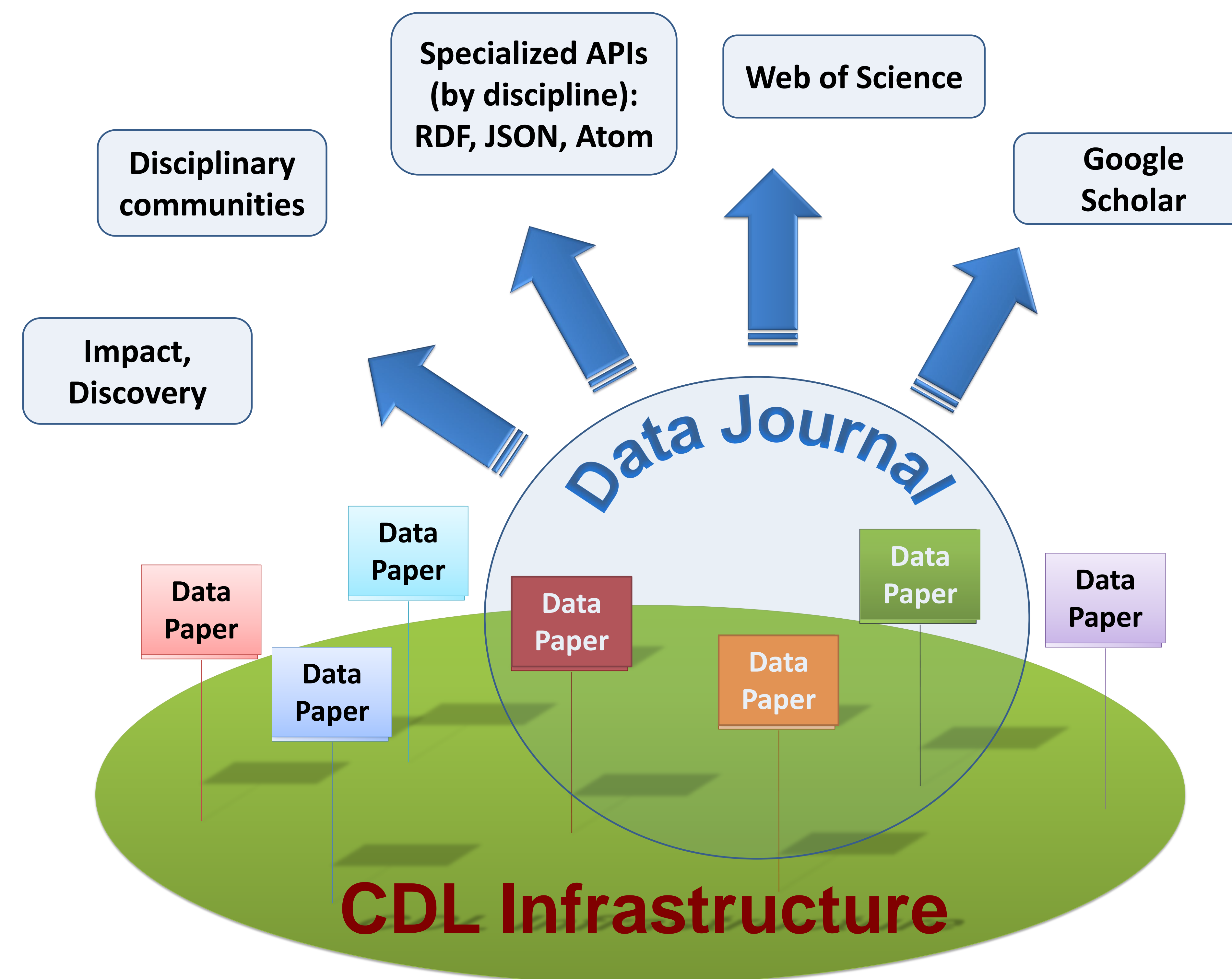
More about the California  
Digital Library: <http://cdlib.org>

## Defining “Data Papers” and “Data Journals”

CDL uses the terms “data papers” (datasets, controlled vocabularies, data analysis code / software) and “data journals” (editorially reviewed / thematic compilations of data papers) to model data dissemination workflows they seek to support. Data papers and data journals also represent different tiers of formalism and review:

**Tier 1- Data Papers.** A data paper will consist of a dataset with at least a minimum level of documentation needed to populate metadata fields for curation and use of the DataCite standard. Contributions of data papers will require minimum effort and no editorial review (unless requested). These low barriers to entry will hopefully encourage more researchers to participate in data dissemination and archiving.

**Tier 2- Data Journal.** A data journal will be an overlay journal that aggregates and improves upon select data papers. Editorial review will improve data quality and align datasets to disciplinary standards. Editors will also work with data contributors to improve data documentation to meet disciplinary needs. Researchers may opt to use data journals to participate in a more prestigious, higher impact channel.



## Acknowledgments

This study is part of a broader endeavor exploring data publication processes and workflows, carried out by the Alexandria Archive Institute and funded by a grant from the Alfred P. Sloan Foundation’s Digital Information Technology program.



## Process & Outcomes

### The purpose of editorial processes

To enter the pilot data journal, a data paper will undergo editorial review. Domain experts will check for methodological soundness. Editors will help contributors improve the quality and completeness of data documentation. A data paper will not be rejected on the basis of “significance” because the significance of a dataset may be hard to evaluate and may change over time. Tool kits to support editorial workflows include:

1. Google Refine (data cleaning): This tool helps users check for consistency, offers many edit functions, and logs changes to document edits and allow for roll-back to prior states.
2. Mantis (issue-tracking): Datasets can often be large and complex, with many internal dependencies. Improving data documentation and quality is analogous to debugging software.

### Part of the Web, not just on the Web

The CDL provides infrastructure for data papers and journals, offering archiving, versioning, and persistent identifiers. Editorial review will also align data to domain standards needed for powerful third-party Web Services / APIs (as offered by Open Context). Editorially-supervised standards alignment will also enable use of Linked Open Data methods.

## Conclusions

Editorial processes and the editorial board itself can perform important signaling roles to elevate the prestige and perceived value of data sharing for both contributors and users. The term “data journal” helps communicate these ideas, since researchers are familiar with the role of conventional journals in promoting quality of conventional publications.

Finally, this model offers significant sustainability advantages over discipline-specific data repositories. Data dissemination, semantics, services, and editorial models need to see continued innovation. Sustainability may be too steep a requirement for experimental, grant-funded projects. Individual publishers may come and go, but library infrastructure will curate content. The proposed approach decouples data curation and preservation from small, experimental efforts to explore innovative models of data dissemination.

## Related Literature

*Archaeology 2.0 and Beyond: New Tools for Collaboration and Communication*, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA. Online at: <http://escholarship.org/uc/item/1r6137tb>.

Griffiths, Aaron (2009) The Publication of Research Data: Researcher Attitudes and Behaviour. *International Journal of Digital Curation* 4(1).

Kunze, John A, Patricia Cruse, Rachael Hu, Stephen Abrams, Kirk Hastings, Catherine Mitchell, and Lisa R. Schiff (2011) *Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines*. White paper report for the Gordon and Betty Moore Foundation.