Our Guidelines to Data Publishing

Mercè Crosas, Ph.D.
Twitter: @mercecrosas
Director of Data Science
Institute for Quantitative Social Science, Harvard University

RDA 3rd Plenary, Dublin, March 26, 2014
Policies? Or Guidelines with supporting Tools?

The **Dataverse repository** hosted at Harvard is **open to all researchers and all disciplines**.

Based on almost a decade of data sharing and data infrastructure and tool building, Dataverse has developed a set of **guidelines to standardize Data Publishing**.
Guidelines for Data Publishing*

- A published dataset must contain at a minimum a required subset of citation metadata and one or more data files.

- A dataset is assigned a data citation compliant with the Joint Declaration of Data Citation Principles.

- We encourage and facilitate sharing sufficient information (domain-specific metadata and complementary files) to enable reusability.

- We encourage and facilitate free, accessible data by assigning default open data licenses to both metadata and files, but allowing terms of use and restrictions to files when needed (metadata must remain open).

- Once a dataset is published, it cannot be unpublished. Versioning is supported for changes in files or metadata. Deaccession is supported when a dataset must be retracted, but citation information remains.

- A dataset might be associated to zero, one or more publications.

*Based on Dataverse 4.0 (coming this Spring)
Guidelines for Connecting Journals to Data

We encourage and facilitate integration between Journals and Data through Dataverse by:

- **Supporting multiple workflows:**
  - Seamless integration with Journal Publishing system
  - Direct submission to a Journal Dataverse
  - Listing data citation assigned by Dataverse in article’s references

- **Supporting multiple journal data policies:**
  - Dataset may be either required or **optional (most common)** upon article submission
  - Dataset may be published before, **at the time (most common)** or after article is published
  - Dataset is peer-reviewed during article review (**less common**)
  - Dataset is published even if article is not published (**not common**)