The Data Publishing Landscape

Mercè Crosas, Ph.D.
Chief Data Science and Technology Officer
Institute for Quantitative Social Science
Harvard University

@mercecrosas
The 60s: First social sciences data archives

The 70 and 80s: First life sciences large databases
90s to Now: Growth of domain-specific and generic data repositories

- Dryad (2006)
- Dataverse (2009)
- Figshare (2011)
- Zenodo (2013)

- Pangaea (1995)
- DataOne (2009)
- Dataverse (2013)
- Zenodo (2014)

- Dublin Core (1995)
- DataCite (2009)
- Joint Declaration of Data Citation Principles (2014)

...to more than 1000 Data Repositories

re3data.org

1500 research data repositories ... and counting
Growth in scholarly and peer-reviewed serials

- 80,000 total Scholarly Journals
- 33,000 peer reviewed
“Research data publishing is the release of research data, associated metadata, accompanying documentation, and software code (in cases where the raw data have been processed or manipulated) for re-use and analysis in such a manner that they can be discovered on the Web and referred to in a unique and persistent way.”

Ways of Publishing Data

Journal’s data policy

Scholarly Article

Data in Repository

Data Descriptor or Data Paper

Data in Repository

Published Dataset in Repository

Scholarly Article

Scholarly Article
The FAIR Guiding Principles for scientific data management and stewardship

Findable Accessible
Interoperable Reusable