Data Citation: From Principles to Implementation

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

National Academy of Sciences Journal Summit, March 20, 2014
Decades of research and practices in data citation

Consolidated to a single set of Principles

By a synthesis group representing 25+ organizations

Driven by the premise that:

"sound, reproducible scholarship rests upon a foundation of robust, accessible data"

and

"data should be considered legitimate, citable products of research"
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

Full Principles: https://www.force11.org/datacitation

Endorsement: https://www.force11.org/datacitation/endorsements
The Dataverse generates a Data Citation for each deposited dataset compliant with the Principles:

- **Principle 2**: Credit and Attribution
- **Principle 4, 5, 6**: Unique Id, Access, Persistence
- **Principle 7**: Specificity and Verifiability
- **Principle 8**: Interoperability and flexibility: Dataverse exports all citation metadata in XML, JSON formats
Implementation Suggestions for Publishers

- Upgrade data citation to references section [Principle 1: Importance]
- In article, cite data by claim [Principle 3: Evidence]
- Provide guidelines for authors based on Principles, but customized to each journal [Principle 8: Interoperability and Flexibility]
- Interoperate with, or recommend, Data Repositories compliant with the Principles
- Build tools to access machine-readable metadata from datasets

Want to be involved?

Join the Data Citation Implementation group:

https://www.force11.org/datacitationimplementation
Remaining Challenges

- Challenges of Provenance: what is the chain of ownership and transformations to the data?
- Challenges of Identity: what should be cited? at what level of granularity and versioning for large, dynamic datasets?
- Challenges of Attribution: How do you support attribution for hundreds/thousands contributors?

Altman M., Crosas M., 2014, “The Evolution of Data Citation: From Principles to Implementation” IASSIST Quarterly, In Press