For a more detailed explanation, with justification for our recommendations, see the paper currently in process, "***Achieving human and machine accessibility of cited data in scholarly publications***"

http://peerj.com/preprints/697/

Human & Machine Actionable Data Citations

The (Really) Short Summary:

Step 1: Define groupings of data that you would like to be cited.

(note: this will be the hardest part of the process for some communities)

Step 2: Mint identifiers for each grouping

- Identifiers should be fully qualified URLs.
- If using some local scheme, include the resolver as part of the URL.

Step 3: Create HTML Landing Pages for each grouping.

- Each page should have:
  - Dataset identifier
  - Title
  - Creator
  - Publisher or Contact
  - Release Date or Year
  - Version
  - a description of the dataset
  - License Information
  - Persistence Statement
  - Attribution to Contributors
  - ... whatever else you like

Step 4: Make a machine-readable version of the landing page.

- Use XML, JSON, RDF, or whatever your community prefers.
- Use the W3C DCAT vocabulary for interoperable descriptions, ORCID for contributors, plus any other metadata standards for the data's intended communities.

Step 5: Use Content Negotiation and Web Linking to connect the HTML and machine-readable format:

Content Negotation :

- Serve the machine-readable version if it's requested in an HTTP `Accept` header.
- You can use the `Accept-Language` header if you want to be multi-lingual, too.

Web Linking:

- Return HTTP `Link` headers with URLs to the alternate representations of the landing page.
- On the HTML landing page, include HTML `<link>` elements with the same information.
- For more complex relationships, link to an OAI-ORE resource map.

For a copy of the poster (which has a few references not yet in the paper) or this handout:

http://commons.esipfed.org/node/7768

Joint Declaration of Data Citation Principles

Preamble

Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.

In support of this assertion, and to encourage good practice, we offer a set of guiding principles for data within scholarly literature, another dataset, or any other research object.

These principles are the synthesis of work by a number of groups. As we move into the next phase, we welcome your participation and endorsement of these principles.

Principles

The Data Citation Principles cover purpose, function and attributes of citations. These principles recognize the dual necessity of creating citation practices that are both human understandable and machine-actionable.

These citation principles are not comprehensive recommendations for data stewardship. And, as practices vary across communities and technologies will evolve over time, we do not include recommendations for specific implementations, but encourage communities to develop practices and tools that embody these principles.

The principles are grouped so as to facilitate understanding, rather than according to any perceived criteria of importance.

1. Importance

Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.

2. Credit and Attribution

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

3. Evidence

In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

INTERPRETATION & ANALYSIS

PREAMBLE:

- **Reproducible science relies on knowing the evidence (data) used**
- **Producing data is an important contribution**
- **Citing data is important for the scientific record & potential re-use of data**

PRINCIPLES:

- **This covers citation, not data archiving.**
- **Implementation is a later effort (currently ongoing) & will evolve over time**
- **Practices will vary to fit their community**

1. IMPORTANCE

- The data used as evidence should be given credit for their contribution.
- Communities should consider a person's work in producing good data for others to use when considering tenure, promotion & grants.

2. CREDIT AND ATTRIBUTION

- **There is no simple "author" for data**, and citing a "first results" or "instrument" paper doesn't give proper credit to people who may come in later and give significant contributions to the calibration or other understanding of the data.

3. EVIDENCE

- The **data used to support your research should be cited in the reference list.**
- You should **link the data being used as evidence** near the claim being made; depending on the journal, this may be inline text, a footnote, or a caption to a plot or table.

DC[1]

[1] Data Citation Principles

4. Unique Identification

A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community.

5. Access

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.

6. Persistence

Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.

7. Specificity and Verifiability

Data citations should facilitate identification of, access to, and verfication of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verfiying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.

8. Interoperability and flexibility

Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities.

7. VERSIONING AND GRANULARITY

- If there are formal releases, assign an identifier to each one, so researchers can **cite a specific version**. If it's not available, citations should **include an access date**.
- If you didn't analyze all of the data, **describe what portion you used** (eg, date, spectral or spatial ranges; specific observing modes; or any other filtering or subsetting.)

Thursday @ 1:30pm : Workshop on Dynamic Data Citation, to discuss RDA work on interoperable subsetting standards.

4. UNIQUE IDENTIFICATION

- For this whole system to work, we need **cross-discipline identifiers.**
- **Identifiers should be fully qualified URLs to a resolver for the identifier.**
- DOIs would allow us to use existing bibliographic tools to track the use of data, **reduce the work needed to prepare for Senior Reviews**, and **find uses of our data by other communities.**

5. ACCESS

- Citations do not need to (and should not, in our community) link directly to the data. **DOIs should link to a webpage with information about the data** so that people can make an informed choice before potentially downloading terabytes of data that isn't useful to them.
- These "landing pages" can be updated to **provide links to current documentation, software, related data** (eg. alternate processed forms or from complementary missions), published papers using the data, and whatever metadata is approrpriate for the community.

6. PERSISTENCE

- **Even if the data goes away** (replaced by better data, removed due to security or budget, or lost by accident), **the landing page remains**, so we **do not have a gap in the scientific record.**
- When appropriate, this "tombstone page" should describe why data was removed, and link to possible replacements or alternatives (eg, better calibrated versions).

8. INTEROPERABILITY AND FLEXIBILITY

- Every journal / community cites things a little bit differently, and has different metadata requirements. The data citation community is working towards a **universal framework that each community can extend for their specific needs.**

To endorse the principles, or for more information, visit

`http://force11.org/datacitation`