Credit for Research Artifacts

Mark A. Parsons

Research Artifact Citation Cluster
ESIP Summer Meeting
21 July 2020

What do you do?
Background

• ESIP developed data citation guidelines more than a decade ago and updated them last year:


• ESIP also recently developed guidelines for citation of software and services:


• The basic issues are resolved, but some issues remain:

Issues stem from the definition or intent of citation

Data citation is a **reference** to data for the purpose of **credit attribution** and facilitation of **access to the data**.

Task Group on Data Citation Standards and Practices, CODATA-ICSTI. 2013. “Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data.” Data Science Journal 12(0) (0): CIDCR1–CIDCR75. [https://dx.doi.org/10.2481/dsj.OSOM13-043](https://dx.doi.org/10.2481/dsj.OSOM13-043).

How does this apply to other “research artifacts”? 
Contributor Roles Taxonomy (CRediT)

A taxonomy of roles adopted by many journals for articles. [https://casrai.org/credit/](https://casrai.org/credit/)
Can they apply to other artifacts?

- Conceptualization
- Data curation
- Formal Analysis
- Funding acquisition
- Investigation
- Methodology
- Project administration
- Resources
- Software
- Supervision
- Validation
- Visualization
- Writing - original draft
- Writing - review & editing
Research Artifact Citation Cluster

- The cluster is trying to take a step back and consider the broader issue of addressing all (most) of the concerns embedded in citation of all (most) types of research artifacts.

- We identified 32 different RAs and began to separate the myriad concerns that citation tries to address including reference, credit, reproducibility, impact, tracking provenance, etc.

- We clustered the objects into 9 general categories and began to consider different concerns or use cases.

- At the winter meeting we assessed the “reproducibility use case”— when do different research objects need to be identified to ensure reproducibility or validity of a result?

- In this session, we want to begin to address the “credit use case” — what roles should be credited where for different research objects.
Today

• **GOAL:** Identify whether the roles of CRedit apply and where they should be captured for different research artifacts.

• Artifact Clusters (change your Zoom name to show where you want to go, e.g. “2 - Mark Parsons”):
  2 - Data and related objects — Mark Parsons
  3 - Software and related objects — Dan Katz
  4 - Samples and physical artifacts — Sarah Ramdeen
  5 - Ontologies and vocabularies — Ruth Duerr
  7 - Instruments and facilities — Rama Ramapriyan

• Each team will walk through their items and the possible location for crediting a particular role for a particular artifact. Four options for where the role could be recorded (Do NOT worry about the order of listing the people in the citation or wherever):
  1. In the citation
  2. In the “acknowledgements” (landing page, metadata, paper, etc.)
  3. Elsewhere (deep in metadata or data, etc.)
  4. Not applicable
Process

Record everything in the Credit tab of the Objects and Concerns Spreadsheet: https://bit.ly/2N3SN1e

• Step 1. Work independently and quietly for five minutes. Check to see if you agree with the definition of each artifact. Just a simple yes or no. Start to review the roles and how they apply to the different artifacts. Take notes!

• Step 2. Quick vote on whether each definition is OK. If a strong majority (~2/3) accept the definition go for it. Otherwise try and come to an agreement on a new definition, but do not spend too much time on this.

• Step 3. Walk through each artifact and determine where each role belongs in the 4 categories above. For each artifact take a couple minutes for everyone to quietly enter their opinion in the matrix. Then discuss.

• Step 4. Consider key messages or takeaways you want to present to the broader group