

[January 8, 2015](#)[January 8, 2015](#)

Project CRediT and Contributorship Roles

I've been involved in Project CreDIT (<http://credit.casrai.org>) a bit, since I read an article on it in Nature:

Allen, L. et al. 17 April 2014. Publishing: Credit where credit is due. *Nature* 508, 312–313, [doi \(Digital Object Identifier\):10.1038/508312a](#) (<http://dx.doi.org/10.1038/508312a>)

and wrote to the authors, since it seemed to overlap some of the ideas I'd been thinking about called transitive credit:

Katz, D.S. 2014. Transitive Credit as a Means to Address Social and Technological Concerns Stemming from Citation and Attribution of Digital Products. *Journal of Open Research Software* 2(1):e20, doi:10.5334/jors.be (<http://dx.doi.org/10.5334/jors.be>)

This led to me joining the project working group (<http://credit.casrai.org/working-group/>), participating in a series of phone calls to discuss the specific taxonomy of contributorship roles (<http://credit.casrai.org/proposed-taxonomy/>), and attending a workshop (<http://www.eventbrite.ca/e/project-credit-workshop-tickets-10314211083>) that was intended to promote discussion about and trials of the roles.

The thing that makes this interesting to me is that these roles ideally could be used to describe who contributed and how they contributed to any scholarly product in any domain, for example, papers, software, data, books, physical objects, etc. As you might imagine, it's not immediately clear if the contributorship taxonomy (<http://credit.casrai.org/proposed-taxonomy/>) can be both general enough to cover all products and also useful enough to make people want to use it.

While thinking about this at the workshop, and in some discussions afterwards, I realized that there is a time element that probably needs to be considered. For a paper, at least, I think about the work that is done (and needs to be recognized in the contributorship model) as having three phases:

1. **planning**, including: conceptualization, methodology, resources, funding acquisition, etc.
2. **doing**, including investigation, formal analysis, software (building code for research), writing (original), validation (authors ensure the answers are right), etc.
3. **reporting**, including data curation, software (packaging and archiving for reproducibility), writing (both original and review & editing), validation (others outside the team ensure the answers are right), etc.

Some roles seem to occur in multiple phases, such as software, writing, and validation, which makes me think that these are yet well-enough defined. And this also make me quite strongly think that having software listed under data curation is wrong. This item is too data-centric.

For a project that is aimed at creating software or data, but not at using it to gain new knowledge immediately, most of the roles still make sense, but the definition may be a bit off. For example, what does investigation mean? Perhaps this is planning the software, determining what algorithms to use, etc.?

I can imagine a set of tables (one for each kind of product, e.g., paper, software, data, maybe different for different domains?) would be useful here, where each table has the set of roles as rows, and the phases as columns, and Xs in the cells where that role fits in that phase. More than one X in a row might mean that role isn't sufficiently well described yet. Different Xs in different tables might also mean the roles aren't correctly described yet.

Finally, the workshop included some discussion about how these roles overlapped authorship, and the answer I gave, which seemed somewhat well received, is that they are separate. Contributorship is intended to be a set of observed properties (i.e., researcher X performed role Y during the project), while authorship is a set of cultural values that differ across communities (e.g., should the person who acquires funding for a project be an author?)

Other interesting questions are:

- Does this taxonomy apply across all scholarly domains? (It was originally conceived in a biomedical context.) What about domains such as architecture? Or art?
- Can this taxonomy be used all types of scholarly products? Such as for a survey paper? Or for a research project that produces physical objects?
- Can the amount of role a researcher fulfills be measured? Perhaps as lead & participant, major & minor, or numerically?
- How can this work be combined with the idea of transitive credit?

Published by:



danielskatz

Assistant Director for Scientific Software and Applications at NCSA, Research Associate Professor in CS, ECE, and the iSchool at the University of Illinois Urbana-Champaign; works on systems and tools (aka cyberinfrastructure) and policy related to computational and data-enabled research, primarily in science and engineering [View all posts by danielskatz](#)

2 thoughts on “Project CRediT and Contributorship Roles”

1. Amy Brand

[January 8, 2015 at 9:21 am](#) [Edit](#)

Thanks, Dan. Note we also have software as a separate role, so it is not exclusively under data curation.

[REPLY](#)

1. danielskatz

[January 8, 2015 at 9:41 am](#) [Edit](#)

Thanks Amy.

The separate software role is good, though I would prefer it to be called “Software Development”.

But the role where software is a subset of data curation just doesn’t make sense to me. I would prefer this as “Data/Software Curation”, defined as “Activities to annotate data and/or software (produce metadata), and maintain and archive research data and/or research software code for later re-use.”

And I would probably add a “Data Development” role, defined as “Data assembly (including generation, search and merge of new and existing data sets or subsets), annotation, and scrubbing, as required for the research activity.”

Alternatively, the existing Software role could be merged with my new Data Development role as “Software/Data Development.”

[REPLY](#)