

Attributing work done (Data Attribution)

David Shorthouse, Nicky Nicolson, Andy Bentley

See example [here](#)

Topics

1. Making FAIR data for specimens accessible (Alex Hardisty, Barbara Thiers, Wouter Addink): [Template](#)
2. Extending, enriching and integrating data (Andy Bentley, Jen Zaspel, Keping Ma): [Template](#)
3. Annotating specimens and other data (Joe Miller, Gil Nelson, James Macklin, Rich Rabeler): [Template](#)
4. Attributing work done (Data Attribution) (David Shorthouse, Nicky Nicolson, Andy Bentley): [Template](#)
5. Analyzing/mining specimen data for novel applications (Andrew Young, Libby Ellwood, Anna Monfils, John Bates): [Template](#)

(The first paragraph is a brief description of the category. This guidance will appear in the category selection area, so try to keep it below 200 characters.)

This is category 4 of the consultation on converging Digital Specimens and Extended Specimens towards a global specification for data integration.

Background

People want their efforts to be acknowledged and recognised. Other people want to know who did the work and when, and for that information to be unambiguous. Collections want to gain attribution for their contribution to the scientific endeavor through specimen and data use in end-products of research. Standardized mechanisms and metrics are required to facilitate this.

The goal of this category is to give shape to what it means to assign credit to individuals, organizations, or even software. It requires us to think about motivations, units of work, agencies, ethics, technologies, and standards of practice. Almost all data that flow from sender to recipients contain some form of structured or unstructured attribution. In short, the producers of new, primary data or secondary data products desire acknowledgment.

This category has significant areas of overlap with many, if not all other categories. However, it differs in its focus on a need to establish consensus on what or who are the parties that require attribution, how we uniquely identify those parties such that a token of their identity accompanies the transmission of data and is shared without ambiguity, and establishing who or what is responsible for storing and providing access to attribution data.

It also differs from other categories because this is where we may expound on ethical and legal considerations. We do not count things merely because we can. We must count things with purpose, with assured measures of accuracy, and with transparent mechanisms that detect and react to abuse.

Information resources

- Quentin Groom, Anton Güntsch, Pieter Huybrechts, Nicole Kearney, Siobhan Leachman, Nicky Nicolson, Roderic D M Page, David P Shorthouse, Anne E Thessen, Elspeth Haston, People are essential to linking biodiversity data, Database, Volume 2020, 2020, baaa072, <https://doi.org/10.1093/database/baaa072> (direct <https://academic.oup.com/database/article/doi/10.1093/database/baaa072/6094701>)
- Thessen, A. E., Woodburn, M., Koureas, D., Paul, D., Conlon, M., Shorthouse, D. P., & Ramdeen, S. (2019). Proper Attribution for Curation and Maintenance of Research Collections: Metadata Recommendations of the RDA/TDWG Working Group. Data Science Journal, 18(1), 54. DOI: <http://doi.org/10.5334/dsj-2019-054>

Questions to promote discussion

Group 1 What is an Agent, Who are the Actors, What do they Expect?

1. Who (or what entities) need(s) to be attributed?
2. How do we uniquely identify agents (= people, organizations, software) responsible for executing work?
3. What strategies should be employed to locally disambiguate “strings” to “things” and then share these unique identifiers for agents?

Group 2 What is a Unit of Work Worthy of Credit?

4. What are the activities pre-, during, and post-transcription of specimen labels that constitute work that ought to be attributed?
5. Are some activities more reflective of expertise and should be weighted more than others?
6. What lines of evidence are appropriate and sufficient to correctly attribute work? *i.e.* How do we trust attributions when these might be created on others' behalf?
7. How do we attribute, or provide credit, to the agents responsible for linking entities and bootstrapping the knowledge graph?

Group 3 How do We Measure FAIRly?

8. What measures should be taken to safeguard against misattribution?
9. How do we ensure that propagated attribution, wherever these land, can be amended or corrected?
10. What standards exist in other domains that store and share attribution and how are these implemented?
11. What standards are missing or require adjustment to best store and share attribution data?
12. What are the sociological/ethical/legal pitfalls we need to be sensitive to?

Group 4 How do We Make a Roundtrip for the Attributions?

13. What are the drivers for attributing work and how might meeting these needs contribute to the long-term sustainability of collections or other producers of primary biodiversity data?
14. What new communities of stakeholders might benefit from access to biodiversity data that includes tokens of attribution?
15. What metrics of reuse do we need to connect back to collections to allow for advocacy and attribution and who/what is responsible for gathering those metrics or defining their structure? (genbank sequences, publications, ??)
16. What technologies do we need in order to make these connections and to supplement or enhance some of the social mechanisms currently in use locally in collections, nationally, and internationally?

Actual discussion will happen here