RESEARCH PAPER

Proper Attribution for Curation and Maintenance of Research Collections: Metadata Recommendations of the RDA/TDWG Working Group

Anne E. Thessen[1,2], Matt Woodburn[3], Dimitris Koureas[4], Deborah Paul[5], Michael Conlon[6], David P. Shorthouse[7] and Sarah Ramdeen[2]

[1] Linus Pauling Institute, Oregon State University, Corvallis, OR, US

[2] The Ronin Institute for Independent Scholarship, Montclair, NJ, US

[3] Natural History Museum, London, UK

[4] Naturalis Biodiversity Center, Leiden, NL

[5] iDigBio, Florida State University, Tallahassee, FL, US

[6] University of Florida, Gainesville, FL, US

[7] Agriculture and Agri-Food Canada, Ottawa, ON, CA

Corresponding author: Anne E. Thessen (annethessen@gmail.com)

Research collections are an important tool for understanding the Earth, its systems, and human interaction. Despite the importance of collections, many are not maintained or curated as thoroughly as we would like. Part of the reason for this is the lack of professional reward for collection, curation, or maintenance. To address this gap in attribution metadata, the Research Data Alliance (RDA) and the Biodiversity Information Standards (TDWG) organization co-endorsed a Working Group to create recommendations for the representation of attribution metadata. After 18 months, this Working Group recommended the use of PROV entities and properties to link people (Agent), the curatorial actions they perform (Activity), and the digital or physical objects they are curating (Entity). Assigning a Role to an Agent is optional. These recommendations are discussed in the context of the RDA, TDWG, and existing standards. Future work includes adapting these recommendations to the specific needs of TDWG and developing a pilot application in collaboration with ORCID and the Data Futures project.

Keywords: attribution; metadata; ontology; Research Data Alliance; Biodiversity Information Standards; TDWG; collections; curation

Introduction
Background and Rationale
Research collections are an important tool for understanding the Earth, its systems, and human interaction (Suarez and Tsutsui, 2004). These collections are very diverse and can include preserved natural history specimens, archeological artifacts, minerals, or historical documents, to name just a few. Maintaining and curating these collections requires a large investment of time and money by institutions and many individuals (Keene, 2012). Knowledge is created from collections by many individuals over time, building on the work of others. For maximum efficiency, work needs to be shared broadly, recorded permanently, and tasks not repeated unnecessarily. Unfortunately, the current research cyberinfrastructure does not support this level of efficiency.

Despite the importance of collections, maintaining them and curating them to keep them up to date often remains challenging for many reasons (e.g., funding and staffing needs) (Rouhan *et al.,* 2017).

Despite the importance of collections, many are not maintained or curated as thoroughly as we would like (Suarez and Tsutsui, 2004). We suggest a major contributing factor for this maintenance backlog is a lack of professional reward for curatorial actions. Most researchers who are qualified to curate a collection are consumed by activities that reap professional reward, such as writing publications and grants. Proper methods of attribution (at the individual and institutional level) are very important for incentivizing digitization, mobilization, and sharing of data deriving from collections (physical and digital). One strategy for elevating the academic value of curatorial actions is to create the necessary infrastructure that captures the breadth of activities undertaken by curatorial staff. Several programs exist for aggregating metrics for research products other than publications, such as ImpactStory (Priem and Piwowar, 2012), OpenVIVO (Ilik *et al.*, 2018), Bloodhound[1], and Altmetric. Thus, there is already infrastructure in place for aggregating these data, if the e-infrastructure for creation of these data is available. What is currently lacking is a standard to best express the actions taken by agents when curating physical and digital collections.

Significant investment has been made in creating the necessary components of the infrastructure that integrate data across a wide variety of disciplines. Many of these components are lists, repositories, or other structures that must be populated with data either by a person or algorithmically (Cachuela-Palacio, 2006; Pyle and Michel, 2008; Boyle *et al.*, 2013; Pyle, 2016). Even an automatically-created data set will require some degree of human curation to ensure quality. Often, very little can be completed without initial work by a person to create reference material. This human-component is a major bottleneck. Thus, existing infrastructure for collective resources is not being populated with data and thus is not maximally useful. One way to widen the bottleneck is to create professional incentives for researchers to contribute to maintaining and curating collections. If people could receive professional credit, ideally recognized by their administrators and funding bodies, they would prioritize these traditionally unrewarded tasks. Unfortunately, a unified mechanism to manage information about curatorial actions does not yet exist.

In order to address this problem, we worked with the Research Data Alliance (RDA) and the Biodiversity Information Standards (TDWG) group to engage a community of users and to develop the metadata standards that accurately describe attribution. The RDA is a group of over 7000 people from 137 countries who meet every six months to develop and adopt infrastructure that promotes data-sharing and data-driven research (Berman, Wilkinson and Wood, 2014). RDA provides a neutral space for working groups to develop recommendations on an 18 month timeframe. Recommendations are developed within the context of Working Groups (WG) and Interest Groups (IG) that form to address a specific problem. TDWG is a scientific and educational association that fosters collaboration among the creators, managers, and users of biodiversity information. TDWG supports interest groups in the development, ratification, and maintenance of standards for biodiversity information. Endorsement of these recommendations by both organizations was crucial for community adoption and long term support of the results. This paper will present the recommendations of the working group developed during four RDA biannual meetings and two TDWG annual meetings. They represent the completion of the 18 month development period within RDA and potentially the beginning of the standard ratification process within TDWG.

These recommendations were developed to record the attribution metadata associated with curation and maintenance of research collections, whether they be physical or digital objects. The schema was designed to be adopted as part of existing data models and workflows used by stewards of these collections, e.g., museums. It assumes the pre-existence of a collections management system that includes within it a means to track research objects and record curator identities through a system of unique identifiers. These recommendations are intended to fit within the context of existing, domain-specific vocabularies for recording various types of metadata.

Results
Recommended Schema
This Working Group recommends a very basic, three-axiom, schema based on PROV entities and properties shown in **Figure 1** (and demonstrated in the PROV-O documentation) (Belhajjame *et al.*, 2013).

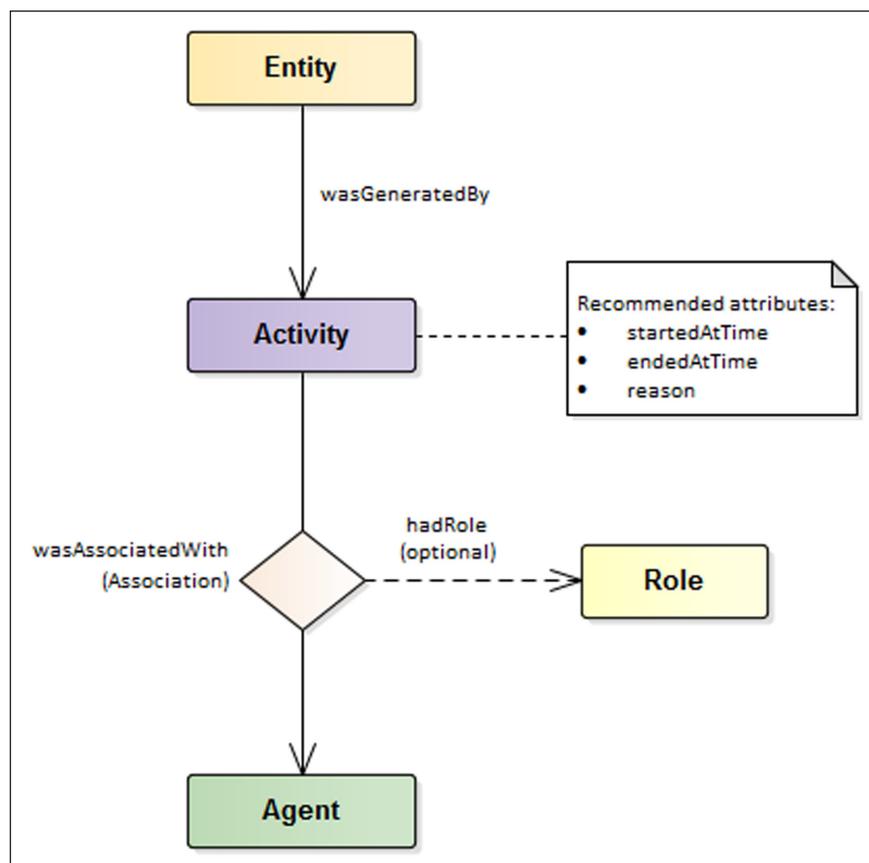[1] Proof-of-concept in beta development.

Figure 1: Recommended Schema. This working group recommends a basic schema linking Entities to Activities and Activities to Agents. Roles are optional.

The key elements of the model for attribution are:

```
Entity wasGeneratedBy Activity
Activity wasAssociatedWith Agent
```

with some additional attributes assigned to the Activity class:

```
Activity has attribute DateTime
Activity has attribute Reason (added as comment)
```

The Entity is the curated data object, whether it be a piece of metadata or a physical object. The Activity is the actual curation activity, such as making a correction or transformation. The Agent is the person performing the curation activity. Every Activity will have a DateTime stamp and a Reason it was performed (optional). The above axioms state that an Entity "wasGeneratedBy" an Activity. The Activity "wasAssociatedWith" an Agent, who performed the Activity. An Activity can be related to an Agent using one of two properties. The first is "wasAssociatedWith" and the second, "qualifiedAssociation", allows for the assignment of a Role. Assigning a Role to the Agent is optional in this recommendation, but specific reifications of this recommendation, such as PROV, may require it. If no role is to be assigned, then "wasAssociatedWith" should be used. This ontology design pattern is very similar to work done by Cox and Car (Simon J.D. Cox and Nicholas J. Car, 2015).

Each specific Entity, Activity, and Agent should be represented by a unique, persistent identifier (McMurry *et al.*, 2017). We recommend the use of IGSN for physical objects (Lehnert *et al.*, 2006), ORCID for people (Haak *et al.*, 2012), and DOI for digital objects (Paskin, 2009) wherever possible. The adoption of IGSN for biological specimens is still being discussed and these recommendations will defer to the future community decision. As such, GUIDs or equivalent standards may be used in place of IGSN. Activities can be identified internal to the curation management system in place. All Activities, Entities,

and Agents should be instances of a PROV Activity class, a PROV Entity class, and a PROV Agent class, respectively. If the appropriate class does not exist as a subclass, users should work with the VIVO (Mitchell *et al.*, 2011)[2] community to request the new VIVO subclass which can be mapped to PROV. Anyone can request that a new term be added to VIVO or raise any other issue via GitHub (*OpenRIF community*, 2018) and join the active ontology-improvement discussion group wiki (*Ontology Improvement Task Force – VIVO – DuraSpace Wiki*, 2018). DateTime should be represented as xsd (CCYY-MM-DDThh:mm:ss[Z|(+|–)hh:mm]).

Justification

The above recommendations are based on an existing ontology, PROV (Belhajjame *et al.*, 2013), that is part of a broader world of interconnected ontologies and vocabularies that are in use and have active community support. The pattern is simple enough to be repurposed in multiple disciplines and on physical and digital objects, yet still conforms to existing semantic frameworks.

This schema supports the following queries identified as important by the use cases:

1. Show me all the Activities performed by Kenji on 16 Sept 2013.

```
SELECT ?agent ?activity ?startdate ?enddate
WHERE
    {
        ?agent foaf:givenName "Kenji" .
        ?activity prov:wasAssociatedWith ?agent .
        ?activity prov:startedAtTime ?startdate .
        ?activity prov:endedAtTime ?enddate .
        FILTER (?startdate >= "2013-09-16T00:00:00+05:30"^^xsd:dateT
        ime) .
        FILTER (?enddate <= "2013-09-16T23:59:59+05:30"^^xsd:dateTime)
    }
```

2. What Activities have been performed on this metadata record? When?

```
SELECT ?activity ?startdate ?enddate
WHERE
    {
        ?activity prov:used | prov:generated :enhanced_image_meta-
        data_record .
        ?activity prov:startedAtTime ?startdate .
        ?activity prov:endedAtTime ?enddate .
    }
```

3. Which Agents have worked with this digital image?

```
SELECT DISTINCT ?agent
WHERE
    {
        ?activity prov:wasAssociatedWith ?agent .
        ?activity prov:used | prov:generated :enhanced_digital_image .
    }
```

---

[2] Users should add new classes through VIVO instead of PROV. These ontologies are linked, but PROV is meant to be high-level and not edited by the community of users, unlike VIVO.

4. What Role did Kenji play in this image resubmission?

```
SELECT ?agent ?activity ?role
WHERE
    {
        ?agent foaf:givenName "Kenji" .
        ?activity prov:wasAssociatedWith ?agent .
        ?activity prov:qualifiedAssociation ?association .
        ?association prov:hadRole ?role .
        ?association prov:agent ?agent .
        FILTER (?activity = :image_resubmission) .
    }
```

Example from Use Cases
Example RDF Turtle representations and diagrams of three use cases are included. One is presented below and two additional examples are given in Appendix A. Examples of how these recommendations could fit in with the larger landscape of existing relevant ontologies and vocabularies are also represented. The RDF Turtle representation can also be found in the project GitHub repository (*TDWG attribution*, 2019).

Use case: Digital record curation
Michael (a researcher) notices that a specimen has an incorrect digital lat/long record. Michael reports the error to Sarah (a data curator), who corrects the record in the database (**Figure 2**).

Attribution:
· Michael should receive attribution for reporting the error
· Sarah should receive attribution for correcting the digital record
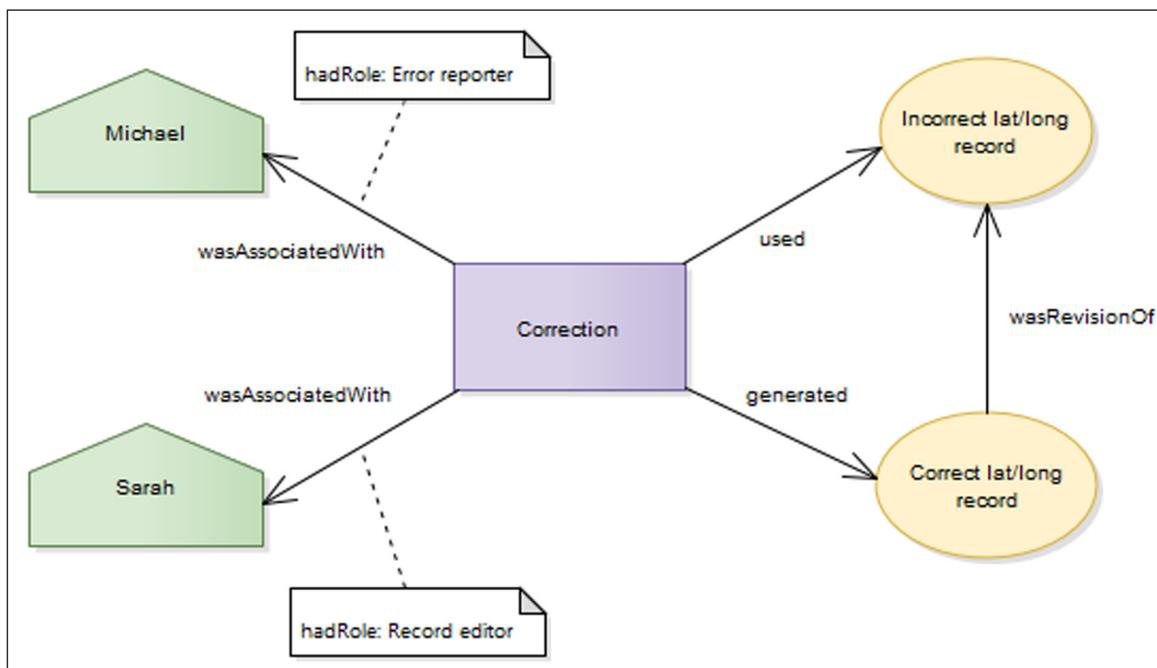

Figure 2: Example of Digital Record Curation. Michael and Sarah are both associated with the correction of a digital metadata record as an error reporter and a record editor, respectively. The correction activity is also associated with the incorrect record, which was used to create the correct record.

RDF/Turtle representation

```
@prefix :          <http://example.org/> .
@prefix dct:    <http://purl.org/dc/terms/> .
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd:    <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix prov:   <http://www.w3.org/ns/prov#> .
@prefix foaf:   <http://xmlns.com/foaf/0.1/> .
@prefix vivo:   <http://vivoweb.org/ontology/core#> .


:activityReason a rdfs:Class .


# Agents
:michael
    a prov:Person, prov:Agent ;
    foaf:givenName    "Michael" ;
    vivo:orcidId       "http://orcid.org/NNNN-NNNN-NNNN-NNNN" ;
.
:sarah
    a prov:Person, prov:Agent ;
    foaf:givenName    "Sarah" ;
    vivo:orcidId       "http://orcid.org/NNNN-NNNN-NNNN-NNNN" ;
.


# Contributor roles
:error_reporter a prov:Role .
:record_editor a prov:Role .


# Entities
:incorrect_lat_long_record
    a prov:Entity ;
    prov:Value      "-89.747988,43.138092" ;
    dct:references  <https://doi.org/XX.XXXX/XXXXXXX> ;
.
:correct_lat_long_record
    a prov:Entity ;
    dct:references      <https://doi.org/XX.XXXX/XXXXXXX> ;
    prov:Value          "43.138092,-89.747988" ;
    prov:wasRevisionOf  :incorrect_lat_long_record ;
.


# Activities
:correction
    a prov:Activity ;
    prov:startedAtTime    "2017-11-21T18:42:13-04:00"^^xsd:dateTime ;
    prov:endedAtTime    "2017-11-21T18:42:15-04:00"^^xsd:dateTime ;
    prov:used                   :incorrect_lat_long_record ;
    prov:generated              :correct_lat_long_record ;
```

```
    prov:wasAssociatedWith    :michael ;
    prov:wasAssociatedWith    :sarah ;
    :activityReason               "Incorrect longitude and latitude on
                                  the digital record." ;

    # Role association
    prov:qualifiedAssociation [
        a prov:Association ;
        prov:agent   :michael ;
        prov:hadRole :error_reporter ;
    ];

    prov:qualifiedAssociation [
        a prov:Association ;
        prov:agent   :sarah ;
        prov:hadRole :record_editor ;
    ] ;
.
```

Discussion and Conclusion
Relationship to Other RDA Recommendations
A discussion of how these recommendations fit in the larger community of RDA Working Groups and Interest Groups can be found in Appendix B.

Relationship to Existing Standards
PROV-O: An ontology for describing provenance. These recommendations use design patterns from this ontology, ensuring compatibility (Belhajjame *et al.*, 2013). The use of PROV means that these recommendations are compatible with VIVO (Mitchell *et al.*, 2011) and BCO (Walls *et al.*, 2014). Users can use entities and properties in PROV, VIVO, and BCO to suit specific provenance needs that are out of scope for these recommendations. For example, linking a transformed image to its original image can be done using PROV derivedFrom.

SESAR/IGSN: A system of identifiers and metadata for physical samples. These recommendations include the use of IGSN as identifiers for physical objects where possible (Lehnert and Arko, 2016). IGSN provides for recording the collector of a sample. The use of IGSN for biological specimens is still being discussed within the biodiversity community (Hobern, Hahn and Robertson, 2018).

TaDiRAH: A vocabulary focused on digital research in the Humanities. This vocabulary contains relevant terms such as "Annotating", "Cleanup", and "Editing" that could be used as an Activity, but is specific to the Humanities (Borek *et al.*, 2014; Perkins *et al.*, 2014). Users should draw terms from a relevant vocabulary or add the terms they need to an existing vocabulary.

CRediT: A vocabulary of contributor roles in research. CRediT is a high-level researcher role vocabulary supported by CASRAI (Brand *et al.*, 2015). If a Role is to be assigned to an Agent, it should come from a controlled vocabulary, such as CRediT; however CRediT is very high-level and may not have the needed terms for collection, curation, and maintenance. Users should draw terms from a relevant vocabulary or add the terms they need to an existing vocabulary.

OpenRIF/VIVO-ISF: An ontology for representing contributor roles, activities, and relationships in clinical research. VIVO (Mitchell *et al.*, 2011) is compatible with PROV (Belhajjame *et al.*, 2013). VIVO might be a good adopter if "Curation" is added as a subclass of "Process". One important point to remember is that PROV is a W3C recommendation, while VIVO is an OBOFoundry ontology (*The Open Biological and Biomedical Ontology (OBO) Foundry*, no date). The critical difference between PROV and VIVO is in the Role class. In VIVO, the Role is unique to the Agent while in PROV, the Role is a separate class that can be assigned to multiple Agents. The consequences of choosing PROV or VIVO should be carefully weighed by each adopter, but will be less of an issue if Role is not used. A PROV Agent would be equivalent to a Person, Group, or Organization in VIVO (or rather, a FOAF Agent). A PROV Activity would be equivalent to an Event from the event ontology. A PROV Entity would be any OWL Thing. **Figure 3** is an attribution model proposed, but not
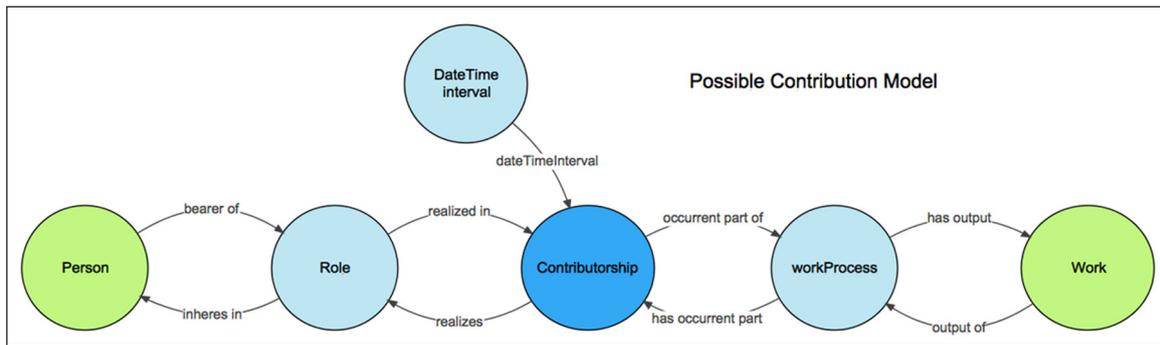
Figure 3: Proposed VIVO Contribution Model. Unlike roles in the PROV model, roles the VIVO model
inhere in the Agent/Person and are realized in the Activity/Work Process.

yet implemented, in VIVO. The person (or Agent) is the bearer of a Role which may have any of the CRediT
types. The Role is realizedIn an occurrentPart (here called Contributorship and not directly represented in
the recommendation) of a workProcess (or Activity) which has output Work (or Entity). The person "partici-
pates_in" (RO_0000056) the work process (not shown). A datetime can be added to the contributorship to
constrain the time of a person's contribution as shown above, but also to the work process to indicate start
and end times for that process (not shown).

Darwin Core: A data standard for biodiversity. Darwin Core does not currently have an extension for
describing curation of objects (Wieczorek *et al.*, 2012). This recommendation will form the foundation of
future work to develop this extension. There is already some demonstrated community support for this work
(Shorthouse, 2017, 2018).

COPDESS: Data publication standards in Earth Science. COPDESS has committed to using IGSN and ORCID,
as these recommendations suggest (Lehnert *et al.*, 2016).

Data Cite: Data publication standards. Data Cite only allows use of DOI (Brase, 2009). These recommenda-
tions suggest using DOI for digital objects where possible.

Biological Collections Ontology (BCO): Ontology for describing the collection and treatment of biological
samples. This ontology describes some activities that could be considered curatorial, such as the analysis and
treatment of biological samples, but is less concerned about attributing those actions to an individual (Walls
et al., 2014). BCO and PROV are compatible. Many of the process classes in BCO could serve as Activities in
PROV.

Future Work and Implementation

These recommendations represent the results of an 18 month collaboration between RDA and TDWG.
After 18 months, RDA requires production of a set of recommendations, but continues support for
working groups in maintenance mode as they work with organizations that wish to adopt their rec-
ommendations. While this manuscript represents the end of active development within an RDA
working group, it also represents the beginning of refinement of these recommendations within a
TDWG interest group. These recommendations will form the basis of a proposal to create an exten-
sion to the Darwin Core standard (Wieczorek *et al.*, 2012; Shorthouse, 2017). Refinement of these
recommendations will continue as adopters come online and within the TDWG standards ratification
environment.

Future work that will not necessarily take place within the context of this RDA/TDWG group, but will be
important for adoption includes:

- **Formal adoption of a specific persistent identifier for specimens by the biological collections
  community (such as IGSN (Hobern, Hahn and Robertson, 2018)).** The biodiversity informat-
  ics community has been struggling with the adoption of identifiers for specimens for many years
  with limited success (Page, 2008; Guralnick *et al.*, 2014, 2015). A new approach is in development for
  uniquely and persistently identifying specimens without a universal identifier system. This method
  uses the specimen metadata/identifier graph as an identifier (Thessen *et al.*, 2018). We will explore this
  method for use in combining sources of attribution metadata without duplicating specimen records.
- **The expansion of existing vocabularies and ontologies to include needed terms.** The addition of
  activity classes to VIVO is in early discussions and will likely include a collection and a curation class.

These discussions will take place in the GitHub repository within the TDWG-sponsored Attribution Interest Group (*TDWG attribution*, 2019).
- **The development of a pilot application for displaying curation activities on an ORCID profile.** The pilot application is in early phases and will demonstrate the flow of attribution metadata from a collections data manager (Bloodhound) to a data aggregator (ORCID).
- **The quantification of the impact of this new standard on collections management.** Studies have shown some impact of robust attribution on changing incentives and research practices (Piwowar, Day and Fridsma, 2007; Piwowar and Vision, 2013; Bornmann and Leydesdorff, 2014; Friesike and Schildhauer, 2015), but these data are hard to get without a dedicated study. Our future plans include applying for funding to explore the effect of improved curation and maintenance attribution on the practice of collections management.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Attribution Metadata Standard and Use Case Examples. DOI: https://doi.org/10.5334/dsj-2019-054.s1
- **Appendix B.** Relationship to Other RDA Recommendations, Working Groups, and Interest Groups. DOI: https://doi.org/10.5334/dsj-2019-054.s2

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Belhajjame, K, et al. 2013. *PROV-O: The PROV Ontology.* Lebo, T, Sahoo, S and McGuinness, D (eds.). W3C.

Berman, F, Wilkinson, R and **Wood, J.** 2014. 'Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance'. *D-Lib Magazine*, 20(1/2). DOI: https://doi.org/10.1045/january2014-berman

Borek, L, et al. 2014. 'Methods-TaDiRAH-Taxonomy of Digital Research Activities in the Humanities'. DARIAH-DE and DiRTdirectory. Available at: http://tadirah.dariah.eu/vocab/?tema=73.

Bornmann, L and **Leydesdorff, L.** 2014. 'Scientometrics in a changing research landscape: bibliometrics has become an integral part of research quality evaluation and has been changing the practice of research'. *EMBO reports*, 15(12): 1228–1232. DOI: https://doi.org/10.15252/embr.201439608

Boyle, B, et al. 2013. 'The taxonomic name resolution service: an online tool for automated standardization of plant names'. *BMC bioinformatics*, 14: 16. DOI: https://doi.org/10.1186/1471-2105-14-16

Brand, A, et al. 2015. 'Beyond authorship: attribution, contribution, collaboration, and credit'. *Learned publishing: journal of the Association of Learned and Professional Society Publishers*, 28(2): 151–155. DOI: https://doi.org/10.1087/20150211

Brase, J. 2009. 'DataCite – A Global Registration Agency for Research Data'. In: *2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*. IEEE, 257–261. DOI: https://doi.org/10.1109/COINFO.2009.66

Cachuela-Palacio, M. 2006. 'Towards an index of all known species: the Catalogue of Life, its rationale, design and use'. *Integrative zoology*, 1(1): 18–21. DOI: https://doi.org/10.1111/j.1749-4877.2006.00007.x

Friesike, S and **Schildhauer, T.** 2015. 'Open Science: Many Good Resolutions, Very Few Incentives, Yet'. In: Welpe, IM, et al. (eds.), *Incentives and Performance: Governance of Research Organizations*, 277–289. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-09785-5_17

Guralnick, R, et al. 2014. 'The Trouble with Triplets in Biodiversity Informatics: A Data-Driven Case against Current Identifier Practices'. *PloS one*, 9(12): e114069. DOI: https://doi.org/10.1371/journal.pone.0114069

Guralnick, RPP, et al. 2015. 'Community Next Steps for Making Globally Unique Identifiers Work for Biocollections Data'. *ZooKeys*, 494: 133–154. DOI: https://doi.org/10.3897/zookeys.494.9352

Haak, LL, et al. 2012. 'ORCID: a system to uniquely identify researchers'. *Learn. Publ.*, 25(4): 259–264. DOI: https://doi.org/10.1087/20120404

Hobern, D, Hahn, A and **Robertson, T.** 2018. 'Options to Apply the IGSN Model to Biodiversity Data'. *Biodiversity Information Science and Standards*, 2: e27087. DOI: https://doi.org/10.3897/biss.2.27087

Ilik, V, et al. 2018. 'OpenVIVO: Transparency in Scholarship'. *Frontiers in Research Metrics and Analytics*, 2: 12. DOI: https://doi.org/10.3389/frma.2017.00012

Keene, S. 2012. *Managing conservation in museums.* 2nd, Revised. Routledge. DOI: https://doi.org/10.4324/9780080510866

Lehnert, K and **Arko, R.** 2016. 'The IGSN Experience: Successes and Challenges of Implementing Persistent Identifiers for Samples'. *EGU General Assembly Conference Abstracts*, 18: 10798.

Lehnert, K, et al. 2016. 'COPDESS (Coalition for Publishing Data in the Earth & Space Sciences): An Update on Progress and Next Steps'. *EGU General Assembly Conference Abstracts*, 18: 16120.

Lehnert, KA, et al. 2006. 'The Digital Sample: Metadata, Unique Identification, and Links to Data and Publications'. In: *American Geophysical Union, Fall Meeting 2006. AGU Fall Meeting Abstracts.*

McMurry, JA, et al. 2017. 'Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data'. *PLoS biology*, 15(6): e2001414. DOI: https://doi.org/10.1371/journal.pbio.2001414

Mitchell, S, et al. 2011. 'The VIVO ontology: enabling networking of scientists'. *ACM Web Conference*, 14–17.

Ontology Improvement Task Force − VIVO − DuraSpace Wiki. 2018. Available at: https://wiki.duraspace.org/display/VIVO/Ontology+Improvement+Task+Force (Accessed: 24 August 2018).

OpenRIF community. GitHub. Available at: https://github.com/openrif/community (Accessed: 24 August 2018).

Page, R. 2008. 'Biodiversity informatics: the challenge of linking data and the role of shared identifiers'. *Briefings in bioinformatics*, 9(5): 345–354. DOI: https://doi.org/10.1093/bib/bbn022

Paskin, N. 2009. 'Digital Object Identifier (DOI ®) System'. In: Bates, MJ and Maack, MN (eds), *Encyclopedia of Library and Information Sciences, Third Edition*, 1586–1592. CRC Press. DOI: https://doi.org/10.1081/E-ELIS3-120044418

Perkins, J, et al. 2014. 'From DiRT Categories to TaDiRAH, a Methods Taxonomy for Digital Humanities'. *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2014*, 181–183.

Piwowar, HA, Day, RS and **Fridsma, DB.** 2007. 'Sharing detailed research data is associated with increased citation rate'. *PloS one*, 2(3): e308. DOI: https://doi.org/10.1371/journal.pone.0000308

Piwowar, HA and **Vision, TJ.** 2013. 'Data reuse and the open data citation advantage'. *PeerJ*, 1: e175. DOI: https://doi.org/10.7717/peerj.175

Priem, J and **Piwowar, H.** 2012. *The launch of ImpactStory: using altmetrics to tell data-driven stories, Impact of Social Sciences.* Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2012/09/25/the-launch-of-impactstor/ (Accessed: 24 August 2018).

Pyle, RL. 2016. 'Towards a Global Names Architecture: The future of indexing scientific names'. *ZooKeys*, (550): 261–281. DOI: https://doi.org/10.3897/zookeys.550.10009

Pyle, RL and **Michel, E.** 2008. 'ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information'. *Zootaxa*, 1950(1): 39–50. DOI: https://doi.org/10.11646/zootaxa.1950.1.6

Rouhan, G, et al. 2017. 'The time has come for Natural History Collections to claim co-authorship of research articles'. *Taxon. International Association for Plant Taxonomy*, 66(5): 1014–1016. DOI: https://doi.org/10.12705/665.2

Shorthouse, D. 2017. 'Proposed Extension to Darwin Core for People and their Roles in the Curation of Physical and Digital Objects'. In: *Pensoft Publishers*, e19829. Pensoft Publishers. DOI: https://doi.org/10.3897/tdwgproceedings.1.19829

Shorthouse, D. 2018. *Agents Actions: Proposed Darwin Core Archive Extension.* Available at: https://github.com/dshorthouse/agents_actions (Accessed: 24 August 2018).

Simon, JDC and **Car, NJ.** 2015. 'PROV and real things'. *21st International Congress on Modelling and Simulation (MODSIM2015)*, 21.

Suarez, AV and **Tsutsui, ND.** 2004. 'The Value of Museum Collections for Research and Society'. *Bioscience*, 54(1): 66–74. DOI: https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2

TDWG attribution. Github. Available at: https://github.com/tdwg/attribution (Accessed: 1 May 2019).

The Open Biological and Biomedical Ontology (OBO) Foundry. (no date). Available at: http://www.obo-foundry.org/.

Thessen, AE, et al. 2018. '20 GB in 10 minutes: a case for linking major biodiversity databases using an open socio-technical infrastructure and a pragmatic, cross-institutional collaboration'. *PeerJ Computer Science.* PeerJ Inc., 4: e164. DOI: https://doi.org/10.7717/peerj-cs.164

Walls, RL, et al. 2014. 'Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies'. *PloS one*, 9(3): e89606. DOI: https://doi.org/10.1371/journal.pone.0089606

Wieczorek, J, et al. 2012. 'Darwin Core: An evolving community-developed biodiversity data standard'. *PloS one*, 7(1): e29715. DOI: https://doi.org/10.1371/journal.pone.0029715