The Others

Ted Habermann

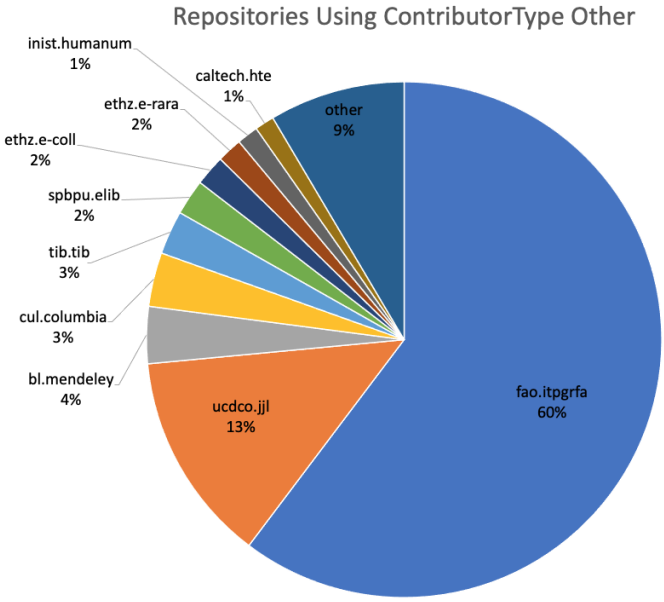
Blog post published January 10, 2022 in *Blog - Metadata Game Changers*
https://doi.org/10.59350/jsksv-g6245

Abstract

::::: {#block-4ce0cb5334f301eba820 .sqs-block .html-block .sqs-block-html block-type="2" border-radii="{"topLeft":{"unit":"px","value":0.0},"topRight":{"unit":"px","value":0.0},"bottomLeft":{"unit":"px","value":0.0},"bottomRight":{"unit":"px","value":0.0}}"} Connecting research is a critical element of the DataCite vision and two kinds of connections are supported by DataCite metadata.

Keywords

Information Systems (Subfield)



Recommended citation

Habermann, T. (2022, January 10). The Others. *Blog - Metadata Game Changers*. https://doi.org/10.59350/jsksv-g6245

Copyright

Copyright © Ted Habermann 2022. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

::::: {#block-4ce0cb5334f301eba820 .sqs-block .html-block .sqs-block-html block-type="2" border-radii="{"topLeft":{"unit":"px","value":0.0},"topRight":{"unit":"px","value":0.0},"bottomLeft":{"unit":"px","value":0.0},"bottomRight":{"unit":"px","value":0.0}}"}

Connecting research is a critical element of the DataCite vision and two kinds of connections are supported by DataCite metadata. First, identifying a dataset or other resource with a DOI makes it possible to connect it to journal articles, data management plans, software, and any of the other DataCite resource types as well as to other objects in the growing web of scholarly research. A second, equally important, set of connections are those made to people and organizations that contribute to the creation and maintenance of the resources described in DataCite. These contributor connections are critical in helping to build active and interactive research communities (Habermann, 2021).

One of the important features of contributors in DataCite (and in many other metadata dialects) is that they include a type element that makes it possible to recognize different types of contributions (Table 1). Like any generic vocabulary, this one focuses on broadly applicable contribution types, i.e. it is missing roles that can be important in specific disciplines or situations. Other lists that have been proposed include the high-level Contributor Roles Taxonomy (CRediT), which is being developed as a NISO standard, and the detailed OBO Foundry Contributor Role Ontology.

:::::

ContactPerson	DataCollector	DataCurator
DataManager	Distributor	Editor
HostingInstitution	Other	Producer
ProjectLeader	ProjectManager	ProjectMember
RegistrationAgency	RegistrationAuthority	RelatedPerson
ResearchGroup	RightsHolder	Researcher
Sponsor	Supervisor	WorkPackageLeader

Table 1. DataCite Contributor Types

::::: {#block-yui_3_17_2_1_1641829001196_20623 .sqs-block .html-block .sqs-block-html block-type="2" border-radii="{"topLeft":{"unit":"px","value":0.0},"topRight":{"unit":"px","value":0.0},"bottomLeft":{"unit":"px","value":0.0},"bottomRight":{"unit":"px","value":0.0}}"}

Contributors are optional in the DataCite schema and exist in ~25% of the records. If they exist, the contributorType must be included. The DataCite Metadata Working Group is currently considering improved alignment of the DataCite contributorTypes with other community lists. The differences mentioned above make improving alignment a challenging task. Expanding the DataCite list has come up as a potential part of a solution. For example, expansion could be done by adopting the CRediT terms wholesale, but it seems worthwhile to explore current usage and understand DataCite user needs as part of evaluating possibilities.

DataCite Others

I queried DataCite for resources that use the 'Other' contributor type (termed "Others" for simplicity) as this might indicate that metadata creators were lacking necessary contributorTypes. The query result includes a list of the top ten users of Other. Table 2 shows these repositories along with the number of occurrences of Other they include and the number of their records that include the Other contributor type. For example, the fao.itpgrfa repository includes 1,050,331 occurrences of Other in 99% of their records. Other repositories include many occurrences of Other but in a smaller one portion of their records (cul.columbia). Together these ten repositories are responsible for ~91% of the occurrences of Other contributor types in DataCite.

:::::

Repository ID	Repository Name	# Others / %
fao.itpgrfa	International Treaty on Plant Genetic Resources for Food and Agriculture	1,050,331 / 99%
ucdco.jjl	University College of Dublin (UCD) Digital Library	230,171 / 99%
bl.mendeley	Mendeley Data	62,148 / 51%
cul.columbia	Columbia University Libraries	58,894 / 9%
tib.tib	TIB Hannover	47,806 / 66%
spbpu.elib	Electronic Library of Peter the Great Polytechnic University	39,010 / 99%
ethz.e-coll	ETH Zürich Research Collection	33,398 / 45%
ethz.e-rara		27,455 / 31%
inist.humanum	Huma-Num	23,336 / 24%
caltech.hte	Caltech High Throughput Experimentation	21,215 / 70%

Table 2. Top 10 users of contributorType=Other in DataCite metadata with number of occurrences and % of records.

::::: {#block-yui_3_17_2_1_1641829001196_21476 .sqs-block .html-block .sqs-block-html block-type="2" border-radii="{"topLeft":{"unit":"px","value":0.0},"topRight":{"unit":"px","value":0.0},"bottomLeft":{"unit":"px","value":0.0},"bottomRight":{"unit":"px","value":0.0}}"}

Table 3 shows the counts of various contributorTypes observed in samples of 500 records from each of the top 10 repositories along with the % of observed contributorTypes that are Other. A 100% in the % column means that Other was the only contributorType in the sample.

The largest collection of Others (60% of the total in DataCite) is in the International Treaty on Plant Genetic Resources for Food and Agriculture repository. Table 3 shows that the fao.itpgrfa sample (of 500 records) includes 500 Others, probably one / record, and no other contributor types. Examining the metadata shows that these contributors are overwhelmingly organizations

like Centro Internacional De Mejoramiento De Maíz Y Trigo (https://ror.org/03gvhpa76) that are also listed as publishers in the metadata.

The next largest collection of Others (13% of the total in DataCite) is from the University College of Dublin (UCD) Digital Library. In this case the sample includes 1318 Others, indicating multiple Others in some, likely most, records. The second common contributorType in this repository is Funder which occurs in nearly every record (488/500).

Other repositories make up ~25% of the Other contributor types in DataCite. Several groups of repositories show similar patterns of contributor types, described in Table 4.

:::::

Repository	ContactPerson	DataCollector	DataCurator	DataManager	Editor	Funder	HostingInstitution	Oth
bl.mendeley								**76**
ethz.e-coll								**55**
fao.itpgrfa								**50**
ethz.e-rara								**21**
cul.columbia								**7**
spbpu.elib					39			**11**
ucdco.jjl		1			2	448	3	**13**
caltech.hte								**45**
inist.humanum	498	40	31	101	9			**30**
tib.tib				500			500	**37**
Total	**498**	**41**	**31**	**601**	**50**	**488**	**503**	**97**

Table 3. Counts of contributorTypes in each of the top 10 Other repositories.

::::: {#block-yui_3_17_2_1_1641829001196_22303 .sqs-block .html-block .sqs-block-html block-type="2" border-radii="{"topLeft":{"unit":"px","value":0.0},"topRight":{"unit":"px","value":0.0},"bottomLeft":{"unit":"px","value":0.0},"bottomRight":{"unit":"px","value":0.0}}"}

Conclusion

The DataCite contributorType list identifies twenty generic types of contributors that can be identified in DataCite metadata (Table 1). The contributorType "Other" is included for contributors that do not fit into one of the twenty types. Understanding the usage of the contributorType "Other" is important as part of the ongoing development of the DataCite Metadata Schema.

DataCite serves over 2000 repositories. Over 90% of the usage of "Other" occurs in ten repositories (Table 2). In fact, two repositories are responsible for 80% of the usage, suggesting that the current contributorTypes serve the community well in most cases. Random samples of 500 records from each of the ten repositories showed several interesting behaviors (Table 4).

:::::

Repositories	Pattern
bl.mendeley, ethz.e-coll, ethz.e-rara	Other is the only contributorType in the sample and the contributors are primarily individuals.
caltech.hte	The same 19 individuals occur in all 500 records, 10 as Researchers and 9 as Other.
spbpu.elib	Other is the most common contributorType in the sample and contributors are mostly departments in Peter the Great Polytechnic University (the location of this repository).
tib.tib	Other is one of several contributorTypes and contributors are mostly organizations outside of TIB.
cul.columbia	Other is the only contributorType in the sample and the contributors are primarily seminars that generated resources.
inist.humanum	The sample includes a variety of contributorTypes, Others are individuals and position titles in Spanish.

Table 4. Repositories and contributorType pattern

::::: {#block-yui_3_17_2_1_1641829001196_61430 .sqs-block .html-block .sqs-block-html block-type="2" border-radii="{"topLeft":{"unit":"px","value":0.0},"topRight":{"unit":"px","value":0.0},"bottomLeft":{"unit":"px","value":0.0},"bottomRight":{"unit":"px","value":0.0}}"}

The last column in Table 4 shows the % of all contributor types in each repository that are Other. In half of these repository samples, Other is the only contributorType used (100% in last column) and that only two of the repositories use more than three contributorTypes. These could be artifacts of the relatively small random selections, but that seems unlikely in so many repositories. The pattern of Other being the only contributorType suggests that in some repositories the selection of the Other contributorType is integrated into an automated record creation process that reflects a policy decision that only one contributorType is acknowledged in the metadata.

Only one contributorType that is not included in the current DataCite vocabulary was found in over 17,000 records. Funder is used as a contributorType by the University College of London in over 97% of their sample with most of the funders being Ireland Department of Arts, Culture. These records are using Version 3 of the DataCite schema which does not include the FundingReference object that replaced the Funder contributorType during 2016. This repository is the largest single user of this deprecated type.

Examining the usage of contributorType across the DataCite repository shows that Other is a rather rare contributorType with 91% of the usage occurring in 10 repositories. Repositories that do use Other tend to show consistent usage patterns across their metadata. In most of the top 10 repositories, Other is the only contributorType used in the repository.

DataCite metadata evolution is driven by the international community of DataCite members and users. Are there contributorTypes that you need that are not included in the list?

::::::