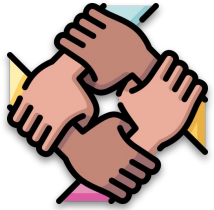
Introduction to Metadata in Dataverse Repositories

For Researchers and Support Staff

November 15, 2023

Julian Gautier
IQSS, Harvard University

Introduce yourself in chat
Post questions in Q&A

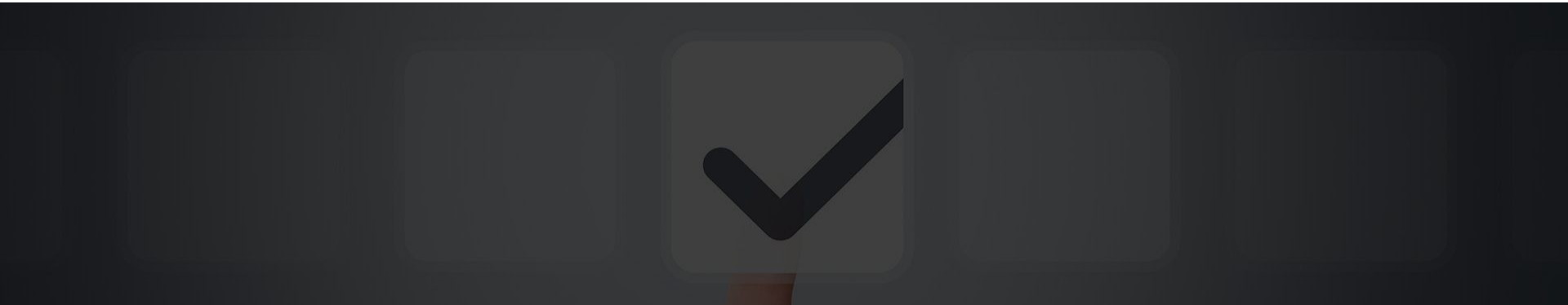


Review Contributor Covenant Code of Conduct

- Use welcoming and inclusive language
- Respect differing viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show empathy towards other community members

What I'll be covering

1. Introduction
2. About metadata in Dataverse repositories
3. Searching and browsing
4. What is "good" metadata?
5. Questions, survey, upcoming webinars



Introduction

By the end of the webinar…

You should be able to use Dataverse's metadata features to make data more FAIR

I should be able to help improve our webinars, like this one, based on your feedback

How this webinar was designed

Conversations with you (mostly with support staff)

Conversations with curation team at Harvard Dataverse

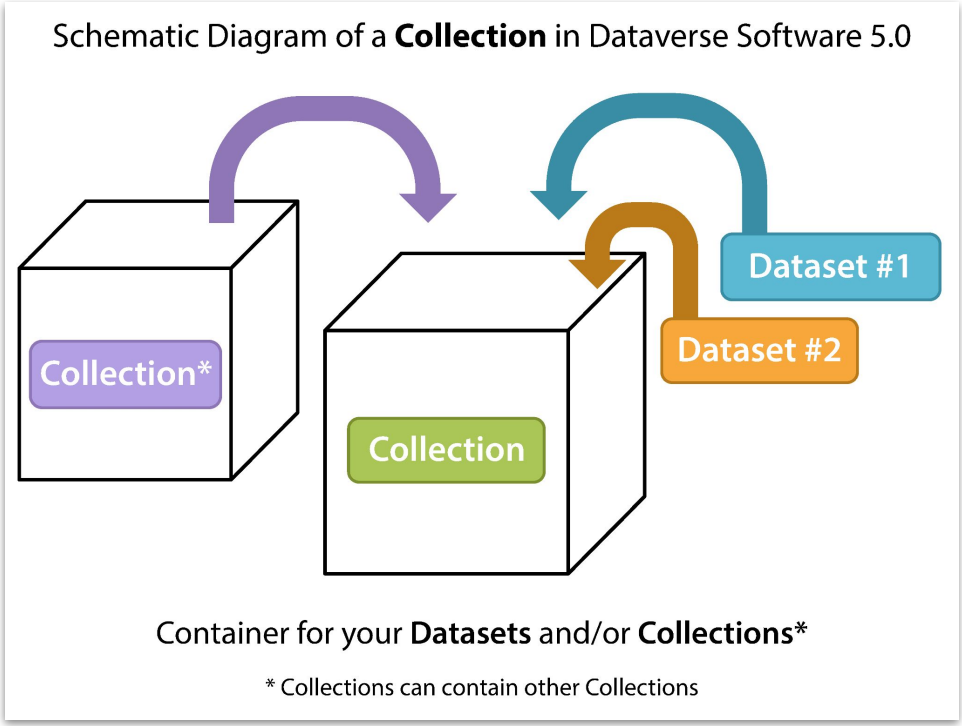
Reviews of support emails and others recorded conversations about metadata

A web survey

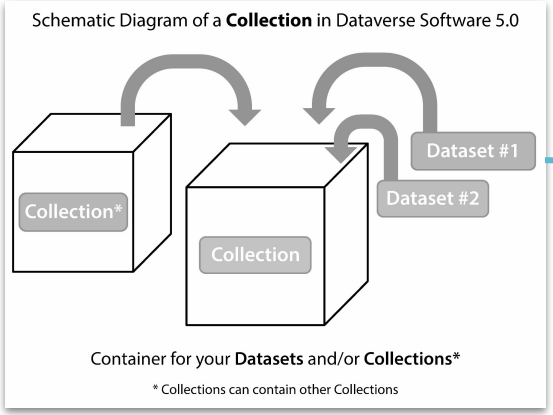
About me

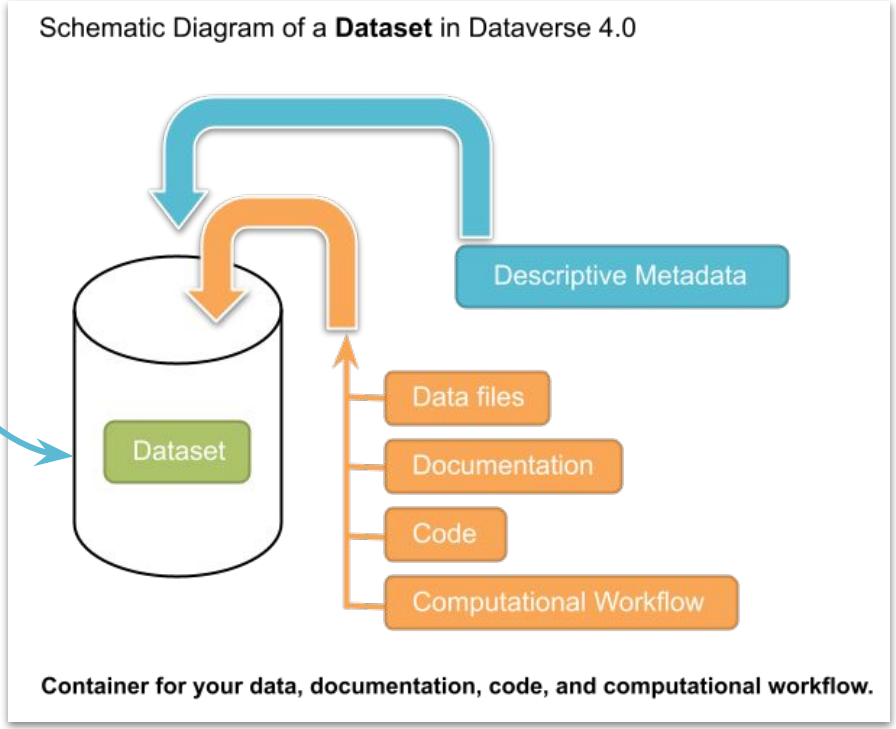
- Working on Dataverse for ~7 years
- User experience researcher, informing design with qualitative research (e.g. interviews, usability testing, surveys) and quantitative research (e.g. analytics)
- Working on metadata design
- Working on curation and support team at Harvard Dataverse

About metadata in Dataverse repositories



About metadata in Dataverse repositories





Metadata you can use to describe

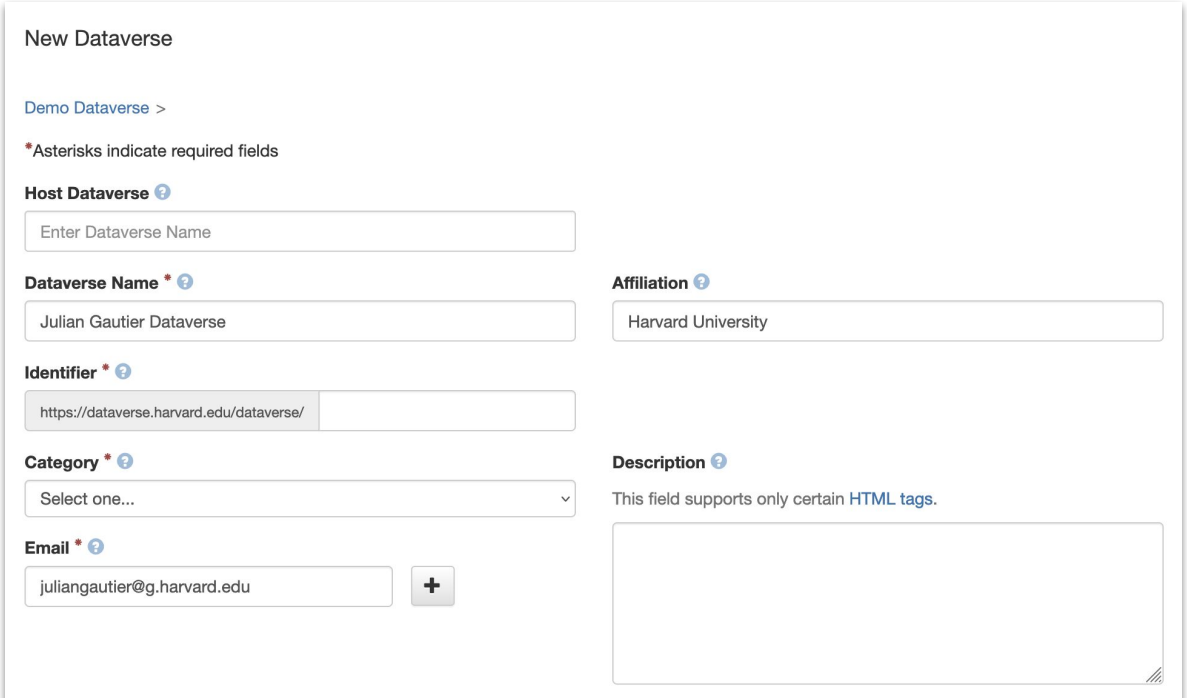
- Collections
- Datasets
- Files

Metadata that Dataverse creates automatically

- Such as persistent IDs, publication dates, version numbers
- Metadata saved from "ingested" tabular data

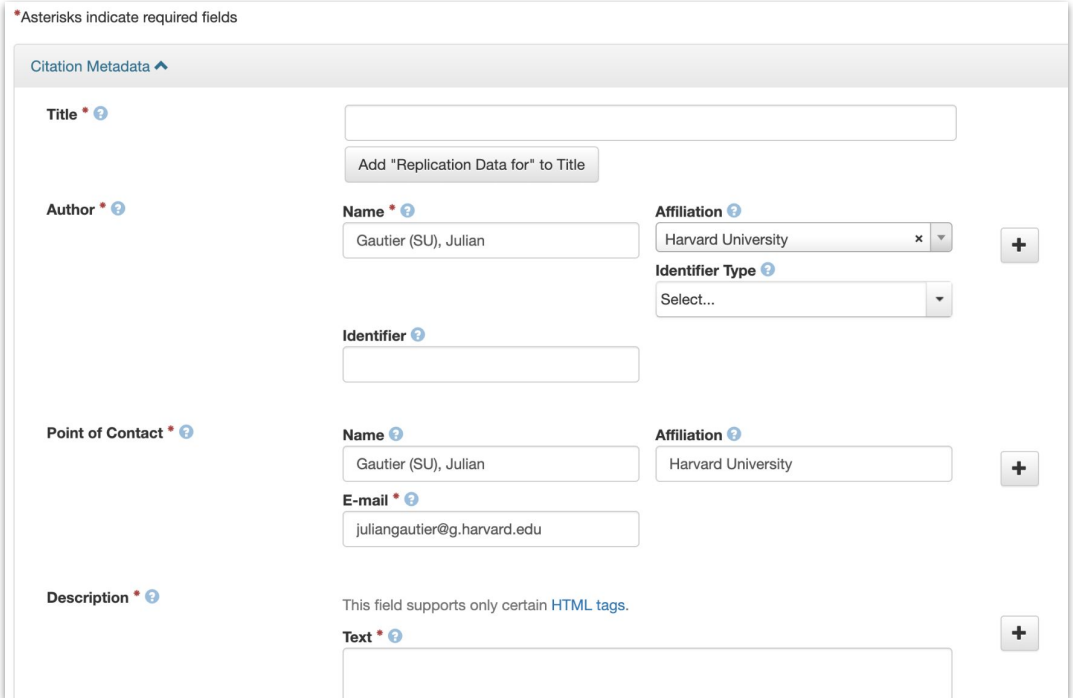
Metadata you can use to describe collections

- Names
- Affiliations
- Identifiers of collection URLs
- Categories
- Emails
- Descriptions



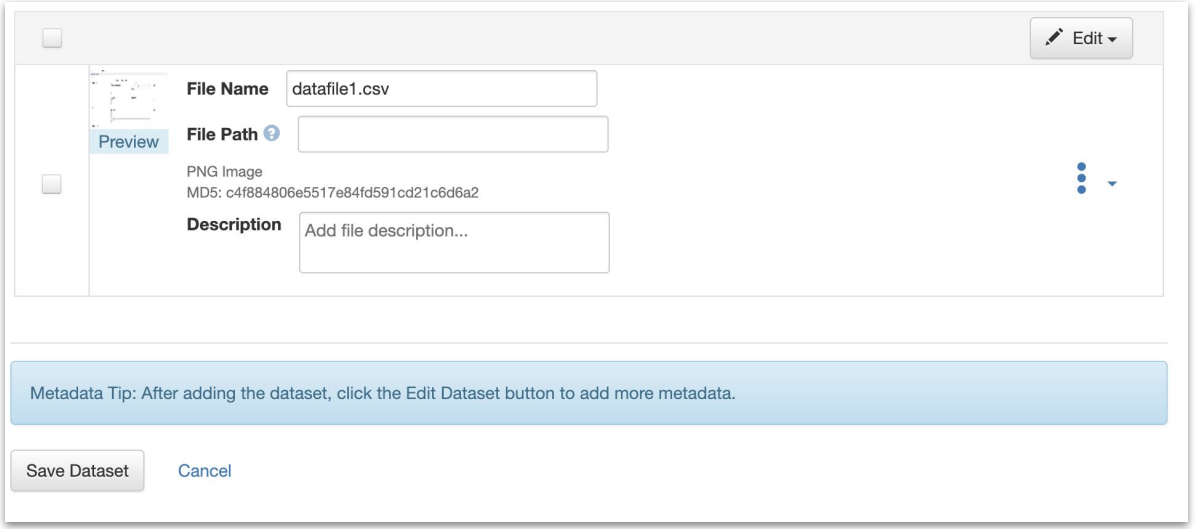
Metadata you can use to describe datasets

Using many fields shown in collapsible panels, such as “Citation Metadata” and “Terms”



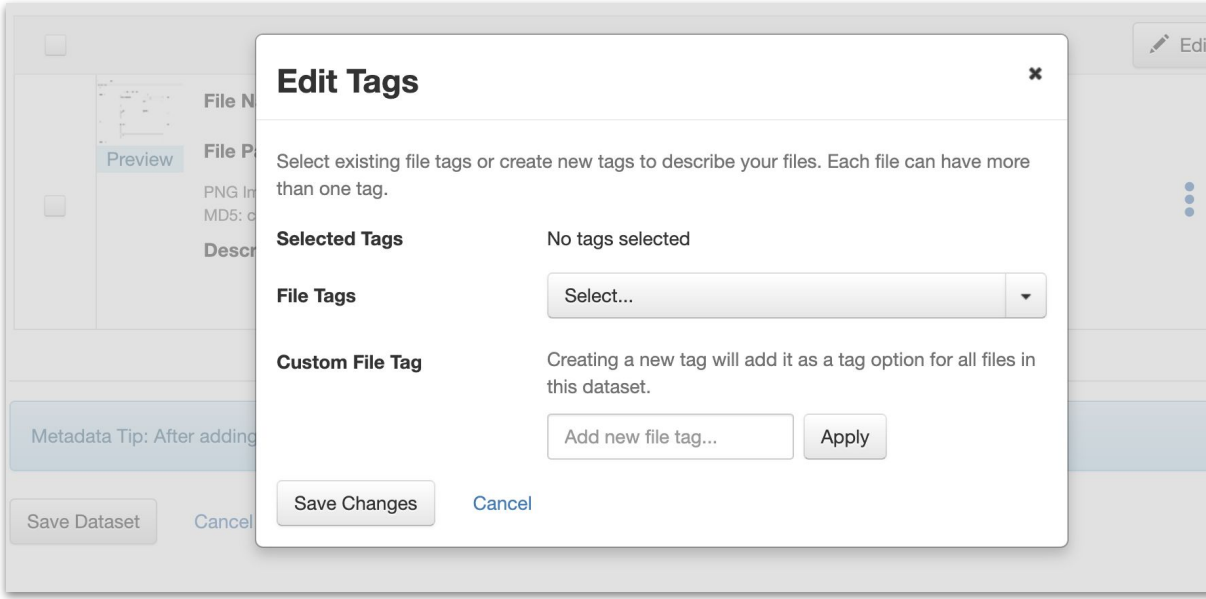
Metadata you can use to describe files

- File names
- File paths
- Descriptions



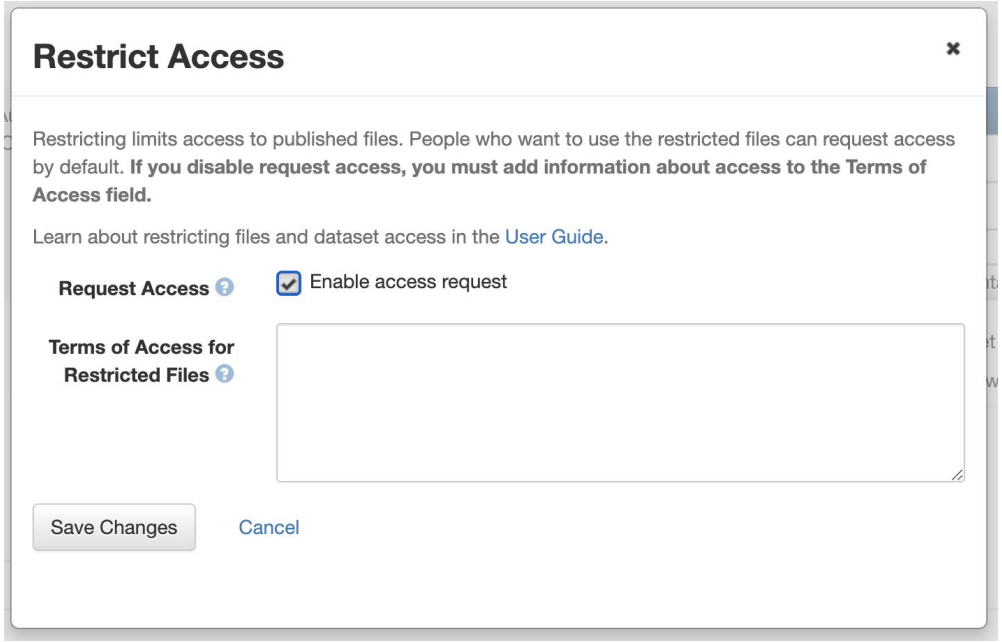
Metadata you can use to describe files

- Tags
- Restrictions
- Embargo dates and reasons



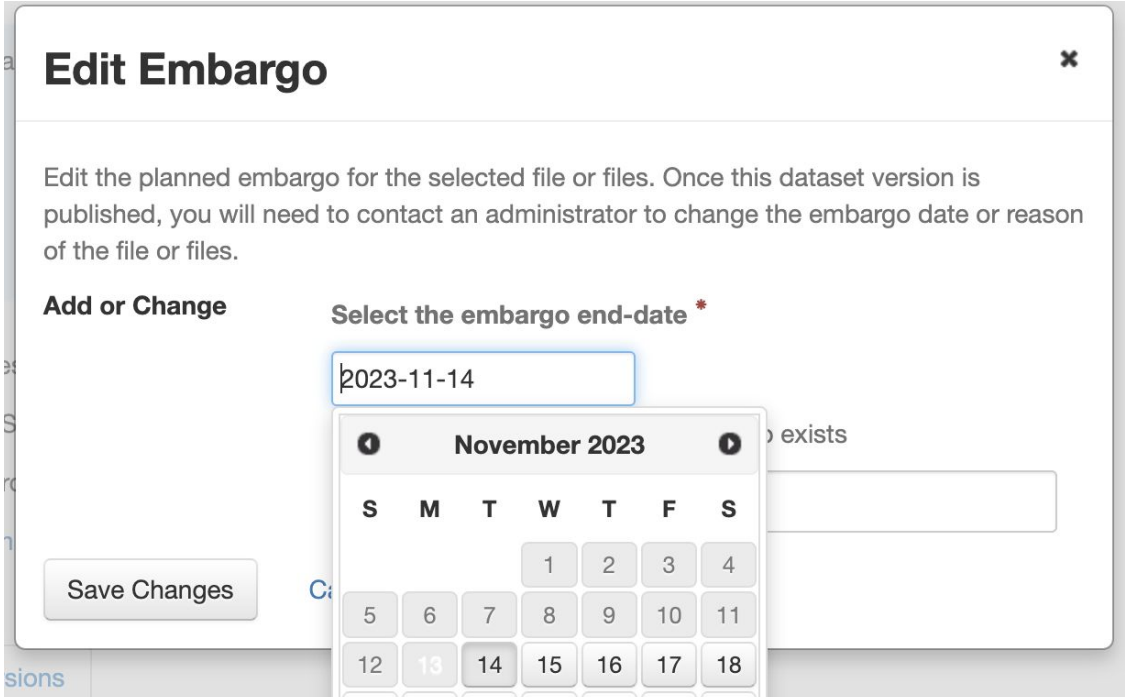
Metadata you can use to describe files

- Tags
- Restrictions
- Embargo dates and reasons



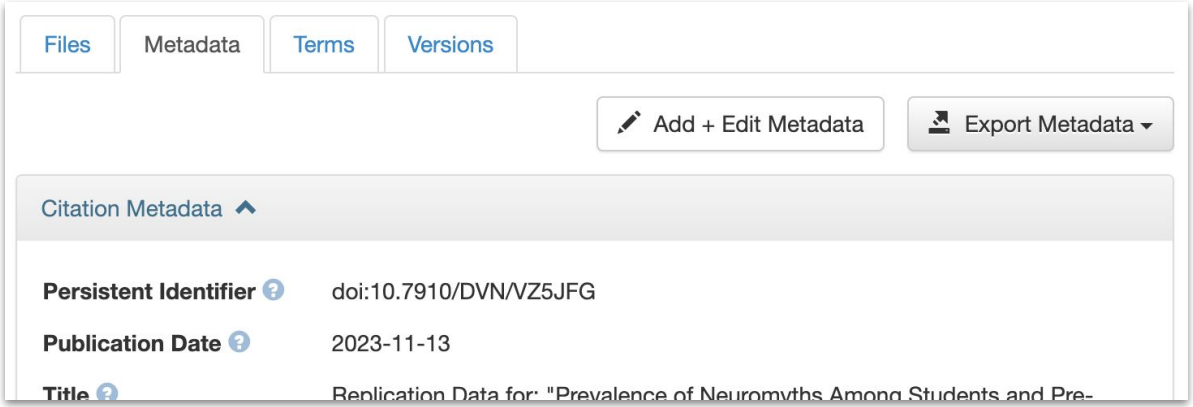
Metadata you can use to describe files

- Tags
- Restrictions
- Embargo dates and reasons



Metadata Dataverse creates for datasets

- Persistent IDs
- Creation dates
- Publication dates
- Version numbers
- Sum of file downloads
- Sum of page views (in repositories with Make Data Count enabled)



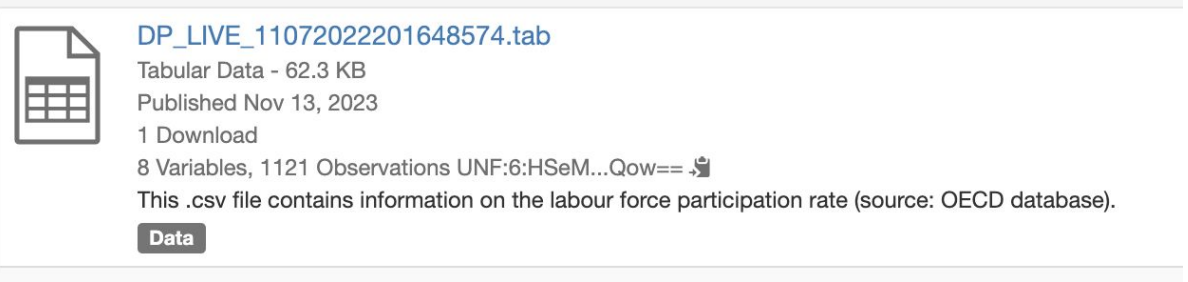
Metadata Dataverse creates for datasets

- Differences among versions

	Dataset Version	Summary	Contributors	
☐	12.0	**Citation Metadata:** Description (1 Changed); Notes (Changed); **Additional Citation Metadata:** (3 Changed); **Files (Added: 98; Removed: 84);** View Details	Julian Gautier	202
☐	11.0	**Citation Metadata:** Description (1 Changed); Notes (Changed); **Additional Citation Metadata:** (4 Changed); **Files (Added: 84; Removed: 57);** View Details	Julian Gautier	202
☐	10.0	**Citation Metadata:** Notes (Changed); Description (1 Changed); **Additional Citation Metadata:** (2 Removed, 2 Changed); **Files (Added: 57; Removed: 50);** View Details	Julian Gautier	202
☐	9.0	**Citation Metadata:** Description (1 Changed); **Additional Citation Metadata:** (1 Added); **Files (Added: 50; Removed: 37);** View Details	Julian Gautier	202
☐	8.1	**Citation Metadata:** Description (1 Changed); View Details	Julian Gautier	202
☐	8.0	**Citation Metadata:** Description (1 Changed); **Additional Citation Metadata:** (1 Added, 1 Changed); **Files (Added: 1; Replaced: 1);** View Details	Julian Gautier	202
☐	7.0	**Citation Metadata:** Description (1 Changed); **Files (Added: 35; Removed: 36);** View Details	Julian Gautier	202

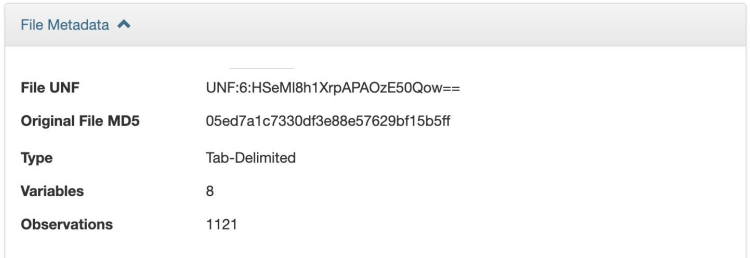
Metadata Dataverse creates for files

- File types
- Byte sizes
- Creation dates
- Publication dates
- Checksums (MD5s)
- File paths
- Counts of downloads



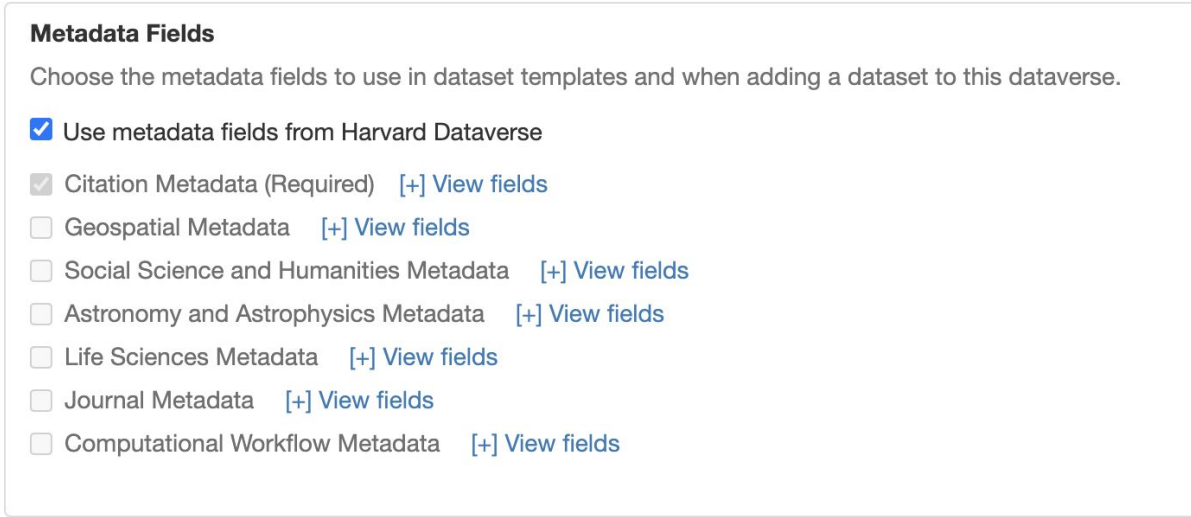
Metadata Dataverse creates for ingested tabular files

- Counts of variables and observations
- Variable statistics
- Checksums (UNFs)



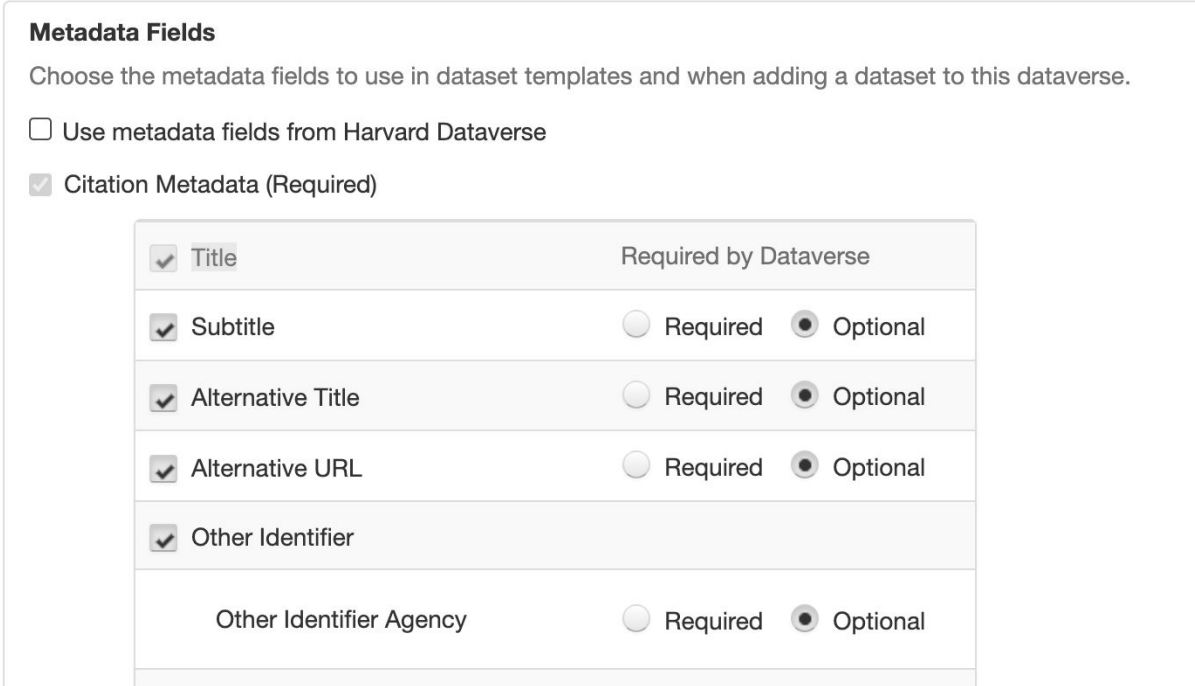
Metadata blocks

- Enabling and configuring
- Editing and creating



Metadata blocks

- Enabling and configuring
- Editing and creating

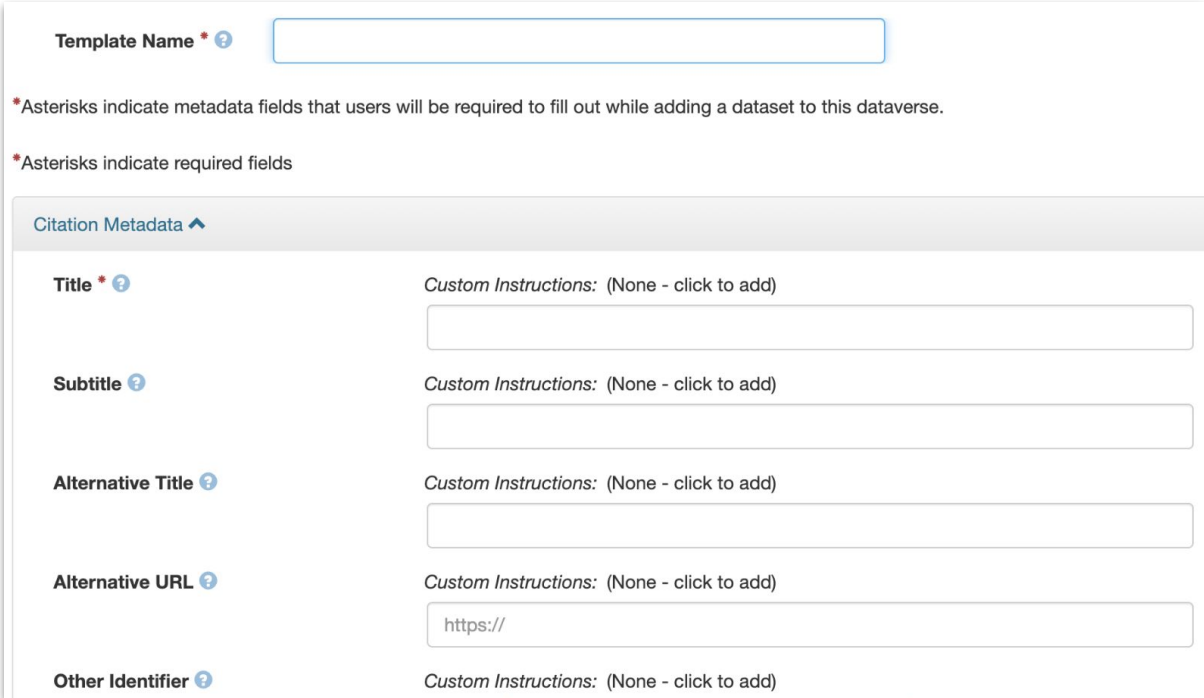


Metadata blocks

- Enabling and configuring
- Editing and creating

Dataset templates

- Copying metadata to multiple datasets
- Creating instructions



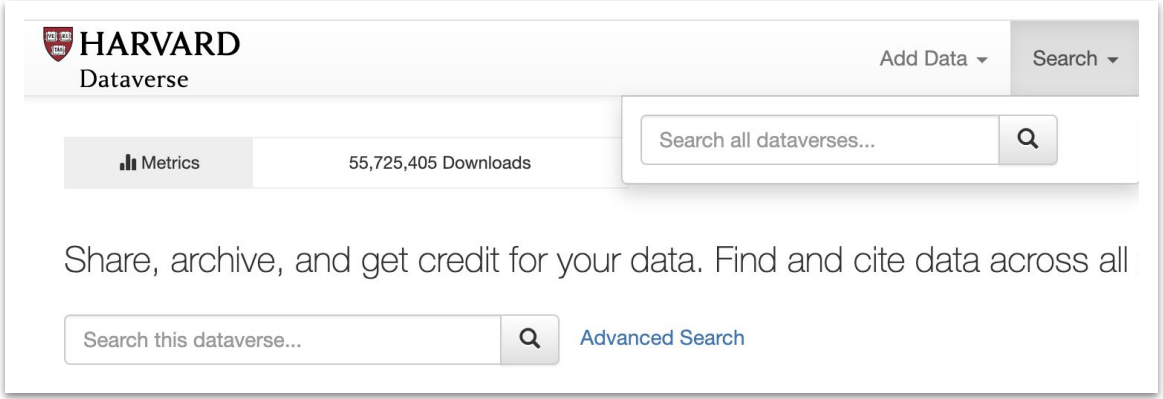
Searching and browsing

Searching in collections

Searching in datasets for files

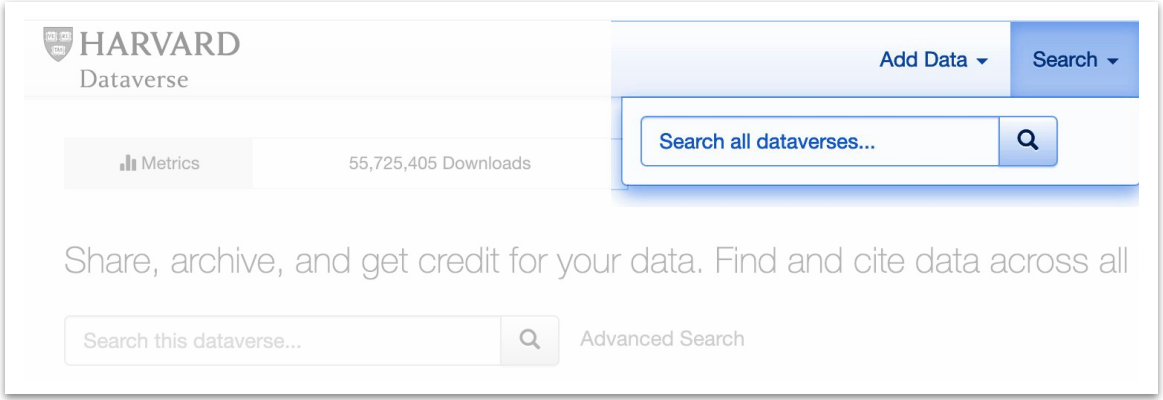
Searching in collections

- Using the general search boxes
- Using advanced search
- Filtering with search facets
- Sorting



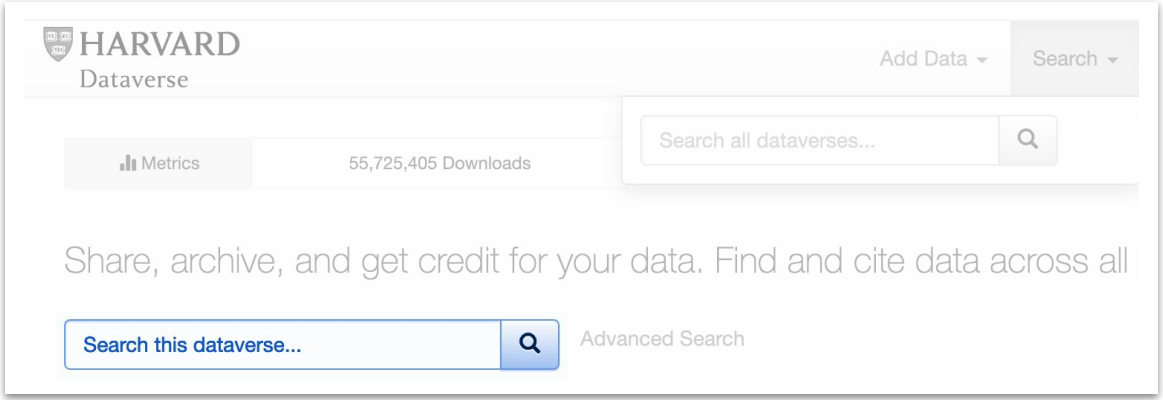
Searching in collections

- Using the general search boxes
- Using advanced search
- Filtering with search facets
- Sorting



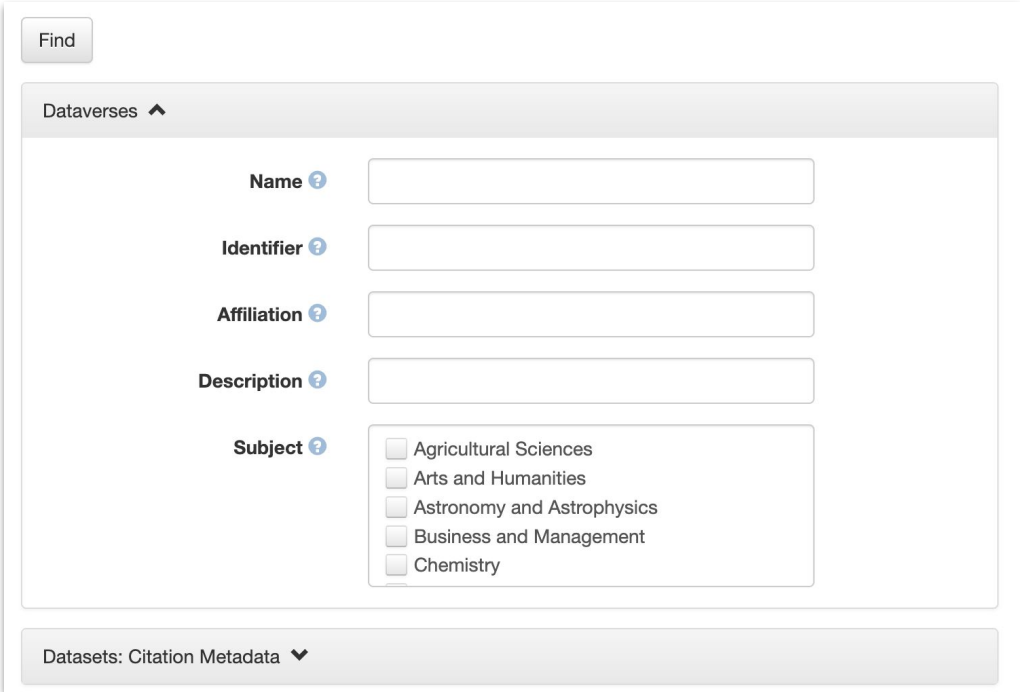
Searching in collections

- Using the general search boxes
- Using advanced search
- Filtering with search facets
- Sorting



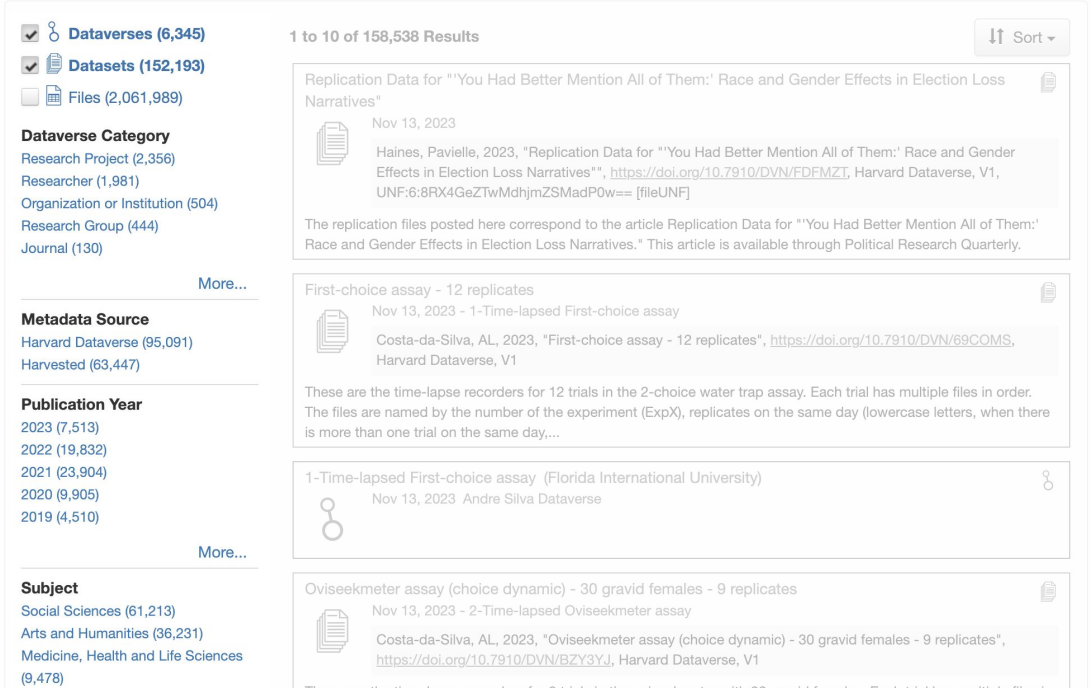
Searching in collections

- Using the general search boxes
- Using advanced search
- Filtering with search facets
- Sorting



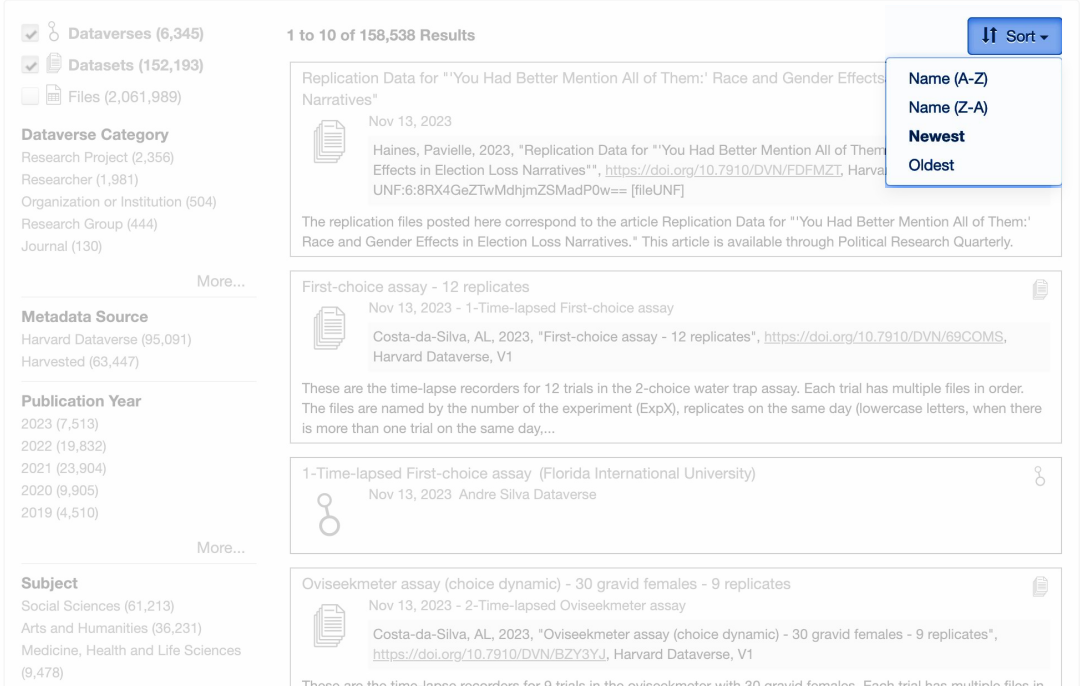
Searching in collections

- Using the general search boxes
- Using advanced search
- Filtering with search facets
- Sorting



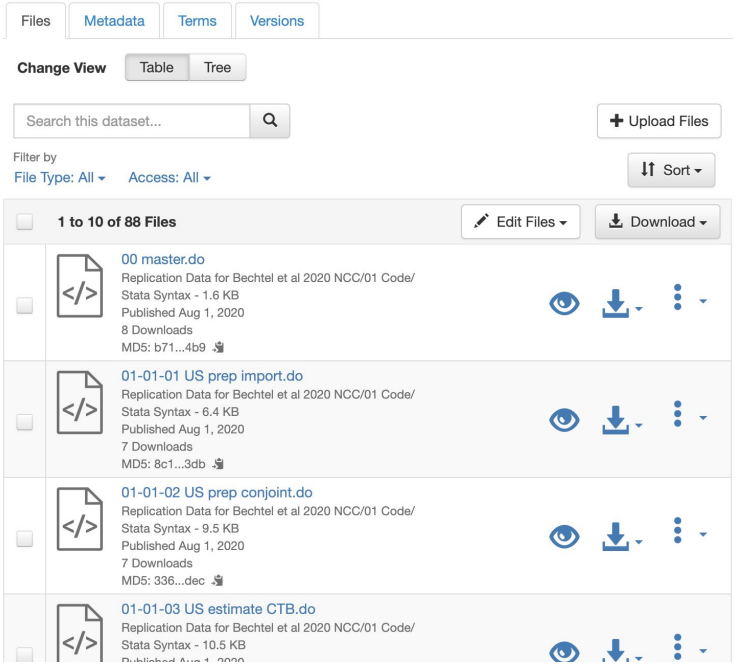
Searching in collections

- Using the general search boxes
- Using advanced search
- Filtering with search facets
- Sorting



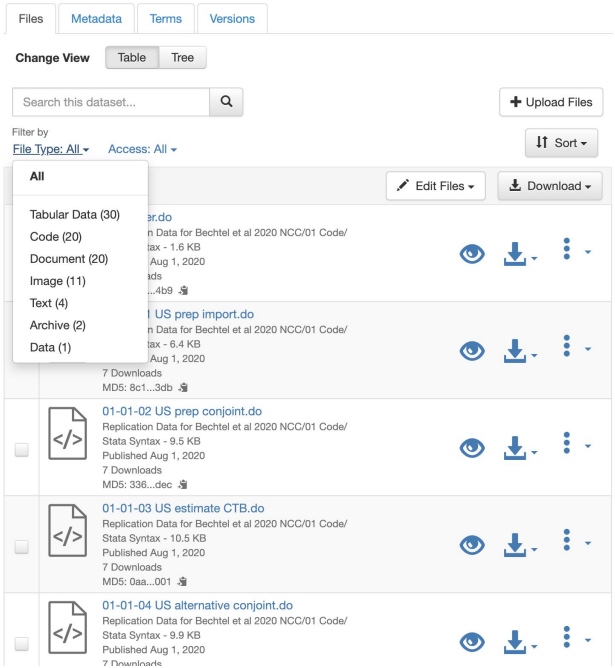
Searching in datasets for files

- Using the general search box
- Filtering with search facets
- Sorting
- Viewing files in folders



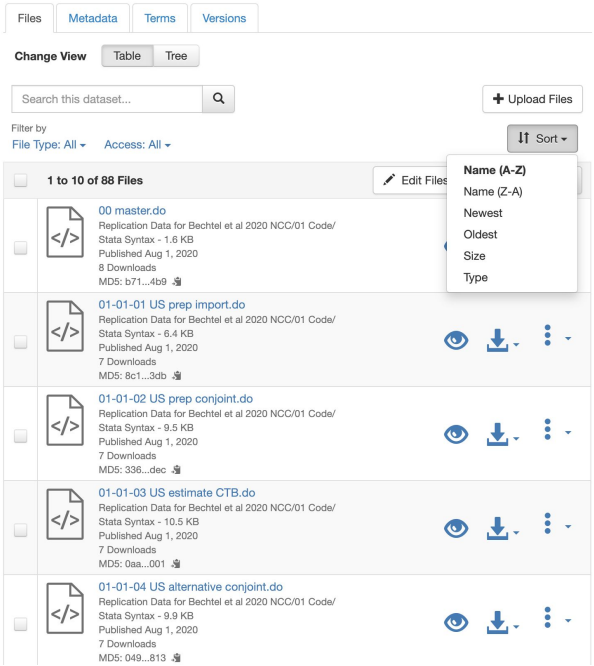
Searching in datasets for files

- Using the general search box
- Filtering with search facets
- Sorting
- Viewing files in folders



Searching in datasets for files

- Using the general search box
- Filtering with search facets
- Sorting
- Viewing files in folders



Searching in datasets for files

- Using the general search box
- Filtering with search facets
- Sorting
- Viewing files in folders

What is “good” metadata?

Metadata that helps others find datasets

Metadata that helps others re-use datasets

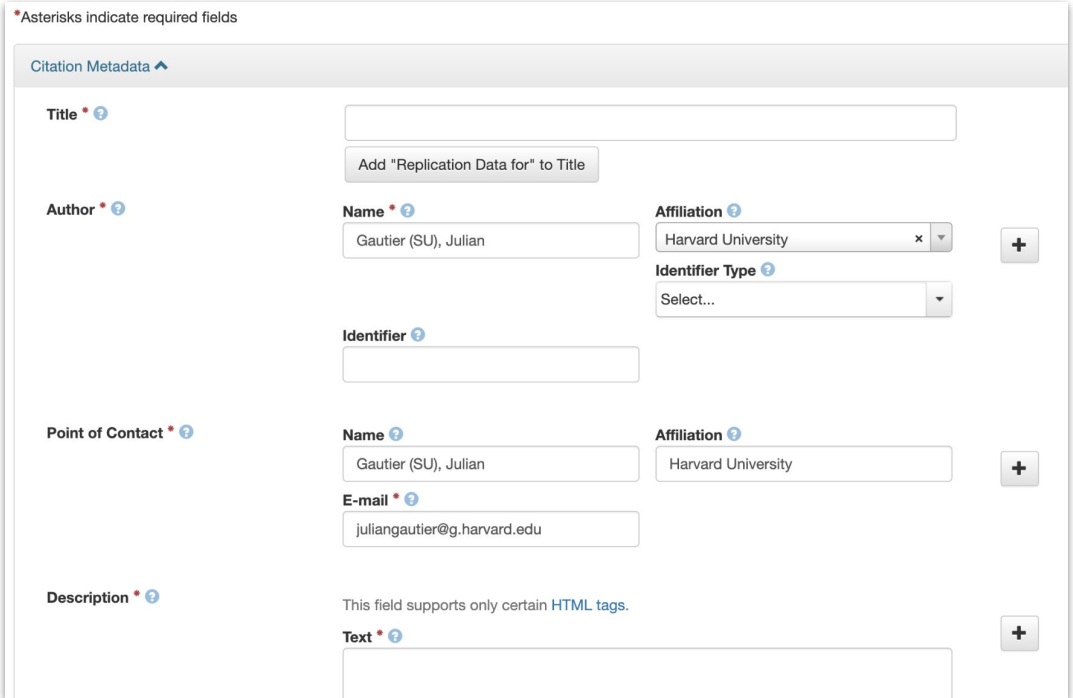
Examples of well-described datasets

What is “good” metadata?

Metadata that helps others find datasets

Metadata fields that discovery systems use when returning search results

Metadata that leverages how people look for data



What is “good” metadata?

Metadata that helps others find datasets

Metadata fields that discovery systems use when returning search results

Metadata that leverages how people look for data

Metadata that helps others find datasets

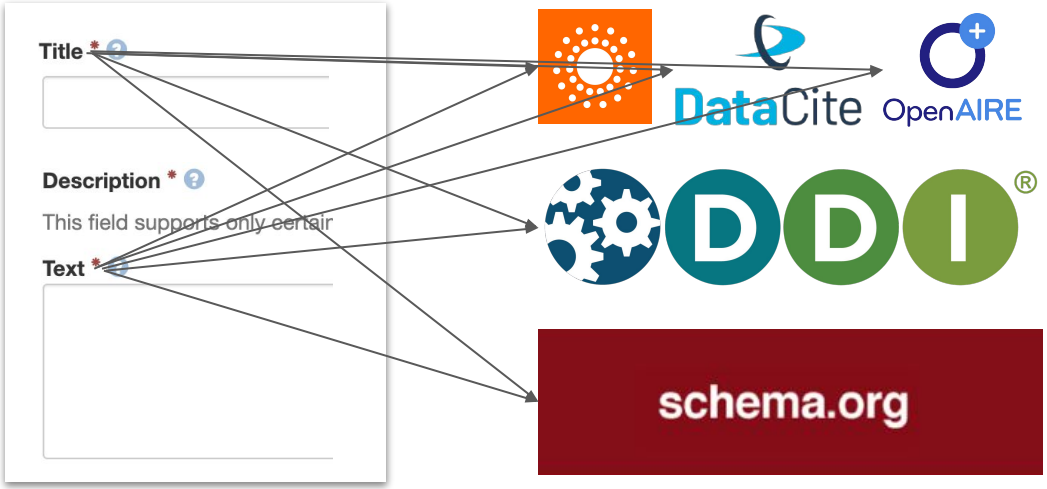
Metadata fields that discovery systems use when returning search results

Metadata that leverages how people look for data

Metadata that helps others find datasets

Metadata fields that discovery systems use when returning search results

Metadata that leverages how people look for data



What is “good” metadata?

Metadata that helps others find datasets

Metadata fields that discovery systems use when returning search results

Metadata that leverages how people look for data

“Everything you always wanted to know about a dataset: Studies in data summarisation”

https://doi.org/10.1016/j.ijhcs.2019.10.004

Laura Koesten, Elena Simperl, Tom Blount, Emilia Kacprzak, Jeni Tennison. Everything you always wanted to know about a dataset: Studies in data summarisation, International Journal of Human-Computer Studies, Volume 135, 2020, https://doi.org/10.1016/j.ijhcs.2019.10.004

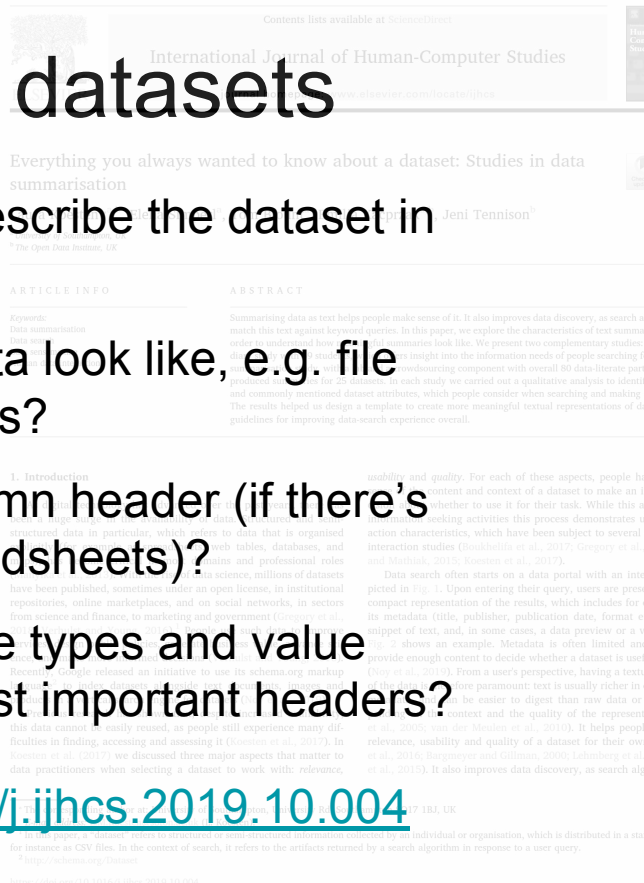
What is “good” metadata?

Metadata that helps others find datasets

Metadata fields that discovery systems use when returning search results

Metadata that leverages how people look for data

1. How would you describe the dataset in one sentence?
2. What does the data look like, e.g. file formats, data types?
3. What are the column header (if there’s tabular data/spreadsheets)?
4. What are the value types and value ranges for the most important headers?



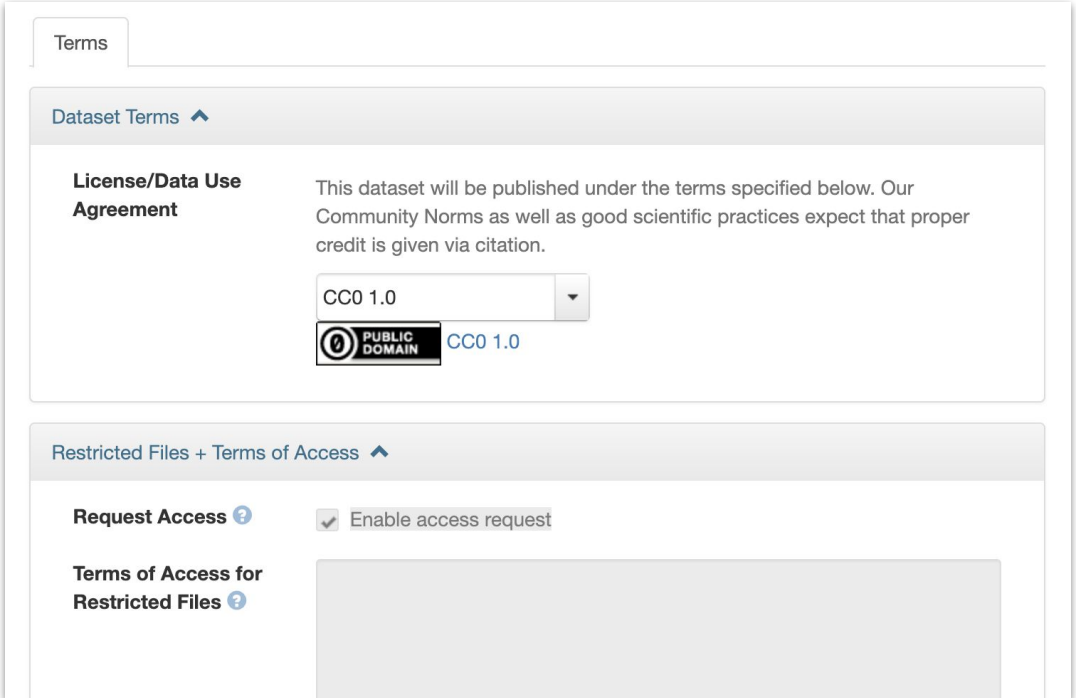
https://doi.org/10.1016/j.ijhcs.2019.10.004

Metadata that helps others re-use datasets

Terms of use and access metadata, e.g. CC0

Links to other resources, e.g. “Related Publications” metadata

Readme files



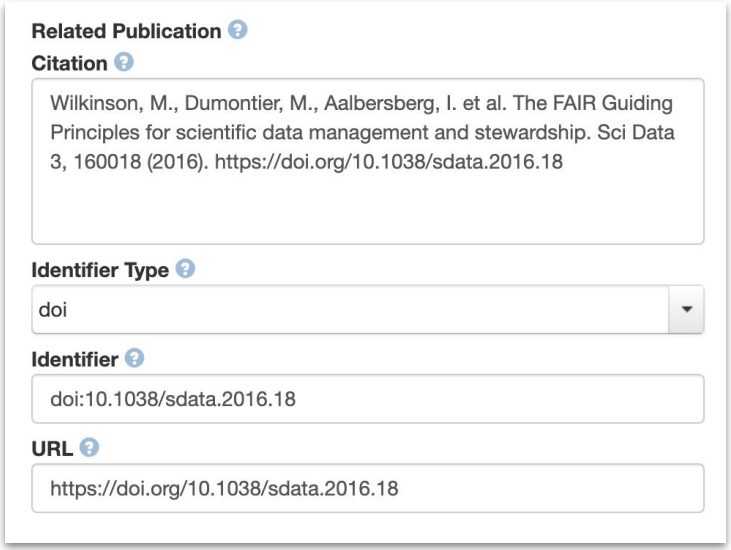
What is “good” metadata?

Metadata that helps others re-use datasets

Terms of use and access metadata, e.g. CC0

Links to other resources, e.g. “Related Publications” metadata

Readme files



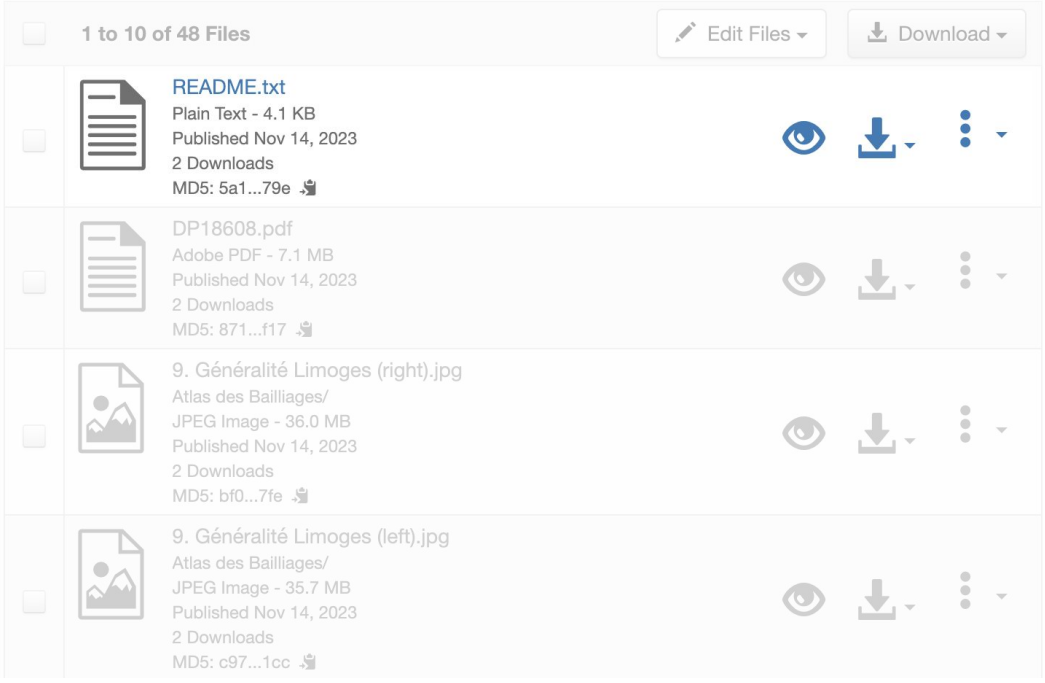
What is “good” metadata?

Metadata that helps others re-use datasets

Terms of use and access metadata, e.g. CC0

Links to other resources, e.g. “Related Publications” metadata

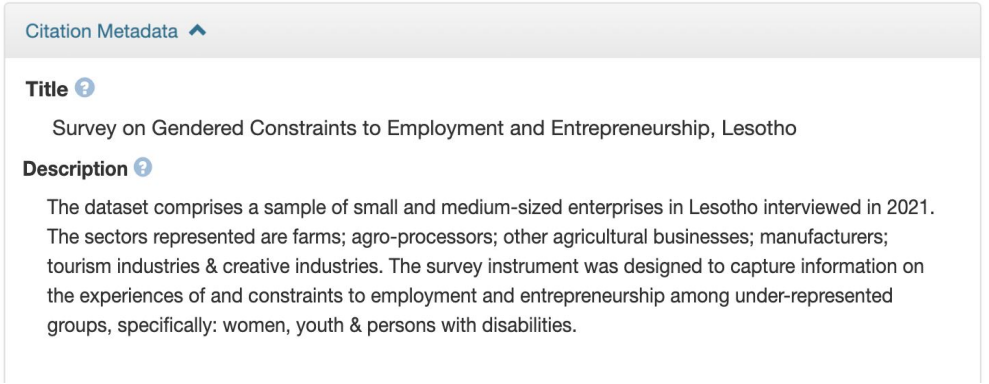
Readme files



What is “good” metadata?

Examples of well-described datasets

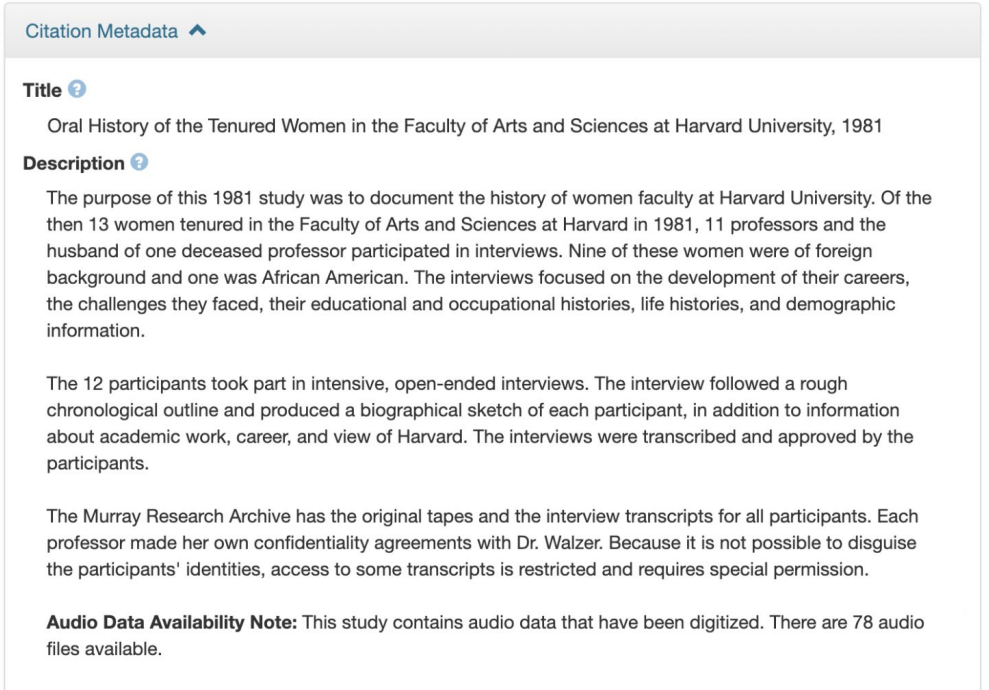
International Food Policy Research Institute



What is “good” metadata?

Examples of well-described datasets

Murray Research Archive



Citation Metadata

Title

Oral History of the Tenured Women in the Faculty of Arts and Sciences at Harvard University, 1981

Description

The purpose of this 1981 study was to document the history of women faculty at Harvard University. Of the then 13 women tenured in the Faculty of Arts and Sciences at Harvard in 1981, 11 professors and the husband of one deceased professor participated in interviews. Nine of these women were of foreign background and one was African American. The interviews focused on the development of their careers, the challenges they faced, their educational and occupational histories, life histories, and demographic information.

The 12 participants took part in intensive, open-ended interviews. The interview followed a rough chronological outline and produced a biographical sketch of each participant, in addition to information about academic work, career, and view of Harvard. The interviews were transcribed and approved by the participants.

The Murray Research Archive has the original tapes and the interview transcripts for all participants. Each professor made her own confidentiality agreements with Dr. Walzer. Because it is not possible to disguise the participants' identities, access to some transcripts is restricted and requires special permission.

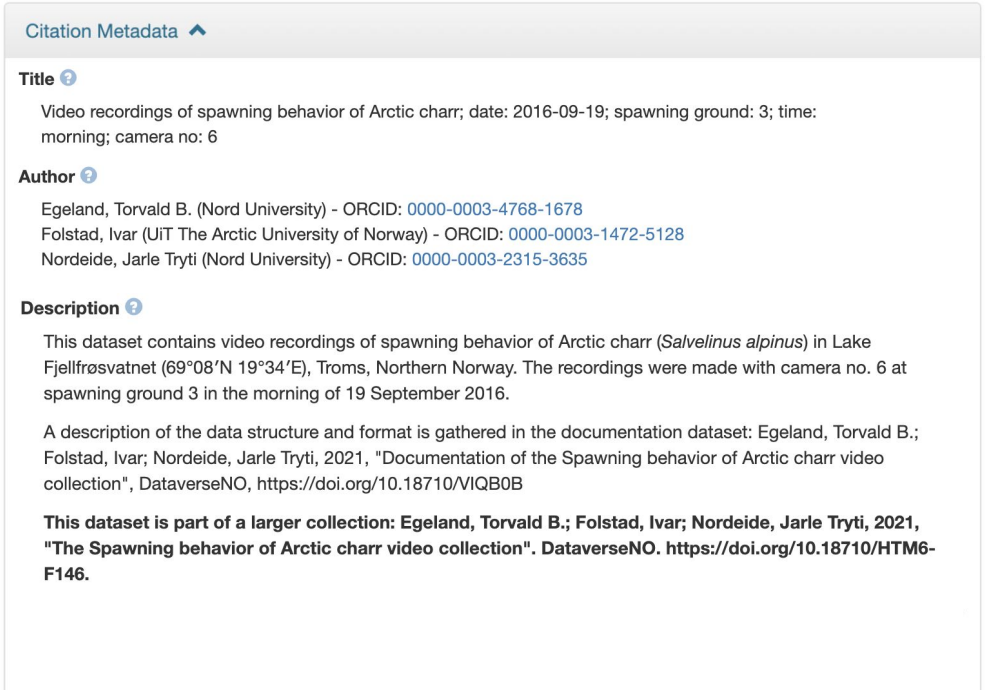
Audio Data Availability Note: This study contains audio data that have been digitized. There are 78 audio files available.

What is “good” metadata?

Examples of well-described datasets

Nord University’s collection in DataverseNO

https://doi.org/10.18710/D1HDHZ



Citation Metadata

Title

Video recordings of spawning behavior of Arctic charr; date: 2016-09-19; spawning ground: 3; time: morning; camera no: 6

Author

Egeland, Torvald B. (Nord University) - ORCID: 0000-0003-4768-1678
Folstad, Ivar (UiT The Arctic University of Norway) - ORCID: 0000-0003-1472-5128
Nordeide, Jarle Tryti (Nord University) - ORCID: 0000-0003-2315-3635

Description

This dataset contains video recordings of spawning behavior of Arctic charr (*Salvelinus alpinus*) in Lake Fjellfrøsvatnet (69°08′N 19°34′E), Troms, Northern Norway. The recordings were made with camera no. 6 at spawning ground 3 in the morning of 19 September 2016.

A description of the data structure and format is gathered in the documentation dataset: Egeland, Torvald B.; Folstad, Ivar; Nordeide, Jarle Tryti, 2021, "Documentation of the Spawning behavior of Arctic charr video collection", DataverseNO, https://doi.org/10.18710/VIQB0B

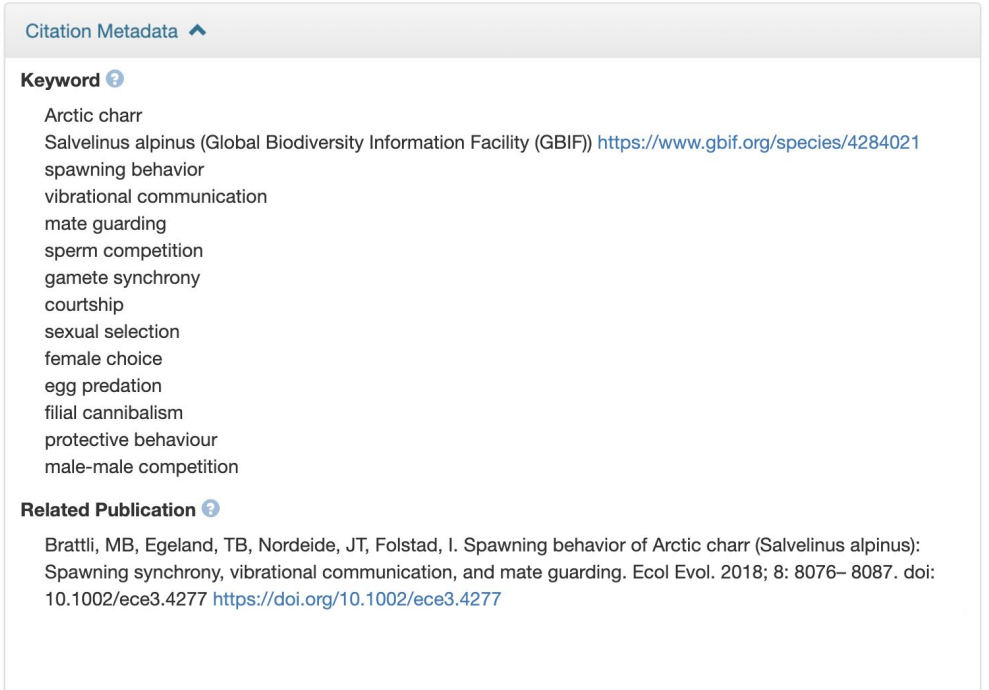
This dataset is part of a larger collection: Egeland, Torvald B.; Folstad, Ivar; Nordeide, Jarle Tryti, 2021, "The Spawning behavior of Arctic charr video collection". DataverseNO. https://doi.org/10.18710/HTM6-F146.

What is “good” metadata?

Examples of well-described datasets

Nord University’s collection in DataverseNO

https://doi.org/10.18710/D1HDHZ


Citation Metadata

Keyword

Arctic charr
Salvelinus alpinus (Global Biodiversity Information Facility (GBIF)) https://www.gbif.org/species/4284021
spawning behavior
vibrational communication
mate guarding
sperm competition
gamete synchrony
courtship
sexual selection
female choice
egg predation
filial cannibalism
protective behaviour
male-male competition

Related Publication

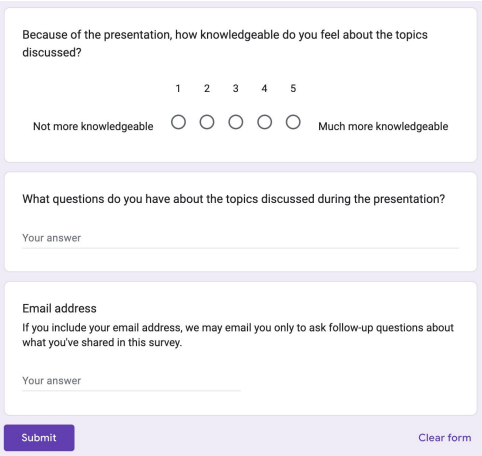
Brattli, MB, Egeland, TB, Nordeide, JT, Folstad, I. Spawning behavior of Arctic charr (Salvelinus alpinus): Spawning synchrony, vibrational communication, and mate guarding. Ecol Evol. 2018; 8: 8076– 8087. doi: 10.1002/ece3.4277 https://doi.org/10.1002/ece3.4277

Questions, survey, upcoming webinars

Q&A

3-question webinar survey

https://forms.gle/Hxyy76fYFJGjFJG78



Upcoming webinars

December 4, 2023
Intermediate Curation on the Dataverse Platform

January 31, 2024
Introduction to Dataverse APIs

2024
Introduction to Metadata in Dataverse Repositories: For Installation Administrators