

Metadata standards in **Dataverse**



Julian Gautier and Gustavo Durand

The Dataverse Project, Institute for Quantitative Social Science (IQSS)
Harvard University

IASSIST 2017 | May 24, 2017

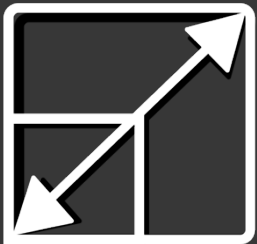


@dataverseorg





Open source data repository
software



Built to support multiple types of
users, data, and workflows



Funded by IQSS and grants, with
partners around the world

Harvard University



23 listed software
installations of various sizes





Harvard Dataverse

Metrics

2,340,571 Downloads



Share, archive, and get credit for your data. Find and cite data across all research fields.

[Find](#)

[Advanced Search](#)

[+ Add Data](#)

☒ [Dataverses \(2,083\)](#)

☒ [Datasets \(73,549\)](#)

☐ [Files \(337,018\)](#)

Dataverse Category

[Research Project \(586\)](#)

[Researcher \(574\)](#)

[Organization or Institution \(170\)](#)

[Journal \(153\)](#)

[Research Group \(26\)](#)

[More...](#)

Metadata Source

[Harvested \(51,644\)](#)

[Harvard Dataverse \(23,988\)](#)

Publication Date

[2015 \(15,613\)](#)

[2011 \(9,627\)](#)

[2012 \(8,152\)](#)

[2016 \(5,032\)](#)

[2009 \(4,020\)](#)

[More...](#)

Subject

[Social Sciences \(31,778\)](#)

[Medicine, Health and Life Sciences \(1,401\)](#)

[Earth and Environmental Sciences \(734\)](#)

[Arts and Humanities \(657\)](#)

[Other \(577\)](#)

[More...](#)

1 to 10 of 75,632 Results

[Sort](#)

[Carolina Plescia 2017, "Replication Data for: Portfolio-specific accountability and retrospective voting: The case of Italy", Harvard Dataverse, V1](#)



Apr 4, 2017 - Italian Political Science Review (IPSR/RISP) Dataverse

Plescia, Carolina, 2017, "Carolina Plescia 2017, "Replication Data for: Portfolio-specific accountability and retrospective voting: The case of Italy", Harvard Dataverse, V1", doi:10.7910/DVN/ITP8S4, Harvard Dataverse, V1

How do voters attribute responsibility for government outcomes when they are the result of a collective decision taken by multiple parties within a coalition government? In this paper we test the argument that in a multiparty coalition system, responsibility attribution should va...

[Walking back to the future: The impact of walking backward and forward on spatial and temporal concepts](#)



Apr 4, 2017

Loeffler, Jonna, 2017, "Walking back to the future: The impact of walking backward and forward on spatial and temporal concepts", doi:10.7910/DVN/FYJ6YT, Harvard Dataverse, V1, UNF:6:iP4Mke9A0lxFU6BvJpj4+Q==

Embodied cognition frameworks suggest a direct link between sensorimotor experience and cognitive representations of concepts (Shapiro, 2011). We examined whether this holds also true for concepts that cannot be directly perceived with the sensorimotor system (i.e., temporal conc...

[Women on the Frontline: Rebel Group Ideology and Women's Participation in Violent Rebellion, Journal of Peace Research 54\(1\)](#)



Apr 4, 2017

Wood, Reed, 2017, "Women on the Frontline: Rebel Group Ideology and Women's Participation in Violent Rebellion, Journal of Peace Research 54(1)", doi:10.7910/DVN/O35OX4, Harvard Dataverse, V2, UNF:6:Wh0vy5Nb9EME0U7uSWORpA==

Replication data for "Women on the Frontline: Rebel Group Ideology and Women's Participation in Violent Rebellion". The Women in Armed Rebellion Dataset (WARD)

[Replication Data for: Unraveling the disease consequences and mechanisms of modular structure in animal social networks](#)



Apr 4, 2017 - Bansal Lab Dataverse

Sah, Pratha, 2017, "Replication Data for: Unraveling the disease consequences and mechanisms of modular structure in animal social networks", doi:10.7910/DVN/WP3BV7, Harvard Dataverse, V2, UNF:6:lpzFUdUw80XMSnUcA2xNaw==

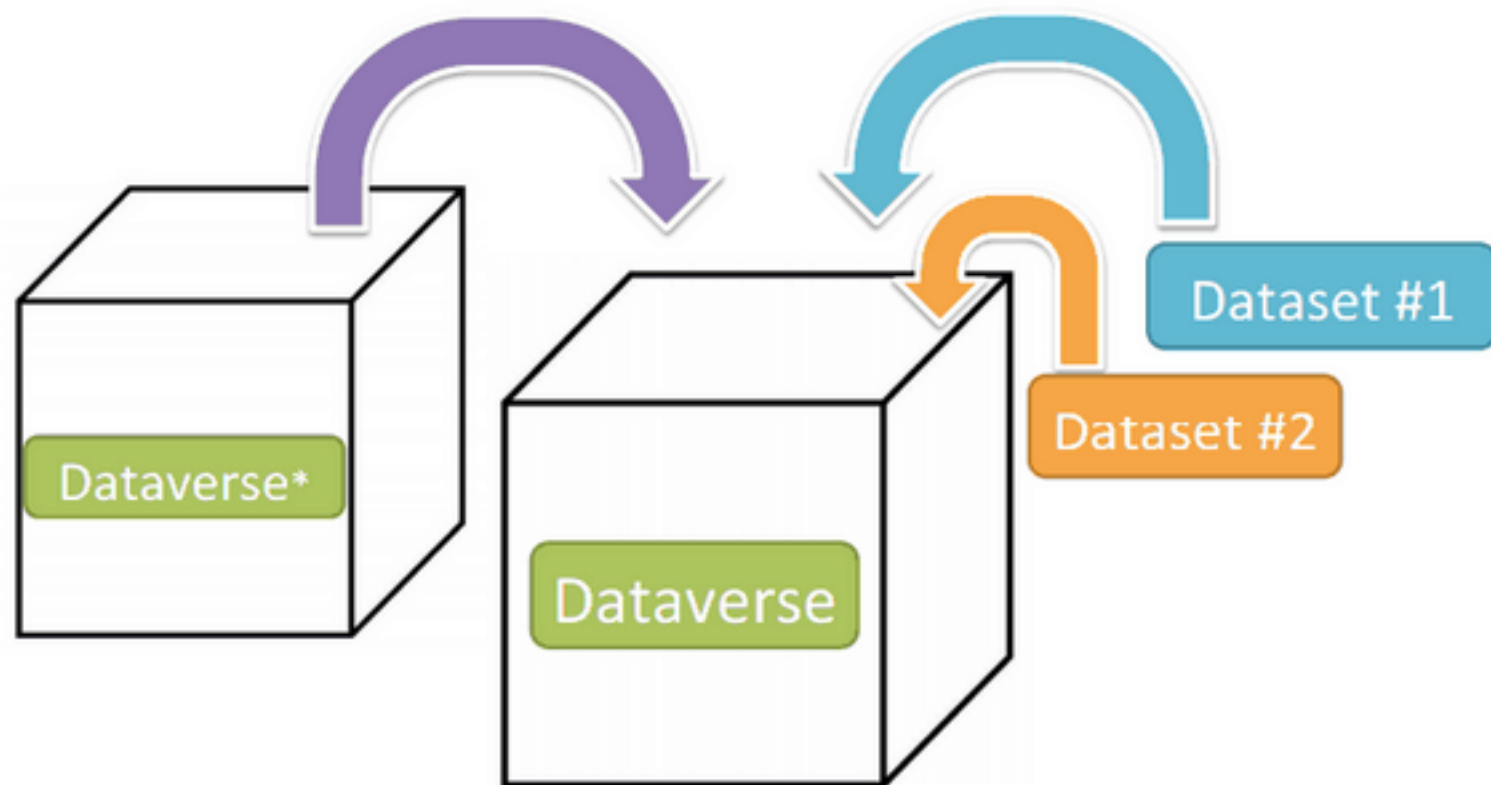
The dataset and link to codes used in our paper "Unraveling the disease consequences and mechanisms of modular structure in animal social

Minimal curation* (must be “data”,
can’t be sensitive data**)

** Preparing curation services with Harvard Library*

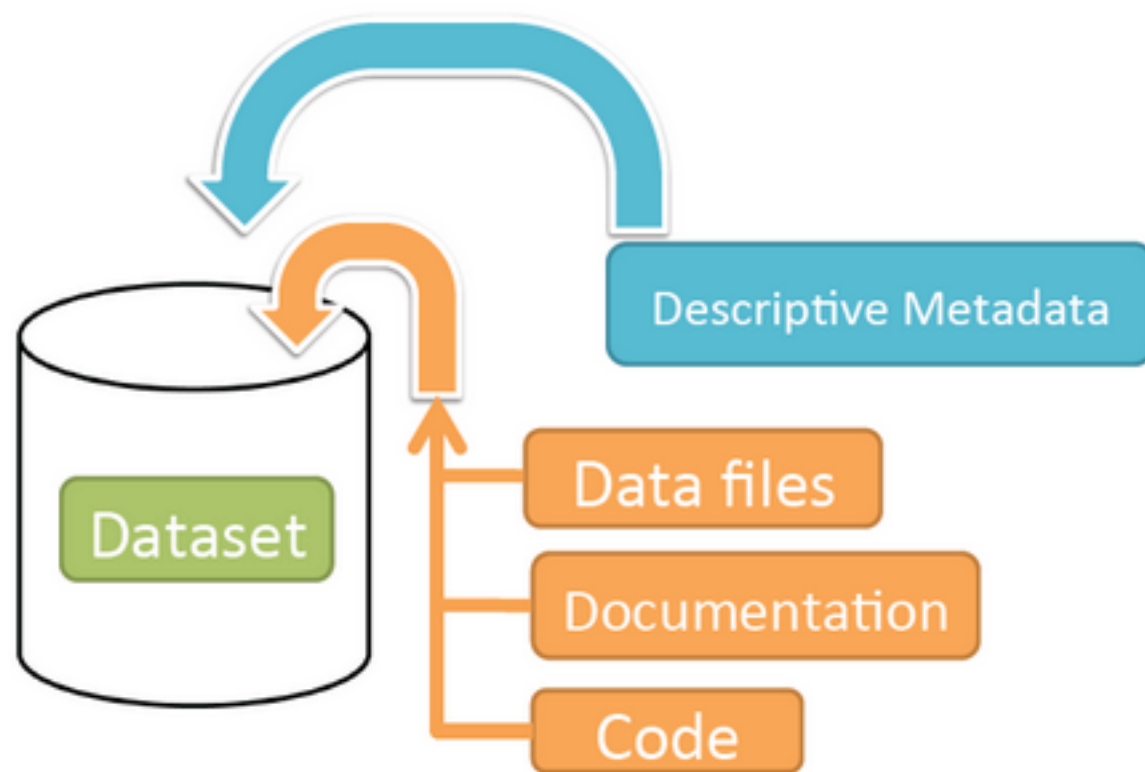
*** Design, development work being done now to
support sensitive data*

Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses**

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

Citation metadata

- Authors
- Title
- Description
- Dates
- Contact
- Subject
- ...

Domain metadata blocks

- Social science
- Geospatial
- Life Sciences
- Journals

Terms of use and access

- Waiver
- Terms of Use
- Confidentiality Declaration
- Terms of Access
- Data Access Place
- ...

File metadata

- File header metadata
- File description
- Variable metadata

DDI XML metadata export

- Dataset level

Files

Metadata

Terms

Versions

Citation Metadata

Dataset Persistent ID	doi:10.5072/FK2/OULBYE
Publication Date	2017-04-03

Export Metadata

Dublin Core

DDI

JSON

```
<codeBook xmlns="ddi:codebook:2_5" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="ddi:codebook:2_5
http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd"
version="2.5">
  <docDscr>
    <citation>
      <titlStmt>
        <titl>Testing tabular file ingest</titl>
        <IDNo agency="DOI">doi:10.5072/FK2/OULBYE</IDNo>
      </titlStmt>
      <distStmt>
        <distrbtr>Demo Dataverse</distrbtr>
        <distDate>2017-04-03</distDate>
      </distStmt>
      <verStmt source="DVN">
        <version date="2017-04-03" type="RELEASED">1</version>
      </verStmt>
      <biblCit>
        Gautier, Julian, 2017, "Testing tabular file ingest",
        doi:10.5072/FK2/OULBYE, Demo Dataverse, V1, UNF:6:InN9v61342Te2eWbTnNsHg==
      </biblCit>
    </citation>
  </docDscr>
  <studyDscr>
    <citation>
      <titlStmt>
        <titl>Testing tabular file ingest</titl>
        <IDNo agency="DOI">doi:10.5072/FK2/OULBYE</IDNo>
      </titlStmt>
      <rspStmt>
        <AuthEnty affiliation="Harvard University">Gautier, Julian</AuthEnty>
      </rspStmt>
      <prodStmt/>
      <distStmt>
        <distrbtr>Demo Dataverse</distrbtr>
        <distDate>2017-04-03</distDate>
      </distStmt>
    </citation>
    <studyInfo>
      <subject>
        <keyword>Other</keyword>
        <keyword>Computer and Information Science</keyword>
      </subject>
      <abstract>Testing tabular file ingest</abstract>
      <sumDscr/>
      <contact affiliation="Harvard University"
        email="juliangautier@g.harvard.edu">Gautier, Julian</contact>
      <depositr>Gautier, Julian</depositr>
      <depDate>2017-04-02</depDate>
    </studyInfo>
    <method>
      <dataColl/>
      <anlyInfo/>
    </method>
  </studyDscr>
</codeBook>
```

DDI XML metadata export

- Variable level for ingested tabular files

Metadata

Versions

File Metadata

Download URL	https://demo.dataverse.org/api/access/datafile/15321
Original File MD5	971e301ae4350a5dcdcf5f8fd00235bb9
UNF	UNF:6:InN9v61342Te2eWbTnNsHg==
Publication Date	2017-04-03

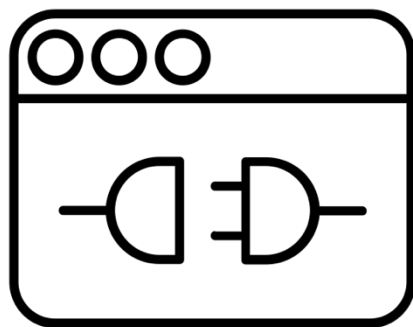
Export Metadata

Dublin Core

DDI

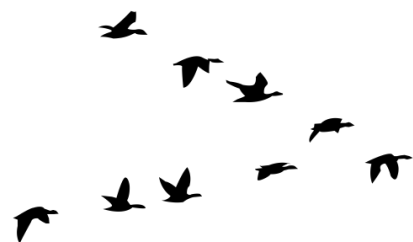
JSON

```
<filetype>text/tab-separated-values</filetype>
</fileTxt>
<notes level="file" type="VDC:UNF" subject="Universal Numeric
Fingerprint">UNF:6:InN9v61342Te2eWbTnNsHg==</notes>
</fileDscr>
<dataDscr>
  <var ID="v187165" name="article_id" intrvl="discrete">
    <location fileid="f15321"/>
    <labl level="variable">article_id</labl>
    <sumStat type="medn">1912.5</sumStat>
    <sumStat type="mode">.</sumStat>
    <sumStat type="stdev">1104.0380428228007</sumStat>
    <sumStat type="mean">1912.5</sumStat>
    <sumStat type="min">1.0</sumStat>
    <sumStat type="vald">3824.0</sumStat>
    <sumStat type="invd">0.0</sumStat>
    <sumStat type="max">3824.0</sumStat>
    <varFormat type="numeric"/>
    <notes subject="Universal Numeric Fingerprint" level="variable"
type="Dataverse:UNF">UNF:6:hQ77fNuUa24FqZSNcuJ20A==</notes>
  </var>
  <var ID="v187164" name="publish_date" intrvl="discrete">
    <location fileid="f15321"/>
    <labl level="variable">publish_date</labl>
    <varFormat type="character"/>
    <notes subject="Universal Numeric Fingerprint" level="variable"
type="Dataverse:UNF">UNF:6:EHT0luqwm7IAQYJshMQUBg==</notes>
  </var>
  <var ID="v187161" name="article_source_link" intrvl="discrete">...</var>
  <var ID="v187160" name="title" intrvl="discrete">...</var>
  <var ID="v187162" name="subtitle" intrvl="discrete">...</var>
  <var ID="v187163" name="text" intrvl="discrete">...</var>
</dataDscr>
<otherMat ID="f15320"
URI="https://demo.dataverse.org/api/access/datafile/15320" level="datafile">...
</otherMat>
<otherMat ID="f15319"
URI="https://demo.dataverse.org/api/access/datafile/15319" level="datafile">...
```



APIs

- Data variable metadata access through APIs
- TwoRavens and Data Explorer



Migration

- Dataset, file level DDI used during Harvard Dataverse for 3.6 to 4.0 migration



Harvesting

- DDI captured when harvesting datasets

Dataverse's approach to metadata standards

- Extensible
- Automated
- Informed by behavior
- Reusable

FGDC.GOV
FEDERAL GEOGRAPHIC DATA COMMITTEE



AIMS AGRICULTURAL
INFORMATION
MANAGEMENT
STANDARDS



UNIFIED ASTRONOMY THESAURUS



DDI Workflows for Dataverse Working Group:

<https://ddi-alliance.atlassian.net/wiki/display/DDI4/DDI+Workflows+for+Dataverse>
(<http://bit.ly/ddi-dv>)

- Some use cases for expanded DDI use in Dataverse:
 - Support for DDI controlled vocabularies
 - More DDI fields
 - Making it easier to import DDI metadata from one system into Dataverse.

The screenshot shows the Atlassian Wiki page for "DDI Workflows for Dataverse". The page is titled "DDI Workflows for Dataverse" and was created by Wendy Thomas, last modified by Amber Leahey on Mar 16, 2017. The page content includes a "Purpose and Goals" section, a "Related Documents" section, and a "Members" table.

Purpose and Goals
To develop use cases for improved support of DDI in the Dataverse repository platform, and, to coordinate development projects that utilize DDI generated by or with Dataverse in support of research data across the lifecycle.

Related Documents

- [Working group activity](#)
- [Draft use cases](#) (please contribute)
- NADDI Conference presentations & meetings (April 5-7th)
- IASSIST BoF session
- Dataverse Community Meeting

Members

Name	Contact
Amber Leahey	amber.leahey@uto
Danny Brooke	dannybrooke@g.h
Johan Fihn	johan.fihn@snd.gu
Olof Olsson	olof.olsson@snd.gu
Wolfgang Zenk-Möltgen	Wolfgang.Zenk-Moeltgen@gesis.o
Esra Akdeniz	Esra.Akdeniz@gesi
Christina Guldfield Madsen	cgm@sa.dk
Anne Sofie Fink	asf@sa.dk

Thanks!

Join the Dataverse community!



#irc

dataverse.org/contact



Dataverse Community Meeting 2017
June 14, 15, 16 at Harvard University



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).