Title

ICPSR Bibliography of Data-related Literature: Data-Linked Publication Citation Files, Full Download, August 1, 2000--June 30, 2025

Project Description

Principal Investigator(s)

Inter-university Consortium for Political and Social Research

Summary

This data file, taken from the database known as the ICPSR Bibliography of Data-related Literature, is a collection of citations to publications. Each citation record in the file represents one publication that is associated with one or more ICPSR studies and/or series of studies in ICPSR's curated study holdings. This dataset contains 119,479 enhanced citation records.

The ICPSR Bibliography was created to be a resource for ICPSR data users, especially as a searchable database for reusable research data discovery via the literature. This dataset is a static representation of that database, downloaded at one point in time. Researchers using this dataset for bibliometric analysis should be aware that the collection is not exhaustive, and it does not claim to contain the universe of findings that could be associated with each study. Additionally, over the years, due to the vagaries of funding, the work of collecting and linking citations was more intensive for some studies than others. Also, some data usage is not well cited in the scholarly literature, and is difficult to find. For those reasons, the collection is skewed in its representation. It may under-report utilization of research data and should only be viewed as a partial statement of its impact.

Before using this dataset to conduct analyses, researchers are encouraged to contact ICPSR staff familiar with the ICPSR Bibliography at [icpsrbibliography@umich.edu].

Funding Sources

The initial creation of the publicly searchable database and back-end architecture of the ICPSR Bibliography of Data-related Literature was funded in 1999 by the National Science Foundation (SES-9977984). Since its launch in 2000, the Bibliography has been supported by the ICPSR membership and by various US federally-funded, foundation-funded, and privately funded projects and topical archives at ICPSR.

Scope of Project

Subject Terms

- Bibliographic resource
- Data-linked scholarly literature
- Social sciences
- Behavioral health sciences
- Linked citation records

Collection Date(s)

8/1/2000 - 6/30/2025

Universe

Data-related literature—that is, journal articles, reports, books, conference presentations, dissertations, and other information objects that meet at least one of the following criteria:

1. Contain analysis of data archived by ICPSR (including publications by the original principal investigator as well as articles containing secondary analysis). These analyses can include the use of the original PI data, data disseminated in printed form or on the ICPSR Web site, or where the exact source cannot be determined from the publication, as long as a version of the data has been archived by ICPSR.
2. Contain discussion, critique, or extension of another's data analysis of a kind that would have required the author to work with the original raw data.
3. Contain discussion, critique, or extension of another's data analysis, which might lead a reader to consult the original raw data.
4. Contain discussion or critique of the data source in general, such as caveats on the use of a particular study that contain more than the simple statement that the data are problematic.
5. Quote the data in charts that are newly compiled by the author (i.e. not charts reproduced from someone else's publication of the data).
6. Contain discussion or critique of the data collection process, study design, or methodology.

Data Types(s)

Citations to publications in the form of comma-separated values (CSV), enhanced with ICPSR study numbers and study titles, and/or series numbers and series titles.

Methodology

The ICPSR Bibliography of Data-related Literature is populated from multiple sources:

1. PIs/Depositors can provide a list of related publications that ICPSR Curators send to the ICPSR Bibliography staff, who add the citations to the Bibliography's database after checking to make sure each publication meets the collection criteria and the citation information is accurate.
2. Data downloaders who follow the terms of use requiring them to make ICPSR aware of any of their publications that analyze the data they accessed. They can email citations directly to the Bibliography at icpsrbibliography@umich.edu.
3. Restricted data users are required to submit annual reports that include a list of their publications and presentations that were based on the data they received from ICPSR. Those annual reports get forwarded to the ICPSR Bibliography staff, who verify and enter the citations into the Bibliography database.
4. ICPSR Curation staff may encounter primary or secondary works when curating a study. They send the citations to the ICPSR Bibliography staff, who verify and add the citations to the Bibliography's database.
5. ICPSR Bibliography staff find the vast majority of the works that are in the ICPSR Bibliography. They do so by detecting data use in the literature--verifying both explicitly cited data use as well as informal and incomplete references. They create queries and set up email alerts in social science and behavioral health literature databases, e.g., Google Scholar, ProQuest, PubMed, and many others. They query these databases for individual studies using search interfaces, or for the full study collection using publisher-provided APIs.

Collection Notes

- This static file will be updated and self-published periodically. The live ICPSR Bibliography of Data-related Literature is updated daily. The current (most up-to-date) version of the collection is searchable via ICPSR's website search interface.

- Researchers using this dataset for bibliometric analysis should be aware that the collection is not exhaustive, and it does not claim to contain the universe of findings that could be associated with each study. Additionally, over the years, due to the vagaries of funding, the work of collecting and linking citations was more intensive for some studies than others. Also, some data usage is not well cited in the scholarly literature, and is difficult to find. For those reasons, the collection is skewed in its representation. It may

under-report utilization of research data and should only be viewed as a partial statement of its impact.

- ICPSR study data are found to be used in a variety of types of peer-reviewed and grey literature. There are 12 RIS types (reference types) in this dataset, labeled: RIS Type. Each is described below:

  1. **Journal Articles** most often come from peer-reviewed journals.
  2. **Preprints** are author-submitted manuscripts that have reached the final stage before undergoing peer review. They are made publicly available on preprint servers. In the Bibliography, a manuscript is categorized as a preprint if:
     a. It is explicitly identified as a preprint and has a registered digital object identifier (DOI); or
     b. It is available on an official preprint server and has been assigned a DOI.
     Having a registered DOI distinguishes a preprint from other manuscripts, indicating the preprint is now part of the scholarly record. If a manuscript lacks a DOI, it is treated as a Document (see below). Note: this reference type was first delineated in the ICPSR Bibliography in December 2024. Therefore, some works that match the definition of a Preprint but were collected prior to that date may exist in the collection categorized as a Document or a Report.
  3. **Books** are authored or edited by one or more people or entities, generally contain multiple chapters, and are distributed by a publisher, either in print or via e-book, or both. If the editor is the author of all of the book's content, the editor appears as the author. If the book is part of a series, the series title is provided, when possible.
  4. **Book Sections** are generally considered chapters in books, authored either by the book's editor or more often, by other people. If the book section is part of a book that is in a series, the series title is provided, when possible.
  5. **Reports** are published works that are not Journal Articles. They are typically issued by government entities, research centers, non-profits, or similar organizations. Reports are often part of a series (e.g., working papers, technical reports) but may also be standalone publications. They include complete publishing information, such as the report series title, publication date, report number, and publisher details.
  6. **Theses** include dissertations, master's theses, and capstone projects (mostly in the form of posters) for master's or PhD-level work. Undergraduate honors theses or posters are not collected. The Type of Work field will be populated with either Dissertation or Thesis.
  7. **Conference Presentations** include both conference abstracts and full papers, as well as posters. Conference proceedings that are published in journals are usually collected as Journal Articles.
  8. **Newspaper Articles** from reputable publications are added when their content is considered to meet the Bibliography's collection criteria.

9. **Magazine Articles** are included if the content is considered to be scholarly and meets the Bibliography's collection criteria.
10. **Electronic Sources** are published on websites and are mostly blogs, but they are not necessarily text-only, e.g., they can be infographics. Often, the Type of Work field will be populated with the specific type.
11. **A/V Materials** were originally considered to be videos, but can include digital slide presentations and podcasts. Often, the Type of Work field will be populated with the specific type.
12. **Documents** are materials typically found online and self-published on authors' websites, ResearchGate, Academia.edu, or university websites. They are characterized by having less formal publication information, are usually in PDF or Word format, and are not formatted for a specific publisher. Documents are often working papers that are not part of a series, or drafts that are not final versions. A subset includes job market papers, which are chapters of dissertations—mainly in Economics—that do not qualify as dissertations themselves. Unlike Preprints, Documents are not hosted on official preprint servers and lack persistent identifiers, such as DOIs.

- The collection contains works that were written by the principal investigators who created the data collections, as well as by secondary authors who reused the data in those collections. There is no distinguisher in this dataset that indicates which record is primary and which record is secondary.

- Not all publications appearing in this collection will use the most recent version of the data. Oftentimes the version was not provided in a publication, or else the version changed after the publication was added to the Bibliography's collection.

- In the collection, there could be citations to publications that most likely analyzed data distributed at ICPSR, but from the information provided this could not be determined absolutely.

- The collection may contain citations to publications that analyzed data very similar to those at ICPSR, but that were clearly retrieved from a different distributor. Such publications are included in the collection for the purposes of edification.

- The ICPSR Bibliography contains citations to publications associated with ICPSR's curated studies only. The ICPSR Bibliography does NOT contain citations related to data projects self-published on ICPSR platforms.

Bibliography Variables

Label	Description
Reference ID	Unique identifier assigned to citation record.
Title	Publication title.
Author(s)	Publication author(s). All authors (last name, first name, initial.) appear in the same field, semicolon delimited.
Study Number(s)	ICPSR study number(s) that has been associated with the publication. (All ICPSR studies are assigned a unique "study number.") All study numbers appear in the same field, semicolon delimited. An ICPSR Study is a collection of data files, and the documentation files that are needed to understand the data files. Studies each have a unique home page on the ICPSR website. Most study home pages display searchable lists of citing publications. Those lists are dynamically compiled from citation records collected in the ICPSR Bibliography of Data-related Literature database.
Study Title(s)	ICPSR study title(s) associated with the study number(s) that has been associated with the publication. All study titles appear in the same field, semicolon delimited. Study Titles(s) appear in the order of the correlating Study Number(s) field. Therefore, the first study number is associated with the first study title, the second study number with the second study title, etc. An ICPSR Study is a collection of data files, and the documentation files that are needed to understand the data files. Studies each have a unique home page on the ICPSR website. Most study home pages display searchable lists of citing publications. Those lists are dynamically compiled from citation records collected in the ICPSR Bibliography of Data-related Literature database.
Study Owner(s)	Acronym representing the name of the owner(s) of the study or studies, in terms of the ICPSR archive that sponsored the creation of the study. All study owner names appear in the same field, semicolon delimited. A partial list of ICPSR archives is available at: Thematic Data Collections at ICPSR [https://www.icpsr.umich.edu/sites/icpsr/find-data/working-together/thematic-collections].

	Study Owner(s) appear in the order of the correlating Study Number(s) field. Therefore, the first study number is associated with the first study owner name, the second study number with the second study owner name, etc.
Secondary Title	If publication is a *Book Section*, this field provides the title of the book. If publication is a *Conference Presentation*, this field provides the name of the conference.
Place of Publication	If publication is a *Book* or *Book Section*, this field corresponds to the publisher's location. If publication is a *Conference Presentation*, this field is the location of the conference.
Publisher	If publication is a *Book* or *Book Section*, this field is the name of the publisher. If publication is a *Conference Presentation*, this field provides the sponsoring organization (often aligning with the conference name). If publication is a *Thesis*, this field is the name of the degree-granting institution affiliated with the item.
Volume	Volume number of publication, if applicable. Sometimes the volume number for *Books* is contained within the title field.
Number	If publication is a *Journal Article*, this field is the journal issue number. If publication is a *Report*, this field is the report number. Report numbers can sometimes instead appear in the secondary title field, however.
Starting Page	Starting page for publication, if applicable.
Ending Page	Ending page for publication, if applicable.
Book Series Title	Title of book series, if applicable.
Type of Work	Free-text field providing more detailed type of work to account for rigidity in RIS Type field, if applicable. Mostly used only to distinguish dissertations from theses, or to specify *Electronic Sources* and *A/V Materials*.
Date Added	Date that the record was added to the ICPSR Bibliography.
URL PDF URL	A URL to a webpage containing the publication, a PDF, or an abstract. Due to the ephemeral nature of URLs,

Abstract URL	there is no guarantee any given URL in the record is still active.
Series Number(s)	If no more specific study number(s) could be associated with the citation record, the ICPSR series number(s) is associated with the citation record. (All ICPSR series have a unique series number assigned by ICPSR.) Some records contain both study numbers and series numbers. All series numbers appear in the same field, semicolon delimited. An ICPSR series [https://www.icpsr.umich.edu/web/ICPSR/search/series] is a named collection of related studies.Typically the studies in an ICPSR series are produced by the same group of investigators, and either explore different facets of the same topic, or repeat the same investigation over time.
Series Title(s)	The ICPSR series title associated with the series number that has been associated with the citation record. All series numbers appear in the same field, semicolon delimited. An ICPSR series [https://www.icpsr.umich.edu/web/ICPSR/search/series] is a named collection of related studies.Typically the studies in an ICPSR series are produced by the same group of investigators, and either explore different facets of the same topic, or repeat the same investigation over time. Series Title(s) appear in the order of the correlating Series Number(s) field. Therefore, the first series number is associated with the first series title, the second series number with the second series title, etc.
Where Found	Free-text field indicating where the citation was found.
Date Published	Date of publication. This can contain "in press" if the citation is for a *Journal Article* and was collected before the final publication details were known, and the record has yet to be updated with the final volume, issue, and page spread information. Note there are "in press" articles in the file, which may now have press details. These will be updated in future releases.
ISSN	ISSN of journal for *Journal Articles*, if applicable.
DOI	Digital Object Identifier of publication, if applicable.

Year Published	Year of publication.
RIS Type	Provides the RIS Type code to indicate type of reference (see Collection Notes for a description of each): ● ADVS = Audiovisual Material ● BOOK = Book ● CHAP = Book section ● CONF = Conference Presentation ● GEN = Document ● ELEC = Electronic Source ● JOUR = Journal Article ● MGZN = Magazine Article ● NEWS = Newspaper Article ● RPRT = Report ● THES = Thesis ● MANSCPT = Preprint
Data Cited by ICPSR	If ICPSR was explicitly cited in publication, this field will be marked "EXPLICIT." Use of an ICPSR DOI by the author of the work is not required, only an indicator that the data were accessed at ICPSR. Note: this field has only been used since September 2019, so this information is not available for the whole collection. Articles entered prior to 2019 have been randomly retrofitted. An absence of an explicit tag does not necessarily indicate ICPSR was not explicitly cited.