

FAIR Open Metadata: A Case Study of RePEc

Anna Oates Schlaack¹  and Christian Zimmermann² 

¹ University of Illinois Urbana-Champaign, ² Federal Reserve Bank of St. Louis

Abstract

This report describes a research project about the use of RePEc metadata and its accordance with FAIR principles. The authors summarize successful aspects of a metadata schema that has underpinned the research landscape in economics for nearly thirty years. This project describes the use cases of 100 studies that have leveraged RePEc metadata and future research that will document the challenges of metadata citation and opportunities to engender open metadata efforts.

1. Introduction

This paper describes an ongoing project that studies how a unique metadata schema has been leveraged by researchers. Specifically, the authors look at RePEc, a metadata project in economics for published research, preprints, persons, and organizations that has been active since 1997. The unique appeal of this metadata led to about 100 scholarly studies leveraging it. The authors investigate why and how these studies use and cite RePEc, chiefly with respect to FAIR principles.

The Barcelona Declaration on Open Research Information was launched in 2024 to make research metadata, or research information, part of the open science movement [1]. Research information adds value beyond research discoverability, as RePEc nicely demonstrates. This research contributes to the dialogue around open metadata, specifically focusing on the gaps and challenges of existing metadata for scholarly reuse. Specifically, this research investigates the gaps and opportunities to improve metadata structures and interoperability for meaningful scholarship, and it identifies requisite qualities of FAIR open metadata.

1.1. Data Citation

Published in 2016, the FAIR principles provide guidance to improve the reuse of digital content, prioritizing findability, accessibility, interoperability, and reusability [2]. Central to reuse is data or metadata citation [3]. Metadata providers want to know where their data is used, notably to advocate for continued resources to sustain their work. While citation of scholarship follows established standards and practices, data citation is haphazard [4], despite longstanding efforts, such as the Joint Declaration of Data Citation Principles or Make Data Count [5], [6], [7], [8], [9]. There is no widely used standard or even recognition across disciplines that data needs to be cited like scholarship, which makes it challenging to document how data is used. This problem is even more acute for metadata. Even if researchers mention how or where they find data, it is not done consistently in a formal and machine-readable approach.

Short Paper

Available Aug 01, 2026

Correspondence to
Anna Oates Schlaack
annaio2@illinois.edu



Copyright © 2026 Schlaack & Zimmermann. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license, which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

1.2. *RePEc*

Ideated at the emergence of the world wide web, Research Papers in Economics (RePEc) was established in 1997 to democratize scholarly communication for researchers in economics. RePEc is a community-led effort that relies upon economic researchers and publishers to maintain a high-quality resource with rich networks that enhance discovery and usability of metadata as a data source for research on economic scholarship and scholarly networks. RePEc data is available independently of the web and further disseminated in major indexes, including EBSCO, EconLit, Google Scholar, OpenAIRE, OpenAlex, among others [10]. While not modeled on the resource description framework (RDF), linked data concepts are integral to RePEc, with unique identifiers and object-oriented modeling. RePEc consists of 13 distinct services that facilitate metadata creation, discovery, and usage tracking [11]. Together, these services create an invaluable resource where researchers can find economic scholarship and discover patterns and networks in their research domain. Of these services, several are key to uncovering relationships and trends in economic research. These are: 1) the primary bibliographic database, 2) the RePEc Author Service, 3) CitEc, 4) Rankings, 5) EDIRC, and 6) RePEc Genealogy.

1.2.1. *RePEc Bibliographic Database*

RePEc's core is the bibliographic metadata, which is provided by more than 2,000 publishers and pre-publication distributors, ranging from major publishers, such as Elsevier, to college and university departmental preprint repositories, such as the Growth Lab at Harvard's Kennedy School or Morocco's Policy Center for the New South. The content providers are responsible for serving their metadata on their servers according to RePEc's metadata standards and delivery protocol, described in section 4.

1.2.2. *RAS (RePEc Author Service)*

RAS is an author-maintained authority service that functions like ORCID in that RAS provides a persistent identifier (PID) and describes a person, their affiliations, and their works. RAS is unique in that it automatically facilitates associating works with people, even when the RAS PID is not provided in publisher metadata, and authors of works provide quality assurance by approving the recommended associated works.

1.2.3. *CitEc*

CitEc is the reference extraction and citation matching RePEc project. It leverages the fact that much of the published literature in economics is available open access as preprints. For gated materials, some publishers openly provide reference metadata within RePEc data or separately.

1.2.4. *Rankings*

A major added value of RePEc for research organizations is ranking information, which is based on RePEc usage through "downloads" and citation metadata. "Downloads" are measured when a user follows a link from a RePEc site. Various impact factors for

journals are also computed. RePEc Rankings have been crucial in encouraging publisher and author participation in contributing open metadata.

1.2.5. EDIRC (*Economics Departments, Institutes and Research Centers in the World*)

EDIRC is a hierarchical index of economic organizations. Like RAS, EDIRC establishes a PID for each organization. In its nature, EDIRC is like the Research Organization Registry (ROR) [12], in that each organization has a PID, and organizations may have parent and child relationships, as well as other information, such as location, website, and field-specific organization type. Because authors identify their affiliations in the RAS, people associated with an organization are then listed in EDIRC. Similarly, publishers may include EDIRC identifiers in their metadata records.

1.2.6. *RePEc Genealogy*

RePEc Genealogy is a crowdsourced effort that enhances RAS by creating an academic family tree for economics. Collected data include the year of graduation for the author's terminal degree, department from which the degree was earned (using EDIRC PID, if available), advisor(s) (using RAS PID(s), if available), and student(s) (using RAS PID(s), if available).

2. Research Projects Using RePEc Data

About 100 studies have leveraged RePEc metadata to various degrees. While not comprehensive, the authors' preliminary analysis identified seven topical categories of these works, including the state of economics research, topics studied in economics, strategic behavior of researchers, dissemination of research, alternative rankings using RePEc data, use of RePEc data to study economists, and use of RePEc statistics as selection criteria (see Appendix B, Table 2 for preliminary counts of studies). The studies leverage various RePEc metadata, including author information, work information, citations and references, and usage statistics and rankings (see Appendix B, Table 3 for more property use details). Two exemplar studies illustrate the unique value of RePEc's metadata model: Aistleitner's 2026 article, "Development and Interdisciplinarity: Additional Insights on the 'Economics Silo'" [13] and Conde-Ruiz et al.'s 2025 article, "Citation Gender Gaps in Top Economics Journals" [14].

Aistleitner evaluates the interdisciplinarity of development studies, leveraging historic publishing patterns of authors who research in the subfield [13]. In addition to a list of development studies articles published between 1980-2023 and citation data from Web of Science, Aistleitner supplements his study with RePEc author metadata, including author full name, the author's complete list of publications represented in RePEc including publication year, and the classification areas in which the author writes as represented by the New Economic Papers (NEP) bibliography. NEP, another RePEc service, disseminates papers and subject-expert editors classify them by subfield. Aistleitner's historical publication analysis would not be possible without RePEc author data.

While female representation in economics has increased over the past 30 years, economics as an area of study and profession remains heavily male dominated. RePEc IDEAS, which covers authors internationally, identifies that fewer than 27% of authors registered in the RAS in 2026 are female [15]. By comparison, the American Economic Association's 2026 gender diversity report, limited to authors based in the United States, shows 40% female representation in academic economics [16]. Given the disparity and impact on research outcomes, gender remains an area of research interest. Conde-Ruiz et al. leverage RePEc data to analyze driving factors for gender gaps in the top 5 economic journals [14]. Their primary dataset comprises articles published between 1999-2023 in the top economics journals (American Economic Review, Quarterly Journal of Economics, Journal of Political Economy, Econometrica, and Review of Economic Studies). They supplement this with RePEc citation data and Journal of Economic Literature (JEL) classification codes recorded in RePEc. Of note, when JEL codes were unavailable for the published article, Conde-Ruiz et al. leveraged RePEc's preprint correlation linking to identify JEL codes in an earlier version of a work, highlighting the value of unique RePEc structures. While gender is central to Conde-Ruiz et al.'s inquiry, neither the publication data nor RAS store gender information. As done in an earlier study, the Conde-Ruiz et al. leverage three different name datasets to assign author gender by the author first name [17]. The authors of this article acknowledge the inherent issues of inferring gender from a name as widely cited across LIS literature, as well as the binary analyses embrittled in economics studies.

3. Challenges of (Meta)data Citation

Citation information is critical to bibliometrics, a well-established area of research and scholarship. With current technology, research citations are easily identified and extracted as machine-readable data. Identifying data citations is trickier, as authors do not consistently cite data and metadata like other research assets. Citations may be hidden in footnotes, appendices, or simple mentions in the text. For RePEc, identifying evidence of data use is easier thanks to its unique name that can be used in a bibliographic keyword search.

There have been myriad interpretations and implementations of FAIR data [18]. While much of the discourse has been dedicated to research datasets, FAIR metadata is less understood and discussed, especially in the context of library resources [19]. Successful efforts of reusing traditional library and research metadata as data, such as Thomas Padilla's IMLS and Mellon Foundation funded "Collections as Data" [20], [21], clearly illustrate the need for libraries and information stewards to be able to find data reuse through metadata citations.

4. RePEc Metadata

Originally developed in 1997, prior to the publication of widely accepted standards, such as the Journal Article Tag Suite (JATS) (2003) [22], Crossref schema (1999) [23], or even Dublin Core (1998) [24], the Research Documents Information Format (ReDIF)

describes research scholarship. The current version of the schema (2015) [25] supports three classes (resource, collection, and tangible) and 9 objects (paper, article, chapter, book, software, archive, series, institution, and person) within those classes. Distinct object types can have shared properties, which are structured as key-value pairs, without any hierarchical relationships within the object type. Metadata is stored and primarily available for reuse as plain text files containing this simple key-value structure. The metadata documentation is not machine readable; however, validation software exists which conforms to the metadata schema [26].

PIDs are used throughout the object templates to facilitate relationships between objects. For example, a resource:article object might reference a tangible:person object. These relationships are core to the added value of RePEc but also introduce challenges to understanding the data. A researcher interested in several intersections of data would require the entire dataset.

While RePEc's core data structure is ReDIF, RePEc includes the bibliographic metadata (or "resource" metadata) in schema.org and <meta> HTML tags on the various discovery platforms (IDEAS and EconPapers). An API is also available and can offer snippets of data intersections. See [Appendix A, Table 1](#) for a complete overview of metadata available for reuse.

Central to this research is the variety of metadata object types leveraged in the studies using RePEc metadata. Evaluating studies using RePEc data, the authors focus on the need for interoperability between metadata object types rather than interoperability between metadata providers. The RePEc data model was built to fully describe persons, their works, their citations, and the organizations to which they belong, and each of the objects link to other object types with PIDs. While major contributors to research information, such as ROR, ORCID, and Crossref, have clearly documented and easily accessible metadata, the systems were designed independently of each other and require notable effort to work with each other.

While RePEc metadata as individual records follow FAIR principles, RePEc metadata as a dataset needs remediation to align, namely, in the accessibility principle of FAIR data. Because RePEc metadata is decentralized, the various metadata types are available on different platforms and in different data formats, requiring the user to normalize the data in the preferred serialization ([Appendix A, Table 1](#)). Nevertheless, more than 100 research works are known to use RePEc metadata, and the rationale for use likely aligns with the community-stewarded nature of RePEc. Often, the researcher is familiar with what the data describes, to the point of being part of it. While RePEc metadata is often available in different data formats or via various platforms, the metadata uniquely links across object types, enabling more expansive data reuse.

5. Future Work

This paper provides the framework and preliminary observations of the authors' larger research that comprehensively investigates more than 100 unique works that are known

to have used RePEc metadata to date. Future research will identify what data sources within the RePEc ecosystem and in addition to RePEc are used (e.g., EconLit), as well as how the metadata is cited. Three goals guide this case study. Firstly, the authors hope to provide recommendations for metadata citation protocols so that uses of collections as data are readily traceable in the scholarly record. The research will evaluate challenges of metadata structure and availability identified in the studies. Finally, the research will identify where metadata interoperability is critical to the research approaches undertaken in the studies. In finding and synthesizing existing literature on this topic, the authors observed that so far none exhaustively described FAIR principles in the context of research information metadata. The authors hope this preliminary research will spark greater interest in the notion of collections as data and what is necessary for users to find, access, and reuse bibliographic and other library resource metadata.

Appendix A. RePEc Metadata Availability

RePEc metadata is available in various formats via several pathways, detailed in [Table 1](#).

Table A1. *How to Get RePEc Metadata [27]*

Metadata Object Type	Availability	Format and Serialization
Resource (paper, article, chapter, book, software)	Publisher sites (ftp, http/https)	ReDIF plain text
	RePEc (ftp)	ReDIF plain text
	OAI-PMH	Academic Metadata Format (AMF) XML, Dublin Core XML
	Rsync	ReDIF plain text, AMF XML, Dublin Core XML
Author/Organization	RePEc API	ReDIF JSON
	RePEc (ftp)	ReDIF plain text
	RePEc API	ReDIF JSON
Citation	CitEc API	Plain XML (includes 4 attributes: date, CitEc URI, citedBy counts, and cites counts), AMF XML, citedBy XML
	RePEc API	ReDIF JSON
Statistics	LogEc (https)	CSV
	RePEc API	JSON (limited dataset)
Person Genealogy	RePEc API	JSON

Appendix B. Preliminary Data Analysis

Table B1. *Distribution of Characteristics (In Progress)*

Categories	Description	Count of unique works
The state of economics research	More generally, studies on the state of research in economics.	16
What topics are studied in economics?	What topics are studied and how these change over time.	7
Strategic behavior of researchers	The strategic behavior of researchers, such as how they cite or choose co-authors.	16
Dissemination of research	How research is disseminated in economics.	16
Alternative rankings using RePEc data	Alternative ways to measure the research output and/or rank authors, institutions, and journals.	31
Use of RePEc data to study economists	Descriptions and studies of the economics profession. Given that RePEc has information about authors, their pedigree, their publications, various aspects of authors' careers have been studied.	10
Use of RePEc statistics as selection criteria	Leveraging RePEc statistics as a selection criterion for the scope of a study. Examples include all authors within a region or field, journals above a threshold, and top institutions or authors.	5
Miscellaneous uses	Papers that do not fit within the categories above.	7
Total		108

Table B2. *Metadata Types Leveraged in Studies (In Progress)*

Class	Properties	Example Uses
Author	Author education	Used to determine pedigree (rank of graduate program) Used to identify academic age, as measured by the year of graduation
	Author name	Used in multiple studies to identify gender as expressed in name interpretation, as gender is not a recordable RePEc attribute (note that the data is available via the RePEc API and is based on name pattern matching using NamSor https://namsor.app/) Namely used to identify associated co-authors
	Authoried works	Used to identify academic age, as measured by the date of the first publication
	Subject classification (JEL, NEP)	Used to determine areas of research
	Gender	Used for gender studies (note that the data is available via the RePEc API and is based on name pattern matching using NamSor https://namsor.app/)
	Affiliation	Used to group people by institution
	Abstract, title, keywords	Used for topical analysis studies
	Journal	Used for impact factor analysis
Works (preprint, article, book, book chapter, software)	Subject classification (JEL, NEP)	Used to determine areas of research
Citations and references		Frequently used in conjunction with works and author information
Usage statistics and rankings	Year	Used for longitudinal studies
	Organization(s)	Used for alternative rankings

References

- [1] “Barcelona Declaration on Open Research Information.” [Online]. Available: <https://barcelona-declaration.org/>
- [2] M. D. Wilkinson, M. Dumontier, and I. Aalbersberg, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, no. 1, p. 160018, 2016, doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- [3] P. Groth, H. Cousijn, T. Clark, and C. Goble, “FAIR data reuse — the path through data citation,” *Data Intelligence*, vol. 2, no. 1–2, pp. 78–86, 2020, doi: [10.1162/dint_a_00030](https://doi.org/10.1162/dint_a_00030).
- [4] G. C. Donovan and M. L. Langseth, “Are researchers citing their data? A case study from the U.S. Geological Survey,” *Data Science Journal*, vol. 23, no. 1, 2024, doi: [10.5334/dsj-2024-024](https://doi.org/10.5334/dsj-2024-024).
- [5] Data Citation Synthesis Group, “Joint Declaration of Data Citation Principles,” FORCE11, 2014. doi: [10.25490/a97f-egykh](https://doi.org/10.25490/a97f-egykh).
- [6] M. Altman and G. King, “A proposed standard for the scholarly citation of quantitative data,” *D-Lib Magazine*, vol. 13, no. 3/4, 2007, [Online]. Available: <https://www.dlib.org/dlib/march07/altman/03altman.html>
- [7] “Make Data Count.” [Online]. Available: <https://makedatacount.org/>
- [8] H. Cousijn, P. Feeney, D. Lowenberg, E. Presani, and N. Simons, “Bringing citations and usage metrics together to make data count,” *Data Science Journal*, vol. 18, no. 1, 2019, doi: [10.5334/dsj-2019-009](https://doi.org/10.5334/dsj-2019-009).
- [9] M. Altman and M. Crosas, “The evolution of data citation: From principles to implementation,” *IASSIST Quarterly*, vol. 37, no. 1–4, pp. 62–70, 2014, doi: [10.29173/iq504](https://doi.org/10.29173/iq504).
- [10] RePEc, “Additional websites using RePEc.” [Online]. Available: <https://repec.org/#further>
- [11] RePEc, “RePEc Services.” [Online]. Available: <https://repec.org/#uses>
- [12] “Research Organization Registry (ROR).” [Online]. Available: <https://ror.org/>
- [13] M. Aistleitner, “Development and interdisciplinarity: Additional insights on the ‘Economics Silo’,” *World Development*, vol. 205, p. 107435, 2026, doi: [10.1016/j.worlddev.2026.107435](https://doi.org/10.1016/j.worlddev.2026.107435).
- [14] J. I. Conde-Ruiz, M. Díaz-Salazar, J.-J. Ganuza, and M. García, “Citation gender gaps in top economics journals,” *SERIEs*, vol. 16, no. 3, pp. 659–697, 2025, doi: [10.1007/s13209-025-00322-z](https://doi.org/10.1007/s13209-025-00322-z).
- [15] IDEAS, “Female Representation in Economics, as of March 2026.” [Online]. Available: <https://ideas.repec.org/top/female.html>
- [16] L. Tesar, “Report of the Committee on the Status of Women in the Economics Profession,” *AEA Papers and Proceedings*, vol. 116, pp. 866–889, 2026, doi: [10.1257/pandp.116.866](https://doi.org/10.1257/pandp.116.866).
- [17] J. I. Conde-Ruiz, J.-J. Ganuza, M. García, and L. A. Puch, “Gender distribution across topics in the top five economics journals: A machine learning approach,” *SERIEs*, vol. 13, no. 1, pp. 269–308, 2022, doi: [10.1007/s13209-021-00256-2](https://doi.org/10.1007/s13209-021-00256-2).

- [18] C. Erdmann, N. Simons, and R. Otsuji, “Top 10 FAIR Data & Software Things.” [Online]. Available: <https://doi.org/10.5281/zenodo.2555498>
- [19] B. Dobreski, H. Moulaison-Sandy, and B. W. Bishop, *How FAIR is MARC? FAIR data principles and bibliographic data*. in Proceedings of the International Conference on Dublin Core and Metadata Applications. 2023. doi: 10.23106/dcmi.953165517.
- [20] “Always Already Computational — Collections as Data.” [Online]. Available: <https://collectionsasdata.github.io/>
- [21] “Collections as Data — Part to Whole.” [Online]. Available: <https://collectionsasdata.github.io/part2whole/>
- [22] ANSI/NISO Z39.96-2015, “JATS: Journal Article Tag Suite, version 1.1,” 2015. [Online]. Available: <https://www.niso.org/publications/z3996-2015-jats>
- [23] H. Atkins *et al.*, “Reference linking with DOIs: A case study,” *D-Lib Magazine*, vol. 6, no. 2, 2000, [Online]. Available: <https://www.dlib.org/dlib/february00/02risher.html>
- [24] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, “Dublin Core Metadata for Resource Discovery,” 1998. [Online]. Available: <https://www.rfc-editor.org/rfc/rfc2413>
- [25] T. Krichel, “ReDIF, version 1,” 2015. [Online]. Available: https://openlib.org/acmes/root/docu/redif_1.html
- [26] I. Kurmanov, “ReDIF-perl Package including rech and rr.pm.” [Online]. Available: <https://hubec.repec.org/ReDIF-perl/>
- [27] “RePEc: Getting the Metadata.” [Online]. Available: <https://ideas.repec.org/getdata.html>