

OpenCitations

DOI (and Beyond) for Publications and Other Citable Research Outcomes

IVOA Northern Spring 2023 Interoperability Meeting
CNR, Bologna, May 10 2023

Silvio Peroni

Director, OpenCitations – @opencitations – <https://opencitations.net>
Research Centre for Open Scholarly Metadata, Department of Classical
Philology and Italian Studies, University of Bologna, Bologna, Italy
silvio.peroni@unibo.it – @essepuntato – <https://essepuntato.it>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA



What is OpenCitations

[OpenCitations](#) an independent, community-led, and not-for-profit Open Science infrastructure organisation

Mission: “to harvest and openly publish accurate and comprehensive metadata describing the world's academic publications and the scholarly citations that link them, and to preserve ongoing access to this information by secure archiving. We provide this information, both in human-readable form and interoperable machine-readable Linked Open Data formats, under open licenses at zero cost and without restriction for third-party analysis and re-use.”

We provide:

- data models: the [OpenCitations Data Model](#) (based on [SPAR Ontologies](#))
- bibliographic and citation databases: [OpenCitations Index](#) (containing citations) and [OpenCitations Meta](#) (containing bibliographic metadata)
- software: [GitHub repository](#) released with open source licenses
- online services: [dumps](#), [REST APIs](#) and [SPARQL endpoints](#)

Citations as first-class object

The citations available in COCI are treated as first-class data entities, with accompanying properties including the citations timespan, modeled according to the OpenCitations Data Model

We launched a system for globally unique and persistent identifiers (PIDs) for bibliographic citations – [Open Citation Identifiers \(OCIs\)](#)

Conventional citation

Identifiers.org
Central registry

Resolution Registry Browse the registry Make a request Also in this section

Namespace: *OCI*

General Information

Name	OCI
Description	Each OCI (Open Citation Identifier) has a simple structure: oci:number-number, where "oci:" is the identifier prefix, and is used to identify a citation as a first-class data entity - see https://opencitations.wordpress.com/2018/02/19/citations-as-first-class-data-entities-introduction/ for additional information. OCIs for citations stored within the OpenCitations Corpus are constructed by combining the OpenCitations Corpus local identifiers for the citing and cited bibliographic resources, separating them with a dash. For example, oci:2544384-7295288 is a valid OCI for the citation between two papers stored within the OpenCitations Corpus. OCIs can also be created for bibliographic resources described in an external bibliographic database, if they are similarly identified there by identifiers having a unique numerical part. For example, the OCI for the citation that exists between Wikidata resources Q27931310 and Q22252312 is oci:01027931310-01022252312. OCIs can also be created for bibliographic resources described in external bibliographic database such as Crossref or DataCite where they are identified by alphanumeric Digital Object Identifiers (DOIs), rather than purely numerical strings.

Journal of Documentation

Setting our bibliographic references free: towards open citation data

Author(s): Silvio Peroni (Department of Computer Science and Engineering, University of Bologna, Bologna, Italy), Alexander Dutton (IT Services, University of Oxford, Oxford, UK), Tanya Gray (Bodleian Libraries, University of Oxford, Oxford, UK), David Shotton (Oxford e-Research Centre, University of Oxford, Oxford, UK)

Abstract: Purpose
- Citation data needs to be recognised as a part of the Commons - those works that are freely and legally available for sharing - and placed in an open repository. The paper aims to discuss this issue.

Citing article

cites

PLOS COMPUTATIONAL BIOLOGY

Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article

David Shotton, Kate Portwin, Graham Stone, Alastair Miles

Published: April 17, 2009 - <https://doi.org/10.1371/journal.pcbi.1000381>

Article Authors Comments Related Content

Cited article

Citation as first-class data entity

has citing article

Journal of Documentation

Setting our bibliographic references free: towards open citation data

Author(s): Silvio Peroni (Department of Computer Science and Engineering, University of Bologna, Bologna, Italy), Alexander Dutton (IT Services, University of Oxford, Oxford, UK), Tanya Gray (Bodleian Libraries, University of Oxford, Oxford, UK), David Shotton (Oxford e-Research Centre, University of Oxford, Oxford, UK)

Abstract: Purpose
- Citation data needs to be recognised as a part of the Commons - those works that are freely and legally available for sharing - and placed in an open repository. The paper aims to discuss this issue.

The Citation



PLOS COMPUTATIONAL BIOLOGY

Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article

David Shotton, Kate Portwin, Graham Stone, Alastair Miles

Published: April 17, 2009 - <https://doi.org/10.1371/journal.pcbi.1000381>

Article Authors Metrics Comments Related Content

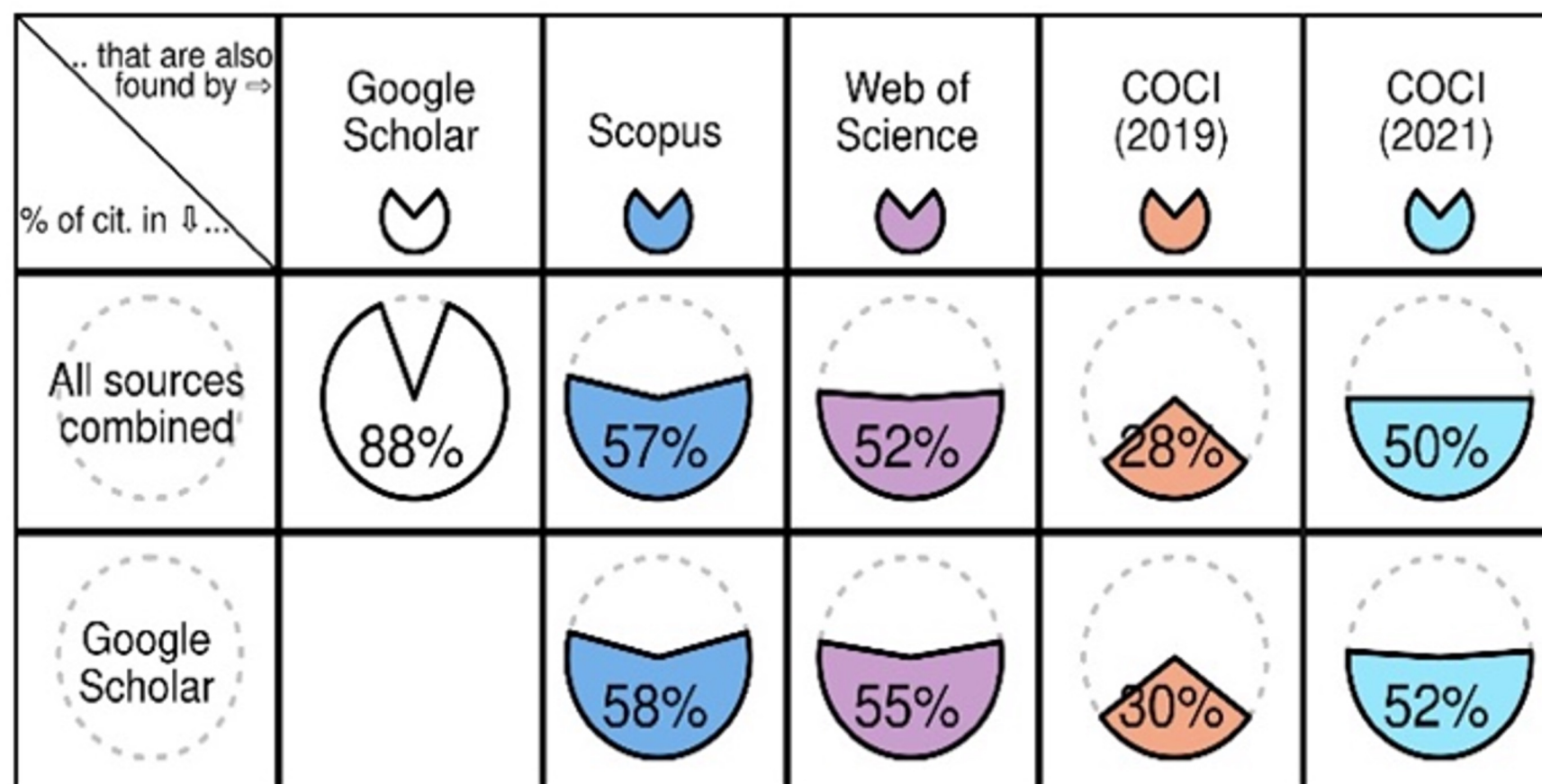
has cited article

Citations in COCI and other indexes

Currently, the principal database containing citations maintained by OpenCitations is [COCI, the OpenCitations Index of Crossref Open DOI-to-DOI Citations](#), which stores citation metadata, together with DOI identifiers for the citing and cited publications

In August 2021, we crossed the significant threshold of one billion citation links in COCI, and we currently host more than **1.46 billion citations**, and is accompanied by DOCI (**169 million citations**, source: DataCite) and POI (**717 million citations**, source NIH-OCC)

A recent independent comparison by Alberto Martín-Martín has shown that the coverage of COCI is approaching parity with those of the main proprietary citation services



Need of going beyond DOIs

Sometimes, bibliographic resources will have been assigned **multiple identifiers**, such as a DOI and a PMID – in such cases, the same citation may be multiply represented in different ways (DOI-to-DOI and PMID-to-PMID) depending on the data source

This duplication poses problems when counting the number of ingoing and outgoing citations of each document, a crucial statistic for libraries, journals, and Scientometrics studies

In addition, the **assignment of globally persistent identifiers to documents is not universal practice** across all scholarly fields: Natural and Social Sciences communities adopt DOIs to a much greater extent than the Arts and Humanities community

In Scopus and the Web of Science Core Collection, almost 90% of the publications in the Sciences and Social Sciences are associated with a DOI, while in the Art and Humanities that figure is only 50%

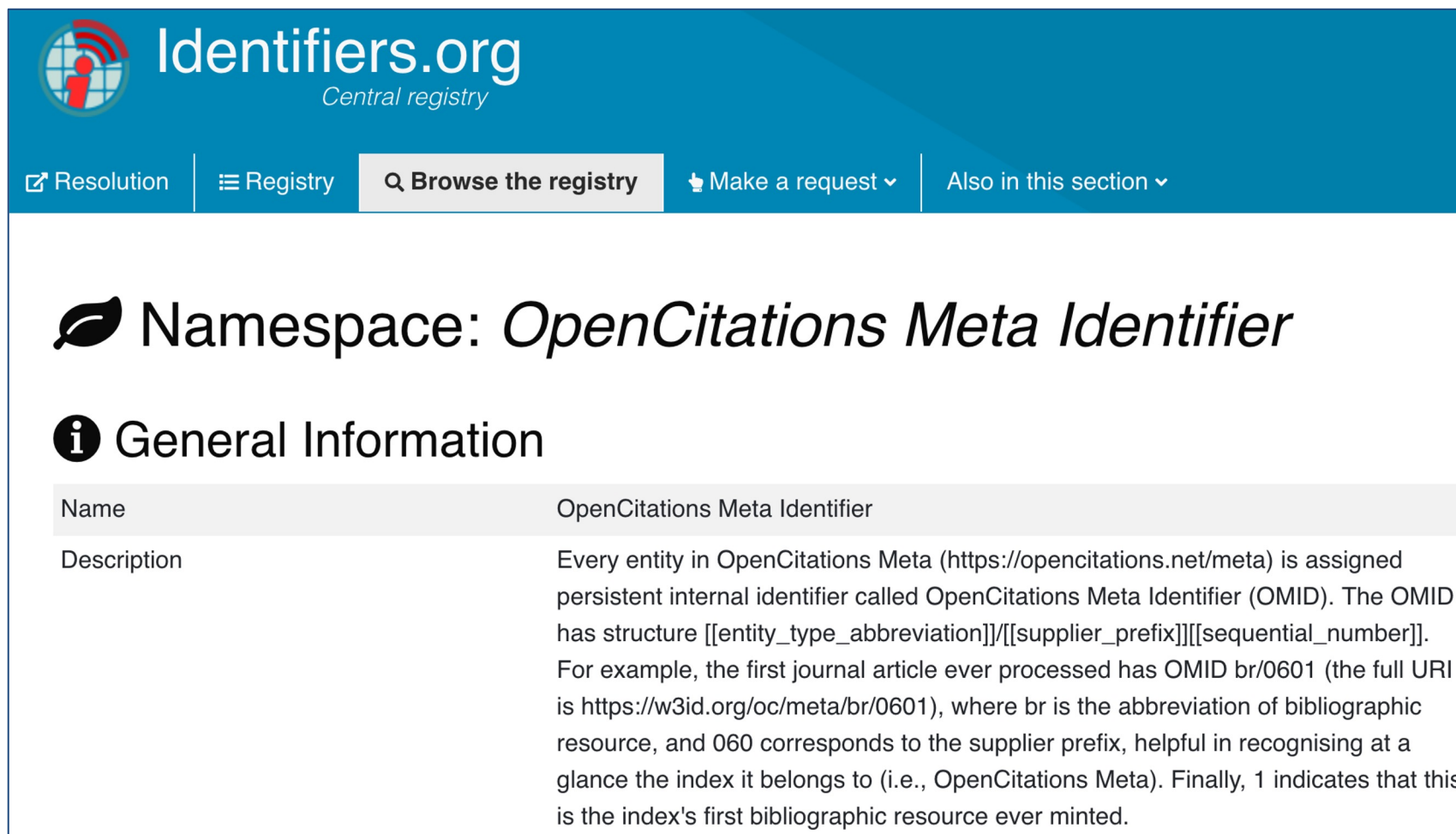
In addition, concerning the Humanities, citations of **ancient primary sources lacking DOIs** (e.g. Aristotle) are required in many fields (e.g. in History)

If a document has no identifier, its metadata **does not respect the FAIR principles** that scholarly digital research objects must be findable, accessible, interoperable and reusable – globally unique and persistent identifier is critical to make metadata findable and accessible

Moreover, a bibliographic resource without an identifier **prevents citation links** involving it from being described

OpenCitations Meta

OpenCitations Meta is our new in-house database containing bibliographic metadata describing all the publications involved in the citations indexed by any OpenCitations Index

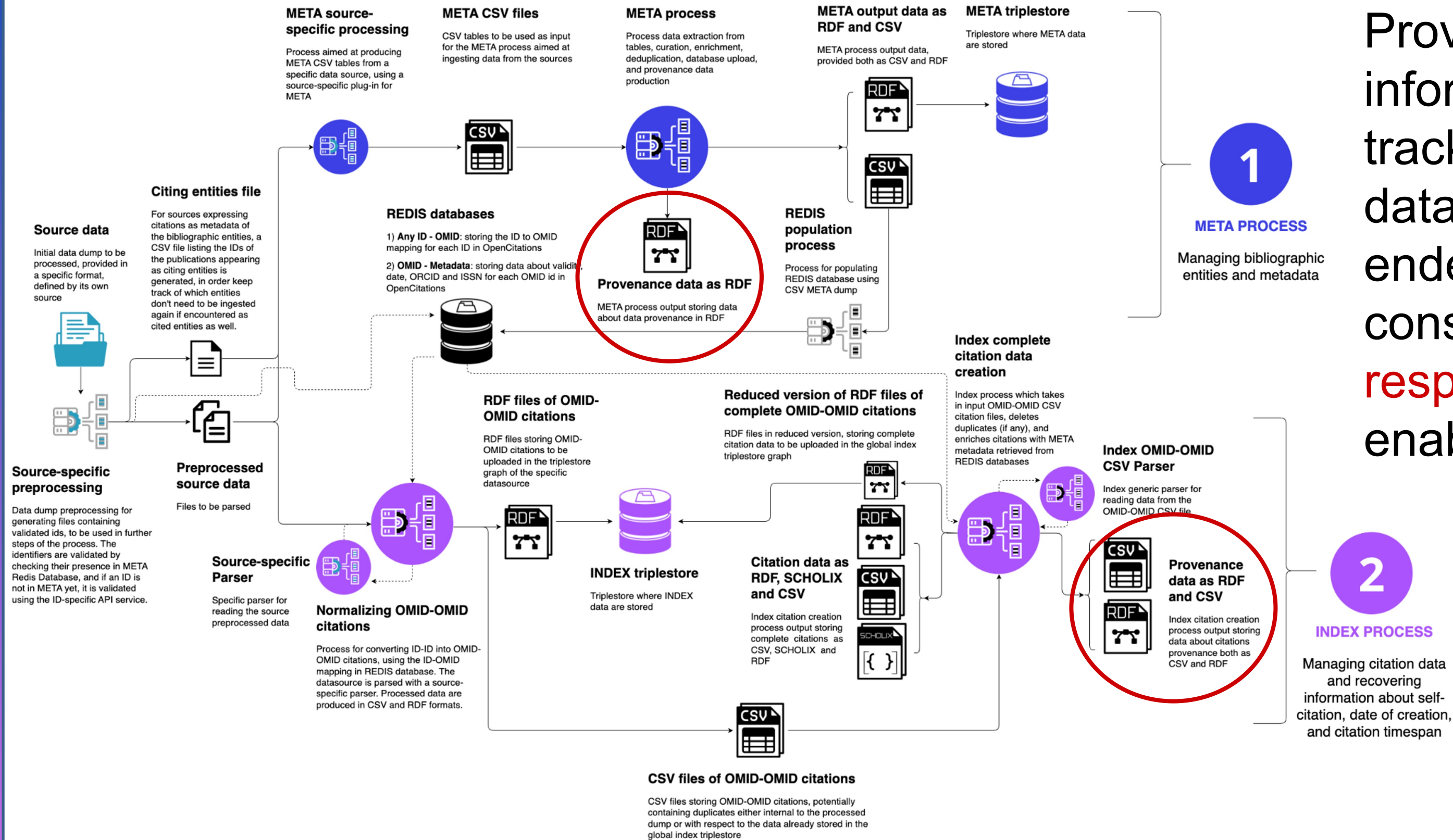


The screenshot shows the Identifiers.org website interface. At the top, there is a blue header with the Identifiers.org logo and the text "Central registry". Below the header, there are navigation links: "Resolution", "Registry", "Browse the registry", "Make a request", and "Also in this section". The main content area displays the "Namespace: OpenCitations Meta Identifier" and a section for "General Information".

Name	OpenCitations Meta Identifier
Description	Every entity in OpenCitations Meta (https://opencitations.net/meta) is assigned persistent internal identifier called OpenCitations Meta Identifier (OMID). The OMID has structure <code>[[entity_type_abbreviation]][[supplier_prefix]][[sequential_number]]</code> . For example, the first journal article ever processed has OMID <code>br/0601</code> (the full URI is https://w3id.org/oc/meta/br/0601), where <code>br</code> is the abbreviation of bibliographic resource, and <code>060</code> corresponds to the supplier prefix, helpful in recognising at a glance the index it belongs to (i.e., OpenCitations Meta). Finally, <code>1</code> indicates that this is the index's first bibliographic resource ever minted.


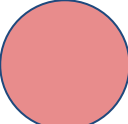
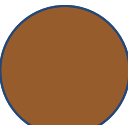
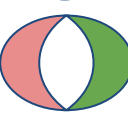
OpenCitations Meta enables the assignment of new globally persistent identifiers (PIDs), namely **OpenCitations Meta Identifiers (OMIDs)**, to bibliographic resources, including those that have not been separately assigned external persistent identifiers, e.g. DOIs, by their publishers


OpenCitations Ingestion workflow

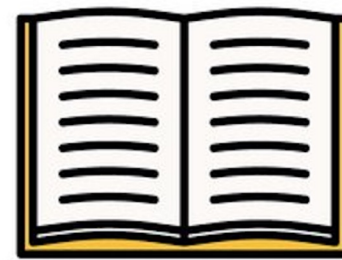


Provenance information and tracking change of data is a crucial endeavour to consider: it is about **responsibility** and it enables **trust**

How OpenCitations approach provenance

-  adding data
-  removing data
-  remaining data
-  delta:
data added & removed

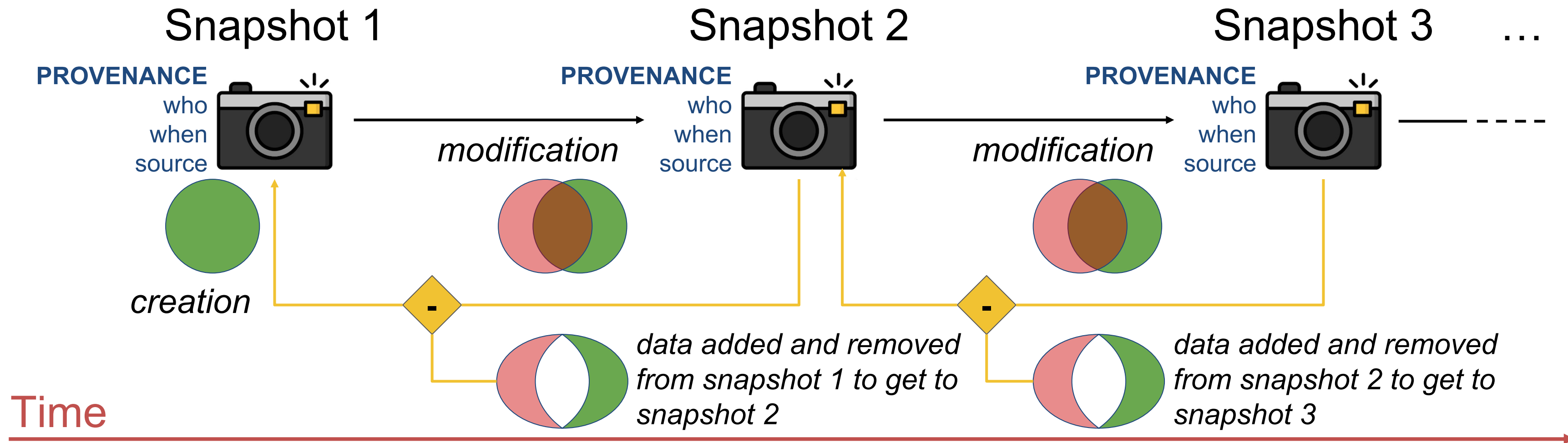
 snapshot: all the data having such an entity as subject at a given time



An entity

(e.g. a journal article)

To keep track of the history of an entity we need to store all the data of its last snapshot plus all the deltas built by modifying the previous snapshots





**Thank you
for your attention**