In-Text Reference Pointer Identifier: Definition

Authors

David Shotton   Oxford e-Research Centre
University of Oxford, Oxford, United Kingdom
david.shotton@opencitations.net
http://orcid.org/0000-0001-5506-523X

Marilena Daquino   Digital Humanities Advanced Research Centre
Department of Classical Philology and Italian Studies
University of Bologna, Bologna, Italy
marilena.daquino2@unibo.it, marilena.daquino@opencitations.net
http://orcid.org/0000-0002-1113-7550

Silvio Peroni   Research Centre for Open Scholarly Metadata
Department of Classical Philology and Italian Studies
University of Bologna, Bologna, Italy
silvio.peroni@unibo.it, silvio.peroni@opencitations.net
http://orcid.org/0000-0003-0530-4305

Abstract

This document defines the In-Text Reference Pointer Identifier (InTRePID).

Cite this as

David Shotton, Marilena Daquino and Silvio Peroni (2020). In-Text Reference Pointer Identifier: Definition. Figshare. DOI: https://doi.org/10.6084/m9.figshare.11674032

Preamble

The key words "MUST" and "MUST NOT" in this definition are to be interpreted as described in [Request For Comments 2119](), meaning that the item or condition referred to is (or is not) an absolute requirement of the specification.

Preliminaries

A **bibliographic citation** is a conceptual directional link from a citing bibliographic entity to a cited bibliographic entity, for the purpose of acknowledging or ascribing credit for the contribution made by the author(s) of the cited entity. The citing and cited entities may be scholarly publications, online documents, blog posts, datasets, or any other authored entities capable of giving or receiving citations

An **open bibliographic citation** is one for which metadata are openly available, as defined in *Silvio Peroni, David Shotton (2018). Open Citation: Definition. Figshare. DOI: https://doi.org/10.6084/m9.figshare.6683855*.

An **Open Citation Identifier** (**OCI**) is a globally unique persistent identifier (PID) for the identification of an individual open bibliographic citation, as defined in *Silvio Peroni, David Shotton (2019). Open Citation Identifier: Definition. Figshare. DOI: https://doi.org/10.6084/m9.figshare.7127816*.

A **bibliographic reference** (aka "a reference") is a text string, typically found within the reference list of a citing bibliographic entity, that **references** a cited bibliographic entity, and that **relates to** the citation that is instantiated between the citing entity which includes the bibliographic reference and the cited entity which that bibliographic reference references. An OCI uniquely identifies such a citation, for which open metadata are available in a particular database.

An **in-text reference pointer** (often loosely called an 'in-text citation') is a textual element within the text of a citing bibliographic entity that **denotes** a single bibliographic reference within the reference list of that citing bibliographic entity, and that **relates to** the citation that is instantiated between the citing entity which includes the in-text reference pointer and the cited entity which is referenced by the bibliographic reference denoted by the in-text reference pointer.

For example, if the sixth bibliographic reference in the reference list of a citing article is as follows:

4. Ko AI, Reis MG, Ribeiro Dourado CM, Johnson WD Jr, Riley LW (1999). Urban epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group. Lancet 354: 820-825.

the in-text reference pointer denoting this bibliographic reference may be manifested in the text in a number of ways, including:

- numerically, for example by using the numeral 6 in superscript or between square brackets, thus: "[6]"; or
- by use of the authors' family names (or the first author's family name followed by *et al*.) and the publication year in parentheses, thus: "(Ko *et al*. 1999)".

There may be one or more in-text reference pointers in the text of the citing bibliographic entity denoting a particular bibliographic reference in the citing entity's reference list, and (if more than one) these may be located within the same or different sections of the body text, for example one in the Methods section and another in the Discussion.

A particular in-text reference pointer may be grouped together with one or more other in-text reference pointers denoting different bibliographic references, thus forming an **in-text reference pointer list** at a particular location within the body text, e.g. "[6, 11-18]", or "(Ko *et al*. 1999, McBride *et al*. 2005).

The scope of this document is the definition of one particular identifier, the **In-Text Reference Pointer Identifier** (**InTRePID**), which is being used to provide identifiers for individual in-text reference pointers located within the text of a citing bibliographic entity. Since InTRePIDs are based on OCIs, an understanding of the structure of an OCI is a necessary prerequisite for understanding the definition of the InTRePID presented in this document.

Rationale

While an OCI uniquely identifies a citation between a citing bibliographic entity and a cited bibliographic entity for which metadata are available in a particular database containing open citation data, it does not provide access to any information about where within the citing entity such a citation is generated, or the reasons for including different in-text reference pointers denoting the same bibliographic reference at different points in the text.

However, a bibliographic reference and an in-text reference pointer that denotes this bibliographic reference within a citing bibliographic entity actually relate to **two different citation instances** between the same two documents, since the function they convey is different.

In the case of the bibliographic reference, no *function* (i.e. the reason why an author cites another work) is usually specified – the only purpose of the citation that is instantiated in this case is that of indicating the presence of a direct link between the citing and cited entities. Such basic 'reference citation' instances have been used in all citation indexes created to date, including COCI (the OpenCitations Index of Crossref open DOI-to-DOI citations).

However, in the in-text reference pointer case, the situation is more nuanced, since here both the function of the citation (even if not explicitly defined) and also the location of the in-text reference pointer are crucially important, and both are specific properties, not of the 'reference citation', but of the distinct 'in-text reference pointer citation' which is instantiated as a result of the author's inclusion of an in-text reference pointer at that place in the text. Thus an in-text reference pointer within a description of use of a chromatographic procedure in the Methods section of a publication may have the function of citing the original publication of that procedure (with the property *cito:usesMethodIn*, if using CiTO, the Citation Typing Ontology), while a separate in-text reference pointer in the Discussion section of the publication denoting the same reference may be in the context of a general discussion of methodology (with the property *cito:discusses*, if using CiTO, the Citation Typing Ontology).

Thus, an article can have multiple citations between the same two documents, determined by the bibliographic reference that is denoted by in-text reference pointers throughout the text. The fact that the citing and cited documents are the same does not mean that the citations are the same. Indeed, knowledge of the citing and cited entities alone is not enough to claim that two citations are actually the same citation.

We have a similar situation when we consider links in web pages. If a web page contains two links to the same target page at two distinct locations within the text, such links are different, despite the fact that the source and target web pages are exactly the same for both links. Since citations, as we have defined them, are *conceptual* relationships, distinct citations between the same citing and cited sources may (and probably do) involve different functions.

Now, considering a dataset filled with these two types of citations (i.e. those that are instantiated because of the inclusion of bibliographic references within the citing bibliographic entity, and those that are additionally instantiated because of the inclusion of in-text reference pointers denoting those bibliographic references), one can choose to consider only those citations related to the bibliographic references, only those citations related to the in-text reference pointers, or both sets.

To permit the description of the locations within a citing entity from which a particular citation is generated, we need to identify each individual in-text reference pointer uniquely within the citing bibliographic entity, using a novel individual globally unique persistent identifier. It is for this purpose that the InTRePID has been created. The use of InTRePIDs to permit the unique identification of individual in-text reference pointers, and thereby to record metadata specific to the individual citations related to these in-text reference pointers and their locations, provides a new method to facilitate in-depth scholarship on in-text reference pointer locations and citation functions, and to permit fine-grained analysis of the relationships between publications.

Definition of an In-Text Reference Pointer Identifier

The **In-Text Reference Pointer Identifier** (**InTRePID**) is a globally unique persistent identifier (PID) for the identification of an individual **in-text reference pointer** relating to an open bibliographic citation. Since the InTRePID contains the numerical part of the OCI identifying the citation to which the InTRePID relates, an InTRePID can be assigned for any in-text reference pointer that relates to an open citation for which a valid OCI has been assigned.

The In-Text Reference Pointer Identifier (InTRePID) system, the assignment of individual InTRePIDs, and the resolution service for InTRePIDs, are the work of OpenCitations, a scholarly infrastructure organization dedicated to open scholarship and the publication of open bibliographic and citation data.

Each InTRePID has a simple structure: the lower-case letters "intrepid" followed by a colon, followed by two parts separated by an oblique stroke, thus:

```
intrepid:<oci-numerals>/<ordinal>-<total>
```

Precisely, an InTRePID is defined as follows (in Backus-Naur form):

```
<intrepid>      ::= "intrepid:" <oci-numerals> "/" <ordinal> "-" <total>
<oci-numerals>  ::= <identifier> "-" <identifier>
<identifier>    ::= <prefix> <any_number>
<prefix>        ::= "0" <pos_number> "0"
<ordinal>       ::= <pos_number> | <pos_number> <any_number>
<total>         ::= <pos_number> | <pos_number> <any_number>
<any_number>    ::= "0" | <pos_number> | <any_number> <any_number>
<pos_number>    ::= "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9" |
                    <pos_number> <pos_number>
```

Since the InTRePID specification derives from the OCI specification and requires use of a valid OCI to be itself valid, **both** `<identifier>`s within `<oci-numerals>` MUST use the same `<prefix>`. In addition, the `<ordinal>` MUST NOT be greater than the `<total>`.

For example, `intrepid:070433-070475/4-6`, `intrepid:070433-070556/1-1`, and `intrepid:070562-070584/2-2` are all valid InTRePIDs.

However, `intrepid:0301-05018/1-2` is not a valid InTRePID, since one is not permitted to combine identifiers having different prefixes within an OCI. Neither is `intrepid:070433-070475/7-6` a valid InTRePID, since the first number following the "/" (the `<ordinal>`) must not be greater than the second number (the `<total>`).

Meaning

From a database containing metadata describing in-text reference pointers, each InTRePID identifying an in-text reference pointer within that database MUST carry the following pieces of information:

1. the identities of the citing and cited bibliographic entities comprising the citation to which the in-text reference pointer relates to;
2. the ordinal position of the in-text reference pointer among that group of in-text reference pointers all denoting the same single bibliographic reference, counting from the beginning of the citing bibliographic entity within which they are located;
3. the total number of in-text reference pointers within the citing bibliographic entity denoting that bibliographic reference.

Within the structure of the InTRePID:

```
intrepid:<oci-numerals>/<ordinal>-<total>
```

the first requirement mentioned above is encoded in the `<oci-numerals>` part of the InTRePID, while the second and third requirements are encoded in the parts `<ordinal>` and `<total>` respectively.

The first part of the InTRePID (`<oci-numerals>`) is the numerical part of the OCI (i.e. omitting its prefix "oci:") that uniquely identifies the particular open citation to which the in-text reference pointer relates. It is composed of two sequences of numerals separated by a dash (thus *aaa-bbb*), which identify the citing and cited bibliographic entities, respectively, as specified in *Silvio Peroni, David Shotton (2019). Open Citation Identifier: Definition. Figshare. DOI: https://doi.org/10.6084/m9.figshare.7127816*.

The second part of the InTRePID is also composed of two numbers separated by a dash (namely `<ordinal>` and `<total>`).

The first of these numbers, `<ordinal>`], identifies the nth occurrence of an in-text reference pointer within the text of the citing bibliographic entity relating to the open citation identified by the OCI, numbering from the start of the document, e.g. "3" for the third in-text reference pointer that denotes the particular bibliographic reference within the bibliographic entity's reference list that instantiates the citation identified by the OCI.

The second of these numbers, `<total>`, defines the total number of in-text reference pointers within the citing bibliographic entity, of which the in-text reference pointer identified by the InTRePID is a group member, that all denote this particular bibliographic reference.

For example, `intrepid:070562-070606/1-1` identifies the first and only in-text reference pointer, "15", which is found in the introductory section of the article having identifier 070562 (https://w3id.org/oc/ccc/br/070562), namely

> Maria Gaudino *et al.* High Pathogenicity of Nipah Virus from *Pteropus lylei* Fruit Bats, Cambodia. Emerging Infectious Diseases. 2020; 26(1): 104–13. https://dx.doi.org/10.3201/eid2601.191284

This in-text reference pointer identified by the InTRePID denotes the fifteenth bibliographic reference in the reference list of Gasparyan *et al*. (2015), namely

> 15. Enchéry F, Horvat B. Understanding the interaction between henipaviruses and their natural host, fruit bats: Paving the way toward control of highly lethal infection in humans. International Reviews of Immunology. 2017; 36(2): 108–21. https://dx.doi.org/10.1080/08830185.2016.1255883

that identifies the cited journal article identifier 070606 (https://w3id.org/oc/ccc/br/070606).

The OCI for the open citation to which this InTRePID relates is `oci:070562-070606` [1].

---

[1] In the example, "070" is the citation data supplier prefix identifying the OpenCitations Citations in Context Corpus as the database providing the bibliographic metadata describing the citing and cited bibliographic entities involved in the citation identified by the OCI, as fully explained in *Silvio Peroni, David Shotton (2019). Open Citation Identifier: Definition. Figshare. DOI: https://doi.org/10.6084/m9.figshare.7127816*. The list of currently supported citation data suppliers and their OCI prefixes is maintained by OpenCitations, and is available at http://opencitations.net/oci. This prefix list is updated every time the OCI Resolution Service is expanded to handle the open citations from a new bibliographic database.

Software

OpenCitations provides an In-Text Reference Pointer Identifier Resolution Service that takes an InTRePID and returns information about the identified in-text reference pointer and related metadata such as its textual location within the citing bibliographic entity. The service is available at https://w3id.org/oc/intrepid, and it is currently in beta testing.

In addition to that, a script has been developed for validating a given InTRePID and for retrieving the data related to the in-text reference pointer identified by such InTRePID. The script, called `intrepid.py`, is available in the OpenCitations GitHub repository.