

ARTICLE
EXCERPTED
FROM:

ISQ

INFORMATION STANDARDS QUARTERLY

SUMMER 2010 | VOL 22 | ISSUE 3 | ISSN 1041-0031

SPECIAL ISSUE: ENHANCED JOURNAL ARTICLES

ARTICLE-LEVEL
ENHANCEMENTS IN
THE HUMANITIES
AND SOCIAL SCIENCES

HOSTING
SUPPLEMENTARY
MATERIAL

ARCHIVING
SUPPLEMENTAL
MATERIALS

DOIS FOR JOURNALS:
LINKING AND BEYOND



Patricia FeENEY

PATRICIA FEENEY

DOIs for Journals: Linking and Beyond

As most readers of ISQ probably know, Digital Object Identifiers (DOIs) are alphanumeric strings assigned to digital objects. Each DOI is unique and, once assigned to an item, remains a constant locator, not changing even as object moves from URL to URL. DOI names are assigned to a range of content but have been most readily embraced by the world of scholarly publishing and by researchers looking for consistent links to mutable resources.

The DOI system is managed by the International DOI Foundation (IDF), an organization that provides oversight to DOI registration agencies and maintains the DOI resolver. CrossRef, a non-profit membership organization dedicated to promoting collaboration between scholarly publishers, is the official registration agency for scholarly materials including journals, books, reports, and conference proceedings, and has registered over 41 million DOIs on behalf of our members. The vast majority (over 36.5 million) of CrossRef DOIs have been assigned to journal articles. This article will focus primarily on CrossRef's implementation of the DOI, with some coverage of how other organizations are delivering DOI-linked content in ways that enhance journal articles.

From an end user perspective, DOIs are used primarily in citations, both in print and online. The most recent edition of the *Publication Manual of the American Psychological Association* recommends that authors include DOIs in their references, allowing researchers to easily locate a cited item by clicking on (or in the case of print typing in) simple DOI links. DOIs for journal articles, books, and conference proceedings have become the standard persistent identifier for most scholarly publishing disciplines. CrossRef DOIs are primarily assigned to individual articles, but publishers opt to assign DOIs on a broader level, using DOIs to link to title and issue

level pages, as well as tables of contents. An example of this is [http://dx.doi.org/10.1002/\(ISSN\)1522-2454](http://dx.doi.org/10.1002/(ISSN)1522-2454), a title-level DOI that links to the home page of *Vakuum in Forschung und Praxis* on Wiley InterScience.

Each CrossRef journal DOI must link to a response page containing bibliographic information and a means to access full text—a DOI does not grant access to content, instead it provides a publisher-approved route to accessing full text. Most DOIs link to text-based journal content, but a DOI can link to alternate formats as well. *The Journal of Visualized Experiments*, a peer reviewed video journal for biological research, assigns DOIs to what are essentially video articles (example: <http://dx.doi.org/10.3791/1733>). Emerging formats present challenges on many fronts, but from an identifier perspective all formats are the same, provided the metadata describes the object.

The rules for creating DOIs are defined in the standard, *Syntax for the Digital Object Identifier* (ANSI/NISO Z39.84). To create a DOI, publishers obtain a DOI prefix from CrossRef, assign individual DOIs to digital objects, and deliver XML-encoded metadata to the CrossRef database. The CrossRef system in turn registers the deposited DOI and URL with the IDF. The deposited metadata consists of basic citation information that can be used to identify and describe a digital

CONTINUED »



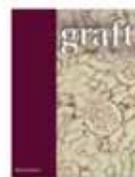
As most readers of ISQ probably know, Digital Object Identifiers (DOIs) are alphanumeric strings assigned to digital objects. Each DOI is unique and, once assigned to an item, remains a constant locator, not changing even as object moves from URL to URL.

```

<citation_list>
  <citation key="1">
    <issn>0386-2615</issn>
    <author>Takata Seiji</author>
    <volume>32</volume>
    <issue>4</issue>
    <first_page>162</first_page>
    <cYear>2005</cYear>
  </citation>
  <citation key="2">
    <doi>10.1007/BF01969578</doi>
  </citation>

```

Figure 1: Sample citation deposit



Graft: Organ and Cell Transplantation is no longer in publication. Online access to this title is now provided through the archiving organizations listed below.

Spreading of T Cell Responses to Autoantigens after Allogeneic Transplantation and Its Potential Involvement in the Rejection Process

Gilles Berthou
Graft (2003), 6(1):16
doi:10.1177/1522162802239753

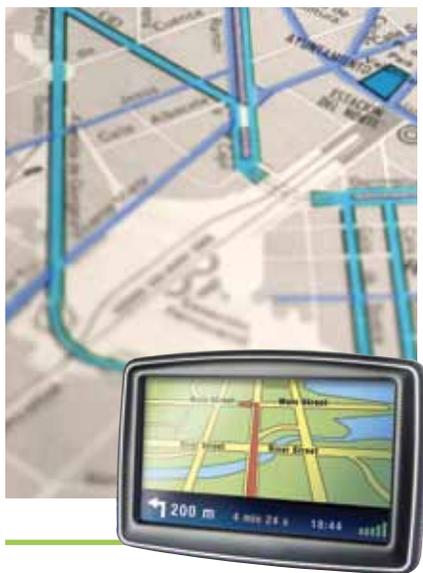
[Article Available - Portico]

[Free Article - Eprints]
[Free Article - Stanford University]



Figure 2: Sample Multiple Resolution page from Graft

Source: <http://dx.doi.org/10.1177/1522162802239753>



Reference linking benefits publisher members by driving traffic cross-publisher, and gives end users a reliable route to finding cited articles online.

CONTINUED »

object. No full text, abstracts, or other content is deposited—only data necessary to describe and locate the item is necessary. A journal article deposit, for example, contains bibliographic metadata such as the journal title, ISSN, volume, issue, page numbers, article title, and author names, as well as other identifying data such as internal publisher identifiers, codens, title abbreviations, language used, and contributor roles.

DOIs and Reference Linking

CrossRef is very much a collaborative effort. Publisher members commit to depositing and maintaining DOIs for all online journal content, but members also commit to querying the CrossRef system to harvest DOIs deposited by other members. The retrieved DOI links are then included in their reference lists published online. This practice, known to CrossRef members as reference linking, is an integral part of CrossRef as an organization. The reference linking process is powered by the metadata submitted with each DOI. Reference linking benefits publisher members by driving traffic cross-publisher, and gives end users a reliable route to finding cited articles online.

Cited-by Linking

CrossRef members may also participate in cited-by linking, an optional service that allows publishers to display citations from other publications that cite their content, providing an easily implemented way to display cross-publisher citations. Participating publishers must include citation metadata for reference lists within their article DOI deposits (see Figure 1), and in turn are able to query the reference lists of other publications. The citations are submitted as XML metadata or as already-deposited DOIs. This service is currently only available for journal content—almost 16 million

Components comprise an ever expanding assortment of data types, ranging from figures and tables to images, video, audio, and PowerPoint presentations. They allow publishers to create durable links to figures, tables, and supplemental content that can be easily updated.

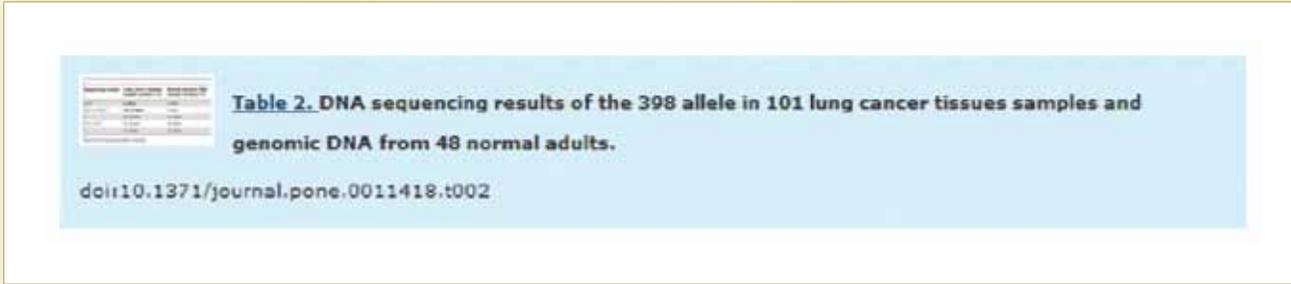


Figure 3: Component DOI
Source: <http://dx.doi.org/10.1371/journal.pone.0011418>

journal DOIs (or 34%) have at least one cited-by link. The cited-by linking network only accesses the data of participating members, but the number of participants grows constantly.

Multiple Resolution

CrossRef DOIs conventionally link the user to a single source of material but, in select circumstances, an item might exist in multiple locations or formats. The DOI specification supports a practice called multiple resolution in which multiple URLs may be attached to a single DOI. As implemented by CrossRef, instead of delivering the user directly to content, the DOI resolves to an interim page containing citation metadata and multiple links to an item. This feature has been enthusiastically adopted by members who co-publish journals, as it allows them to dually host an authoritative version of an article.

A recent focus on preserving online journal content has resulted in cooperative efforts between archiving institutions to preserve and provide continuing access to titles that have ceased publication and are no longer maintained by the original publisher. The multiple resolution process allows DOIs assigned to these journals to resolve to multiple hosts, allowing end users to choose between the archiving organizations that host the content. DOIs are currently assigned to *Auto/Biography* and *Graft*, originally published by SAGE, and *Brief Treatment* and *Crisis Intervention* from Oxford University Press (OUP), both of which have been archived by Portico and CLOCKSS (see Figure 2).

DOIs for Supplemental Content

Publishers are increasingly delivering supplemental journal content online, and DOIs can be assigned to supplemental materials as well. Publishers generally use two content types to link to supplemental materials: components and datasets. Supplemental materials are typically not cited on their own and as such aren't discoverable by querying the CrossRef system, but assigning DOIs allows publishers to easily create and update durable links to content that otherwise might not survive platform migrations and ownership changes. Other registration agencies facilitate assigning DOIs to supplemental content as well, particularly data not provided by the publisher such as datasets, videos, maps, and raw scientific data.

Components

Components comprise an ever expanding assortment of data types, ranging from figures and tables to images, video, audio, and PowerPoint presentations. They allow publishers to create durable links to figures, tables, and supplemental content that can be easily updated. Only a small number (~300,000) of CrossRef's 41 million+ DOIs are components, but the number grows daily. The CrossRef definition of component is fairly loose, viewing them as a container element and allowing the publisher to determine how their supplemental material is classified. Consequently required component metadata is simple, consisting of sparse metadata describing the content and file type. The metadata perhaps most relevant to the

CONTINUED »

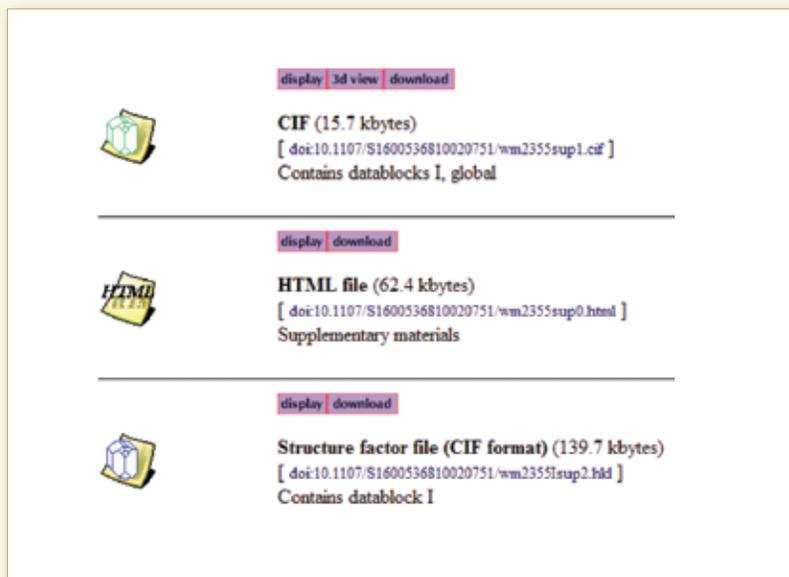


Figure 4: Multiple component DOIs

Source: <http://dx.doi.org/10.1107/S1600536810020751>

CrossRef collects a number of dataset DOIs but they are also increasingly being registered by organizations devoted to delivering datasets and other types of raw scientific data.

CONTINUED »

component is that of the item the component is supplementing, also known as the parent DOI. Other component metadata consists of an item description, format (or file type), and of course the DOI and URL. Optional elements include item titles, contributor information, and publication dates.

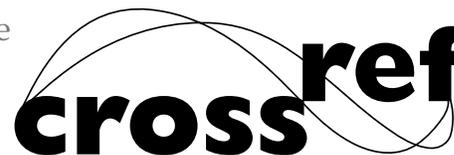
A component must be associated with a parent DOI that has been created for a CrossRef content type (journal, book, conference proceeding, technical report, working paper, standard, dissertation, or dataset). The majority of deposited components are associated with journal articles. Although components have not been widely adopted across the membership, several CrossRef members have successfully integrated them into their content. The Public Library of Science uses components to link to tables and figures for their journal PLoS ONE. The tables and figures appear within the text in both the print and online versions of an article, with the DOI listed below an image thumbnail (see Figure 3). This component DOI links directly to the full-sized table or figure.

The International Union of Crystallography (IUCr) uses components to provide durable links to crystallographic information files (CIFs) and other supplemental materials in a variety of formats, as well as including a DOI directed at an HTML page containing all supplemental material for an article (see Figure 4). IUCr also includes the component DOI within the HTML version of the article.

Datasets

DOIs may also be assigned to datasets, a content type dedicated to database records. Datasets typically exist as stand-alone databases, but individual dataset records or a database as a whole may be used to supplement journal article data. CrossRef collects a number of dataset DOIs but they are also increasingly being registered by organizations devoted to delivering datasets and other types of raw scientific data. Dataset providers and journal publishers are able to provide

In the coming months CrossRef will be launching a new project, CrossMark, that will allow users to retrieve information about publisher-maintained versions of a document—including the status of a document, publisher metadata, and of course the CrossRef DOI assigned to the document.



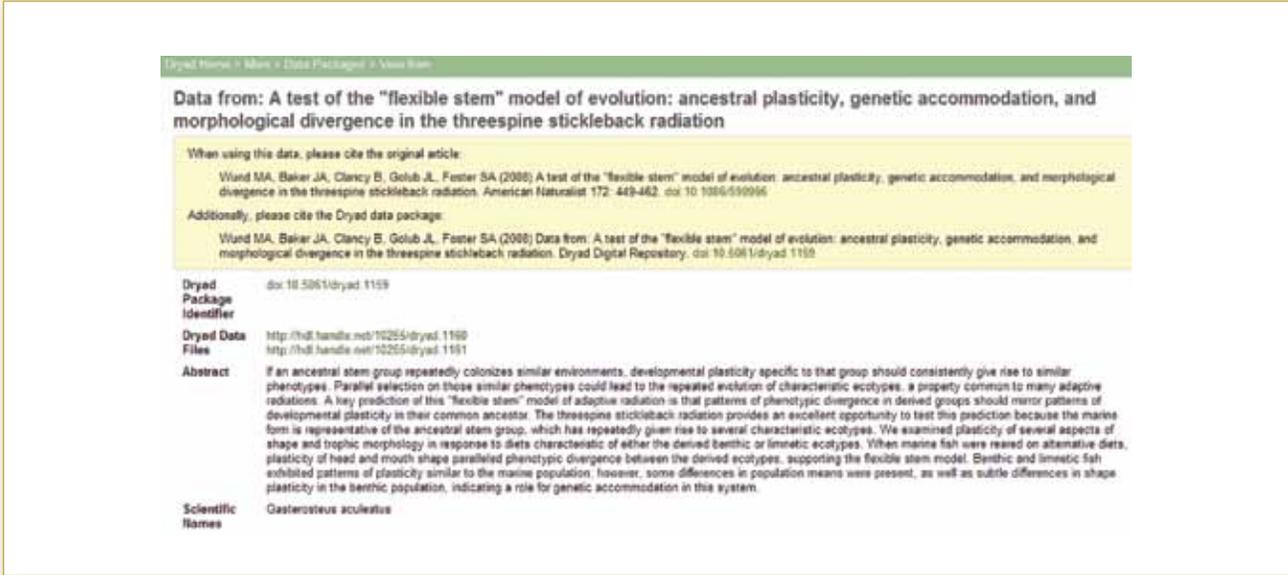


Figure 5: Dataset DOI landing page with original journal article citation and DOI
 From source: <http://datadryad.org/handle/10255/dryad.1209>

lasting links to supplemental content by cross-linking between hosted datasets and journal content.

DOI linking between datasets and journal content is nascent but off to a promising start. One example of this is Dryad, a newish data repository focused on evolutionary biology and ecology that partners with a number of major evolutionary biology and ecology-centric publications. Dryad lists among its goals to “preserve all the underlying data reported in a paper at the time of publication, when there is the greatest incentive and the ability for authors to share their data.” Accordingly, they register DOIs for the datasets they deliver. When compiled in conjunction with a published article the dataset landing page provides DOI links to the parent article, as shown in Figure 5, an example of data supporting a paper published in *Molecular Ecology*.

In the Dryad example, the parent article does not provide a link back to the dataset. Collaboration between dataset providers and journal publishers does exist, as evidenced by a dataset hosted by PANGAEA, an open access network for geo-scientific and environmental data. A journal article DOI link is provided in the citation, and the response page for the article DOI contains a link to PANGAEA supplementary data.

DOI registration for journal articles has become an accepted practice, as have CrossRef enhancements such as reference and cited-by linking. More and more publishers struggling to represent their supplemental data online are using DOIs for linking. Recent efforts to assign DOI links to raw data are encouraging and more reciprocal linking can be expected in the future, as can other DOI-related enhancements. In the coming months CrossRef will be launching a new project,

CrossMark, that will allow users to retrieve information about publisher-maintained versions of a document—including the status of a document (withdrawn, corrected, enhanced, etc.), publisher metadata, and of course the CrossRef DOI assigned to the document. The CrossMark process will be enabled in part by the metadata deposited with CrossRef DOIs. Although this project is still in the planning stages, it’s a promising sign of how DOI use can evolve. | SP | doi: 10.3789/isqv22n3.2010.06

PATRICIA FEENEY <pfeeney@crossref.org> is the Metadata Quality Coordinator at CrossRef.

CrossRef website
www.crossref.org

CrossMark
www.crossref.org/crossmark.html

CLOCKSS
www.clockss.org

Dryad
www.datadryad.org/factSheet

PANGAEA
www.pangaea.de

Portico
www.portico.org

Publication Manual of the American Psychological Association
www.apastyle.org

Syntax for the Digital Object Identifier (ANSI/NISO Z39.84)
www.niso.org/standards/z39-84-2005/



RELEVANT LINKS