

Component DOIs Revisited

Martin Fenner

July 9, 2015

Four years ago I wrote a blog post about component DOIs. It is time to revisit the topic, in particular since our approach to citing data associated with a publication has changed since 2011.

Component DOIs are explained in the CrossRef Help System:

DOIs may be assigned to items that are part of a journal article, book chapter, or any other content item. A component would typically be a figure, table, or image which is part of or referred to by the parent item. Assigning a DOI to a component allows direct linking to the component item.

Component DOIs are DOIs, i.e. persistent identifiers that link directly to the resource in question, e.g. a figure in a publication. The component DOI for a figure in a PLOS paper used in the 2011 post still works as expected, despite changes to the URL of the journal landing page.

The problem with component DOIs is the problem with DOIs in general: there is basic functionality common to all DOIs, and there are additional services specific to subgroups of DOIs. This confuses users - in particular since there is no easy way to immediately see what kind of DOI they have in front of them - and in the case of component DOIs there is one important feature missing.

DOIs are assigned by registration agencies (CrossRef and DataCite are the most relevant ones for scholarly content), and these RAs have built different services around DOIs, e.g. different ways to describe and search the metadata (title, authors, etc.) associated with a DOI. Component DOIs are again different, the most important difference is that in the CrossRef implementation they are not discoverable by querying the CrossRef system (Feeney, 2010). Component DOIs are also always associated with a parent DOI (for the article, book, etc.). Although this is the expected behaviour, we shouldn't expect component DOIs to always look like an extension of the parent DOI, as in `10.1371/journal.pone.0006022.g002` used in the example above.

In essence, a component DOI is a **DOI light**. We can use them for persistent linking, but we can't use them for discovery via the CrossRef Metadata Search (and by extension other indexing services). A common use case for component

DOIs is supplementary information in a journal article. Content in supplementary information files is already much harder to find than content in the body of an article, using component DOIs instead of regular DOIs makes the content again harder to find.

All of this might not have been much of an issue when I wrote the 2011 post, but making the data underlying a publication publicly available and discoverable is increasingly becoming something that funders, publishers and institutions expect. Most of these data are not deposited in dedicated data repositories, but in supplementary information files (for PLOS articles published since March 2014 this is true for more than 50% of papers). Using regular DOIs for supplementary information files with proper metadata and proper inclusion in indexing services will make it easier to find, access and reuse these data.

Unfortunately that still leaves us with the problem that the supplementary information files then will have CrossRef DOIs, whereas data repositories typically use DataCite DOIs, so that we need to search for these datasets in two different places. But that is material for another post.

References

Feeney, P. (2010). DOIs for Journals: Linking and Beyond. *Information Standards Quarterly*, 22(3), 27. <https://doi.org/10.3789/isqv22n3.2010.06>