

# Digital Object Identifiers: Stability for citations and referencing, but not proxies for quality

 [blogs.lse.ac.uk/impactofsocialsciences/2015/04/23/digital-object-identifiers-stability-for-citations/](https://blogs.lse.ac.uk/impactofsocialsciences/2015/04/23/digital-object-identifiers-stability-for-citations/)

4/23/2015

*What exactly is a Digital Object Identifier (DOI) and how does it help in the management and long-term preservation of research? [Laurence Horton](#) explains the basic structure and purpose of a DOI and also points to some limitations. DOIs are not the only way of providing fixed, persisting references to objects, but they have emerged as the leading system.*

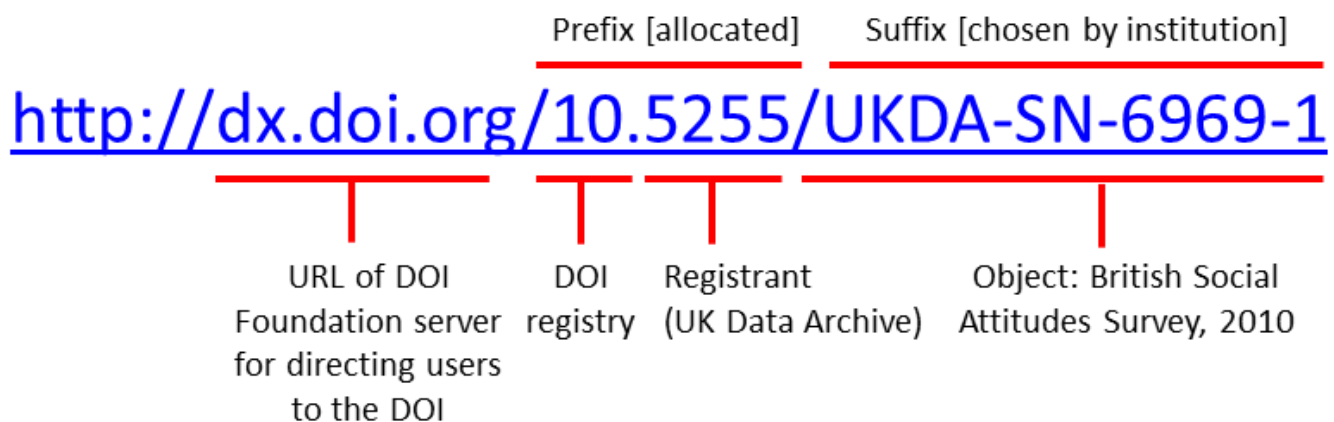


A DOI is a Digital Object Identifier. It is an online reference (digital), pointing to (identifying) a resource (object). The DOI system links, through a directory, references and web addresses of an object to a “landing” page providing information on access and metadata about that object — at a [minimum](#) [PDF] its creator, title, publisher, year of publication, and DOI. This allows DOIs to provide a stable, persistent, resolvable reference taking users to an object, even if web addresses or other references to the location of an object, or its content, change.

DOIs appeared with the new millennium, and there are now over 100 million assigned. The [International DOI Foundation](#) governs DOIs and regulates them to an [ISO standard](#). Registration Agencies like [DataCite](#) or [CrossRef](#) make up the foundation and provide the structure supporting DOIs. Allocation Agents, who are members of Registration Agencies, manage assigning DOIs to objects. Clients, like universities, sign a contract and pay an annual fee to agents to become “registrants” and create, or “mint”, DOIs. When minted, DOIs are registered with the Foundation whose directory then points associated web addresses to the landing page

Objects need not be digital to have a DOI — they can be physical, like a book. Nor need they be static — objects can change over time, like a dataset. If web addresses or the object content significantly changes, clients must update the DOI record so the Foundation’s directory continues pointing users to the landing page.

Let us illustrate DOIs using the dataset downloaded most by LSE staff and students from the UK Data Service, the [British Social Attitudes Survey, 2010](#).



DOIs combine a prefix and suffix. The prefix is fixed and standardised. The “10” identifies the link as a DOI, followed a four-digit number showing the registrant who minted it, so a DOI prefixed 5255 always comes from the UK Data Archive. The registrant defines the suffix. Here, the UK Data Archive uses its own sequential numbering system but it could use longer or shorter strings of numbers, letters, or both. The “1” at the end is the UK Data Archive’s indicator this is a first edition of the data set.

Anything can have DOI as long as it has a digital landing page. Indeed, DOI's may be the only thing shared by Watson and Crick's outline of DNA published in *Nature* ([10.1038/171737a0](https://doi.org/10.1038/171737a0)) later recognised with a [Nobel Prize](#), and the film *Holiday on the Buses* ([10.5237/A929-C667](https://doi.org/10.5237/A929-C667)) described as “[absolutely abysmal](#)” by *Radio Times*. Also, if you only have the prefix and suffix in a reference, copy and pasting into Google or most reference manager software also “resolves” the DOI and retrieves its metadata.



Image credit: [Hypertext Editing System by Greg Lloyd 1969](#) (Wikimedia CC-BY 2.0)

### How does it fit into Research Data Management?

DOIs are an investment in making data citable, elevating it to the status of a research output with reuse equating to citation. In a world dependent on publishing and being cited, if your data is available, discoverable and citable then people will discover it and [it will be cited](#). DOIs are also flexible. Depending on the policy of the registrant, they can be allocated to datasets, variables, documentation, and different versions of datasets, not just publications.

What DOIs are not is a symbol of data quality. You can attempt to define “quality” but the problem is using DOIs as a proxy. Just because something has a DOI does not mean it is good — just watch *Holiday on the Buses*. Also, reading the International DOI Foundation [handbook](#) does not produce a mention of quality. Identification, yes. Resolution, yes. Management, yes. Quality, no. We must not start using tools designed for one end to another.

### What does it do for preservation?

We can start (and it is a start, there is still [lots to address](#)) bringing [stability](#) to data referencing by using DOIs. In the past, referencing was simpler: you cited something by describing its print location — author, title, publication, volume, and page numbers. These days it can be complicated. Websites, databases, audio, video, blogs, social media, software, eThis, and iThat, the research world just does not exist only on paper. Also, while the internet is not a “[series of tubes](#)”, it does “[rot](#)”. Websites change addresses, servers get switched-off, resources significantly change, and when that happens without care, [original resources and references disappear](#). For example, it does not take long in the reference section of Wikipedia articles to come across links to pages that are dead, broken or dangling. It is irritating, but if you are a legal scholar, when URLs cited in court judgements [no longer work](#) it is a fundamental problem.

DOIs are [not the only way](#) of providing fixed, persisting references to objects, but have emerged as the leading system. Because of the infrastructure underpinning DOIs — the technology, financial commitment and willpower behind the system — objects with DOIs are discoverable, citable and offer long-term reassurance that will remain the case.

*Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.*

## **About the Author**

**Laurence Horton** is Data Librarian at the London School of Economics and Political Science. He is responsible for Research Data Management support in the School. He can be found on Twitter [@laurencedata](#).

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.