What is a DOI?

From the DOI Handbook:

A DOI is a digital identifier of an object, any object — physical, digital, or abstract. DOIs solve a common problem: keeping track of things. Things can be matter, material, content, or activities.

A DOI is a unique number made up of a prefix and a suffix separated by a forward slash. This is an example of one: 10.1000/182. It is resolvable using the doi.org proxy server by displaying it as a link: https://doi.org/10.1000/182.

Designed to be used by humans as well as machines, DOIs identify objects persistently. They allow things to be uniquely identified and accessed reliably. You know what you have, where it is, and others can track it too.
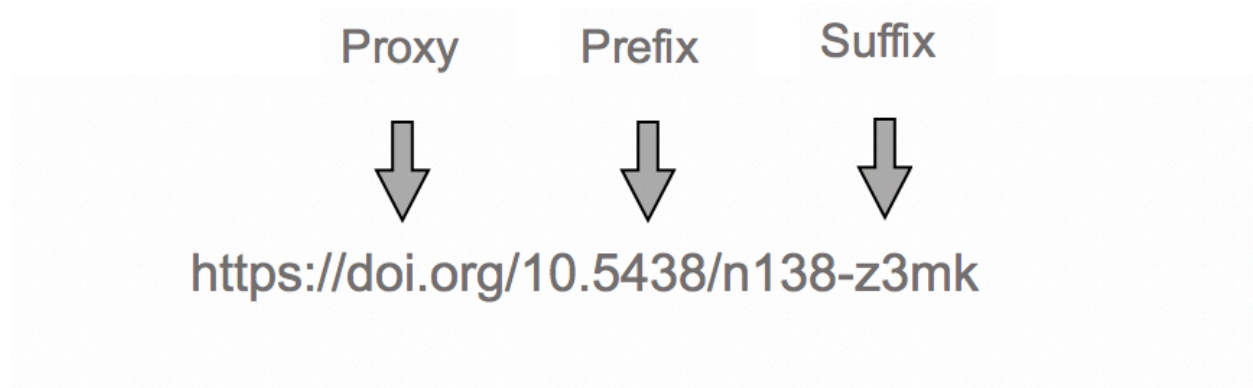

DOI Foundation Logo

DOIs are an example of Persistent Identifiers (PIDs). DOIs have been adopted by the scholarly communication community as, almost, the 'default' identifier for publications. In the last years, they have been adopted as identifiers for data publication, enabling data citation and reuse.

DOI names are provided by DataCite and other DOI registration agencies, coordinated by the International DOI Foundation (IDF). Out of the current 12 DOI registration agencies, most of them assign DOI names to scholarly content.

A DOI name is divided into three parts, separated by the first two slashes ('/'):

Proxy

The proxy is not part of the DOI specification, but almost all scholarly DOIs that users encounter today will be expressed as HTTP URLs. DataCite therefore recommends in the DOI display guidelines that all DOIs are displayed as permanent URLs, consistent with the recommendations of other DOI registration agencies, e.g. the Crossref DOI display guidelines.

Prefix

A DOI prefix always starts with '10.' and continues with a number (e.g. '10.1234' or '10.20865'). The DOI prefix is used as a namespace so that DOIs are globally unique without requiring global coordination for every new identifier. Prefixes are managed in the handle system and all DOI prefixes are numbers without any semantic meaning. One lesson learned with persistent identifiers is that adding meaning to the identifier (e.g. by using a prefix with the name of the data repository) is always dangerous, because – despite best intentions – all names can change over time.

Suffix

A DOI suffix can be (almost) any string of characters and symbols. This provides flexibility, but can be problematic. There are a few things to consider when you create a suffix:

- A DOI is also a URL. The list of characters allowed in a URL is limited so including a character in the DOI suffix that is not allowed in a URL can break the functionality of the DOI.

- Semantic meaning (human readable information) in the DOI suffix is problematic because the meaning can change over time but a DOI string cannot be changed or deleted once it has been registered in the Global Handle server.
- Version information in the suffix is misleading and may lead users to think that a "V2" in the suffix means the DOI will automatically resolve to version 2 of the underlying content. Versioning in the suffix is not a functionality of the DOI system. See versioning documentation for best practice recommendations.

🚧

Reserved characters should not be included in a DOI suffix

These reserved characters should not be included in a DOI suffix: `; | / | ? | : | @ | & | = | + | $ | ,`

The easiest and recommended option is to use a randomly generated suffix. The auto-generated DOI strings use `a-z` and `0-9`. They avoid `i`, `l`, `o` as they are easily mixed up with `0`, `1`. We group the suffix into blocks of 4, separated by a hyphen. You can generate a random suffix in both Fabrica and the API and your DOI will look something like this: `10.5438/9te8-5h68`.

👍

Auto-generate a suffix

The auto-generate DOI name functionality means the DOI suffix is generated automatically and will look something like this: `10.15138/33bv-s284`

If you choose not to use this option, remember:

- The DOI suffix must be unique within each prefix. The optimum length of a DOI suffix is 6–10 characters.
- Only use `a-z`, `0-9` and `-` in a DOI suffix. Other characters might have special meaning or will be escaped. DOI suffixes are not case sensitive.
- Avoid human-readable information in a DOI suffix because any meaning may change over time.

Keep in mind that in contrast to HTTP URIs – which are case-insensitive – https://doi.org/10.60787/AFRISCHOLAR.31780 and https://doi.org/10.60787/afrischolar.31780 are the same DOI. All DOIs are converted to upper case upon registration and DOI resolution, but DOIs are not consistently displayed in such a way.

More information about suffixes:

- DataCite blog post
- Crossref blog post
- Reserved characters in Elasticsearch
- Technical implementation for generating random DOI strings