ETH-Bibliothek
E-Publishing

ETH Zurich
Rämistrasse 101
8092 Zurich

Phone +41 44 632 21 35
doi@library.ethz.ch
www.doi.ethz.ch

ETH Zurich DOI Desk

DOI Registration Manual

Date 26/09/2014
Version 1.1
Authors Barbara Hirschmann, with contributions from Matteo Corti and Angela Gastl

Version Control

Version	Date	Changes	Author
1.0	26/09/2014		Barbara Hirschmann
1.1	13/11/2015	Metadata property Resource Type changed to mandatory; Default values for missing mandatory properties	Barbara Hirschmann

Table of Contents

Part I: Guidelines for Digital Object Identifiers....3
1 Syntax....3
2 Granularity....5
3 Display guidelines....6
Part II: DOI Management Tool....8
4 Interface specifications....8
5 Metadata specifications....11
6 Instructions for manual registration....16
7 Instructions for the handling of duplicates and deleted objects....19

Part I: Guidelines for Digital Object Identifiers

1 Syntax

Based on the specifications laid out in the „DOI Handbook"[1] of the International DOI Foundation (IDF) as well as technical requirements defined by the DataCite Metadata Store[2] the structure of a DOI name has to follow a set of basic rules.

Each DOI name is made up of a **DOI prefix** and a **DOI suffix** separated by a forward slash. Two examples:

DOI name	Prefix	Suffix
10.1000/123456	10.1000/	1237456
10.1000/ISBN1-900512-44-0	10.1000/	ISBN1-900512-44-0

Prefix

- The ETH Zurich DOI Desk assigns one or more prefixes to each DOI client.
- Usually each client application (platform, repository, journal,…) that registers DOIs has its own prefix.
- All DOI prefixes begin with "10." This is how DOIs can be distinguished from other identifiers within the Handle system[3].

Suffix

- The suffix is an alphanumeric string.
- It is recommended that only the following characters are used within a DOI name
    - 0-9
    - a-z
    - A-Z
    - - (dash)
    - . (dot)
    - _ (underscore)
    - + (plus)
    - : (colon)
    - / (slash)
- It is the client's responsibility to assure that every DOI name is unique. Possible methods for guaranteeing uniqueness within the suffix are
    - to integrate a sequential number,
    - to integrate an existing identifier issued by the database that generates the DOI name.
- There is no defined limit on the length of the suffix. However, we recommend to use a maximum of 50 characters.
- The DOI name is case-insensitive.
- Other identifiers (such as ISBN numbers, internal database codes etc.) can be integrated into the suffix.
- The DOI Desk and the DOI client will agree upon the syntax of the DOI suffix and document it in the service level agreement signed by both parties.

[1] DOI Handbook, Version 5 (April 2012), http://doi.org/10.1000/182
[2] DataCite API v2 for datacentres, https://mds.datacite.org/static/apidoc
[3] http://www.handle.net

- It is common practice to use dots (.) to label different versions of an object and slashes (/) to describe parts of a whole.

  Examples

DOI	10.1234/abc123
DOI of a new version	10.1234/abc123.1
DOI of an excerpt	10.1234/abc123/2

2 Granularity

According to the DOI Handbook[4] of the IDF a DOI name can be assigned at any level of granularity that a registrant deems to be appropriate.

The decision should be based on the concept of „functional granularity“: it should be possible to identify an entity whenever it needs to be distinguished.

For example, a DOI name can be assigned to:

- a novel
- a specific edition of that novel
- a specific chapter within that edition of the novel
- a single paragraph
- a specific image
- etc.

Multiple DOI names can be assigned to upper level entities as well as to lower level entities of the same object.

It is important that a DOI should enable a user to cite a digital object as easy as possible and in a scientifically sound manner.

[4] DOI Handbook, Version 5 (April 2012), http://doi.org/10.1000/182

3 Display guidelines

3.1 How to display a DOI

DOIs should always be **displayed as permanent URLs** or resolve to a permanent URL.

Examples:

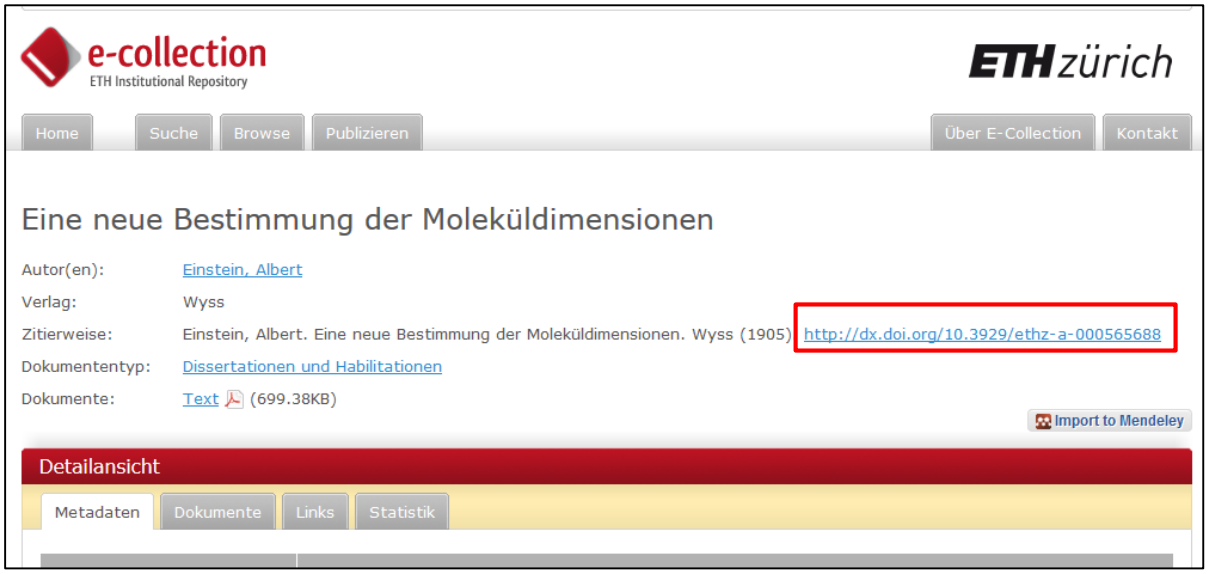
Permanent URL	**http://doi.org/10.3929/ethz-a-000565688**
DOI that resolves to a permanent URL	**10.3929/ethz-a-000565688**

We recommend that DOIs are displayed in this form wherever citation information about a digital object is presented, especially online. In print publications the first form mentioned above should be used.

3.2 Where to display a DOI

DOIs must be **displayed on the client's landing pages**.

Example of a DOI landing page:



We recommend that DOIs are also used, displayed or distributed in the following contexts:

- Title pages, tables of contents, abstracts or full texts of digital objects
- Metadata feeds to third parties
- "How to Cite This" instructions
- Social networking sites

3.3 Short DOIs

The ShortDOI service is an open system that creates shortcuts to DOI names. The service creates short handles and enables permanent links that are very short. This can be important in situations where space is limited.

For example the short form of the DOI http://doi.org/10.3929/ethz-a-010010483 is http://doi.org/scj.

A benefit of the ShortDOI service is that only one ShortDOI is created for each DOI Name and the ShortDOI returns exactly the same results as the DOI name.

Part II: DOI Management Tool

4 Interface specifications

4.1 Background

The DOI Management Tool harvests DOIs from the configured sources (libraries and publishers) and, after validating them, publishes the DOIs to a handle server and to DataCite [1].

4.2 Interface

DOIs are harvested using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) version 2.0 [2, 3]. OAI-PMH uses XML over HTTP. DOI publishers are required to offer an open OAI-PMH interface which delivers metadata in the “simple Dublin Core” format [4-7]. XMLs should be encoded using the UTF-8 character set.

4.2.1 URL and availability

The interface should be reachable with a fix DNS name and always be available. The interface should be able to correctly handle requests with the “ListRecords” OAI-PMH verb (both with or without the “from” parameter). If the “from” parameter is specified only the records newer than the specified timestamp should be delivered.

4.2.2 Data format

Data is delivered as an XML with the XML schema http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd. The following restrictions apply:

4.2.2.1 DOIs (<dc:identifier>)

Tag

DOIs must be enclosed in a <dc:identifier> XML tag. Only one single DOI can be present in a record. The DOI is being recognized as such if

- it is prefixed with the string “doi:” (i.e. doi:10.xxx/yyyy)
- it is post fixed with the string “ / doi” (i.e. 10.xxx/yyyy / doi)
- the prefix is followed by a forward slash (i.e. 10.xxx/yyyy)

DOI Format

DOIs must be composed by a valid prefix, a slash (“/”) and a suffix. Although the format of the suffix is free, DataCite imposes additional restrictions. To be exported to DataCite a DOI must be composed by the following characters only:

- 0-9
- a-z
- A-Z
- - (dash)
- . (dot)
- _ (underscore)
- + (plus)
- : (colon)
- / (slash)

DOIs are by definition case-insensitive. The submission of two DOIs that differ only in the case of the used characters will be rejected.

4.2.2.2 URLs (<dc:identifier>)

URLs must be valid URLs according to the Internet standard STD 66 for the URI scheme [8].

4.2.2.3 Multiple identifiers

Both DOIs and URLs are embedded in a <dc:identifier> tag. If the OAI-PMH interface delivers more than one URL or more than one DOI per record the relevant ones must be distinguishable:

- the main DOI must have the correct prefix
- the main URL must begin with a single previously defined prefix

4.2.2.4 Mandatory attributes

Although the Dublin Core format does not define any mandatory metadata attributes, DataCite requires at least:

- one title (<dc:title>)
- one creator (<dc:creator>)
- one publisher (<dc:publisher>)
- one date (<dc:date>)
- one type (<dc:type>)

See chapter 5 for more information on mandatory and optional metadata attributes.

4.2.2.5 Date (<dc:date>)

Dates (dc:date) must be expressed using either the ISO 8601 standard format (preferred) or the Swiss KIDS format.

4.2.3 Lifecycle

DOIs cannot be deleted. If the related resource is deleted the Landing Page must be kept available.

4.3 Validation

Before submission the delivered XMLs should be validated using the *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) Validator & data extractor* [9].

4.4 Example

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH
  xmlns              = "http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi          = "http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation = "http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-
PMH.xsd"
>
  <responseDate>2012-02-15T07:25:05Z</responseDate>
  <request
    verb           = "ListRecords"
    from           = "2012-02-13T21:12:14Z"
    metadataPrefix = "oai_dc"
    set            = "DOI"
  >
    http://aleph-prod.ethz.ch/OAI
  </request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:eth-bib-pub:EBI01-000342738</identifier>
        <datestamp>2012-02-14T06:26:01Z</datestamp>
        <setSpec>DOI</setSpec>
      </header>
      <metadata>
        <oai_dc:dc
          xmlns:oai_dc       = "http://www.openarchives.org/OAI/2.0/oai_dc/"
```

```
        xmlns:dc          = "http://purl.org/dc/elements/1.1/"
        xmlns:dcterms     = "http://purl.org/dc/dc/terms/"
        xmlns:xsi         = "http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation = http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd
      >
        <dc:type>Printed language material</dc:type>
        <dc:creator>Zingg, André</dc:creator>
        <dc:language>ger</dc:language>
        <dc:identifier>10.3929/ethz-a-000342738 / doi</dc:identifier>
        <dc:title>
          Experimentelle und thermodynamische Untersuchung des Siedegleichgewichts des Systems NH-
3-H-2O bei hohen Drücken
        </dc:title>
        <dc:publisher>Zürich</dc:publisher>
        <dc:date>1985</dc:date>
        <dc:format>175 S.</dc:format>
        <dc:identifier>http://e-collection.ethbib.ethz.ch/show?type=diss&amp;amp;nr=7743</dc:identifier>
      </oai_dc:dc>
    </metadata>
  </record>
 </ListRecords>
</OAI-PMH>
```


4.5 References

1. *DataCite Metadata Store*. Available from: https://mds.datacite.org.
2. Lagoze, C., et al. *The Open Archives Initiative Protocol for Metadata Harvesting*. 2008; Available from: http://www.openarchives.org/OAI/openarchivesprotocol.html.
3. Wikipedia contributors. *Open Archives Initiative Protocol for Metadata Harvesting*. Available from: http://en.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting.
4. Wikipedia contributors. *Dublin Core*. Available from: http://en.wikipedia.org/wiki/Dublin_Core.
5. Kunze, J. and T. Baker, *The Dublin Core Metadata Element Set*, in *IETF RFC 5013* 2007.
6. International Organization for Standardization, *ISO 15836:2009, Information and documentation, The Dublin Core metadata element set*, 2009.
7. National Information Standards Organization, *ANSI/NISO Z39.85 - The Dublin Core Metadata Element Set*, 2007.
8. Berners-Lee, T., R. Fielding, and Masinter, *Internet standard STD 66 : Uniform Resource Identifier (URI): Generic Syntax.* 2005.
9. Banos, V. *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) Validator & data extractor*. 2012; Available from: http://validator.oaipmh.com.

Author of this chapter: Matteo Corti

5 Metadata specifications

The ETH DOI Management Tool harvests the descriptive metadata on the DOI objects from its customers in the format Dublin Core Simple. Dublin Core is also the entry format for manual registration.

The DOI Management Tool then automatically converts the data into the DataCite Metadata Schema 3.1 (schema.datacite.org) and transfers it to the DataCite Metadata Store (mds.datacite.org and search.datacite.org).

As part of this metadata pool, the data can be used freely by third parties for the integration in scientific search engines or the development of other services.

The following table and chapter explain the mapping mechanism implemented by the DOI Management Tool to convert data from Dublin Core format to the DataCite Metadata Schema 3.1.

ID	Dublin Core	DataCite Property	Obligation	Occurrence	Standard value
1	dc:identifier	Identifier	mandatory	2-n	
		identifierType			DOI
2	dc:creator	creatorName	mandatory	1-n	
3	dc:title	Title	mandatory	1-n	
4	dc:publisher	Publisher	mandatory	1	
5	dc:date	PublicationYear	mandatory	1-n	
		Date			
		dateType			Available
6	dc:type	resourceTypeGeneral	mandatory	1	Controlled List
	dc:type	resourceType	optional	0-1	
7	dc:subject	Subject	recommended	0-n	
8	dc:contributor	contributorName	recommended	0-n	
		contributerType			Other
9	dc:language	Language	optional	0-1	
10	dc:format	Size	optional	0-n	
		Format			
11	dc:rights	Rights	optional	0-n	
12	dc:description	Description	recommended	0-n	
		descriptionType			Other

Table 1: Dublin Core - DataCite mapping

5.1 Property definitions and mapping rules

ID 1: Identifier

This property contains the DOI and the URL of the resource. See chapter 4.2 for details.

ID 2: Creator

The main researchers involved in producing the data, or the authors of the publication, in priority order.
May be a corporate/institutional or personal name.
The personal name format should be: family, given.

ID 3: Title

A name or title by which a resource is known.

ID 4: Publisher

The name of the entity that holds, archives, publishes, prints, distributes, releases, issues, or produces the resource.

ID 5: PublicationYear / Date

Definitions – PublicationYear for datasets and publications

According to DataCite, the *PublicationYear* is the year when the data was or will be made publicly available.

In the case of datasets, "publish" is understood to mean making the data available on a specific date to the community of researchers.

- If that date cannot be determined, use the date of registration.
- If an embargo period has been in effect, use the date when the embargo period ends.
- If there is no standard publication year value, use the date that would be preferred from a citation perspective.

Definitions – PublicationYear for digitised versions of physical objects

If the DOI is being used to identify a digitised version of an original item, the recommended approach is to supply the PublicationYear for the digital version and not the original object.

The Title field may be used to convey the approximate or known date of the original object. Other metadata properties available for additional date information about the object include: Subject and Description. However, only Title will be part of the citation.

If the DOI is being used to identify the original object and the publication date of this is unknown or not standard we recommend to use the date that would be preferred from a citation perspective.

Mapping routine

The DataCite Property *PublicationYear* is mandatory and must contain only one year in YYYY format.

However, clients of the DOI Desk can deliver various dates in dc:date. The DOI management tool uses the earliest year and transfers it to DataCite as *PublicationYear*.

If a date is provided including month and day (YYYY-MM-DD), only the year will be delivered as *PublicationYear*.

All interpretable dates will be delivered to DataCite in the *Date* field with the dateType Property „Available“.

Dublin Core Date	DataCite PublicationYear	DataCite Date (dateType=Available)
2000	2000	2000
2004-03	2004	2004-03
2004-03-02	2004	2004-03-02
2004/2005	2004	2004/2005
2004-03/2005-06	2004	2004-03/2005-06
2004-03-02/2005-06-02	2004	2004-03-02/2005-06-02
uuuu	0000	uuuu
2004 **2005**	2004	2004 2005
17uu	1700	17uu
17uu **18uu**	1700	17uu 18uu
1750 **uuuu**	0000	1750 uuuu
notBefore 0885 :: notAfter 0925	0885	notBefore 0885 :: notAfter 0925

Table 2: Mapping table for dc:date values

Date Ranges

Use RKMS - ISO8601 standard for depicting date ranges. Example: 2004-03-02/2005-06-02

Unknown Dates

An unknown date can be expressed by the value “uuuu” which will be exported to DataCite *PublicationYear* as “0000”.

ID 6: ResourceType

It is mandatory to provide one of the following values within dc:type:

- Audiovisual
- Catalog
- Collection
- Dataset
- Event
- Image
- InteractiveResource
- Model
- PhysicalObject
- Service
- Software
- Sound
- Text
- Workflow
- Other

The supplied value can be combined with another free text value which describes the resource type in more detail.

For textual resources, whenever possible, the resourceType “Text” should be used and combined with a term from the CASRAI Output Types list.

Example:
<dc:type>Text</dc:type>
<dc:type>Working Paper</dc:type>

ID 7: Subject

Subject, keyword, classification code, or key phrase describing the resource.

ID 8: Contributor

The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.

ID 9: Language

The primary language of the resource.
Allowed values are taken from IETF BCP 47, ISO 639-1 language codes.
Examples: en, de, fr

ID 10: Size/Format

The dc:format field will be mapped to *Format*, unless it is recognized as *Size*.

Examples for values recognized as *Size*:

- 512 x 256 px.
- 1 Band
- 204 S.
- 212 p.

ID 11: Rights

Any rights information for the resource.

ID 12: Description

All additional information that does not fit in any of the other categories.
May be used for technical information, series or journal titles, abstracts etc.

5.2 Unknown mandatory values

If one of the mandatory properties *Creator*, *Title* or *Publisher* cannot be specified at all use one of the following values to express this fact:

Value	Meaning
(:unac)	temporarily inaccessible
(:unal)	unallowed, suppressed intentionally
(:unap)	not applicable, makes no sense
(:unas)	value unassigned (e.g., Untitled)
(:unav)	value unavailable, possibly unknown
(:unkn)	known to be unknown (e.g., Anonymous, Inconnue)
(:none)	never had a value, never will
(:null)	explicitly and meaningfully empty
(:tba)	to be assigned or announced later
(:etal)	too numerous to list (et alia).
(:at)	the real value is at the given URL or identifier

An unknown date must be expressed by the value “uuuu” which will be exported to DataCite *PublicationYear* as “0000”.

5.3 Missing mandatory values

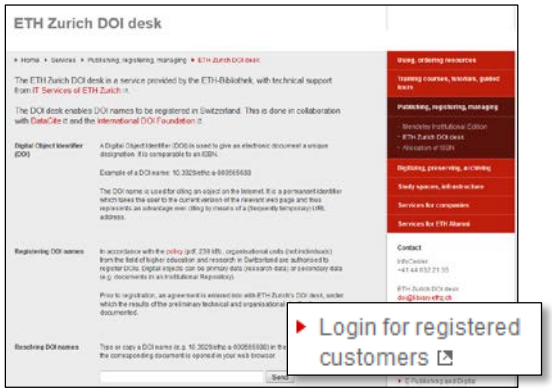
If a DOI is registered missing one of the mandatory values, the DOI Management Tool will produce the following values upon export to the DataCite Metadata Store:

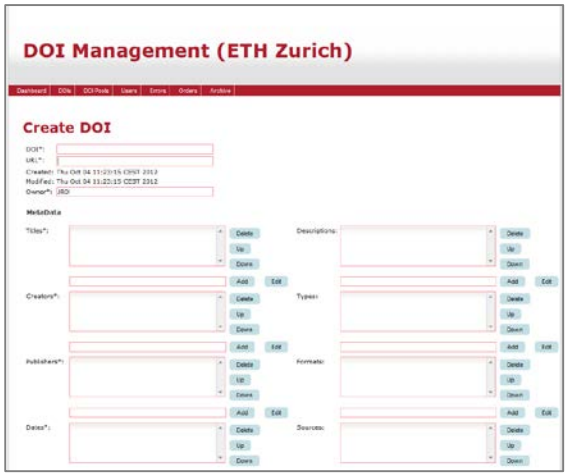
Property	Value
Creator	(:unav)
Title	(:unav)
Publisher	(:unav)
Date	0000
Type	The default value for this field has to be chosen in the DOI Pool configuration from the following list ▪ Audiovisual ▪ Catalog ▪ Collection ▪ Dataset ▪ Event ▪ Image ▪ InteractiveResource ▪ Model ▪ PhysicalObject ▪ Service ▪ Software ▪ Sound ▪ Text ▪ Workflow ▪ Other

A missing creator, title, publisher or date produces an error message in the DOI Dashboard. The alerted client can then update the DOI with the mandatory metadata.

A missing type will not produce an error message, assuming that some clients will opt to not supply type values themselves but always use the default value produced by the DOI Management Tool upon export to DataCite.

6 Instructions for manual registration

	**1 DOI Homepage** Go to www.library.ethz.ch/en/doi-desk → Login for registered customers to access the web interface.
	2 Login Choose your Home Organisation and log in with your usual user name and password. If you do not have an AAI account provided by your institution use the credentials provided by the DOI Desk and login via the “Virtual Home Organisation”.
	3 Dashboard The dashboard shows an overview of all registered and recently updated DOIs.

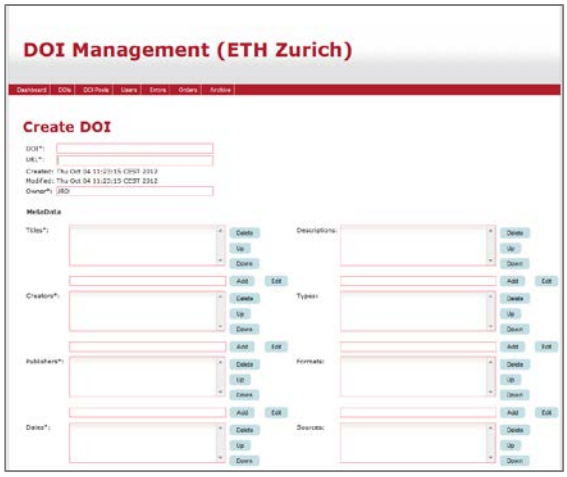


4 Create DOI

Click "Create DOI" in the column “Commands”.

Submit the following information:

- Complete DOI name
- Current URL address



Please also read chapter 5 of this manual which has more detailed information about the meaning and use of the metadata fields!

5 Submit Metadata

The fields *Titles, Creators, Publishers*, *Dates* and *Types* are mandatory. All other fields are optional.
If you would like to submit more than one value for one field (e.g. two or more authors), please add a new row for each value.

Formatting rules:

Creators (=Author): family, given

Dates**:** YYYY, YYYY-MM or YYYY-MM-DD

Languages**:** Language codes from ISO 639-1 (e.g. en, de, fr)

Descriptions**:** This field can be used for submitting information on a series or journal.

Types: One of the following values has to be chosen:

- ✓ Audiovisual
- ✓ Catalog
- ✓ Collection
- ✓ Dataset
- ✓ Event
- ✓ Image
- ✓ InteractiveResource
- ✓ Model
- ✓ PhysicalObject
- ✓ Service
- ✓ Software
- ✓ Sound
- ✓ Text
- ✓ Workflow
- ✓ Other

Click "Submit" to confirm.

6 Update URL

Retrieve your registered DOIs by clicking on the "DOIs" tab in the main menu.

Search for the DOI name and retrieve the entry by clicking on the DOI.

Click „Edit".

Manually change the URL address and / or the metadata.

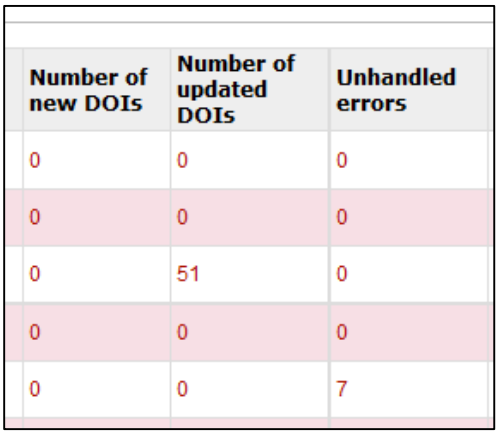
Click „Submit".

7 Search

The menu item „DOIs" leads to a search functionality.

You can e.g. search for DOIs according to the date of registration or modification.

At the bottom right corner you can go to the next page of the results list.



Number of new DOIs	Number of updated DOIs	Unhandled errors
0	0	0
0	0	0
0	51	0
0	0	0
0	0	7

8 Error Monitoring

Please take the time to monitor if your DOI registration has been successful.

Any error caused by wrongly entered metadata will be displayed on the Dashboard in the "Unhandled errors" column.

If you need help in analyzing error messages please contact us at doi@library.ethz.ch.

7 Instructions for the handling of duplicates and deleted objects

In particular cases there might arise a justified need to delete an object that has been attributed a DOI including its Landing Page entirely from the client's publication system[5].
This can be the case for example when a document has to be deleted for copyright related reasons or when an object has been published unintentionally as a duplicate.

Since DOI names should and cannot be deleted at all, the DOI Desk recommends to redirect the DOI name to a Tombstone Page or to another DOI name (in case of duplicates).

7.1 Use cases

7.1.1 Use case 1: An object has been deleted entirely from the client's system (redirect to a Tombstone Page)

If an object referenced by a DOI including the according Landing Page had to be deleted from the client's system, the client has to redirect the according DOI to an information page, a so-called Tombstone Page.

The Tombstone Page is a website that serves the purpose of providing a valid and persistent URL address even after a DOI object has been deleted. It informs that (and why) the object has been removed and provides the person who tried to resolve the DOI name with ways to contact the former publishing agent.

We recommend to use the following Tombstone Page provided by the DOI Desk for this purpose: http://www.doi.ethz.ch/invaliddoi.html.

Any enquiries about deleted DOIs that the DOI Desk receives via this page will be forwarded to the responsible DOI client.

Alternatively a client can, in consultation with the DOI Desk, choose to

- provide their own Tombstone Page for their DOI Pool
- provide an individual Tombstone Page for each invalid DOI

In both cases it is the client's responsibility to guarantee that the URL of their Tombstone Page is persistent.

If an individual Tombstone Page is provided for a single DOI, it should contain information as to why the object is no longer available.
Otherwise the Tombstone Page must at least provide a contact address.

7.1.2 Use case 2: A duplicate has been removed (redirect to another DOI name)

If a record had to be removed from the client's system because it was a duplicate, the client has to redirect the according DOI to the DOI name of the record/object that is still available.

The DOI name of the deleted object will be redirected to the „DOI-URL" of the duplicate object. A "DOI-URL" begins with the prefix http://doi.org/ and ends with the DOI name.

Example: http://doi.org/10.3929/ethz-a-000084852

[5] The system/repository/web interface/database operated by the client; e.g. ETH E-Collection, retro-seals,...

7.2 Instructions

The following instructions apply under the precondition that the affected DOI has been removed irrevocably from the metadata set delivered via the client's interface. This guarantees that the redirection executed via the web interface will not be overwritten during the next harvesting routine.

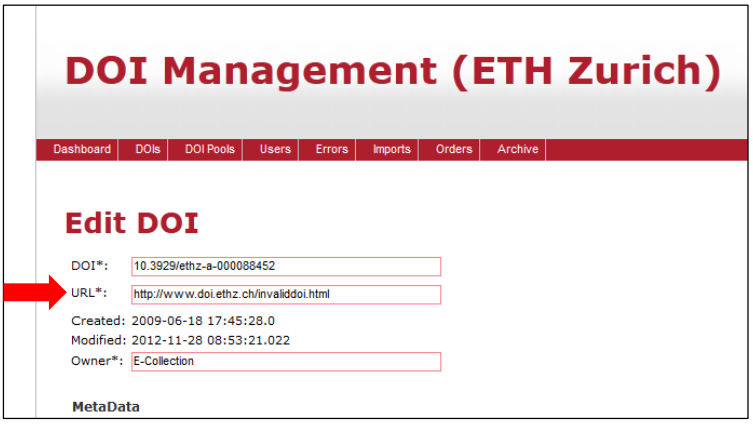
1. Log in to the DOI Management web interface via www.library.ethz.ch/en/doi-desk → Login for registered customers.

2. Search for the inactive DOI via the menu „DOIs".

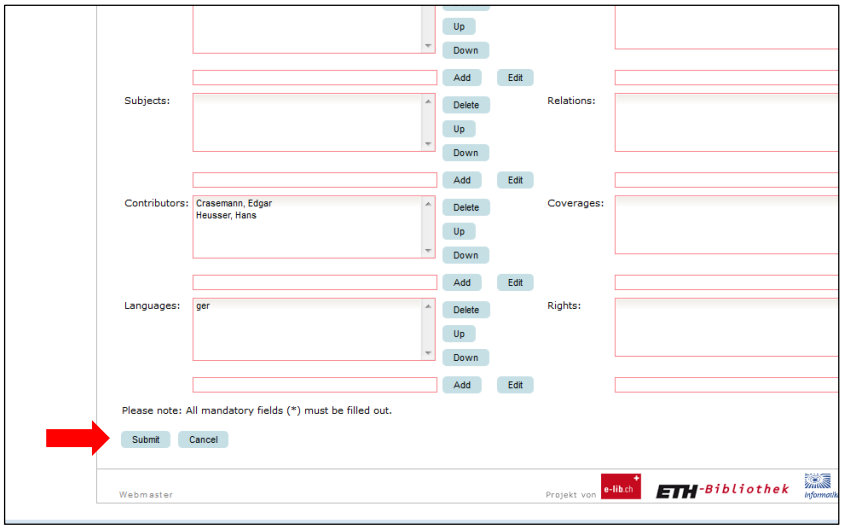
3. Click on the DOI name to show the full record.

4. Click „Edit" to activate the editing mode.

5. Modify the URL by providing
   a. a Tombstone Page, e.g. http://www.doi.ethz.ch/invaliddoi.html
   b. another DOI, e.g. http://doi.org/10.3929/ethz-a-000084852



6. Save the changes by clicking on the „Submit“ button:



According to the *Policy for the registration of Digital Object Identifiers by ETH Zurich* the DOI Desk retains the right to store the metadata belonging to an object even beyond its deletion.

Metadata of deleted objects will neither be deleted from the ETH Zurich DOI Management Tool nor from the DataCite Metadata Store (http://search.datacite.org).