

DataCite Canada

Making research data citable and accessible

Presentation for RDC webinar series
2017-03-23

Karen Morgenroth, Cyndie Found
National Research Council Canada



DataCite consortium

DataCite: Role and mission

- Role:
 - To address the gap between underlying data and the research they support
- Mission
 - To establish easier access to research data
 - To increase acceptance of research data as legitimate contributions to the scholarly record
 - To support data archiving to permit results to be verified and repurposed

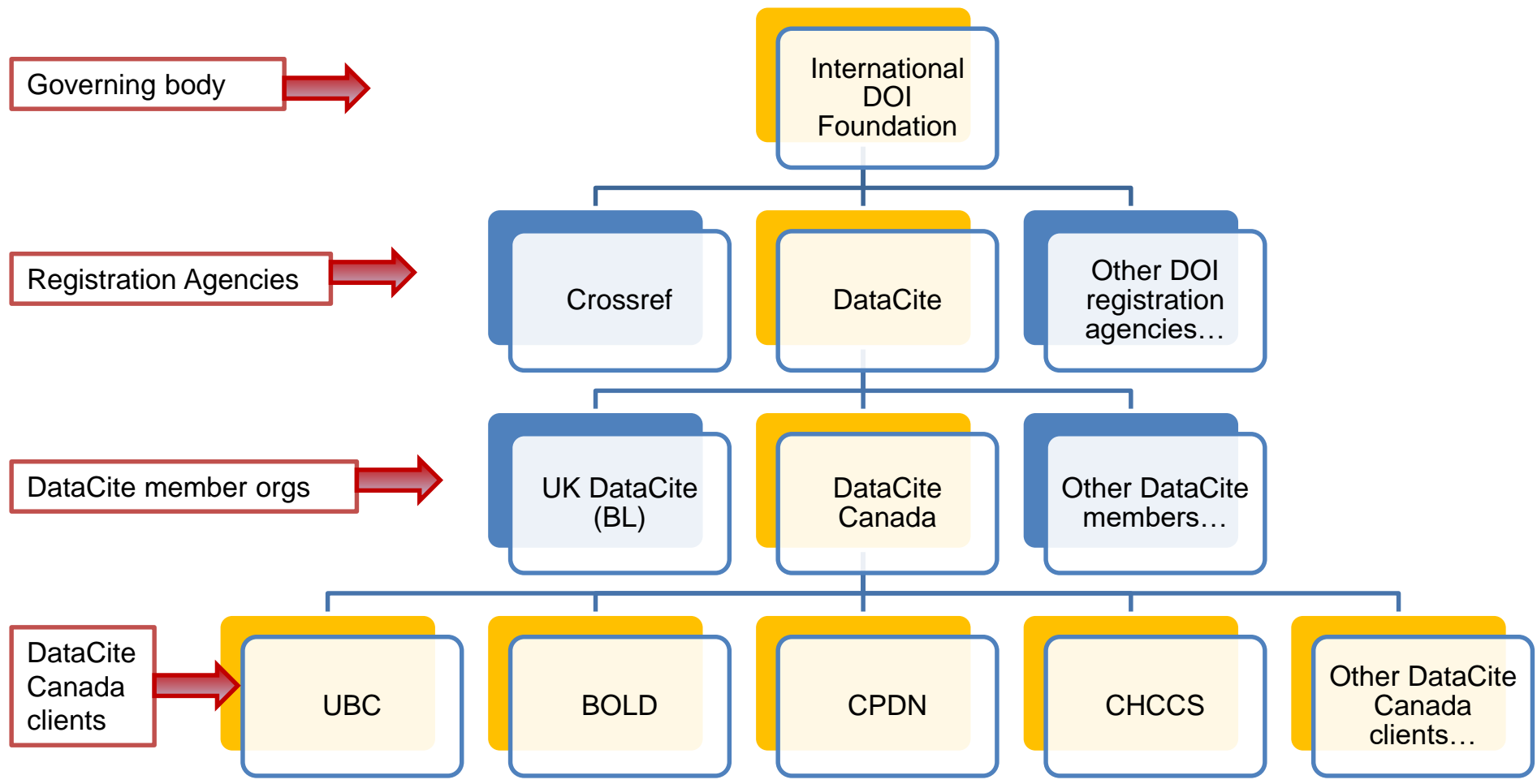
What does DataCite do?

- Maintains the registration and resolution infrastructure
- Maintains a searchable database of metadata
- Establishes and shares best practices

DataCite membership



DataCite and the DOI landscape



DataCite: Integration, collaboration, partnerships

- Data workflow and storage systems (e.g., Dataverse, OJS, DSpace)
- Other registration bodies, identifier systems (e.g., ORCID, Crossref)
- Metrics tools (e.g., Thomson Reuters)
- Best practices for data citation and preservation (e.g., Force11)

DataCite Canada

- Operates through the National Research Council Canada
- Provides
 - Access to DataCite tools and services for minting DOIs
 - Support and guidance
 - A link between DataCite Canada clients and the international research data community

DataCite Canada: Clients

- Canadian, not-for-profit
- Maintain and update metadata
- Store and ensure long-term access to content
- Can mint DOIs for a variety of research objects:

Data sets	Images and video	Physical models
3D models	Theses	Data papers
Small OA journals	Reports	Proceedings

DOIs

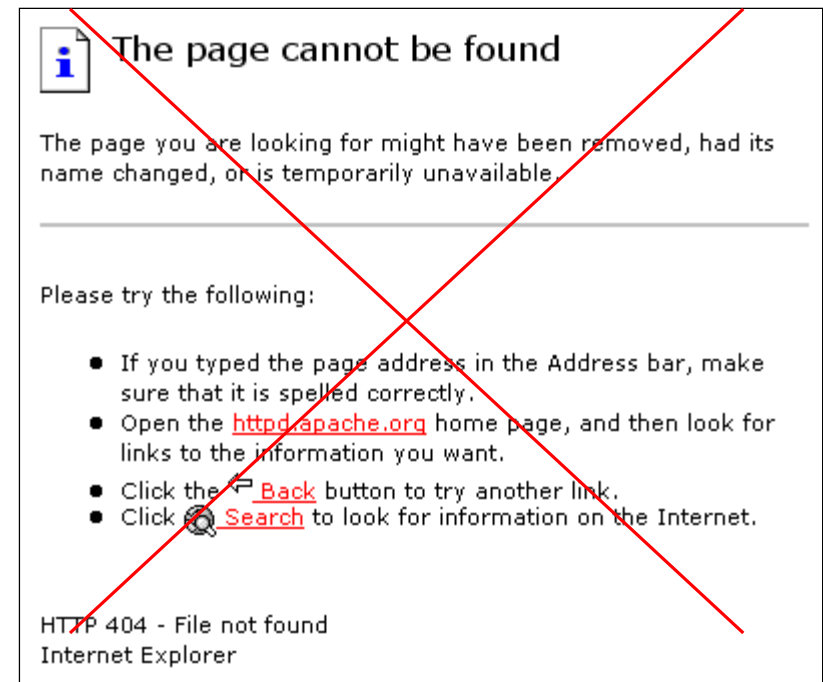


Persistent identifiers (PIDs)

- Long-lasting references to objects
- Digital Object Identifiers (DOIs)

Why DOIs?

- Persistence
- Increased access to data through metadata capture
- Widely used in scholarly publishing
- ISO standard: 26324:2012, Information and documentation—Digital object identifier system





The DOI name

- A serial code that uniquely identifies content in the digital environment
- Provides current information about resources, digital or not
- Consists of a prefix and a suffix

Prefix			/	Suffix
10	.	4224	/	17210689

- Points to a URL

National Re

Programs and service

Home > Publications > NRC

Publications

NRC publications

- Bioenergy Perspective
- Codes Canada publica
- Construction Innovatio
- Construction Technolo
- Energy Storage News
- Canadian National Ma
- Specification (NMS)
- Skygazing
- NRC Publications Arch

National Science Lib

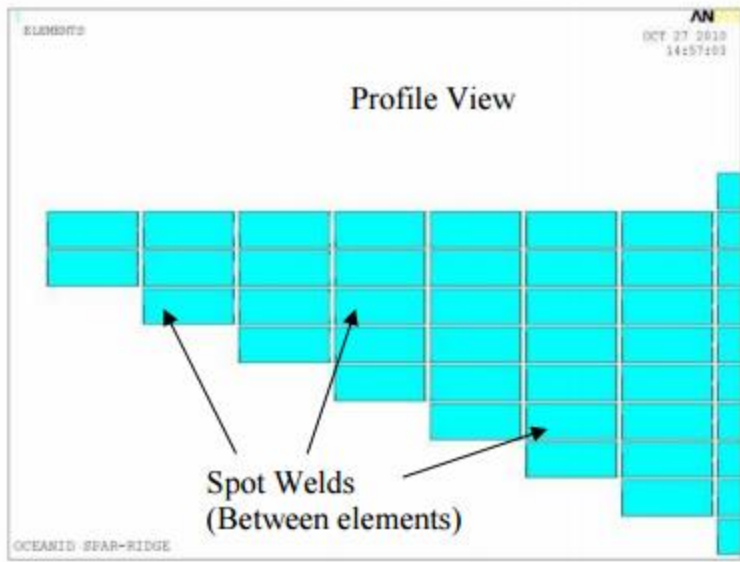
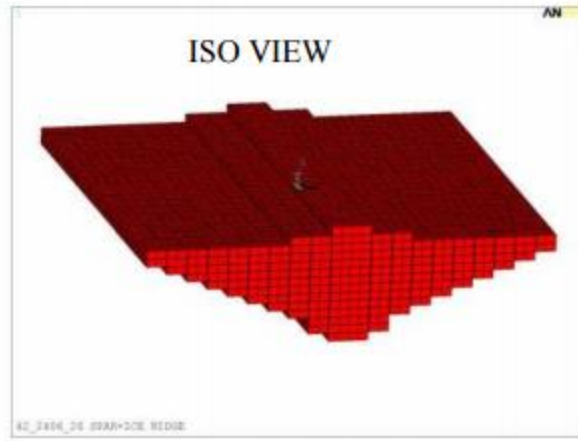
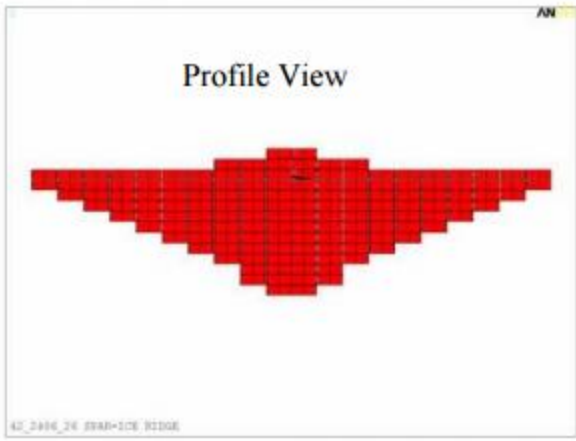


Figure 1: FY Ice ridge configuration

Record identifier 84bae1a3-ca30-4328-9bd3-7ad27ca20700

project
 fiscal year
 Task 1 is to
 (FY) ice
 report next
 tively. The
 cit FEM was
 interactions,

ZA2842: IAB Employee Sample 1975-1990 (Regional File)

Bibliographic Citation | **Content** | **Methodology** | **Data & Documents** | **Errata & Versions** | **Further Remarks**

Publications

Dataset Number of Units: 426914

Number of Variables: 26

Data Type: -

Analysis System(s):

Availability

C - Data and documents are only released for academic research and teaching after the data depositor's written authorization. For this purpose the Data Archive obtains a written permission with specification of the user and the analysis intention.

Download of
Data and Documents

All downloads from this catalogue are free of charge. Data-sets available under access categories B and C must be ordered via the shopping cart. Charges apply! Please respect our Terms of use.



DDI Documents



- Study Description in DDI format DDI-Codebook (2.5)
- Study Description in DDI format DDI-Lifecycle (3.1)



You can order this study via shopping cart
add to shopping cart



General Access to studies and data sets at the GESIS Data Archive for the Social Sciences

Registration metadata

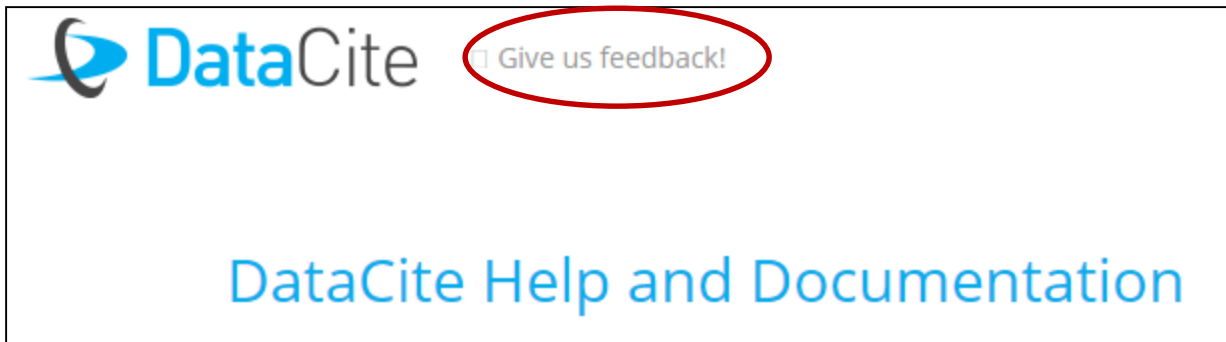
Mandatory	Recommended	Optional
<ul style="list-style-type: none">• Identifier• Creator• Title• Publisher• Publication year• Resource type	<ul style="list-style-type: none">• Subject• Contributor• Date• Related identifier• Description• GeoLocation	<ul style="list-style-type: none">• Language• Alternative ID• Size• Format• Version• Rights

- Current version (4.0): <http://doi.org/10.5438/0012>
- XML examples available

DataCite services

Metadata Store

- Used for minting DOIs and registering metadata
- Two ways to use it:
 - Web interface: <https://mds.datacite.org/>
 - API: <https://datacite.readme.io/docs/api>
- Please visit the new DataCite documentation portal and provide feedback: <https://datacite.readme.io/>



Metadata Store: Web interface

- Used for single or manual operations
- <https://mds.datacite.org/>

▼ Register new Dataset

DOI latency: Be aware that it can take up to 24 hours until a DOI update is globally known. New DOIs should be resolvable after about 5 minutes.

For testing purposes please only use our dedicated test prefix 10.5072

DOI:

Uri:

XML upload: No file chosen

Please select an XML file. It must reference a schema located under the following base URL: <http://schema.datacite.org/meta/>

XML:

Search and discovery



- <https://search.datacite.org/>

A screenshot of the DataCite search interface. The page title is 'DataCite Search'. Navigation links for 'Works', 'People', 'Data Centers', and 'Members' are visible. A 'Sign in' button is in the top right. The search results for 'Arctic shorebird breeding and population monitoring in Kivalliq Region, Nunavut' are displayed. The citation is shown in APA format: 'Rausch, J., Johnston, V., & Martin, J.-L. (2012). Arctic shorebird breeding and population monitoring in Kivalliq Region, Nunavut [Data set]. Canadian Cryospheric Information Network. https://doi.org/10.5443/10622'. A 'Copy to Clipboard' button is next to the citation. At the bottom left, the DOI link 'https://doi.org/10.5443/10622' and a 'Cite' button are highlighted with red circles. A large red circle highlights the citation text and the 'Copy to Clipboard' button.

Citation formatter

- <http://citation.crosscite.org/>

DOI Citation Formatter

Paste your DOI:

For example 10.1145/2783446.2783605

Select Formatting Style:

apa

Begin typing (e.g. Chicago or IEEE.) or use the drop down menu.

Select Language and Country:

en-US

Begin typing (e.g. en-GB for English, Great Britain) or use the drop down menu.

Format

DOI Citation Formatter

Paste your DOI:

For example 10.1145/2783446.2783605

Select Formatting Style:

apa

- academy-of-management-review
- acm-sig-proceedings
- acm-sig-proceedings-long-author-list
- acm-sigchi-proceedings
- acm-sigchi-proceedings-extended-abstract-format
- acm-siggraph
- acs-nano
- acta-anaesthesiologica-scandinavica
- acta-anaesthesiologica-taiwanica
- acta-naturae
- acta-neurochirurgica
- acta-ophthalmologica
- acta-palaeontologica-polonica
- acta-pharmaceutica
- acta-pharmaceutica-sinica-b
- acta-philosophica
- acta-polytechnica
- acta-psychiatrica-scandinavica
- acta-societatis-botanicorum-poloniae
- acta-universitatis-agriculturae-sueciae
- administrative-science-quarterly
- advanced-engineering-materials
- advanced-functional-materials
- advanced-materials
- advances-in-alzheimers-disease
- advances-in-complex-systems
- african-journal-of-emergency-medicine

Statistics

Active filters (clear all): allocator CISTI - National Research Council Canada

Registrations by Allocators

Registrations by Allocators Registrations by Datacentres Registrations by Prefixes Resolutions by Month

December 2016

#	Prefix	Total attempted	Successful	Failed	Total unique DOIs	Unique DOI: successes	Unique DOI: failures	Top 10 DOIs: successes	Top 10 DOIs: failures
48	10.5683 CISTI.OSP	496	494	2	31	29	2	1 10.5683/SP/VoZIFT meta (143) 2 10.5683/SP/DJLHSB meta (128) 3 10.5683/SP/NGMQN X meta (108) 4 10.5683/SP/NGGWE G meta (21) 5 10.5683/SP/XoV8CX meta (14) 6 10.5683/SP/U5EV9 M meta (12) 7 10.5683/SP/GUBPAJ meta (10) 8 10.5683/SP/UPOMF N meta (9) 9 10.5683/SP/VWCDK W meta (7) 10 10.5683/SP/AVBWJ M meta (6)	1 10.5683/SP/DJL meta (1) 2 10.5683/SP/XDZ49 Y meta (1)

CISTI.MACLIB - McMaster University Library	16	0	0	0	16	0	0	100%	0%
CISTI.NRC - National Research Council	4 304	95	27	11	4 302	2	0	100%	9%
CISTI.OSP - Scholars Portal	128	95	31	1	127	1	0	100%	0%
CISTI.OTNDC - Ocean Tracking Network	9	0	0	0	9	0	0	100%	0%
CISTI.PODER - CybelePress	11	2	1	0	11	0	0	100%	0%
CISTI.UAM - University of Alberta Museums and Collections Services (MACS)	7	0	0	0	7	0	0	100%	0%
CISTI.UBC - University of British Columbia	205 695	3 837	116	0	201 723	3 972	0	100%	0%

DataCite Canada client profiles

Client profile: University of British Columbia



Successes	<ul style="list-style-type: none">• DOI minting integrated into Dataverse, DSpace, AtoM and CONTENTdm• Use of persistent identifiers promoted across campus• Development of custom registration interface to simplify small/mid-range batch registration
Challenges	<ul style="list-style-type: none">• Difficult to complete mandatory metadata for older objects (e.g., no clear publication year, creator)• Bulk editing of DOIs time consuming
DOIs minted	205K+

Client profile: Centre for Biodiversity Genomics/ Barcode of Life Database (BOLD)

Successes	<ul style="list-style-type: none">• Integrated DOI registration within in-house systems
Challenges	<ul style="list-style-type: none">• Determining contributors of aggregate datasets assembled by algorithm• Lack of language-specific toolkits that perform batch operations
DOIs minted	31K+

Client profile: NRC Publications Archive

Successes	<ul style="list-style-type: none">• Integrated DOI registration within in-house system• Registration method reusable in multiple collections
Challenges	<ul style="list-style-type: none">• Ability to do quick updates of URLs following migration of the database• Bilingualism and metadata<ul style="list-style-type: none">• No “relationType” value for “translation of”• No easy way to indicate that an object has been created in both official languages
DOIs minted	4K+

Questions?

Contact DataCite Canada at:
NRC.DataCiteCanada.CNRC@nrc-cnrc.gc.ca

Contact DataCite at:
support@datacite.org



Appendices

DataCite Canada: Becoming a client



- Prospective client contacts DataCite Canada via email
- DataCite Canada determines if client meets criteria (e.g., Canadian, not-for-profit), provides client with further info)
- After client decision to proceed, DataCite Canada partnership officer and client work together on agreement
- Once agreement signed, DataCite Canada sets up new account, assigns unique prefix—and the client can begin registering content
- DataCite Canada provides client with services support, documentation

Metadata Store: Web interface general error messages

- **"already exists"**. You have tried to mint a DOI that already exists in the system. Please give a new DOI to this dataset.
- **"XML must be uploaded when minting DOIs"**. You must select the XML document that holds the metadata for your dataset in order to mint the DOI.
- **"DOI does not match"**. The DOI given in your metadata does not match the DOI you have given in the form. Please check that you have typed both correctly and try again.
- **"domain of URL is not allowed"**. You have tried to give a DOI to a landing page outside of your allowed domain(s). To check your allowed domains, go to 'My Account' (top right). To add a new domain to your account, please contact DataCite Canada for assistance.
- **"DOI prefix is not allowed"**. You have tried to mint a DOI to a prefix that is not allowed on your account. To check which prefix(es) you are allowed to use on your account, go to 'My Account' (top right).
- **"may not be empty"**. You must give the URL of the landing page that you want the DOI to point to.
- **"must be a valid URL protocol not allowed"**. You must enter the full URL, including the 'http://'. Please check you have entered the URL correctly, and try again.
- **"Handle not exists"**. If you receive this message while trying to update a DOI it is likely that you have tried to update the DOI too soon after creating it. It takes a few minutes for DOIs to be copied through and recognised in the global registry, and if you try to make changes before that process is finished, you will get an error. Please try to update the DOI again in 10 minutes.
- **"malformed DOI"**. The DOI you have tried to mint doesn't have the correct syntax.

<http://www.bl.uk/aboutus/stratpolprog/digi/datasets/datacitefaq/troubleshooting.html>

Metadata Store: Web interface “xml error:” messages

The XML carrying the metadata for your DOI has to conform to XML standards and specifications given at <http://schema.datacite.org>.

- **“xml error: The entity name must immediately follow the '&' in the entity reference.”** You have an ampersand in your metadata. Make sure you escape or encode any [reserved characters](#) in the body of your metadata. For example, all ampersands in the body of your metadata should be replaced with the text &
- **“xml error: cvc-complex-type.3.2.2: Attribute '[xyz]' is not allowed to appear in element '[abc]”** Check the text in your field tag [xyz], and ensure it is free of typing errors. You may for example have ‘identifierType’ instead of ‘identifierType’
- **“xml error: cvc-complex-type.2.4.a: Invalid content was found starting with element '[xyz]'. One of '{“http://datacite.org/schema/kernel-2.2”:[abc]}’ is expected.”** This message could be caused by one of two problems. 1. Elements in your metadata may not be in the correct order. Check the ordering of your metadata elements against the order in the example metadata given here: <http://schema.datacite.org/>. 2. You may have a typing error in your field tag. Check the text in your field tag [xyz] for errors e.g. you may have “titls” instead of “titles”.
- **“xml error: cvc-enumeration-valid: Value '[something]' is not facet-valid with respect to enumeration '[list of allowed terms]’. It must be a value from the enumeration.”** You have given a Type element a term that is not from the allowed list. There are 8 fields whose content is restricted to a controlled list. These are: contributorType, dateType, descriptionType, identifierType, resourceTypeGeneral, relatedIdentifierType, relationType and titleType. Please check the [metadata schema](#) for the list of allowed terms for these fields.
- **“xml error: cannot find schemaLocation for namespace”.** You have declared the wrong schema in the XML header. Please see <http://schema.datacite.org> for the list of allowed schema.
- **“xml error: Invalid byte [#] of [#]-byte UTF-8 sequence”.** This may occur when your metadata contains accented or special characters, e.g. ü, é or ° and suggests your XML file has not been properly saved in UTF-8 encoding. Re-save the XML file in UTF-8 and try again.
- **“Property xml threw exception; nested exception is javax.validation.ValidationException: ParseError at [row,col]:[2,6] Message: The processing instruction target matching “[xX][mM][lL]” is not allowed”.** You have formatted your XML with spaces between the elements. Recreate your metadata without these and try again.

<http://www.bl.uk/aboutus/stratpolprog/digi/datasets/datacitefaq/troubleshooting.html>



Building DOI names

- DataCite recommends using only the following characters in DOI names (<https://mds.datacite.org/static/apidoc>):

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

- The DOI suffix:
 - Is assigned by clients
 - Must be unique to its prefix
 - Is flexible: can be a sequential number or a legacy identifier; can contain human-readable parts
 - Can have many nodes, separated by a dot
 - Can be any length; for practical reasons, a limited number of characters is recommended