BRD Implementing DataCite DOI support for non-EZID customers in Dataverse

Document Approval	
Requirements Gathering Deadline	**January 7, 2016**
Approved by Executive Committee & Stakeholders	E.C., January 13, 2016
Update 02/09/2016 In addition to adding DataCite DOI support we are also working on sending additional metadata to both EZID and DataCite in order to make these datasets more discoverable. Phase 1 of the DataCite metadata extension is being tracked here. Also added an "Out of Scope" section below.	

Business Requirement Description

According to DataCite many of their member Data Centers are interested in using Dataverse but they use the main DataCite MDS API (https://mds.datacite.org/static/apidoc) to mint their DOIs rather than through EZID which is what we currently only support. Furthermore, we have several other Dataverse installations (Heidelberg University, Scholars Portal, and University of Alberta) which would like to migrate to 4.x but cannot or do not use EZID. To support these users we will need to add the ability in the Dataverse application to register DOIs through the DataCite MDS API. This is not unprecedented: Peking University has added this functionality in their 4.0 installation and DSpace has added this functionality in their system as well, which is documented here and DataCite has an MDS test instance for us to try it out ahead of time http://test.datacite.org/.

Assumptions, Dependencies, Constraints & Risks

Reserving" DOIs is an EZID functionality not supported by the DataCite MDS API (since EZID never calls the MDS, but does this internally). This is functionality that Dataverse would have to do internally when talking directly to the DataCite MDS. This should be straightforward as a data center normally has full control over every DOI used with a specific DOI prefix. "Reserving" DOIs in the strict sense doesn't exist, it is just a mechanism to ensure we are not using the same DOI twice for different datasets. The easiest implementation would be if Dataverse would generate a UUID internally instead of reserving with EZID, but most data centers probably want DOIs that follow a particular pattern. This should be implemented in some general way for both the EZID API and the DataCite API (as well as extendable to support Handles later).

Peking University (contact: luopc@lib.pku.edu.cn) has setup some integration with DataCite MDS in their Dataverse 4.0 installation (http://opendata.pku.edu.cn/), according to Stephen this is not documented in their public codebase, so we will want to coordinate work with them as

soon as possible to make sure there is no duplication of effort. Peking's code: https://github.com/pengchengluo/Peking-University-Open-Research-Data-Platform

Out of scope: Certain Dataverse installations (Scholars Portal) who have multiple institutions, with each of them having their own DataCite DOI prefix, will need to be able to select from a list of multiple prefixes (set at the Dataverse level) to register DOIs for their datasets (tracked here).

Stakeholders: Executive Committee, DataCite Community (Martin Fenner)

Resources:

- (1) Developer for backend work. Stephen Kraffmiller had originally setup the EZID DOI support so would be the most knowledgeable developer to work on this.
- QA Kevin Condon to test functionality

High-Level Effort Estimates

According to Stephen, if Peking's code matches our business and technical requirements this should take less than a week to be completed. Otherwise, this could take 1-2 days to design the FRD and about 2 weeks of development effort.

External deliverable deadlines: Have FRD completed and report progress and estimated delivery date at DataCite Members Meeting February 25-26, 2016.

References:

https://github.com/IQSS/dataverse/issues/24
https://github.com/IQSS/dataverse/issues/2917
https://github.com/IQSS/dataverse/issues/2943
https://github.com/pengchengluo/Peking-University-Open-Research-Data-Platform
http://test.datacite.org/
https://wiki.duraspace.org/display/~pbecker/DOI+support+using+DataCite