Notes on Data Publishing in JLDG

JLDGOP

Toshiyuki Amagasa, Tomoteru Yoshie, and Hideo Matsufuru

October 17, 2022 ver.1.0

Abstract

This document explains the DOI registration process for Lattice QCD ensembles in Japan Lattice Data Grid. While this note is originally prepared as a contribution to the Notes on Data Publishing in ILDG, we also provide it as a separate document.

4 Data Publishing Practice in JLDG

JLDGOP, the administration group of Japan Lattice Data Grid [1], started investigations toward the DOI publication of QCD ensembles on JLDG in 2015, following the consensus formed in the ILDG workshop held in April 2015. In Japan, DOI registration is managed by Japan Link Center (JaLC) [2]. JaLC has specified "Guidelines for Registering DOIs for Research Data" [3] that describes conditions required for DOI registration of research data. JLDGOP decided to make the DOI registration business to be established as an activity of Joint Institute for Computational Fundamental Science (JIC-FuS) [4], and the DOI Preparation Subcommittee launched under the JICFuS Steering Committee in 2016.

The DOI Preparation Subcommittee formulated two policy documents: "DOI registration policy for public QCD data", and "JLDG Public Data Management Policy with DOIs", which are summarized in the following paragraphs. These policies have been approved by the JICFuS Steering Commit-

tee, followed by the establishment of the DOI Operations Subcommittee[1], which started official registration service in October 2018. The Center for Computational Science, University of Tsukuba [5] is registered as a member of JaLC to apply for DOI registration under the above policies and handles the registration of DOIs. The first DOI-registered QCD ensemble on JLDG is a set of configurations on $32^3 \times 64$ with 2+1 flavors generated by PACS-CS Collaboration [6]. The latest list of DOI registered QCD ensembles is available in [7].

Persistent Access "JLDG Public Data Management Policy with DOIs" specifies the measures to guarantee the persistent access to public QCD data as follows.

1. The JICFuS DOI Operations Subcommittee operates at least one additional system, independent from the main one. The system is made accessible from the Internet to store backups of public QCD data, and provides access to the data in the event that the services of the JLDG's main file sharing system are disrupted for any reason.

2. If the JLDG Administrators Group dissolves for any reason, it delegates the management of public QCD data to some organization in consultation with the relevant organizations.

Registration flow The flow of DOI registration is formulated in "DOI registration policy for public QCD data". The registration process is started by the action of "Contributor", a representative of the "Public QCD Data Creator" who 1) is willing to publish lattice QCD data through ILDG and 2) wishes to register DOIs for public QCD data. The "Contributor" needs to be a JLDG user. The QCD data for DOI-registration are assumed to be already public on ILDG through JLDG. The landing page and DataCite XML are automatically generated from the ILDG Ensemble XML [8] and

---

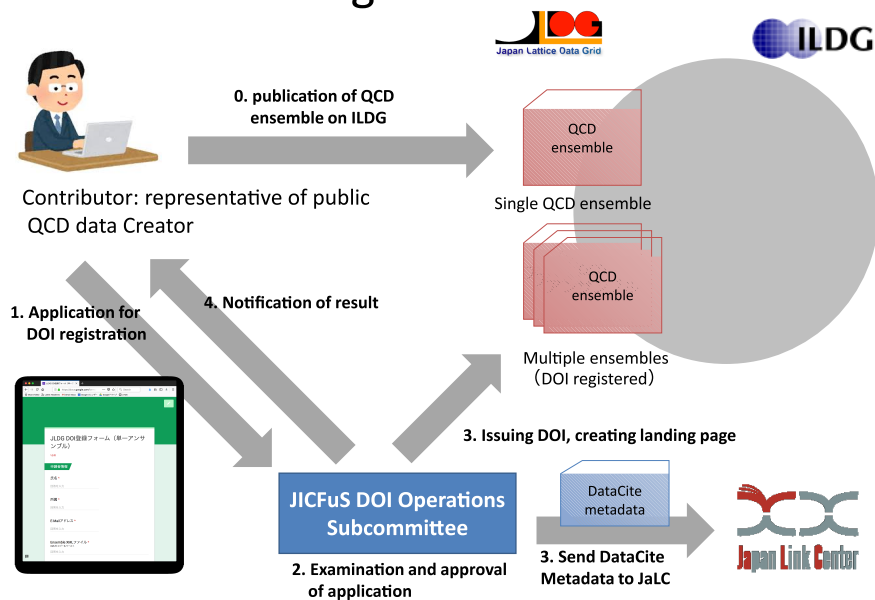[1]The members at the time of July 2022 are Toshiyuki Amagasa, Tomoteru Yoshie, and Hideo Matsufuru.

Figure 1: Flow of DOI registration for QCD data on JLDG.

the information provided at the application. The basic unit for registering DOIs is the QCD ensemble. It is also possible to register a new DOI for a set of the QCD ensembles for which DOIs have been already registered. In this case the license conditions for the original QCD ensembles are inherited.

The DOI registration is processed along the following flow (Fig. 1).

1. **Contributor:** representing Public QCD data creators, submits an application for DOI registration of public QCD data to the JICFuS DOI Operations Subcommittee, together with the information specified in Table 1.

2. **JICFuS DOI Operations Subcommittee:** if the application is accepted, 1) assigns a DOI suffix and create DataCite metadata for DOI registration, 2) creates a landing page for public QCD data, 3) applies for DOI registration to JaLC, 4) notifies the applicant of the completion of the registration when the registration is completed. If the application

For DOI registration for a single QCD ensemble	
Name	required
Affiliation	required
E-Mail address	required
Ensemble XMLfile	required
Data set creator	required
Reference	required
Acknowledgment (funding information, etc.)	
License information	required
Remarks (Contact information, etc.	
For DOI registration for multiple data sets	
Name	required
Affiliation	required
E-Mail address	required
Public QCD data DOI(s) for registration	required
Data set creator	required
Reference	required
Acknowledgment (funding information, etc.)	
Remarks (Contact information, etc.	

Table 1: List of information to be submitted when applying for registration of DOIs using public QCD data.

is not accepted, the applicant is notified that.

Technical details The automatic generation of the DataCite metadata and the landing page is processed as follows. From the target ensemble XML and the CSV record containing auxiliary information (data creator, bibliographic references, research funding, license, etc.), 1) the CSV record is converted to XML format (hereafter called JLDG metadata). 2) DataCite metadata are generated from ensemble XML and JLDG metadata using XSLT (XSL Transformations). While any XSLT implementation can be used, we use Saxon [9]. We also get the list of LFNs (Logical File Names) using ILDG MDC (Meta Data Catalog). 3) The landing page is generated from the DataCite XML and the list of LFNs using XQuery (implemented as a part of Saxon).

DOI format In JLDG, the DOI suffix for each ensemble is determined with the following format:

(Collaboration name).(serial number of 6 digits)

Detailed parameters are not reflected in the suffix. The collaboration name is a mandatory record in the QCDml [8]. For multiple ensemble DOI, the suffix of the form:

(Collaboration name).set(serial number of 4 digits)

is assigned.

References

[1] Japan Lattice Data Grid, `https://www.jldg.org/`.

[2] Japan Link Center, `https://japanlinkcenter.org/`.

[3] Japan Link Center Joint Steering Committee, "Guidelines for Registering DOIs for Research Data" (2015), `DOI:10.11502/rd_guideline_en`.

[4] Joint Institute for Computational Fundamental Science, `https://www.jicfus.jp/`.

[5] Center for Computational Science, University of Tsukuba, `https://www.ccs.tsukuba.ac.jp/`.

[6] Lattice QCD gauge ensemble: JLDG/PACS-CS/RCNF2+1/RC32x64_B1900Kud01370000Ks01364000C1715, `DOI:10.34845/PACS-CS.000001`.

[7] Lattice QCD gauge ensemble: Public dataset DOI index, `https://www.jldg.org/DOI/`

[8] International Lattice Data Grid for Metadata Working Group, QCDml, `https://www2.ccs.tsukuba.ac.jp/ILDG/`.

[9] SAXONICA, `https://www.saxonica.com/welcome/welcome.xml`.