

International Lattice Data Grid: Part I

Hideo Matsufuru for ILDG and JLDG

Computing Research Center,
High Energy Accelerator Research Organization (KEK)
and
Graduate University for Advanced Studies (SOKENDAI)



Contents

- **Contiguous talks by HM and Hubert Simma report**
 - General introduction to ILDG
 - Activity toward ILDG 2.0
 - Metadata aspects
 - Middleware aspects (HS)
 - Status of regional grid: JLDG (HM) and LDG (HS)

Part I.

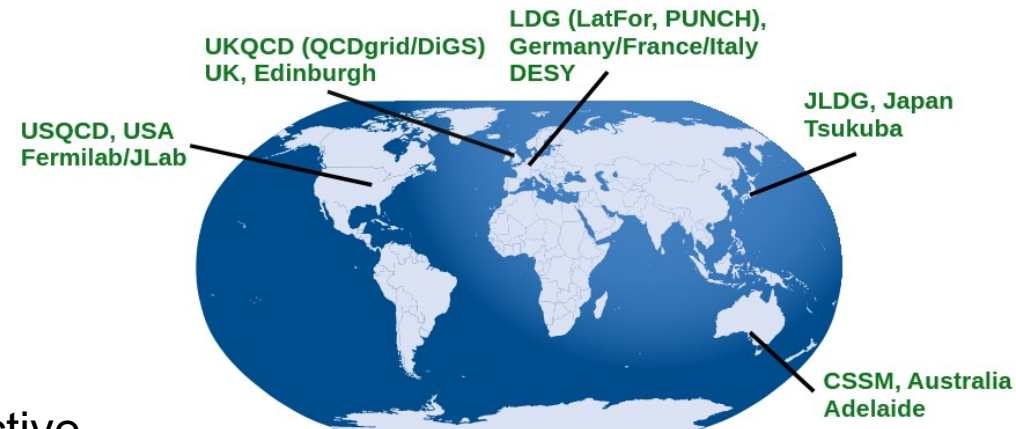
- **Introduction to ILDG**
 - Organization and history
 - Rebooting ILDG under FAIR principle
 - Activity of Metadata Working Group (MDWG)
 - Introduction to QCDml
- **Status of JLDG**



Introduction to ILDG

<https://hpc.desy.de/ildg/>

- **International Lattice Data Grid**
 - **Community-wide initiative to share primary data (configuration)**
 - Proposed in Lattice 2002
 - Implemented in successive years
 - Board and two working groups for Metadata and Middleware
 - Cf. M.G. Becket et al., Comput. Phys. Commun. 182 (2011) 1208
 - **Common format for metadata and configuration files**
 - QCDml (markup language) for ensemble and configuration
 - Standard binary data format with LIME
 - Metadata catalog for search
 - **Middleware: Central Authentication and standardized interfaces (API)**
 - **Grid of grids**
 - Autonomous Regional Grids, each is operated in its own (and ILDG compatible) way
 - Implements storage, grid file system, etc., and provide services
 - 5 regional grids
 - Recently LDG and JLDG are most active





Introduction to ILDG

- Organization (→ <https://hpc.desy.de/ildg/>)
- Board
 - Strategic and organizational matters, directioning of Working Groups
 - Two representatives of each RG and conveners of WGs:
 - Y. Aoki, F. Di Renzo, R. Edwards, F. Karsch (former chair), Y. Kuramashi (chair), D. Leinweber, H. Matsufuru (ex-officio), C. McNeile, A. Portelli, H. Simma (ex-officio), J. N. Simone, F. Stokes
- Metadata Working Group (MDWG)
 - Update of metadata schema, support for data publishing
 - G. Andronico, J. Hettrick, G. von Hippel, G. Koutsou, H. Matsufuru (convener), Y. Nakamura, D. Pleiter, H. Simma, J. Simone, C. Urbach, T. Yoshie
- Middleware Working Group (MWWG)
 - Technical aspects of services, infrastructure, user tools
 - Basavaraja BS, O. Kaczmarek, G. Koutsou, D. Pleiter, A. Rago, H. Simma (convener), J. Simone, T. Yoshie

(Members as of Sept. 2024)



Activity toward ILDG 2.0

- Active discussion had resumed in 2022 toward ILDG 2.0
 - F. Karsch, H. Simma, T. Yoshie, PoS LATTICE2022 (2023) 244
 - F. Di Renzo, PoS LATTICE2023 (2024) 112
- Restructuring based on FAIR principles
 - FAIR guiding principles (Wilkinson et al. (2016)) → Hubert's talk
 - Findable
 - Accessible
 - Interoperable
 - Reusable
- Metadata
 - Latest documents available at <https://gitlab.desy.de/ildg/mdwg>
 - Enable markup of new features of recent simulations
 - Specification of schemata of 2.0 are almost fixed, will be available soon
 - Update of data format
 - Enabling more flexible binary data format
 - Data publishing
 - under discussion
 - Prototypes by USQCD and JLDG



Activity toward ILDG 2.0

- **Middleware**
 - Hubert's talk
- **Sharing more ensembles**
 - Data session in Lattice conference
 - Lattice 2022: Cf. G. Bali et al., PoS LATTICE2022 (2023) 203
 - Held at Lattice 2024 (<https://conference.ippp.dur.ac.uk/event/1265/>)
 - Hands-on workshop
 - June 2023 <https://indico.desy.de/event/39311/>
 - participants from 12 collaborations, requests of updating schemata
 - Planned in early 2025



QCDml Concepts

- For interoperability,
 - Data are stored in standardized format
 - Metadata are essential – marked up in searchable format
- Metadata for ensemble and configuration
 - XML: extensible and hierarchical, both human and machine readable
 - Rigorous schema definition (XSD)
 - Powerful query language (XPath)
 - W3C standards
 - QCDml (markup language)
 - Ensemble XML
 - Configuration XML
 - Defined as XML schemata (→ <https://gitlab.desy.de/ildg/mdwg/>)
 - machine readable specification
 - Current public versions: Config: 1.3.1, Ensemble 1.4.8
 - Documents
 - Hands-on slide: <https://indico.desy.de/event/39311/>
 - C.M. Maynard and D. Pleiter, Lattice 2003 (hep-lat/0409055)
 - P. Coddington et al., PoS(Lattice 2007) 048



Introduction to QCDml

- How it looks like – Example of Ensemble XML file (based on Ver.1.4)

```
<?xml version='1.0' encoding='UTF-8'?>
<markovChain xmlns="http://www.lqcd.org/ildg/QCDml/ensemble1.4"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.lqcd.org/ildg/QCDml/ensemble1.4 QCDmlEnsemble1.4.xsd">
  <markovChainURI>mc://JLDG/JLQCD/Nf20verlap/ensemble1</markovChainURI>
  <management>
    <collaboration>JLQCD</collaboration>
    <projectName>two flavor overlap</projectName>
    <archiveHistory>
      <elem>
        <revision>1</revision>
        <revisionAction>add</revisionAction>
        <participant>
          <name>H.Matsufuru</name>
          <institution>KEK</institution>
        </participant>
        <date>2008-03-27T21:39:43+09:00</date>
        <comment/>
      </elem>
    </archiveHistory>
  </management>
  <physics>
    <size>
      <elem>
        <name>X</name>
        <length>16</length>
      </elem>
    </size>
  </physics>
</markovChain>
```




Introduction to QCDml

- Metadata content (Cf. Hands-on slide)
 - Physics
 - Description of action and parameters
 - Algorithm
 - Information about the used algorithm and algorithmic parameters
 - Source code
 - Used code, compile parameters, compilation software
 - Machine
 - Machine used to produce the configuration
 - Data Management
 - Information related to data provenance and checksum
- More examples are available in <https://gitlab.desy.de/ildg/mdwg/>



QCDml 2.0

- QCDml ver.2.0 is almost fixed, will be released soon
 - New xsd (schema) files and documents are in preparation
- New features in ver.2.0
 - Enabling markup of **multiple configs in single config. XML**
 - Enabling **additional info** as top-level element (supposing collaboration specific)
 - **Annotation** for detailed explanation of parameters
 - Extension of actions
 - **QCD + QED** (photon action, coupling to photon)
 - New actions (Moebius domain-wall, exponential clover)
 - More gauge groups (for BSM)
 - Higher fermion representations
 - Enabling more boundary conditions (open, openSF)
 - Allowing usage of ORCID in <participant> element
 - **License information**: reuse conditions
 - <license> now becomes mandatory element
 - Funding information
- Several updates are not backward-compatible
 - need conversion of previous metadata



Japan Lattice Data Grid

Cf. T. Amagasa et al., J. Phys. Conf. Ser. 664 (2015) 042058

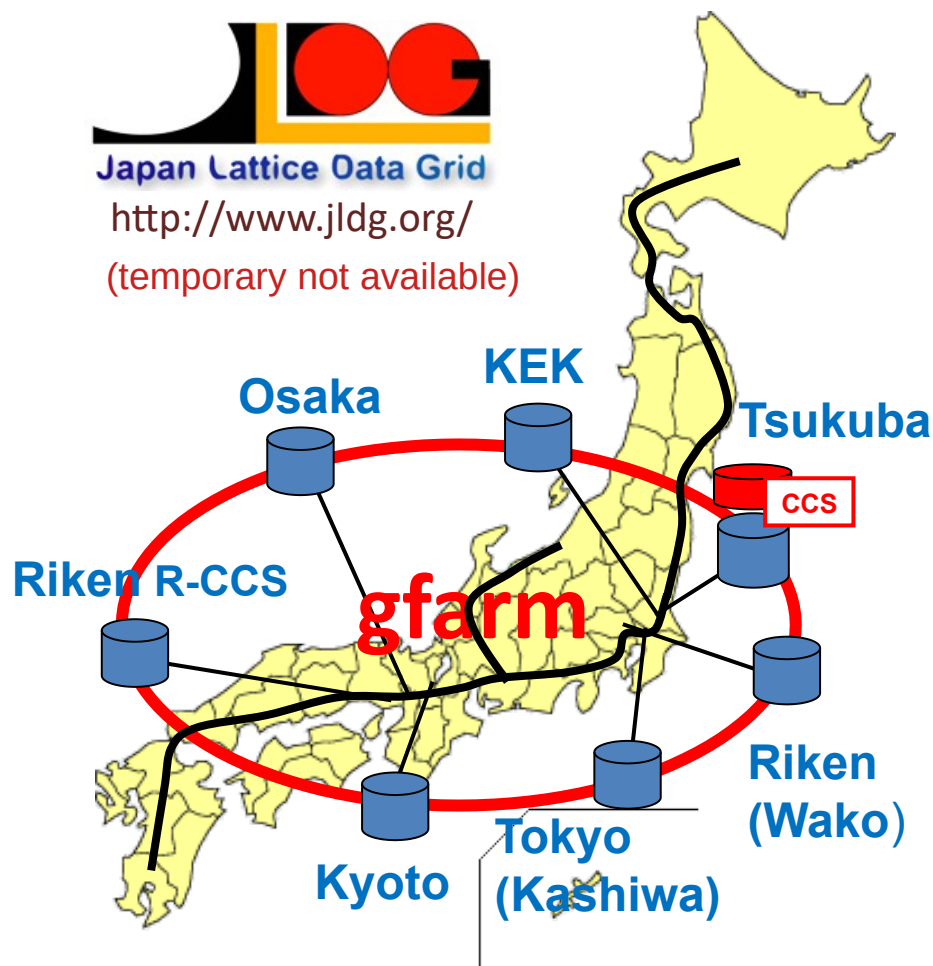
- Data Grid for lattice QCD and related science (nuclear/astrophysics)
 - Sharing public data and fast data transfer within collaborations



Japan Lattice Data Grid

<http://www.jldg.org/>

(temporary not available)

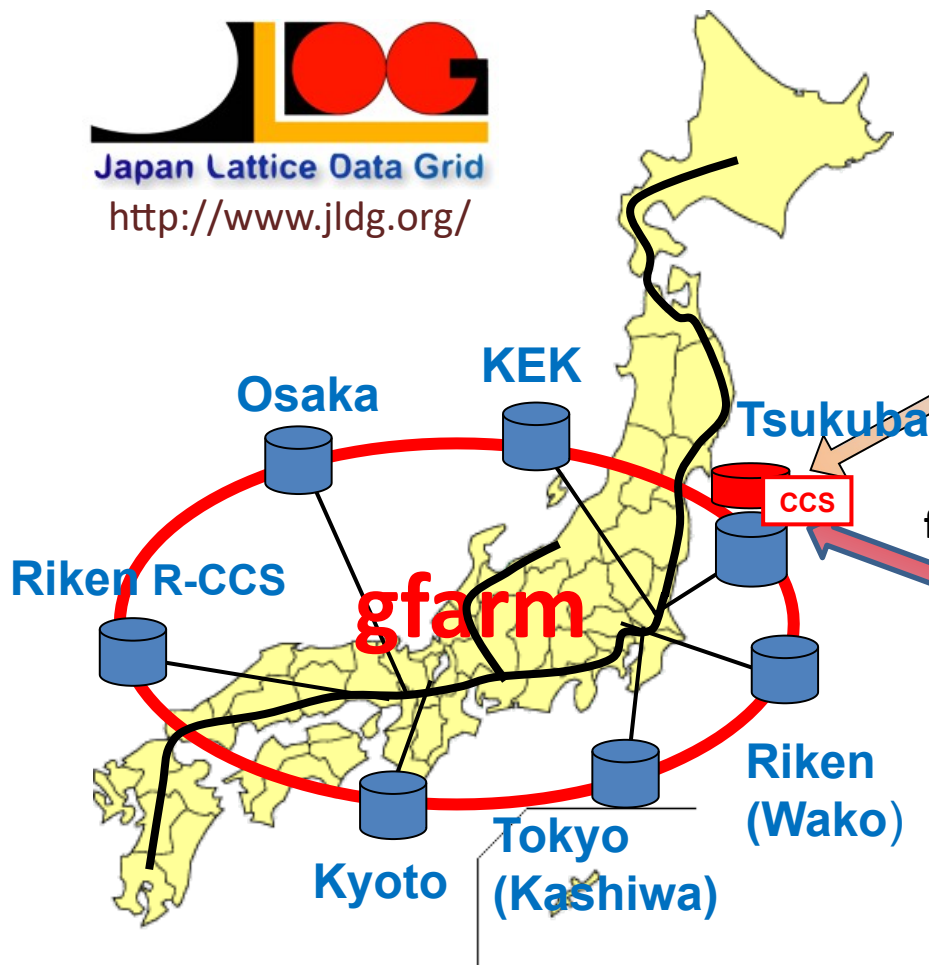


- Operation officially started in May 2008
- Constructed on VPN HEPnet-J/sc
 - On SINET operated by NII (National Inst. of Informatics)
- Gfarm grid file system
 - User can access as if single file system
 - Automatic replication
- Currently 7 sites
- Operated by JLDG team (members from sites and groups)
- Resources (as of Aug 2024)
 - Total storage: 21 PB, 11.3 PB used (87%)
 - Total number of files: >200M
- Regional grid of ILDG



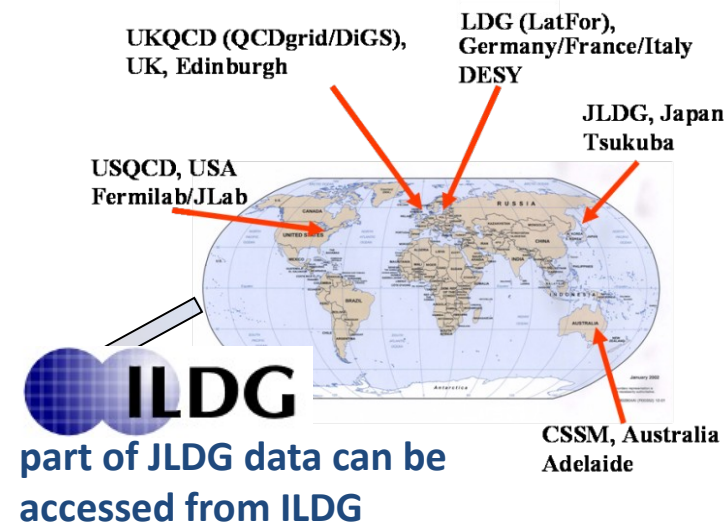
Japan Lattice Data Grid

- Cooperated systems

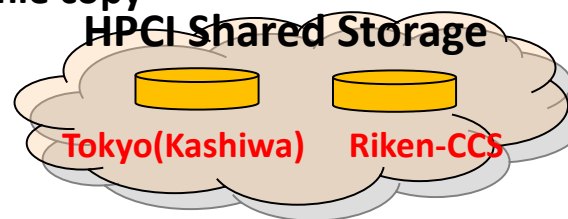


ILDG

- Connected to ILDG at U. Tsukuba



fast parallel file copy



HPCI shared storage

- Cooperated at Tsukuba and RIKEN RCCS
- JLDG is available with HPCI certificate



Japan Lattice Data Grid

- **JLDG Team**
 - Members from sites and groups
 - Members attending meeting regularly:
 - H. Ohno, T. Amagasa, O. Tatebe, T. Yoshie (Tsukuba), H.M. (KEK), T. Doi (RIKEN Wako), H. Togashi, H. Togawa (Osaka), Y. Nakamura (RIKEN Kobe)
- **Activity of JLDG team**
 - Regular meeting once per month
 - Maintenance of sites, resource, users, security, etc.
 - Preparation to token-based authentication
 - Prototype setup of lattice data for data publishing and DOI registration
 - Register DOI for public ensembles of gauge configurations
 - DOI Operations Subcommittee started service in Oct 2018
 - Activity concerning ILDG
 - Contribution to Board / MDWG / MWWG
 - Software development
 - Faceted navigation (T. Amagasa)
 - Markup tool (HM)

backup



Introduction to QCDml

- XML schema is easily parsed by machine → validation of XML
 - But hard to read for human: user friendly tool is desirable
 - Example of markup tool: <https://www.jldg.org/QCDml/>

QCDml Markup Tool

Version 0.9.5++ (4 Dec 2023 candidate for ver.2.0.0)

This tool provides web-based markup for ensemble and configuration XML files based on the QCDml schema, for those who do not want to read an XML schema.

- ▶ Usage:
- ▶ Caution:
- ▶ To do:
- ▶ Useful links

Schema file QCDmlConfig1.3.1.xsd loaded
XML file config_next/c-annotation-algo.xml loaded

```
<gaugeConfiguration xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
  <management >
    ◀ insert <revisions> ? yes /no 
    ◀ insert <reference> ? yes /no 
    <crcChecksum> input ▷  </crcChecksum>
    <archiveHistory>
      <elem >
        ◀ insert <revision> ? yes /no 
        <revisionAction> generate modify </revisionAction>
        <participant >
          <name> input ▷  </name>
          <institution> input ▷  </institution>
        </participant>
        <date> 2023-10-19T11:53:18 modify </date>
        ◀ insert <comment> ? yes /no 
      </elem >
    ◀ insert <elem > ? yes /no 
  </archiveHistory>
</gaugeConfiguration >
```



Activity of JLDG

- DOI registration of data
 - Register DOI for public ensembles of gauge configurations
 - Based on the consensus at ILDG workshop in Apr 2015
 - DOI Operations Subcommittee started service in Oct 2018
 - Established under JICFuS (Joint Institute for Computational Fundamental Science)
 - U. Tsukuba CCS is a member of JaLC (Japan Link Center)
 - Landing page and persistent server at Tsukuba
 - DOI is assigned to the data already public through JLDG

Flow of DOI Registration in JLDG

