The International Lattice Data Grid (ILDG 2.0)

ILDG

Lattice 2023 - Fermilab

Francesco Di Renzo - UNIPR & INFN Parma
on behalf of ILDG Board and Working Groups - [credits: H. Simma]

INFN Istituto Nazionale di Fisica Nucleare

UNIVERSITÀ
DI PARMA

August 4, 2023

Overview

1. Motivation for and Basics of ILDG

2. Progress since Lattice 2022

3. Outlook

ILDG Why at all?

Gauge configurations in Lattice simulations are valuable and expensive in terms of
- computing resources (i.e. energy, CO_2, tax payer's money)
- human effort

ILDG goals

➜ Make these precious raw data sharable, usable, and citable for the community

➜ Ensure basic quality standards for Lattice data

➜ Provide a framework for the organization and putting into practice of solid data management (-plans)

It is fair to say that we have been pioneering the FAIR principles!
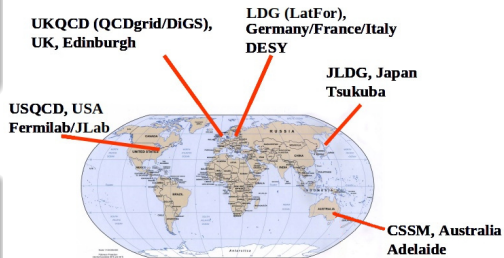. . . and Open Science/Open Data paradigms!

Findable
Accessible
Interoperable
Reusable

Basic Concepts of ILDG

ILDG

- Federation of autonomous "Regional Grids" (RG)
- ... forming a single Virtual Organisation (VO)
- Community-wide agreed standards (QCDml metadata schema, data format, APIs)

UKQCD (QCDgrid/DiGS),
UK, Edinburgh

LDG (LatFor),
Germany/France/Italy
DESY

JLDG, Japan
Tsukuba

USQCD, USA
Fermilab/JLab

CSSM, Australia
Adelaide

Regional Grids

- Organized with individual policies
- ... and implemented with different architectures and technologies
- Autonomous operation of services: Metadata and File Catalogue, Storage Elements, Web Page

Status and Plans at Lattice 2022

The 39th International Symposium on Lattice Field Theory (Lattice 2022)

plenary talk and a lunch meeting on ILDG (and parallel session on Lattice Data)

❏ Increasing interest and need for community-wide data sharing, implementing FAIR principles (Findable, Accessible, Interoperable, Reusable)

❏ Increasing emphasis on Open Science/Open Data paradigms

❏ ILDG 1.0 had severely degraded (services offline or usable only by experts)

❏ Start community effort to re-activate and modernize ILDG (organization, technology, usability, . . . i.e. back to work, together!)

❏ Leverage support and funding of national initiatives for ILDG 2.0 e.g. in Germany (PUNCH4NFDI) and Italy (INFN)

Current Status: Operation of Regional Grids

Resuming services of regional grids

- JLDG (Japan): operational   60 ensembles ($\approx$ 40 k configs) from ILDG 1.0
- LDG (Continental Europe): operational   259 ensembles ($\approx$ 300 k configs) from ILDG 1.0
- UK: in progress
- US: planned

✔ Regional Grids (can) use same Metadata and File Catalogues
- completely re-factored design (through PUNCH4NFDI funding)
- containerized and easy to deploy
- standardized REST API
- support RG-specific configuration and (write-)access policies

PUNCH4NFDI

✘ All regional grids are currently short in available storage space! 🙁

ILDG Hands-on Workshop

- ❒ Hands-on Workshop (June 2023):
    - High number of participants (37 from 12 collaborations)
    - Successful exercises: search, download, markup, upload
    - First test for new middleware (services, clients, containers)
    - Valuable feedback for further development and to support/enable major data providers
    - Material available on https://indico.desy.de/event/39311 (instructions and examples) and https://gitlab.desy.de/ildg/hands-on (user environment and low-level tools)

- ❒ Single contact email address for ILDG (developers, Working Groups, and Board)
    - ✉ ildg-contact@desy.de

- ❒ Information to be collected on web page https://hpc.desy.de/ildg (back in operation! content to be improved, URL might move)

Regular meetings of

☐ **Metadata Working Group (MDWG)**
  - agree on community-wide standards for data description
  - balance scientific with technical requirements
  - specify metadata schema (QCDml) and data formats

☐ **Middleware Working Group (MWWG)**
  - specify interfaces of services to ensure interoperable RGs
  - support RGs by exploring technologies and sharing expertise
  - suggest or develop prototypes of user tools
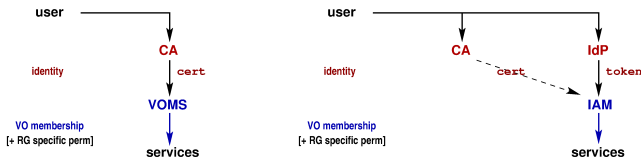
☐ **Board**
  - represent ILDG towards community and service providers
  - decide policies and guidelines for membership and data sharing
  - support regional grids in applying for resources
  - oversee working groups

T. Amagasa, B. B.Sagar,
C. DeTar, B. Joo,
C. McNeile, O. Kaczmarek,
G. Koutsou, H. Matsufuru,
Y. Nakamura, D. Pleiter,
H. Simma, C. Urbach,
O. Witzel, T. Yoshie,
J. Zanotti, . . .

F. Karsch, Y. Aoki,
B. Blossier, F. Di Renzo,
R. Edwards, W. Kamleh,
D. Leinweber, Y. Kuramasi,
A. Portelli, J. Simone

Current Status: Work in Progress

- **Extension** of metadata schema and file format based on user feedback
- Development of user-friendly client tools and graphical interfaces (search, markup)
- Documentation and instructions for non-experts
- Exploring new storage elements (also Cloud)
- New user registration (IAM @ CNAF, with support by INFN)

Identity vetting with (increasingly!) high Level of Assurance is unavoidable!

- Revision of AUP and VO-Policy (see draft on https://hpc.desy.de/ildg/vo_policy)
  Community participation in defining and formulation is needed!

Outlook

It is reasonable to look forward to a renewed, boosted role of ILDG (2.0) due to progress testified by

- renewed interest by many lattice collaborations

- work by re-activated WGs

- ILDG infrastructure (more and more) back in operation

- possibly simplified access and more available documentation (with efforts also by users!)

- community-wide Hands-on Workshop requested soon

There is still work ahead of us, but we can be reasonably optimistic!

Aim by Lattice 2024

New configurations from at least 5 collaborations available through ILDG

. . . to be continued . . .

(lunchtime) **meeting at 12:15**
(DEMO and discussion)

WH2SE Comitium

On Zoom: look for a breakout labeled ILDG