# International Lattice Data Grid

**Status and Progress**

**Carsten Urbach (On behalf of the ILDG Working Groups)**

⟨NuMeriQS⟩

ILDG

UNIVERSITÄT BONN

# International Lattice Data Grid

… what was that again?

- community wide effort

- share valuable gauges

- federation of autonomous grids

- FAIR already 20 years ago

unfortunately, service degraded over the
last decade

# International Lattice Data Grid

… what was that again?

- community wide effort

- share valuable gauges

- federation of autonomous grids

- FAIR already 20 years ago

unfortunately, service degraded over the last decade

**ILDG 2.0**
- resume and upgrade services
- working user tools
- make ensembles citable
- extension of metadata schemata
- extension of file format
- refurbish authentication procedure

# Metadata

**Aim**

- standardised
- unique
- extensible
- flexible

metadata description

# Metadata

## Aim

- standardised
- unique
- extensible
- flexible

metadata description

## XML / XSD

- machine readable
- half-decently human readable
- standardised (**W3C**)
- standard query (search) technology (XPATH)
- software tools available

# Metadata

## Aim

- standardised
- unique
- extensible
- flexible

metadata description

## XML / XSD

- machine readable
- half-decently human readable
- standardised (**W3C**)
- standard query (search) technology (XPATH)
- software tools available

## Working well in ILDG

- ensemble and config schemata
- plus file format

# Infrastructure

**Storage Elements**

**Metadata Catalogue**

**File Catalogue**

**Authentication Service**

# Infrastructure

**Storage Elements**

**Metadata Catalogue**

- needs to be provided / maintained
- mostly by local grids
- global authentication

**File Catalogue**

**Authentication Service**

# Updates and Extensions of Metadata Schemata

- there were some obvious extensions needed
  - e.g. funding reference

- we queried the community for requested changes

- following a (virtual) hands-on workshop last year

$\Rightarrow$ we received a (long) list of requests

- metadata working group proposed solution to (almost) all requests
  [Andronico, Hettrick, von Hippel, Koutsou, Matsufuru, Nakamura, Pleiter, Simma, Simone, Urbach, Yoshie]

- see also the git repo: gitlab.desy.de/ildg/mdwg

# Common Updates and Extensions

**common changes to both config and ensemble XSDs, examples:**

- allow usage of ORCID in the `<participant>` element
- ⇒ alternative to specifying name and institution

- added `<annotation>` element in different locations
- ⇒ allows to provide additional optional information

- added free sub-tree `<additionalInfo>` at top level

# Ensemble Schema Updates: License

- provide a usage license as part of the metadata

- required for complying to FAIR principles

- `<license>` now a **mandatory** element in ensemble description

- different ways to specify a license
    - standard license
    - custom license

- combined with possibility to specify embargo times for which all rights are reserved

- including also an `<acknowledgement>` element

## Ensemble Schema Updates: Funding

- `<fundingReferences>` included now in the schema
- ⇒ allows to mark-up one or more funding references
- different formats allowed
  - possible to specify e.g. award number

- subset of DataCite schema
  datacite.org/

# Ensemble Schema Updates: New Actions

- additional gauge groups $\rightarrow$ quite general now

- higher fermion representations as new quark actions

- open and openSF boundary conditions

- Möbius domain wall fermions

- QED gauge actions

- exponential clover actions

- fermions coupled to photon fields

# Configuration Schema Updates: Multiple Configurations

- in particular finite $T$: combine several configurations into one container binary file

- replace `<markovStep>` by `<markovSequence>`

- a `<markovSequence>` can contain several `<markovStep>` elements

- a `<markovSequence>` corresponds one-to-one to one binary file

- the `<markovChainURI>` now sub-element of `<markovSequence>`

- `<dataLFN>` is now the first top-level element

- of course: separate `<crcCheckSum>` and `<avePlaquette>` for each binary record

# User Registration and Authentication (IAM)

- VOMS (virtual organisation membership services) ceased operation beginning of July

- **new ILDG membership registration**
⇒ now via IAM (identity and access management) service at CNAF (INFN)
  iam-ildg.cloud.cnaf.infn.it/
  MoU between ILDG and CNAF

- GRID certificates can be used to log in to the IAM

- alternative provided by EDUGAIN
  edugain.org

- depends on your institution's identity management

# OpenIDC and Tokens

- CNAF IAM can be used combined with OIDC (open ID connect) service (see e.g. `oidc-agent`)

- `oidc-agent` will authenticate against IAM and provide a so-called token
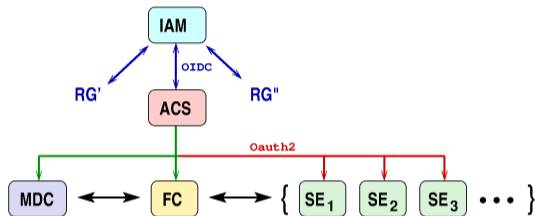


[figure by H. Simma]

- the token can be used to authenticate against the file and metadata catalogue servers

- also provides a very elegant mechanism for fine-grained access control

**Only thanks to the work by H. Simma and Basavaraja BS!**

# Why do we need authentication at all?

> **Storage Providers**
> - require restricted access
> - storage and bandwidth expensive

# Why do we need authentication at all?

### Storage Providers
- require restricted access
- storage and bandwidth expensive

### Users
- need to respect rules

# Why do we need authentication at all?

**Storage Providers**
- require restricted access
- storage and bandwidth expensive

**Users**
- need to respect rules

**The Lattice Community**
- enforce embargoes
- restrict write access

# Metadata and File Catalogues

- the metadata and file catalogues required a complete re-implementation
  (REST API, simple containerised deployment, etc.)

- we could acquire funding through DFG funded PUNCH4NFDI for a software developer
  → poster by G. Pederiva

- **Basavaraja BS** and **H. Simma** have essentially finished this critical overhaul

- new implementation general and flexible

⇒ usable by other communities and for other schemata!

- new implementation hopefully scalable (PostgreSQL)

- prototypes of new services are up and running for LDG and JLDG

# Client Tools

- proof of concept command line tool (`try-client`)
  gitlab.desy.de/ildg/hands-on/try-client

- web interface for simple operations
  hpc.desy.de/ldg/metadata_indices/

- GUIs in preparation for generation of markup and complex search queries
  → poster by G. Pederiva

- web interface by JLDG
  www.jldg.org/QCDml/

# Client Tools

- proof of concept command line tool
  (`try-client`)
  gitlab.desy.de/ildg/hands-on/try-client

- web interface for simple operations
  hpc.desy.de/ldg/metadata_indices/

- GUIs in preparation for generation of
  markup and complex search queries
  → poster by G. Pederiva

- web interface by JLDG
  www.jldg.org/QCDml/

**The new tokens allow**

- access possible both with

  - `gfal2` tools, e.g. `gfal-copy`

  - plain `curl` commands

- for `gfal2`: a docker container with the
  relevant software is provided
  see
  gitlab.desy.de/ildg/hands-on/material

# Summary and Outlook

**ILDG 2.0 is close to being fully operational**

- overhaul of infrastructure and schemata nears completion
- first uploads ongoing

**Future**

- community wide hands-on workshop (in the planning)
- further (faster) schema extension when needed
- improve client tools

# Summary and Outlook

**ILDG 2.0 is close to**

- overhaul of infr
- first uploads or

**Challenges**

- sustain the software development expertise
- acquire sufficient storage space

**community wide effort required!**

**Future**

- community wide hands-on workshop (in the planning)
- further (faster) schema extension when needed
- improve client tools

# Summary and Outlook

**ILDG 2.0 is close to**

- overhaul of infr
- first uploads on

**Future**

- community wide hands-on workshop (in the planning)
- further (faster) schema extension when needed
- improve client

**Challenges**

- sustain the software development expertise
- acquire sufficient storage space

**community wide effort required!**

**Thanks for your interest!**