university of groningen

ILDG and its European Regional Grid: State, Related Initiatives, and Outlook

Dirk Pleiter

FSE | BI | CS

04 April 2025

Overview

1. ILDG
2. European Open Science Cloud (EOSC)
3. ILDG Regional Grid in Europe
4. Summary and Conclusions

Content

ILDG Mission

- The International Lattice Data Grid (ILDG) is a community-driven initiative of theoretical physicists to enable the sharing of primary data from numerical simulations in Lattice QCD based on the FAIR principles
- Working towards this goal requires
  - Standards and policies
  - Federated digital infrastructure services and resources
  - Organisational structures
- Realisation based on a system-of-systems infrastructure (“regional grids”)

Organisational Structures

- ILDG Metadata Working Groups (MDWG)
  - Agrees on community-wide standards for the description of digital objects
  - Specifies metadata schemata (QCDml) and data formats
- ILDG Middleware Working Group (MWWG)
  - Specifies interfaces of services to ensure interoperable regional grids
  - Supports the implementation of regional grids
  - Suggest or develop prototypes of user tools
- ILDG Board
  - Represents ILDG towards community and service providers
  - Decides on policies and guidelines for membership and sharing of digital objects
  - Supports regional grids in applying for resources
  - Oversees working groups

Standardisation Efforts (1/2)

- QCDml
  - Critical for the findability of digital objects
  - Ensures unique description of digital objects
  - Aims for extensibility and forward-compatibility
  - Implemented by two XML schemata
    - Ensemble metadata schema
    - Configuration metadata schema
  - Significant changes in QCDml 2.0 (not backward compatible) including
    - Extension of the supported lattice actions and gauge groups
    - Additional metadata: license specification, references to funding

Standardisation Efforts (2/2)

- ILDG File Format
  - Critical for interoperability and reusability
  - Recent backward compatible updates including
    - Packing of multiple gauge configurations
    - Support of gauge groups other than SU(3)
- ILDG REST API
  - Interface for accessing ILDG catalogue services using web-based protocols

Services

- Metadata catalogue services (MDC)
  - Catalogue for storing, finding, and downloading metadata
  - One catalogue per regional grid (possibly based on the same implementation)
- File catalogue (FC)
  - Catalogue for managing the mapping between Logical File Names and URIs pointing to the location of digital objects
  - One catalogue per regional grid (possibly based on the same implementation)
- Identity and Access Management (IAM) services
  - Global service for managing ILDG membership and access attributes
  - Connects to the global eduGAIN IAM infrastructure for federating identity services

ILDG versus Regional Grids

- Role of ILDG
  - Governance at the global level
  - Defining global standards and policies
  - Manage the IAM services
- Role of the Regional Grids
  - Governance at the regional level
  - Defining regional policies
  - Manage and operate catalogue services (MDC, FC)
  - Possibly organise and manage storage resources

Content

EOSC Mission

[EC]

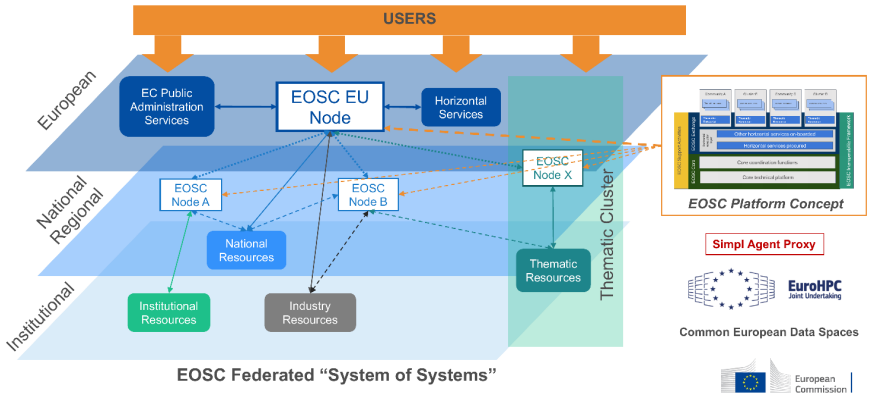
- "The ambition of the EOSC is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes."
- "The EOSC ultimately aims to develop a 'Web of FAIR data and services' for science in Europe upon which a wide range of value-added services can be built."

Tripartite Governance

- European Commission (DG RTD and CNECT)
- EOSC Association
- EOSC Steering Board
  - EC Expert Group with members representing EU Member States and countries associated to Horizon Europe

EOSC Federation (1/2)

[P. Szegedi et al.; 2024]



EOSC Federation (2/2)

- Responsibilities of the EOSC Federation bodies [A. Götz et al.; 2025h]
  - Develop federation strategy
  - Define membership criteria
  - Manage federation services (IAM, resource catalogues, ...)
  - Develop and decide on technical standards and requirements as well as policies
  - Define SLAs
  - Implement monitoring capabilities
  - Governance at federation level
- Access through the EU EOSC Node portal:
  `https://open-science-cloud.ec.europa.eu/`

EOSC Federation: Expected Outcomes

- Selected expected outcomes relevant for ILDG: [A. Götz et al.; 2025h]
  - Facilitated research reproducibility
  - Increased collaboration, community and knowledge sharing
  - Increased standardisation and interoperability
  - Improved research integrity
  - Increased robustness and trustworthiness

EOSC Node (1/2)

- Nodes are implemented and operated by a single organisation or a consortium of organisations
- Nodes need to comply with decisions, rules and policies of the EOSC Federation
- Core capabilities of an EOSC Node: [A. Götz et al.; 2025h]
  - Resource catalogue and registry services
  - IAM services
  - Helpdesk
  - Service monitoring
  - Service and research product accounting
  - Order management
  - ...

EOSC Node (2/2)

- Generic node capabilities: [A. Götz et al.; 2025h]
  - Data transfer
  - Notebooks
  - Compute and storage resources
  - File sync & share
- Node Exchange: Resources shared with the EOSC Federation

EOSC Federation: Interoperability Framework

- Key components
  - EOSC Interoperability Guidelines
  - EOSC Interoperability Registry
  - EOSC Interoperability Framework Governance
- Selected guidelines [EOSC Resource Hub]
  - AARC Blueprint Architecture 2019 (AARC-G045)
  - EOSC Data Transfer: Architecture and Interoperability Guidelines

EOSC: FAIR Data Repositories

- Data repositories are part of the EOSC Nodes
- EOSC aims to measure the FAIRness of the data repositories
- Requirement of having Persistent IDentifier (PID)
  - Guidelines for creating a user tailored EOSC Compliant PID Policy [R. van Horik et al.; 2024]
  - Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC) [EC; 2020]
- Recommendation to aim for repository certification
  - CoreTrustSeal
  - Nestor Seal
  - ISO 16363 certification
- The metadata must be harvestable via OAI-PMH protocol

Content

Digression: Virtual Organisations (VO)

- **Virtual Organisation** = A set of individuals and/or institutions defined by sharing rules [I. Foster et al.; 2001]
  - VOs vary tremendously in their purpose, scope, size, duration, structure, community, and sociology
- Model that has for long be successful for grid infrastructures:
  - A (research) community organises itself as VO
  - Digital infrastructures provides resources and services to the VO
  - The VO manages the use of these resources and services

Roles: Defining Policies

- **Rationale**: Policies allow to define how a VO organises itself, which allows to create trust in the VO and establish a contact point for the VO
  - VO – Digital infrastructure stakeholders
  - VO – Funding organisations
- Possible policy topics
  - Internal governance
  - Data management policies
  - Resource management policies

Roles: Resource Allocation

- **Rationale**: Reduce the burden of digital infrastructure providers by managing the provided resources
- Prerequisites:
  - Resource allocation policies
  - Ability to show best-possible use of the provided resources

Roles: Storage Resources

- **Rationale**: Applying for storage resources and/or requesting specific data management services is more likely to be successful when applying as a community
- Prerequisite: Ability to demonstrate impact
  - Showcase research output of a community
  - Demonstrate better use of data that has been generated using expensive resources
- Possible data management services requirements
  - Long-term storage, certified data repositories
  - Suitable data management interfaces
  - Data transfer services

Roles: Organise Funding

- **Rationale**: The regional grid is an infrastructure effort that requires funding
  - Examples:
    - Funding for maintaining and further developing services (including the necessary software)
    - Funding for storage resources
- Funding opportunities
  - EC funding: Difficult
  - National funding: To be explored
    - Possible opportunities related to the establishment of national EOSC Nodes
  - Thematic funding and funding related to research infrastructures

Roles: Liaise with Other Initiatives

- **Rationale**: Leveraging the potential of digital infrastructure initiatives requires visibility as a community
  - Examples: EOSC, EuroHPC

Roles: Communication

- **Rationale**: Success of a sustainable regional grid strongly depends on its visibility
- Challenge: Organise continuous heartbeats in relevant communication channels

Organisational Structure

- Organisational structure
  - General decision-making body (board or council)
  - Working groups
  - People with special roles (e.g., communication officer)
  - Possibly special committees (e.g., resource allocation)
- Representation in general decision-making body
  - All countries with lattice research groups
  - All collaborations
- Aim on consensus-based decision-making
  - Possible exception: special committees

Content

Summary and Conclusions

- Key steps towards ILDG 2.0 have been accomplished
- There are many connections to EOSC and possibly synergies to exploit
- A regional grid for Europe needs to be established
  - Opportunities for generating added value for lattice research groups in Europe can be identified
  - Challenge: Dependence on voluntary efforts