EERAdata

TOWARDS A FAIR AND OPEN DATA ECOSYSTEM IN THE LOW-CARBON ENERGY RESEARCH COMMUNITY

August Wierling
Western Norway University of Applied Sciences

How EERAdata can collaborate with and contribute to EOSC

1 What is EOSC ?
2 Task force
3 Workshops by EOSC-Nordic
4 Discussion

1.What is EOSC ?

The European Open Science Cloud (EOSC) is an environment for hosting and processing research data to support EU science.

The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes.

- Seamless access
- FAIR (Findability, Accessibility, Interoperability and Reusability) management
- Reliable reuse of research data and all other digital objects produced along the research life cycle (e.g. methods, software and publications) “

Source: https://eosc-portal.eu/about/eosc

1.What is EOSC

Home » Use Cases » Community Use Cases

Community Use Cases

This page provides examples of "EOSC in practice" use cases or success stories that highlight how EOSC services and resources can support the daily work of researchers and innovators. Please scroll down to make the use cases appear.

If you wish to share your EOSC use case, please fill out the webform.

Run4science.org - Measuring environmental and biodiversity data... while running!

We want citizens to measure their environment by using smartphones. Most of the citizen

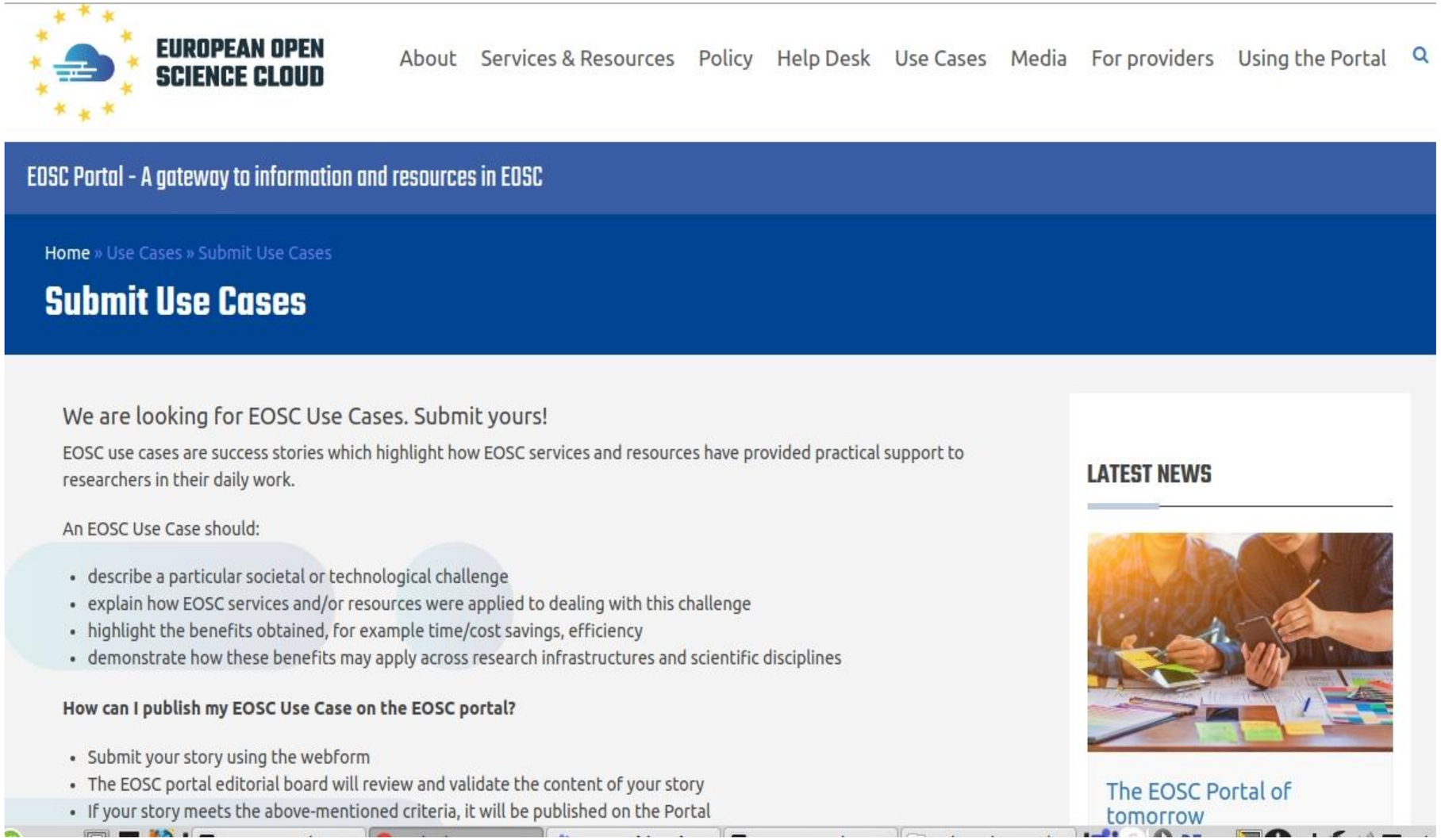
Kampal Artificial Intelligence for rare disease diagnosis

In the context of the EOSC-hub project, Kampal Data Solutions is benefitting from storing the healthy and ill patients' registries to a database on EOSC infrastructure.

Guardomic• bot mitigation engine

Web services owners struggle daily to protect their websites from bot traffic and their users from fraudulent digital ads or cryptocurrency web mining.

1.What is EOSC ?



2. EOSC task forces

- AG Implementation of EOSC
  - TF PID Policy and Implementation
  - TF Researcher Engagement and Adoption
  - TF Rules of Participation Compliance Monitoring
- AG Metadata and Data Quality
  - TF FAIR Metrics and Data Quality
  - TF Semantic Interoperability
- AG Research Careers and Curricula
  - TF Data Stewardship Curricula and Career Paths
  - TF Research Careers, Recognition, and Credit
  - TF Upskilling Countries to Engage in EOSC
- AG Sustaining EOSC
  - TF Defining Funding Models for EOSC
  - TF Long-Term Data Preservation
- AG Technical Challenges on EOSC
  - TF AAI Architecture
  - TF Infrastructure for Quality Research Software
  - TF Technical Interoperability of Data and Services

Source: https://eosc.eu/news/eosc-association-task-forces-community-agenda

2. Presentation at TF FAIR Metrics and Data Quality

High quality data in the energy domain

Reflections from EERAdata

by August Wierling, Valeria Jana Schwanitz, HVL

01.03.2022

2. Presentation at TF FAIR Metrics and Data Quality

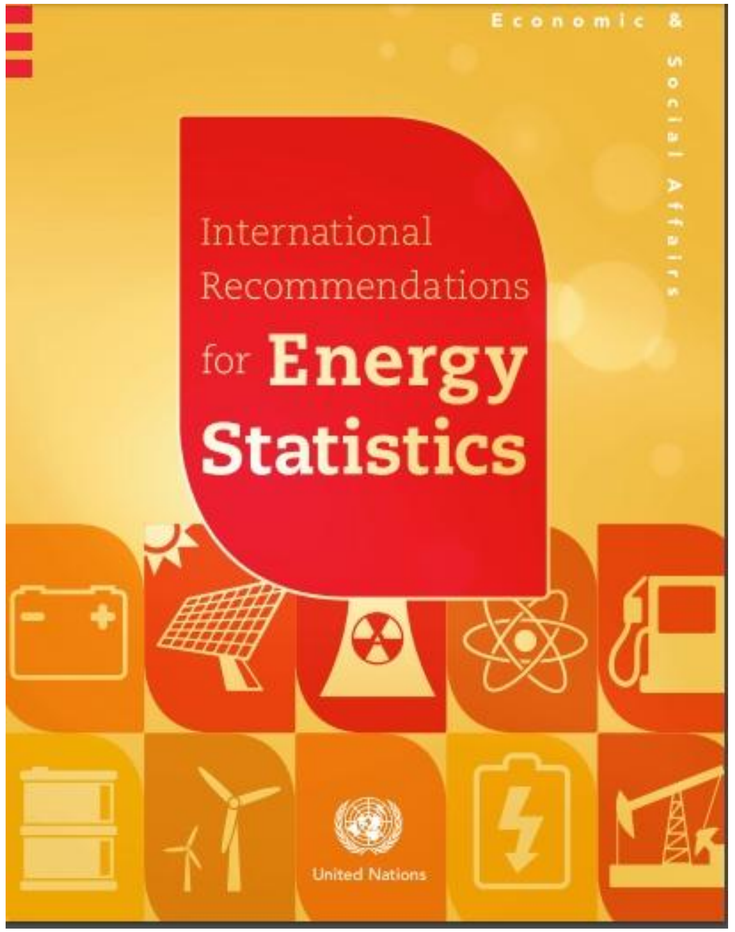
Data quality defined in terms of its “fitness for use”

Dimensions of Quality:

- Relevance
- Accuracy and Reliability
- Timeliness and Punctuality
- Coherence and Comparability
- Accessibility and Clarity

Trade-offs between the dimensions:

- Accuracy vs. timeliness
- Confidentiality vs. transparency



2. Presentation at TF FAIR Metrics and Data Quality

Completeness: Customer data is often collected in bulk. It is often difficult to ensure that all the elements of each piece of data is present in the master data pool. Businesses should identify the key data elements that are required for analysis and ensure that those elements are intact.

Consistency: Utilities must ensure that data stored in separate locations are in sync and consistent. This means democratising access to data across the enterprise and having regular checks to ensure that the data is up to date in all locations.

Conformity: Data is sometimes collected and stored in inconsistent formats and data types. This requires businesses to standardise data input methods, structures, and configurations. Ensuring that each record conforms to the same standards reduces errors and allows data to be extracted easily at any point.

Accuracy: Utility companies have reported that at least 20% of their customer data is inaccurate. This means that the data stored in the business database does not accurately reflect its value in real life. This makes it difficult for businesses to derive any insight as the base data for analysis cannot be trusted.

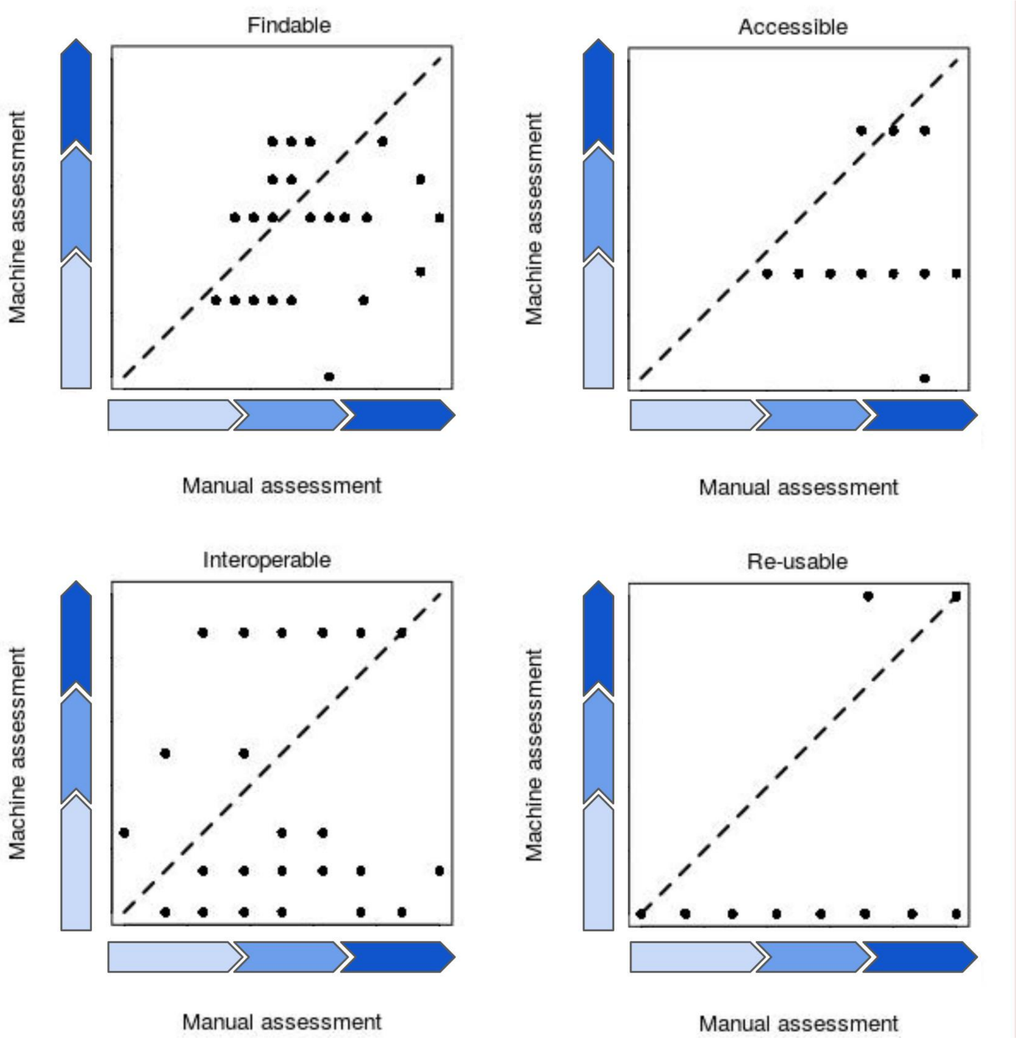
Integrity: Integrity is often used interchangeably with quality. However, the integrity of data refers to the completeness of the relationships that exist between data sets. This can occur when certain data elements are duplicated within the data set, leading to values that are “orphaned” or not linked to the right customer.

Timeliness: Utilities must ensure that their data is truly reflective of real-time customer habits and business operations. High-quality data is complete and available for analysis at a moment’s notice.

Source:
https://www.smart-energy.com/industry-sectors/data_analytics/the-ultimate-guide-to-managing-utility-data-quality/

2. Presentation at TF FAIR Metrics and Data Quality

Question:
Did machine-based assessment lead to unexpected results ?



2. TF Semantic Interoperability

Examples for problems to be addressed by means of case studies:

- Data – homogeneous entities and heterogeneous formats/languages (O1.1, O1.4)
- Data – human understanding and consumption (O2)
- Data – Data model versioning (O1.1, O1.3)
- Query – Federation with heterogeneous access points (O1.3, O1.5)
- On-the-fly ontology (term) integration – user-driven ontology (O1.4, O2)
- Guide users to add data in a machine-actionable way (O2)
- Interoperability of electronic lab notebooks with digital infrastructures (O1.3, O1.4)
- Assisted specification of data provenance metadata (O1.1)
- Combine discoverability across domains and findability in domains concerning (O1.2, O1.3)
- Develop visualisation to navigate through and communicate about the KOS universe (O1.2)
- Long-term archiving of semantic artefacts (cost-benefit analysis) (O1.6)
- Provenance - Audit trail of ownership (history) and transformation of open metadata with DP's (O1.4, O1.5)

Taken from: charter document at
https://www.eosc.eu/sites/default/files/tfcharters/eosca_tfsemanticinteroperability_draftcharter_20210614.pdf

3. EOSC-Nordic

EOSC NORDIC

About EOSC-Nordic | News and Articles | Events | Newsletter | Knowledge Hub

Contribute +

Event

Workshop on National Policies Relevant to EOSC Deployment

Join the workshop on "National Policies Relevant to EOSC Deployment – Status, Gaps & Steps towards Harmonisation" in Strasbourg on 4th May!

Read more

3. Webinar programme in the past years

26.11.2020 STEP 1 Global Unique Identifiers for Datasets (F1)

03.02.2021 STEP 2 Metadata (F3)

29.04.2021 STEP 3 generic metadata standards

07.10.2021 STEP 4 domain-specific metadata webinar

08.02.2022 STEP 5 Value and limitations of FAIR assessment tools

3. Discussion about assessment tools

GO FAIR meeting invitation:

“Key FAIR Evaluation stakeholders together to discuss workflows for metadata gathering. This is prompted by the highly divergent scores currently being issued by the various FAIR assessment tools. “