Challenges in Open Research Data

Isabel Bernal

CHIST-ERA Open Science Advisory Board

Project Seminar 2022

28 March 2022, online
COMMON CHALLENGES TO OPEN RESEARCH DATA

- Insufficient rewards and incentives for data creators (misalignment across different evaluation/funders systems)
- Perceived loss of competitive advantage by data creators (free riders risk)
- Perceived loss of control by data creators in cross-border data access, sharing and reuse
- Complex net of legal frameworks to comply with (copyright, sui generis database directive, patents, trade secrets, GDPR, Public sector Open data Directive...)
- Ethical/privacy considerations
- Fear for data breaches (confidentiality, security, reuse violations..)
- Sustainability/governance issues and/or insufficient alignment with FAIR principles in some data infrastructures/repositories
- Insufficient data skills/competences amongst scientific community to fully implement standards and benefit from advanced technologies (computation, processing, storage...), understanding of FAIR Principles, data management, metadata...
- Significant cost on data creators (organizations, researchers, services...) to implement risk management approach/anonymization/facilitate and engage stakeholders...
- Regulation of public-private partnerships

CHALLENGES TO MAXIMIZE REUSE OF OPEN RESEARCH DATA

- Data Quality issues (relevance, accuracy, completeness, timeliness): data quality is a challenging concept: data can be good for certain applications but not for others
- Semantic Interoperability: lack of common standards, proliferation of incompatible standards/formats..
- Reusability: cost and effort to adopt and maintain community standards broadly
- Long term sustainable business models and funding for open data provision
- Engagement of users communities to maximize reuse and innovation: existing barriers in:
  1. Availability and Access ( eg APIs with limited documentation/usability/flexibility)
  2. Findability (eg poor metadata)
  3. Usability ( eg data fragmentation, licenses)
  4. Understandability (eg poor metadata, nonuser friendly interfaces)
  5. Linking and combining data (eg format issues, legal issues)
  6. Comparability/compatibility (same as above)
  7. Interaction with data creators
  8. Feedback mechanisms