



# B2SAFE: Practical Policies

Willem Elbers (MPI-TLA)  
2<sup>nd</sup> EUDAT Conference



# Policy Collecting

- Collected policy natural language descriptions from several projects and institutes
  - (EUDAT core communities), IN2P3, REPLIX, PanData, ILL



- Total of 53 policies
- Classified according to type (hard/soft) and risk.
- First selection round:
  - Integrity checking
  - Format conversion

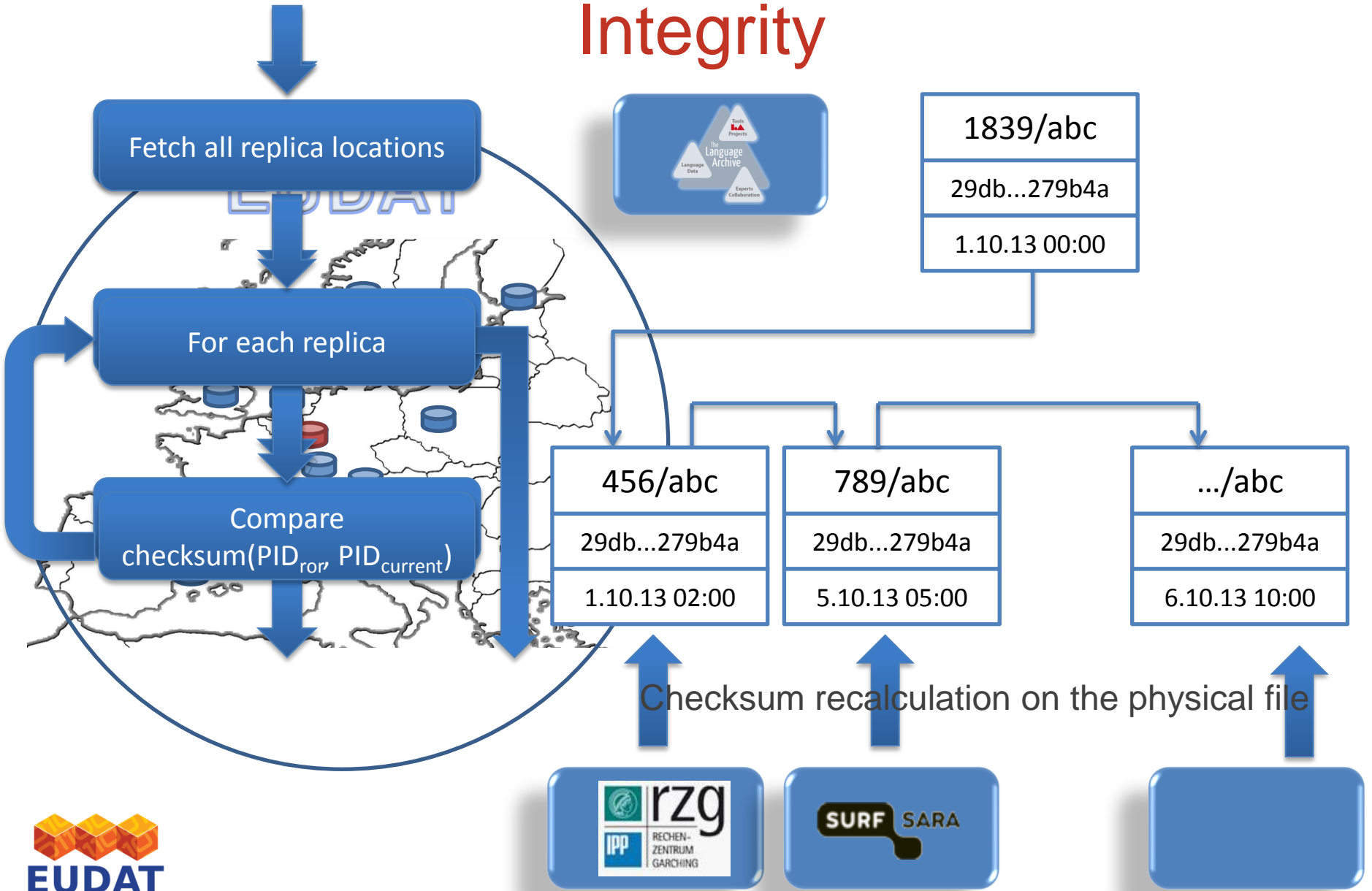
# Integrity

- Verify that:
  - At least  $n$  replicas of the digital object exist
  - The replica checksums match the checksum of the repository of record object
- Requirements:
  - Original Digital Object as input **Use the EUDAT**
  - Reference **Persistent Identifier Structure**
  - Checksum stored for each replica
- We can't rely on on-demand checksum computation
  - Checksum must have a timestamp associated

# Integrity

- Three policies needed
  - Policy to compute checksum
    - Running at the EUDAT data centers
    - At a set periodic interval
  - Policy to verify the number of replicas
  - Policy to verify the replica checksums
- Visualization of the results for the community data managers
- Integration into the Data Policy Manager

# Integrity



# Format Conversion

- Domains: format verification versus format conversion
- On ingest the RoR is responsible for format verification and conversion
- Over time a format can become obsolete
  - A format obsolescence strategy must be in place
- Where does the conversion take place?
  - The end user converts the data before or during the ingestion manually (out of EUDAT scope)
  - The conversion is a part of ingestion workflow on EUDAT server side (in EUDAT scope).
  - Some format after some time becomes obsolete and all replicas in that format shall be converted (in EUDAT scope).

# Questions



Thank you for your attention