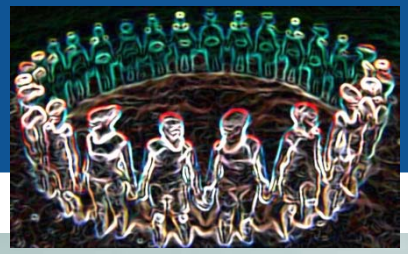


European Commission

CALL 3

MANAGING, PRESERVING AND COMPUTING WITH BIG RESERACH DATA

1. Managing, preserving and computing with big research data

Development and deployment of integrated, secure, permanent, on-demand service-driven and sustainable e-infrastructures for scientific computing and data

TOWARDS GLOBAL DATA E-INFRASTRUCTURES RESEARCH DATA ALLIANCE

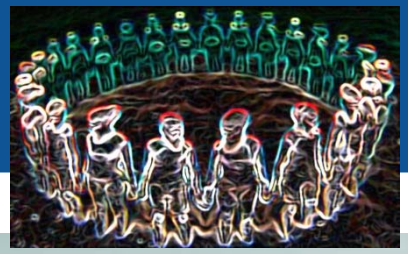
2. e-Infrastructure for Open Access

Robust e-infrastructure supporting Open Access policies in Europe, providing reliable and permanent access to digital scientific records

E-INFRASTRUCTURES FOR OPEN ACCESS

3. Towards global data e-infrastructures

Consolidating Europe's contribution to the Research Data Alliance (RDA) and ensuring that RDA serves to foster research data interoperability at global level



European Commission

Call 3 Topic 1: Managing, preserving and computing with big research data

- A federated pan-European data e-infrastructure
  - **cost-effective and interoperable solutions for data management and long term preservation**
- Services for quality and reliability
  - **including certification mechanisms and services**
- Federating data management and curation tools and services
  - **on the basis of an open architecture**
  - **support development of Data Management Plans**
- Large scale virtualisation of data/compute centre resources

On Long Term Data Preservation

- Data Preservation is useless without **knowledge preservation**
- Data Preservation will become useless without **provenance information**
- We need to integrate tools to assure this in Data Preservation:
  - Identify/define/(certify?) "Indicators"
    - Typically "variables" extracted/derived from a defined (complex) data set
    - Relevant for experts: PUBLICATIONS
    - Relevant for non experts: REPORTS
  - Establish/analyze the full (digital/non-digital)-chain for indicators:
    - Basic, Complementary and Derived Data, Software, and…
    - "Knowledge" (in brain? recipes? context?)
  - Implement the "validation" of knowledge and provenance
    - Tests: "simple" distributions comparison ("control plots")
    - Digital keys/signature (on data to assure provenance)
    - "Measured Values" preservation: trace down to the basic observation
      - example: assign an ID to each "measurement" , a PID to each "collection"
      - example: check ID-final indicator correlations with changes in software
      - (not uncorrelated with bit preservation)

On Long Term Data Preservation

- Examples (our limited experience):
  - HEP events
    - Event by event stability analysis
  - **Environmental observations**
    - Cross-correlations/cross-checks and impact on models towards IPBE
  - Genetic data
    - Exploit intrinsic provenance (a digital key DNA coded)
- Objectives:
  - **PROVIDE LONG TERM VALIDATION TOOLS FOR DATA ACROSS RESEARCH AREAS**
  - Deeper understanding of our data towards better preservation
  - Preservation of knowledge
    - Learn towards preserving “brain” knowledge in-silico (too ambitious?)