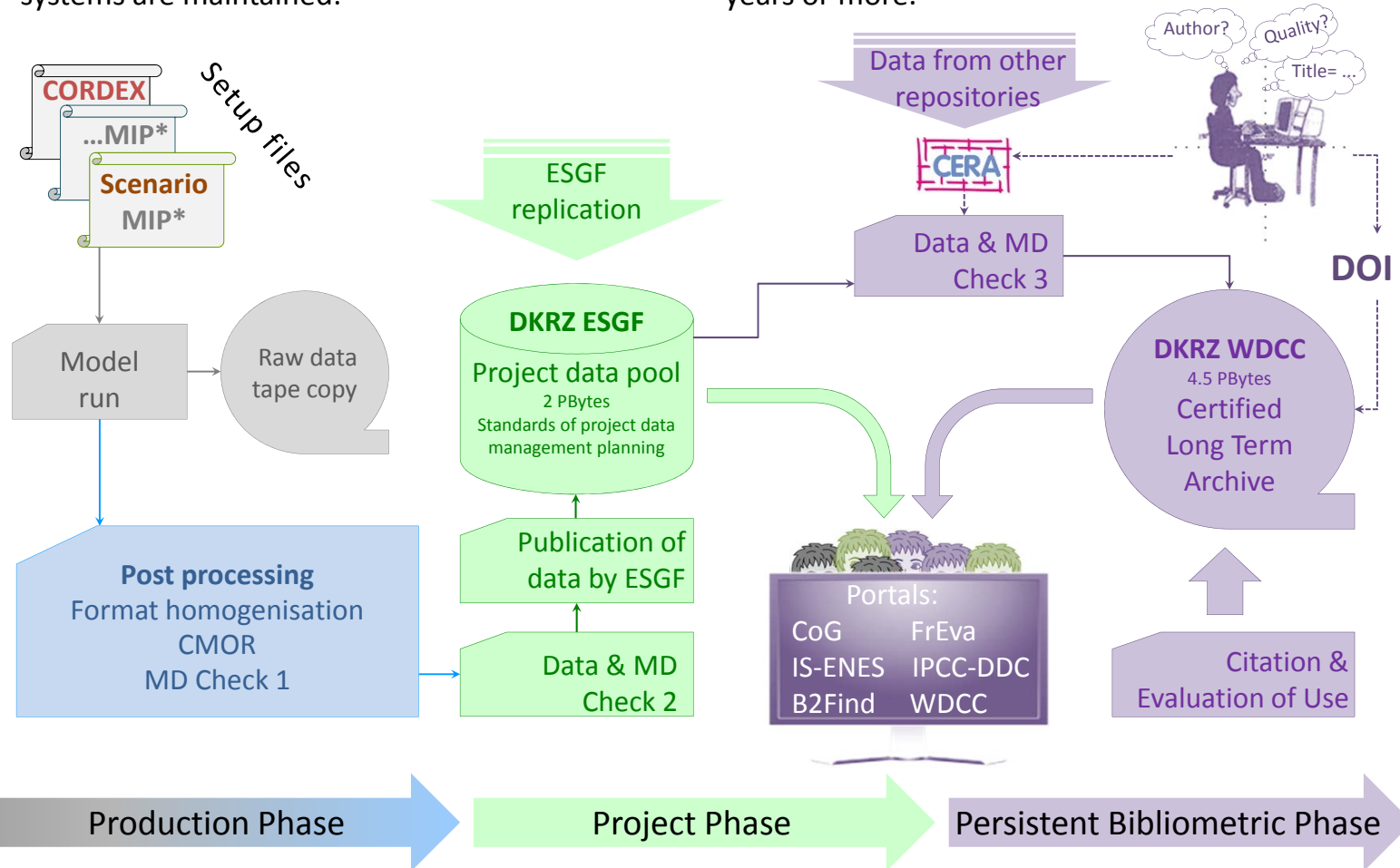


WDCC and ESGF –

The Double Structure of the Digital Data Storage at DKRZ

In many scientific projects, the data repositories face different requirements during the running project and later during the Long Term Archiving (LTA). This is why at DKRZ, two different storage systems are maintained:

The DKRZ **ESGF** data node is used for data storage and dissemination during the project runtime. **WDCC's** CERA database is especially designed for the following long term archival storage of ten years or more.



Data sources

The storage systems of DKRZ get climate model data from many different sources – be it from the in-house computer systems, be it by replication from the ESGF community, or be it from single customers. The latter come from all parts of climate research with a focus on climate model data.

DKRZ ESGF

The Earth System Grid Federation* is a global collaboration of distributed data nodes with large quantities of disk storage. It is designed for efficient data access and dissemination via Graphical User Interfaces, via OPeNDAP server, and via scripts. During the community review data still can be enhanced by new versions or withdrawn. Project specific data quality assurance procedures are supported.

WDCC

Long Term Archiving at WDCC* means curated storage on tapes. For efficient data access a 1.5 PB cache is installed. – All data have extensive meta data for use by third parties that have to pass comprehensive quality checks. This is additionally supported by best practice information on citation and identifiers, e.g. DOIs (Digital Obj. Identifiers). Data use and citation can be monitored. For the data persistency in case of crises, arrangements still need to be established.

* Different Model Intercomparison Projects

* See esgf.llnl.gov

* See www.wdc-climate.de