PID Collections: DKRZ use case

Tobias Weigel, Stephan Kindermann, Michael Lautenschlager
(DKRZ)
The use case at a glance

- Data management for international climate model intercomparison project (WCRP CMIP6)
  - Federated infrastructure, developer community, new governance mechanisms
- Automatic generation of PIDs for files and their aggregations
- Smart tools that exploit additional information
How are the collections used for CMIP6 compatible with those used in other communities?

Can we build tools and services that work across communities, saving effort for everyone?

Data object types:
- netcdf files (primarily); XML metadata, errata documentation, workflow output

Key requirements include:
- PIDs are the primary addressing mechanism (collections and items)
- Solution independent from concrete storage of objects and metadata
- Independence from particular service provider models
- CRUD + L operations on objects within a collection
- Collection templates
- Batch operations on multiple collections and their items
- Ability to search, filter and aggregate items across collections
- Future: Support for controlled vocabularies