PERSISTENT IDENTIFIER SERVICES

PROGRESS UPDATE & FUTURE ROADMAP

ESGF F2F Workshop,
Washington, DC, December 2016

Tobias Weigel
Stephan Kindermann
Merret Buurman
Katharina Berger
Guillaume Levavasseur
...
2016 Progress & Accomplishments

PID related tasks worked on:
• Task 15.1 PID service performance and reliability tests
• Task 15.3 Initial deployment of PID infrastructure components and publication workflow tests
• Task 15.6 PID service integration (GUI, publisher)
• Task 15.9 Integrate replication and versioning with PID infrastructure as part of ESGF CMIP6 publication workflow

Accomplishments:
• Developed scalable PID service (distributed message queue + handle system + landing pages + publisher client)
• PID client documentation and PID SW CI and testing environment
• PID service integrations
  -- COG: pid links, pid for data card
  -- Publisher: versioning, replica PID references, unpublishation
  -- Errata service connection
2016 Missed Milestones (if any)

No missed milestones

Open points:

• deployment process
  -- project dependent installation (pip, part of CMIP6 handler,..)

• PID configuration step (security credential exchange) supervisioning (CDNOT)

• CMIP6 PID publications to production ESGF message queue (rabbit-mq exchanges at PCMDI and IPSL)
2017 Roadmap

Planned Tasks:

#1 - high: Operational deployment of RabbitMQ nodes at several sites

#2 - high: CMIP6 publication tests (with PID integration) from pilot sites (PCMDI, BADC, IPSL) – February 2017

#3 - medium: PID configuration process for CDNOT

#4 - medium: Develop a first pilot of a PID CLI tool
   -- info display e.g. new versions, replica info,..

#5 - medium: Develop a concept for enhancing the data cart service
   -- improvements based on feedback of current solution
   -- future roadmap (cog, cli,..)
Additional Resources Needed (if any)

• No additional resources needed

• Hardware:
  -- 1 rabbit mq server (VM, rabbit mq) (DKRZ)
  -- 1 rabbit mq server (VM, rabbit mq) (IPSL)
  -- 1 rabbit mq server (VM, rabbit mq) (PCMDI)
  -- 1 (PID + rabbit) server (physical) (DKRZ)

• Stable operations required (24/7)
Community Feedback from Survey (if applicable)

- Reliability and resilience of resources, support for data access and download: Main motivators for PID usage
- Good UI design: Needs to be observed for data cart / CLI tool
- How are data used, for which purposes: Data tracking through PIDs – more work required (external use, callbacks)
Collaborations Needed (if any)

• Collaboration with other ESGF working teams
  ‣ Publication Team

• Collaboration with outside entities
  ‣ RDA:
    -- PID collection WG
    -- PID record profiles
    -- PID ESGF processing use case (→ ISO study group)

• EUDAT:
  -- operational policies for PID infrastructure
  -- PID record profiles

• MPCDF (Germany): PID tests