Federated Quality Control Procedure for CMIP5

M. Stockhause, H. Höck, M. Kurtz,
M. Lautenschlager, F. Toussaint

EGU 2012, 25.04.2012

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

Outline

- CMIP5 technical infrastructure
- CMIP5 quality control workflow
- Federated quality control concept and implementation within CMIP5
- CMIP5 quality control experiences and status

CMIP5 Technical Infrastructure

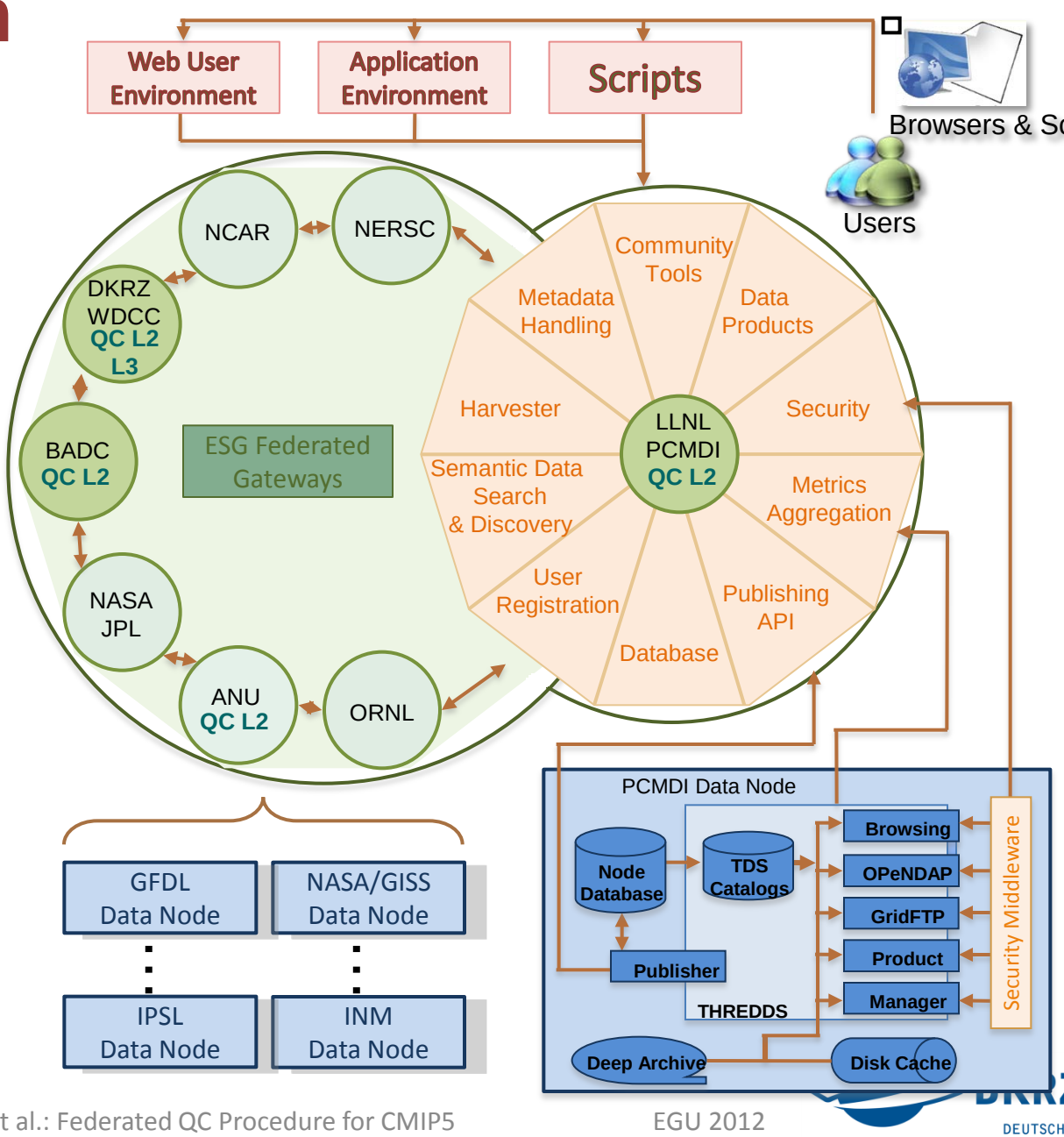
Earth System Grid (ESG)

CMIP5 Infrastructure

PCMDI
Data infrastructure and security – ESG

BADC
Metadata infrastructure - CIM / Metafor

WDCC
Quality Control / DOI data publication



http://earthsystemgrid.org
http://pcmdi7.llnl.gov

DEUTSCHES KLIMARECHENZENTRUM

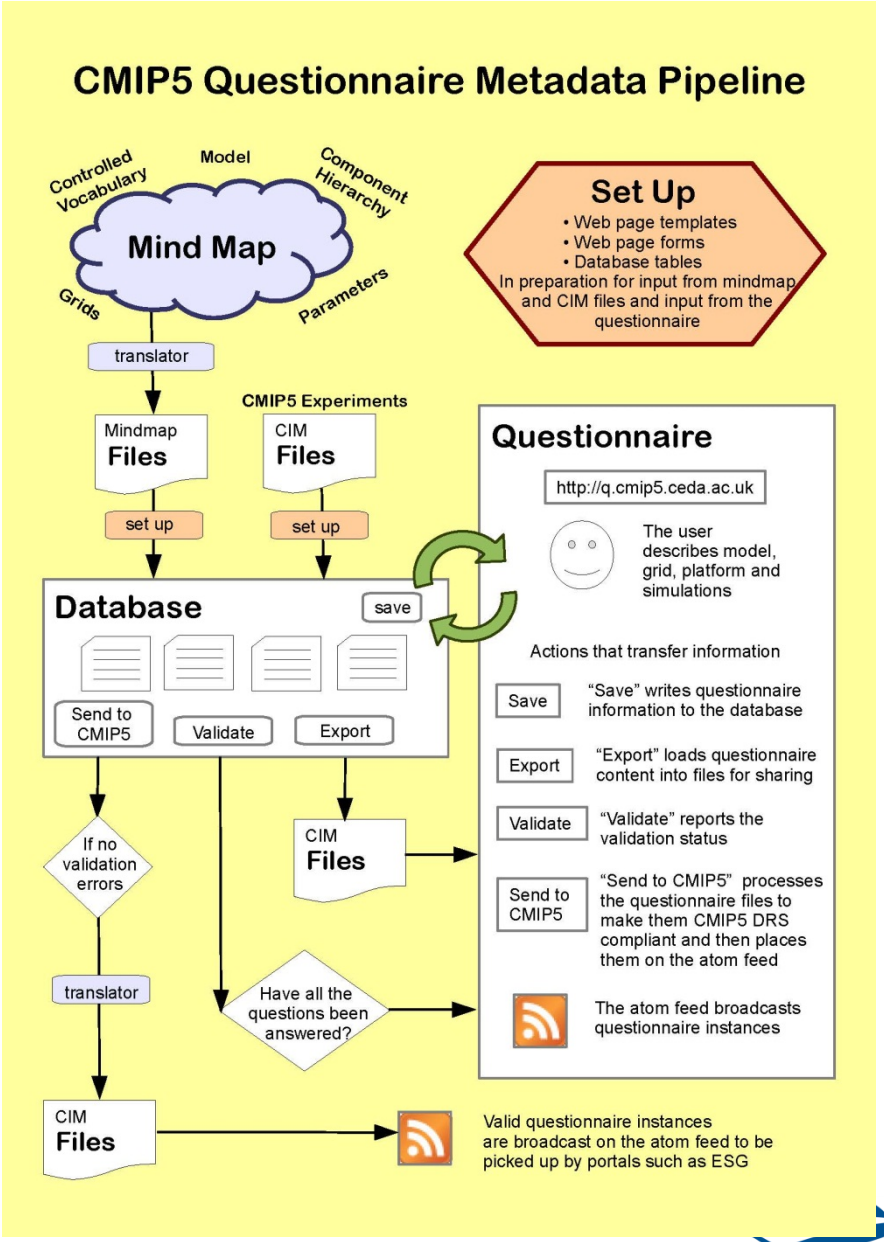
CIM / Metafor

CMIP5 Infrastructure

PCMDI
Data infrastructure and security – ESG

BADC
Metadata infrastructure - CIM / Metafor

WDCC
Quality Control / DOI data publication



http://metaforclimate.eu

DKRZ DEUTSCHES KLIMARECHENZENTRUM

CMIP5 Quality Control Workflow

Quality Level in CMIP5

Quality Level is assigned to CMIP5 experiments

- **Quality Level 1 – all data:**
  separate technical checks of data and metadata

- **Quality Level 2 – output1 data:**
  consistency checks – QC tool of MPI-M QC used

- **Quality Level 3 / DataCite DOI – output1 data:**
  cross- and double-checks of data and metadata (TQA)
  DataCite (datacite.org) DOI data publication process

Data Changes *without* user notification

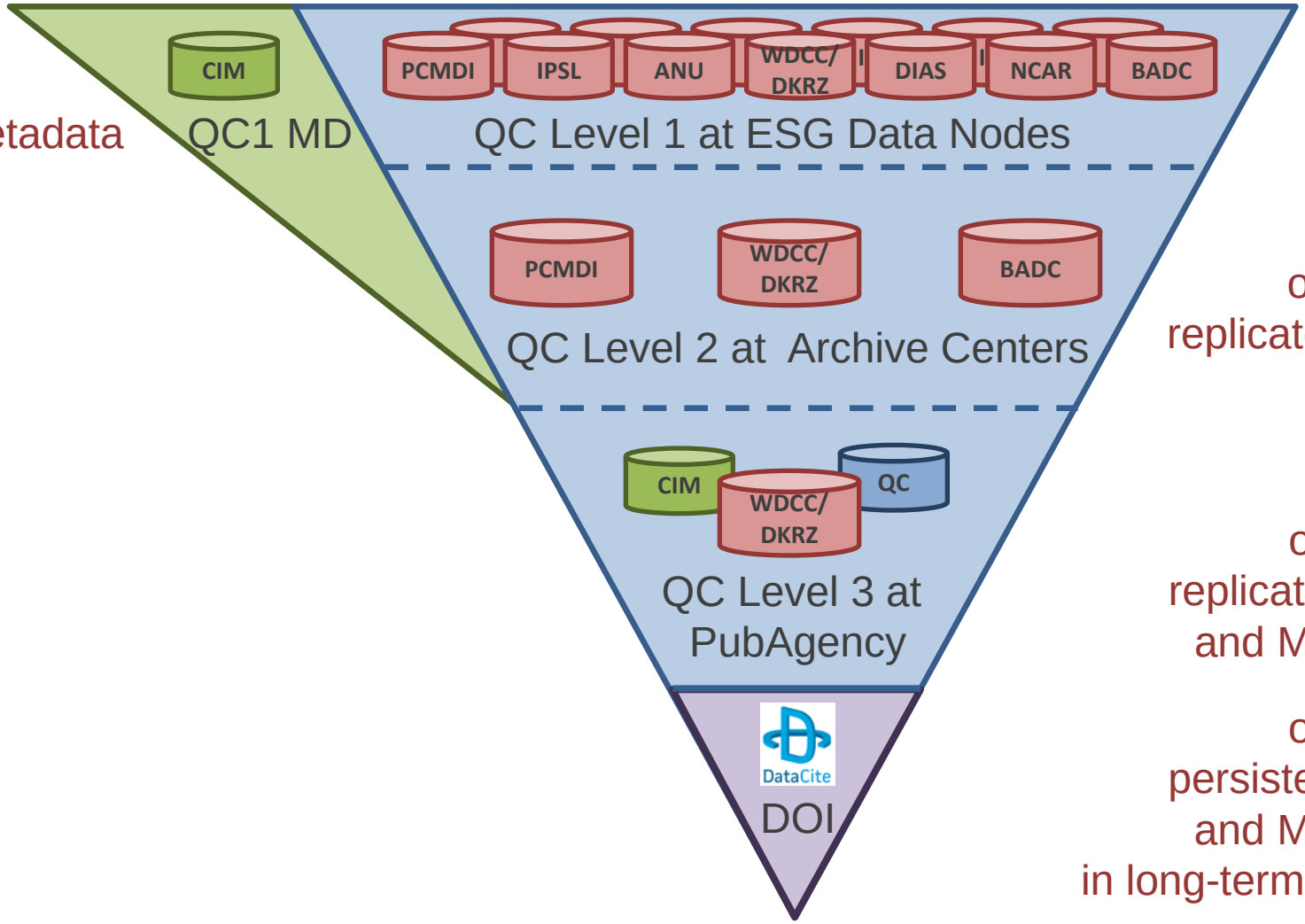
Data Changes *with* user notification

Data Persistence & Data Citation by DOI

e.g. DOI:10.1594/WDCC/CMIP5.NCCNMpc
(DOI Resolver: http://dx.doi.org/10.1594/WDCC/CMIP5.NCCNMpc)

DKRZ DEUTSCHES KLIMARECHENZENTRUM

CMIP5 Quality Control

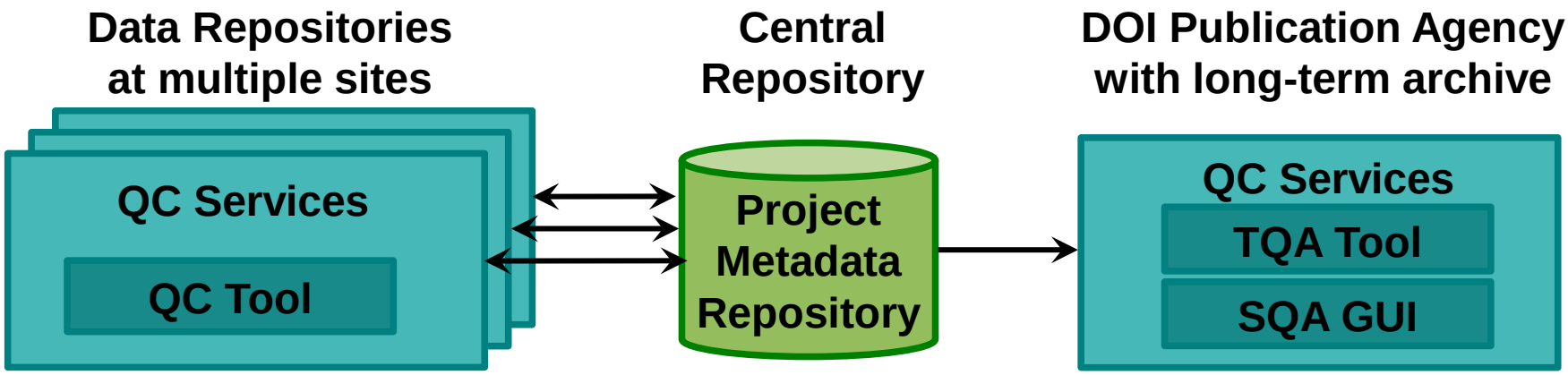


DKRZ DEUTSCHES KLIMARECHENZENTRUM

Federated Quality Control
Concept and CMIP5 Implementation

DKRZ
DEUTSCHES
KLIMARECHENZENTRUM

Federated Quality Control Approach



QC checks at Data Repositories:

QC Tool embedded in a QC Service layer, which supports the interaction with the Central Repository and QC result analysis.

Central Repository for Project Metadata:

Storage of QC metadata along with other metadata on data and additional information

CMIP5: QC, CIM questionnaire and ESG datanode metadata

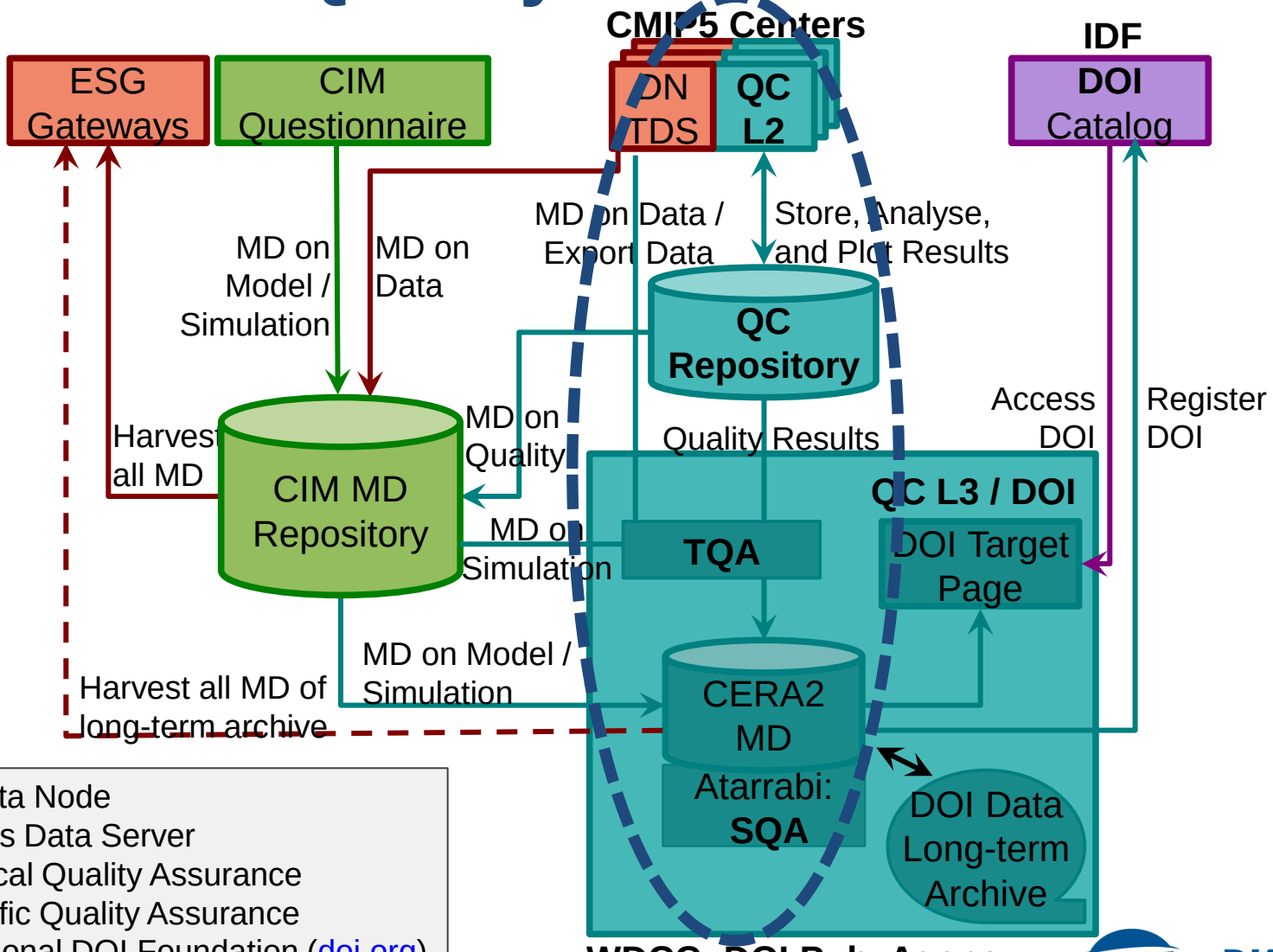
Long-term Archive and DOI data publication:

Project metadata access for cross-checking (TQA Tool) using QC Services

DOI data publication and long-term archiving after author approval (SQA GUI)

DKRZ DEUTSCHES KLIMARECHENZENTRUM

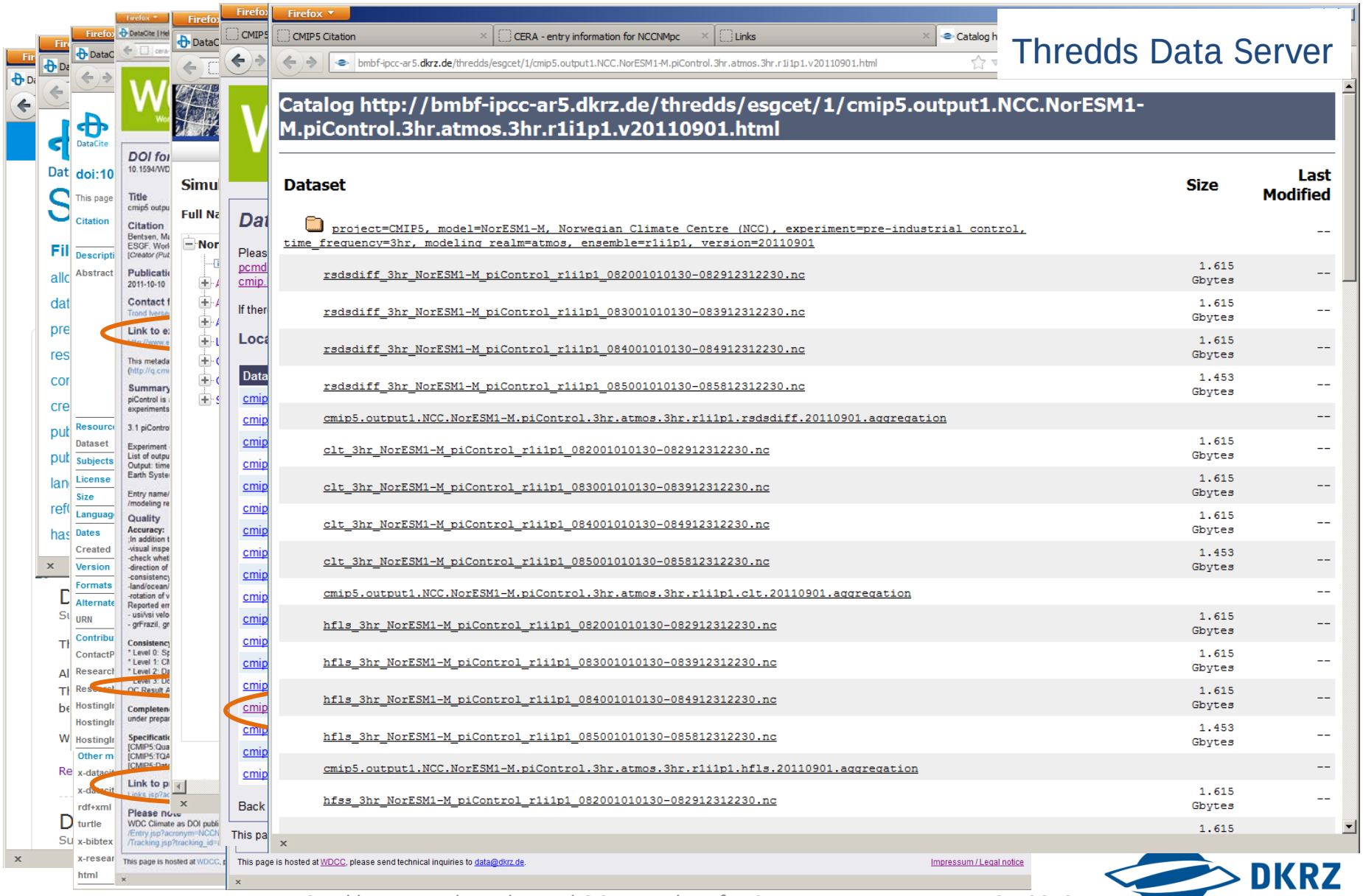
Federated Quality Control in CMIP5



DN: ESG Data Node
TDS: Thredds Data Server
TQA: Technical Quality Assurance
SQA: Scientific Quality Assurance
IDF: International DOI Foundation (doi.org)

DKRZ DEUTSCHES KLIMARECHENZENTRUM

What is a data DOI?



DKRZ DEUTSCHES KLIMARECHENZENTRUM

Experiences and Status of the CMIP5 Quality Control

CMIP5 QC Experiences (1)

Modification of QC Concept for CMIP5 implementation:

- Project Metadata Repository falls in 3 parts: TDS metadata on data, CIM metadata on model/simulation, and QC metadata

  -> QC L3 cross-checks had to access 3 metadata sources

- slow data replication due to narrow band widths: QC L2 checks are performed at data nodes in parallel to data replication

  -> QC L2 Managers at CMIP5 Centers need to coordinate the QC on the data of their responsibility at CMIP5 data nodes,

  QC L2 assignment by QC L2 Managers at the CMIP5 Centers

  -> QC L3 had to add cross-checks on consistency

Higher degree of local distribution of QC L2 checks and increased complexity of QC L3 cross-checks

CMIP5 QC Experiences (2)

Problems within the technical infrastructure of CMIP5:

- All infrastructure components are still improved
  -> QC L3 has to deal with changing interfaces to the data and metadata infrastructure components
- DRS IDs (Data Reference Syntax) are not enforced in usage
  -> QC L3 has to perform different mappings of DRS IDs used in the different technical components to the documented ones and keep them up to date
- QC has not been integrated in the ESG gateways, yet
  -> Set-up of additional QC Services for data replication and users, e.g. services for quality status for ESG publication units, QC results, data citation, and quality status services

Additional services for scientists and technicians and increased complexity of QC L3 cross-checks

CMIP5 QC Experiences (3)

Problems outside of the technical infrastructure:

- Delay in the CMIP5 data delivery / ESG publication
- Data is revised several times and ESG published as new versions
- Delay in the documentation of model and simulations in the CIM questionnaire

Delay in the DOI data publication after QC L3

Status of the QC for CMIP5

QC Status of CMIP5 Experiments (20.04.2012):

- Quality Control 1: **775** Experiments
- Quality Control 2: **252** Experiments (finished: 66)
- Quality Control 3: **6** Experiments
- DataCite DOI: **6** Experiments

Status: http://cera-www.dkrz.de/WDCC/CMIP5/QCStatus.jsp

Summary

- The federated quality control approach is capable to serve as QC procedure within CMIP5. First data DOIs have been assigned.
- For the implementation within CMIP5, the QC had to increase in complexity, additional services had to be set up to bridge delays in the integration of the QC into the technical infrastructure for CMIP5.
- Delays in data and metadata publication as well as in the data replication causes delays in the DOI data publication (QC).
- Ongoing developments in the data and metadata parts of the technical infrastructure require efforts for QC L3 code adaptations.

Outlook

- Consolidation of QC L2 checker tool and QC services
- Integration of QC into technical infrastructure of CMIP5
- Development of the CIM repository as a real metadata exchange repository for CMIP5
- Central entry to the access of all quality information: QC workflow, findings of the scientific community; add provenance information

Questions / Comments ?

More Information:

CMIP5 QC: http://cmip5qc.wdc-climate.de

QC Services: http://cera-www.dkrz.de/WDCC/CMIP5

WDCC Data Publication System: http://cera-www.dkrz.de/atarrabi2

CMIP5 Project: http://cmip-pcmdi.llnl.gov/cmip5/

Acknowledgements:

Funded by the German Federal Ministry of Education and Research (BMBF)
Contributions of the ESGF group

Publication submitted:

M. Stockhause, H. Höck, F. Toussaint, and M. Lautenschlager:
Quality assessment concept of the World Data Center for Climate and ist application to CMIP5 data
Geosci. Model Dev. Discuss., 5, 781-802, 2012, doi:10.5194/gmdd-5-781-2012.

stockhause@dkrz.de