

# Identifier Infrastructure Usage for Global Climate Reporting

IoT Week 2017, Geneva

Tobias Weigel

Deutsches Klimarechenzentrum (DKRZ)

World Data Center for Climate (WDCC)

# Scientific driver: Global climate modelling



## CMIP Phase 6 (CMIP6)

### Overview CMIP6 Experimental Design and Organization

The overview paper on the CMIP6 experimental design and organization has now been published in GMD (Eyring et al., 2016). This CMIP6 overview paper presents the background and rationale for the new structure of CMIP, provides a detailed description of the CMIP Diagnostic, Evaluation and Characterization of Klima (DECK) experiments and CMIP6 historical simulations, and includes a brief introduction to the [21 CMIP6-Endorsed MIPs](#).

A brief summary can be found in the following overview presentation ([CMIP6FinalDesign\\_GMD\\_161109.pdf](#)) and below. After a long and wide community consultation, a new and more federated structure has been put in place. It consists of three major elements:

1. a handful of common experiments, the DECK (Diagnostic, Evaluation and Characterization of Klima) and CMIP historical simulations (1850 – near-present) that will maintain continuity and help document basic characteristics of models across different phases of CMIP,
2. common standards, coordination, infrastructure and documentation that will facilitate the distribution of model outputs and the characterization of the model ensemble, and
3. an ensemble of [CMIP-Endorsed Model Intercomparison Projects](#) (MIPs) that will be specific to a particular phase of CMIP (now CMIP6) and that will build on the DECK and CMIP historical simulations to address a large range of specific questions and fill the scientific gaps of the previous CMIP phases.

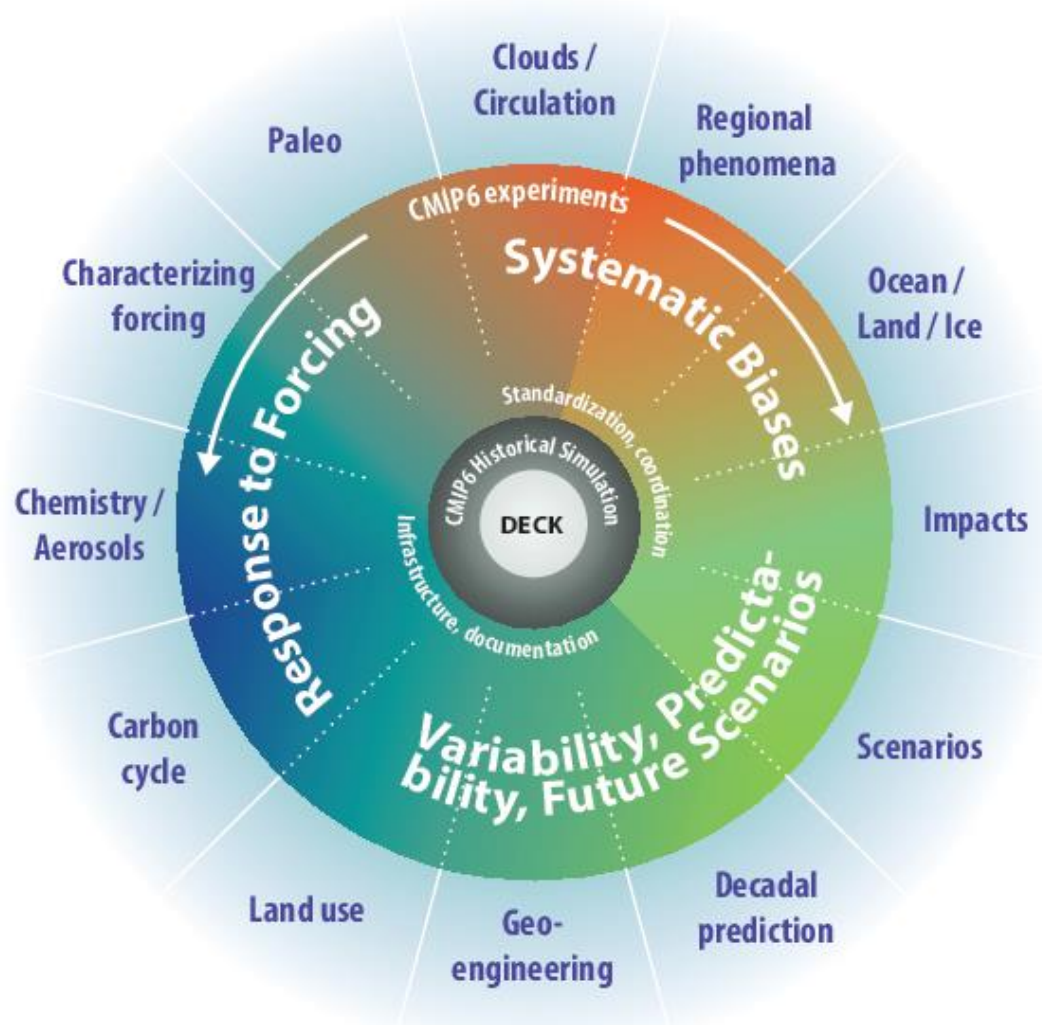
### WGCM

<a href="#">Overview</a>
<a href="#">Members</a>
<a href="#">Meetings</a>
<a href="#">Publications</a>
CMIP
<a href="#">CMIP Panel</a>
<a href="#">CMIP3</a>
<a href="#">CMIP5</a>
<a href="#">CMIP6</a>
Catalogue of MIPs
<a href="#">CMIP6-Endorsed MIPs</a>
<a href="#">Other active MIPs</a>
<a href="#">Former MIPs</a>

◀ Modelling Overview

<https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6>

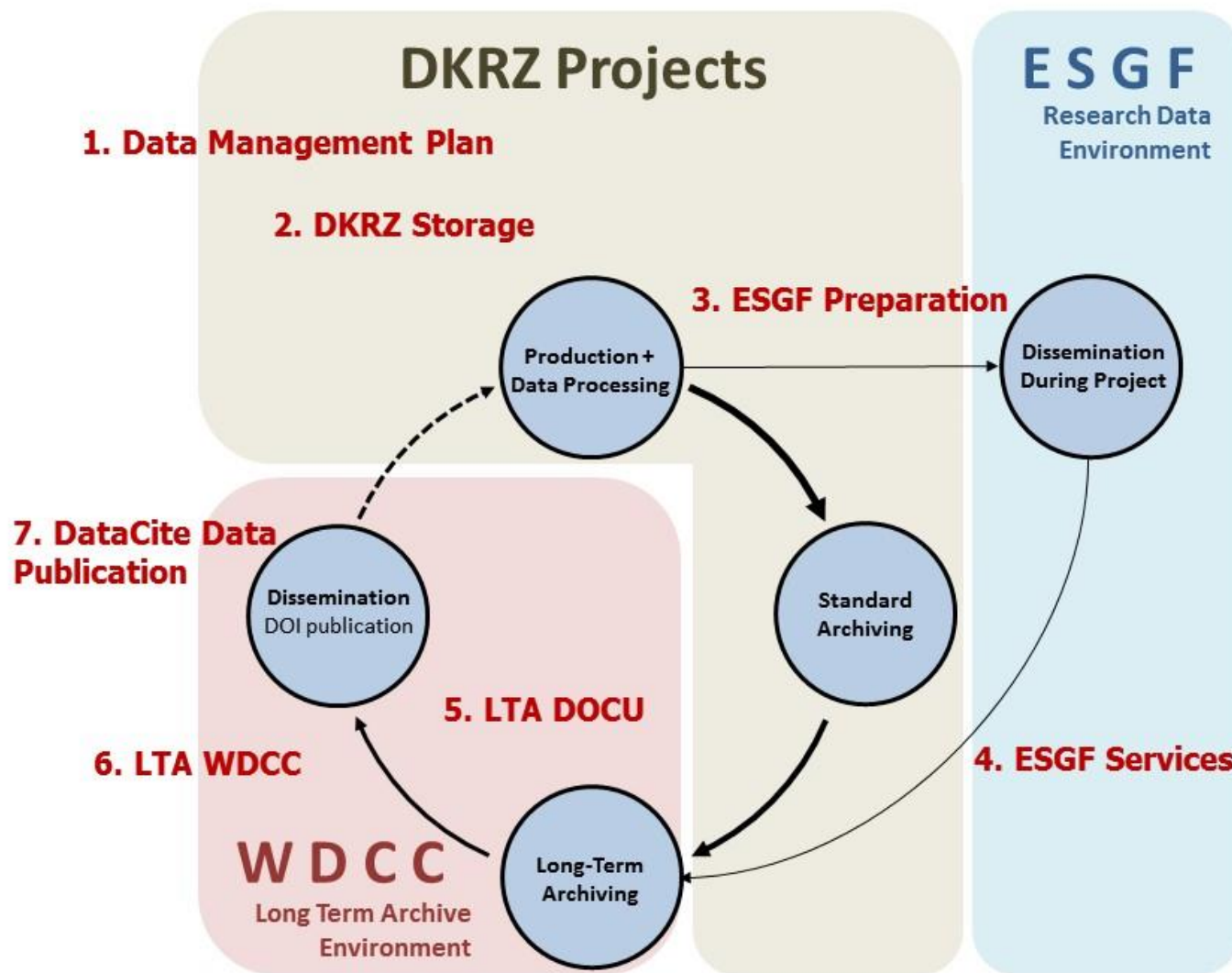
# Scientific driver: Global climate modelling



- Operational phase ca. 2017-2021+
- Community-driven, aligned with IPCC AR6
- Global data volume in order of 100-250 PB
  - full replication impossible!

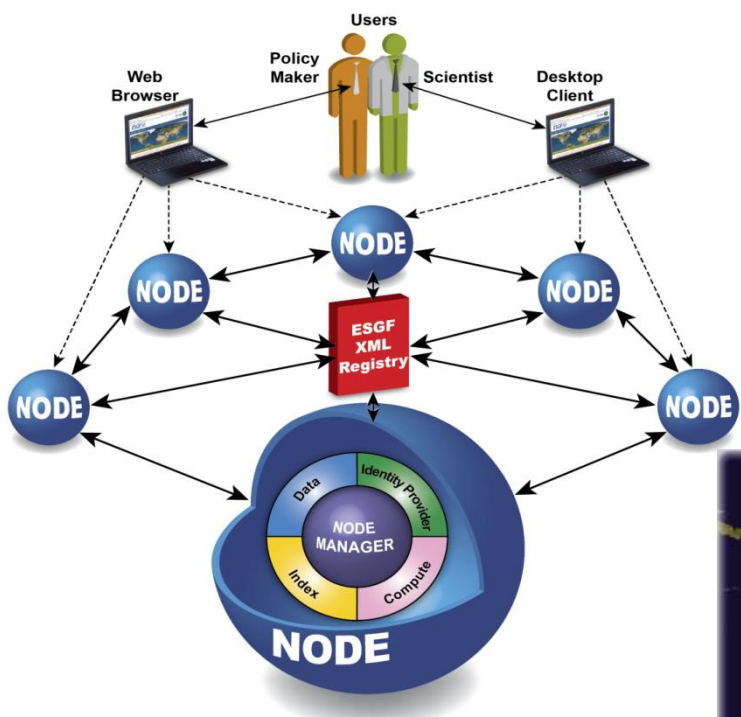
Eyring, Bony, Meehl, Senior, Stevens, Stouffer, Taylor: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev., 9, 1937-1958, 2016. doi:10.5194/gmd-9-1937-2016

# The climate data life-cycle



M. Lautenschlager

# The Earth System Grid Federation



**ESGF**   
 Earth System Grid Federation

<http://esgf.llnl.gov>

<http://esgf-data.dkrz.de>



D. Williams (LLNL); U.S. DOE 2017. 6th Annual Earth System Grid Federation Face-to-Face Conference Report. DOE/SC-0188. U.S. Department of Energy Office of Science



# DKRZ technical infrastructure and ESGF

## HPSS tape

- 190 Pbyte capacity



## Lustre file system

- 54 PByte



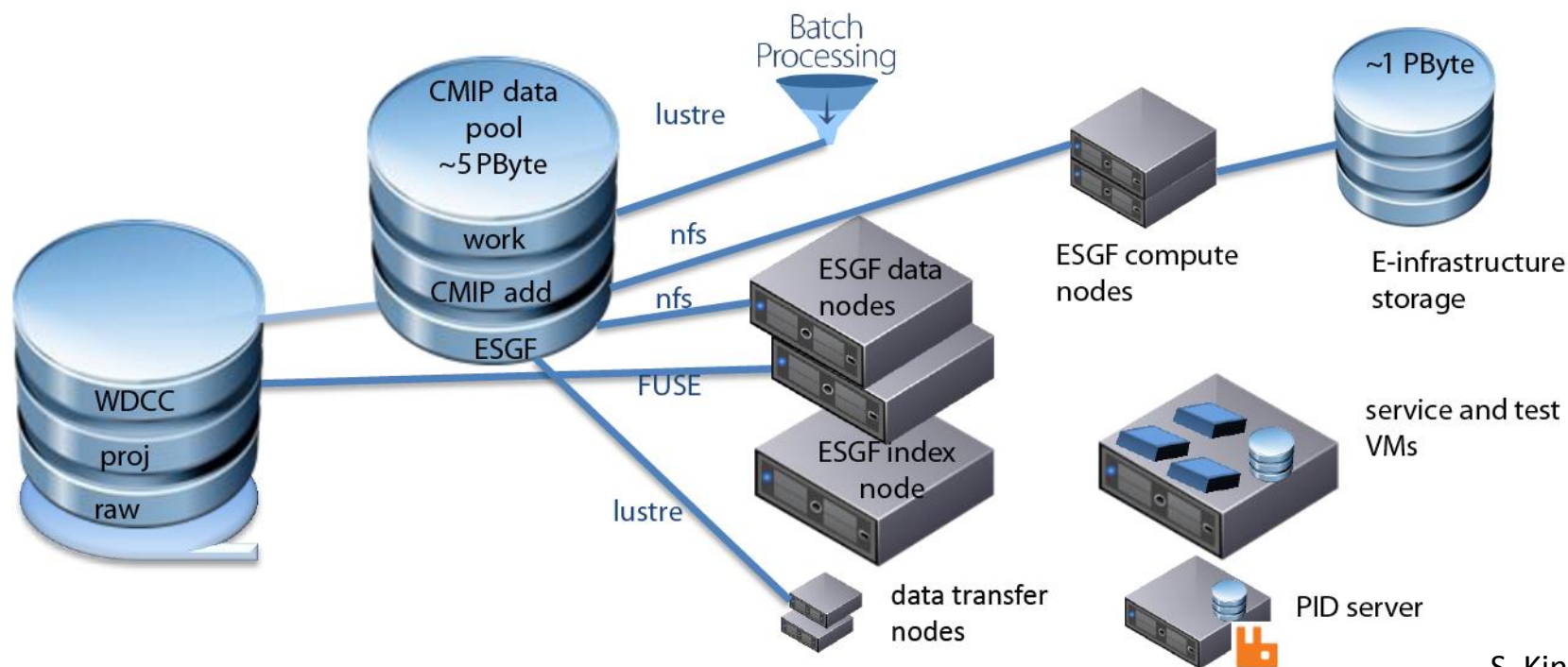
## „Mistral“ HPC

- 3.6 Pflops
- ~100.000 cores



## Compute/ storage cluster

- VM servers, database servers
- Openstack cloud storage

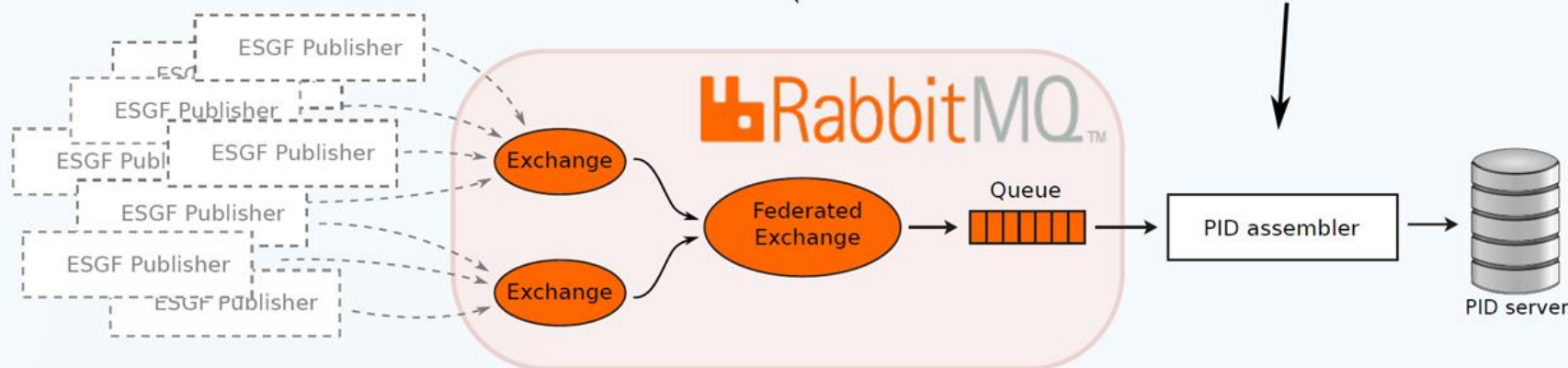


S. Kindermann

# Making it scalable requires additional effort

The PID registration and metadata update tasks are pushed to a message queueing system facilitating high availability and scalability...

...and then processed asynchronously.



Buurman, Weigel, Juckes, Lautenschlager, Kindermann:  
Persistent Identifiers for CMIP6 in the Earth System Grid Federation,  
EGU 2016

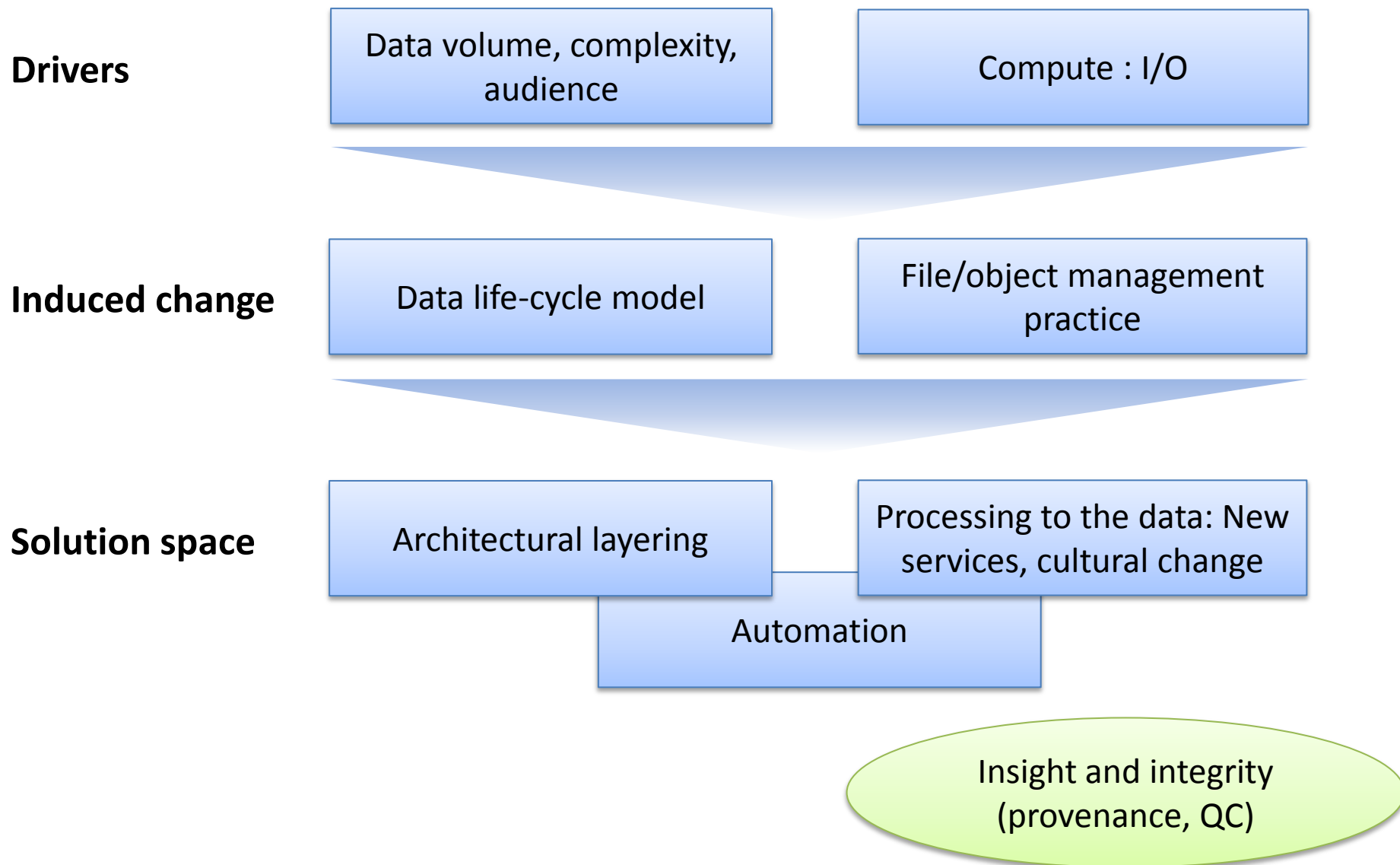
# Properties stored in Handle records for ESGF

Files	Datasets
URL	URL
aggregation_level	aggregation_level
url_replica	replaced_by
tracking_ID	replaces
checksum	errata_IDs
is_part_of	has_parts
DRS_ID	DRS_ID
file_size	
file_name	

**Kernel Information Profiles**



# Why do we care? What is the long-term strategy?



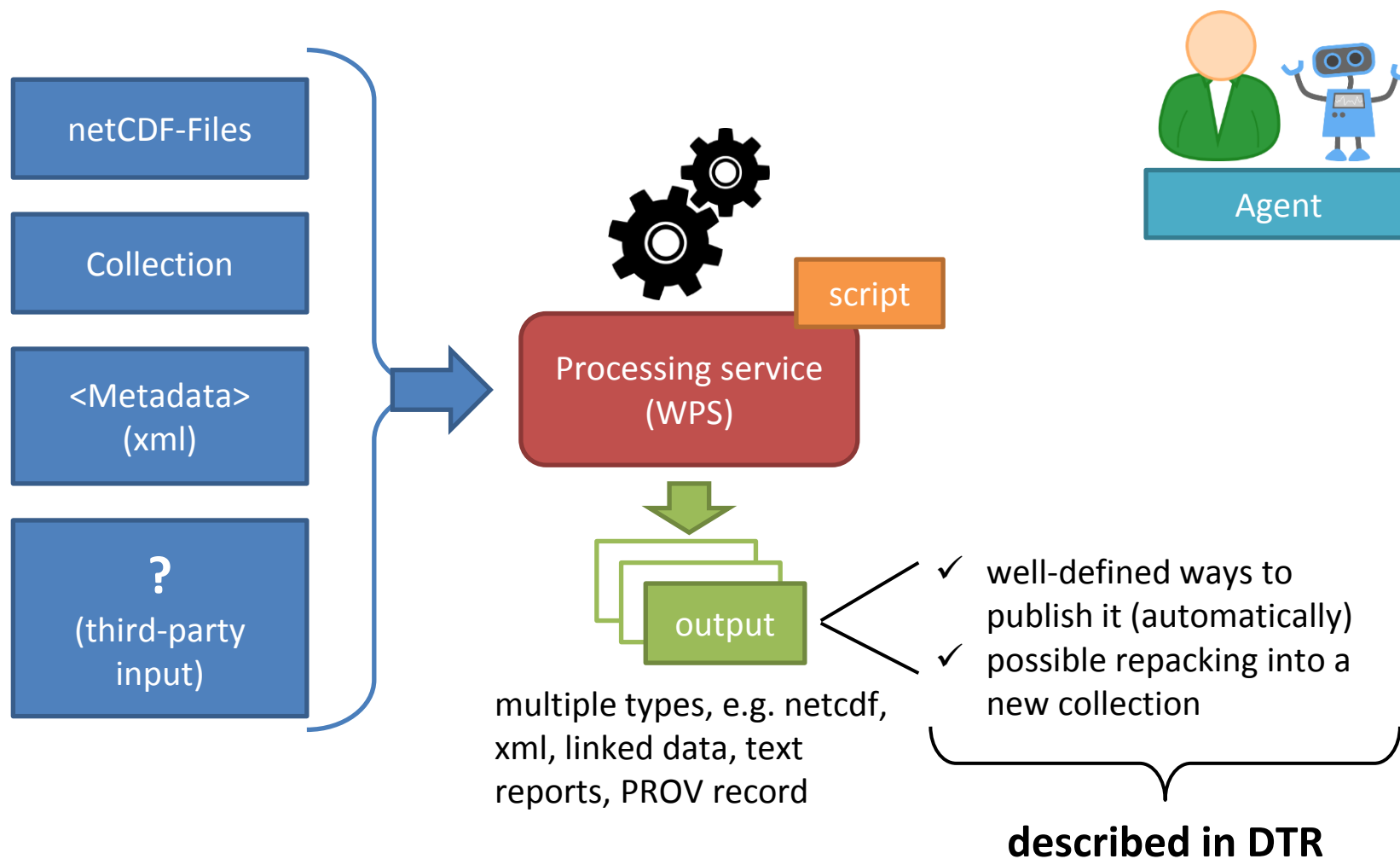
# The users' reality...

```

handle@handle:~/hs/svr_10876
[handle@handle svr_10876]$ ls
admpriv.bin          handle_keystore      sitebndl.zip
admpriv.bin.old      logs                 siteinfo.json
admpub.bin           privkey.bin          txn_id
admpub.bin.old       pubkey.bin           txns
config.dct           serverCertificate.pem webapps
contactdata.dct      serverCertificatePrivateKey.bin webapps-storage
delete_this_to_stop_server serverCertificatePrivateKey.pem webapps-temp
[handle@handle svr_10876]$

```

# Type-Triggered Automated Processing (T-TAP)



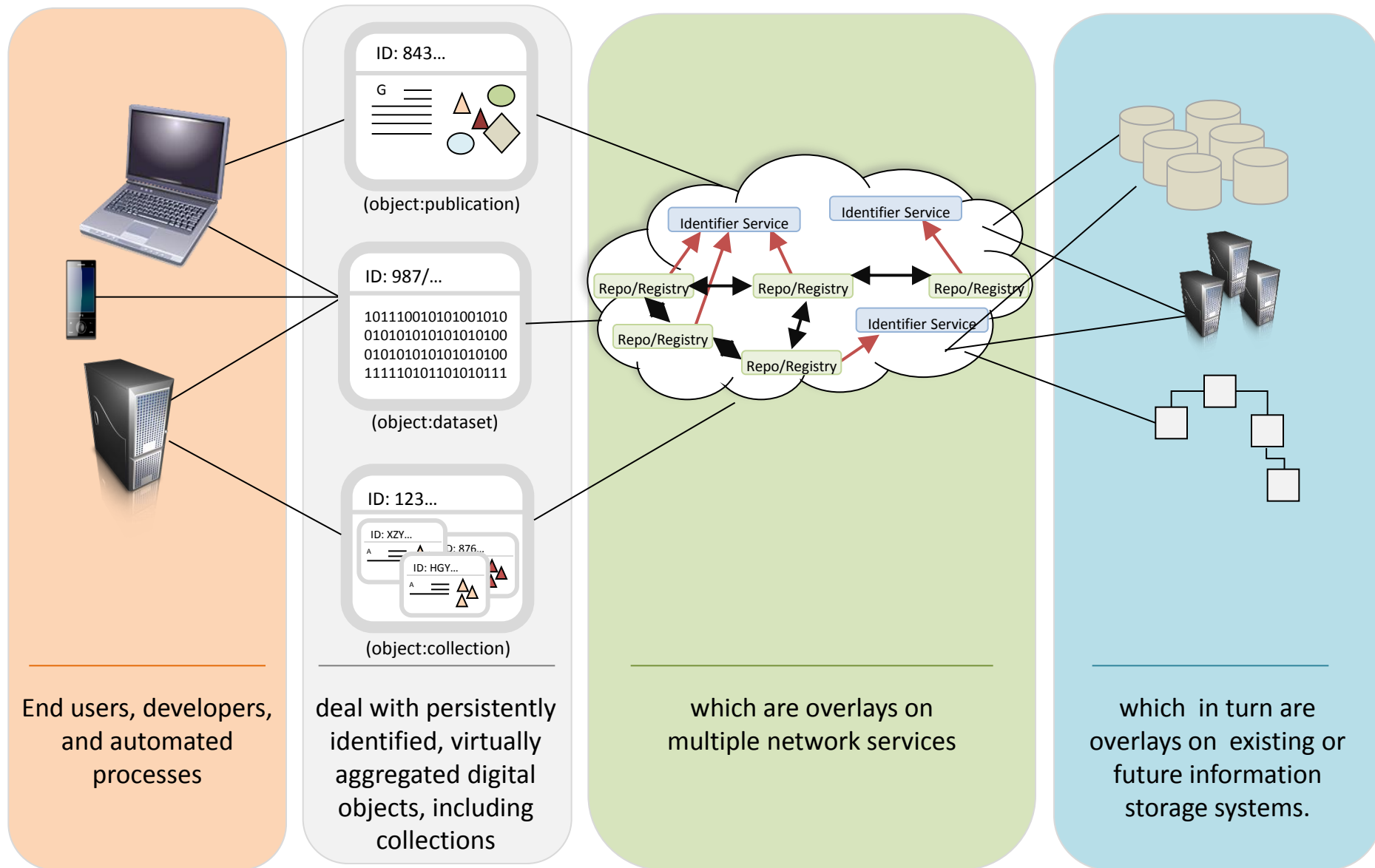
# Data processing perspectives

- Climate data analytics service (for EOSC)
  - Cluster-based, 2 pilot implementations, 2018+



- Copernicus Climate Change Service (C3S)
  - coordinated by ECMWF, operational 2018+
- WPS-based service ecosystem with multiple deployments

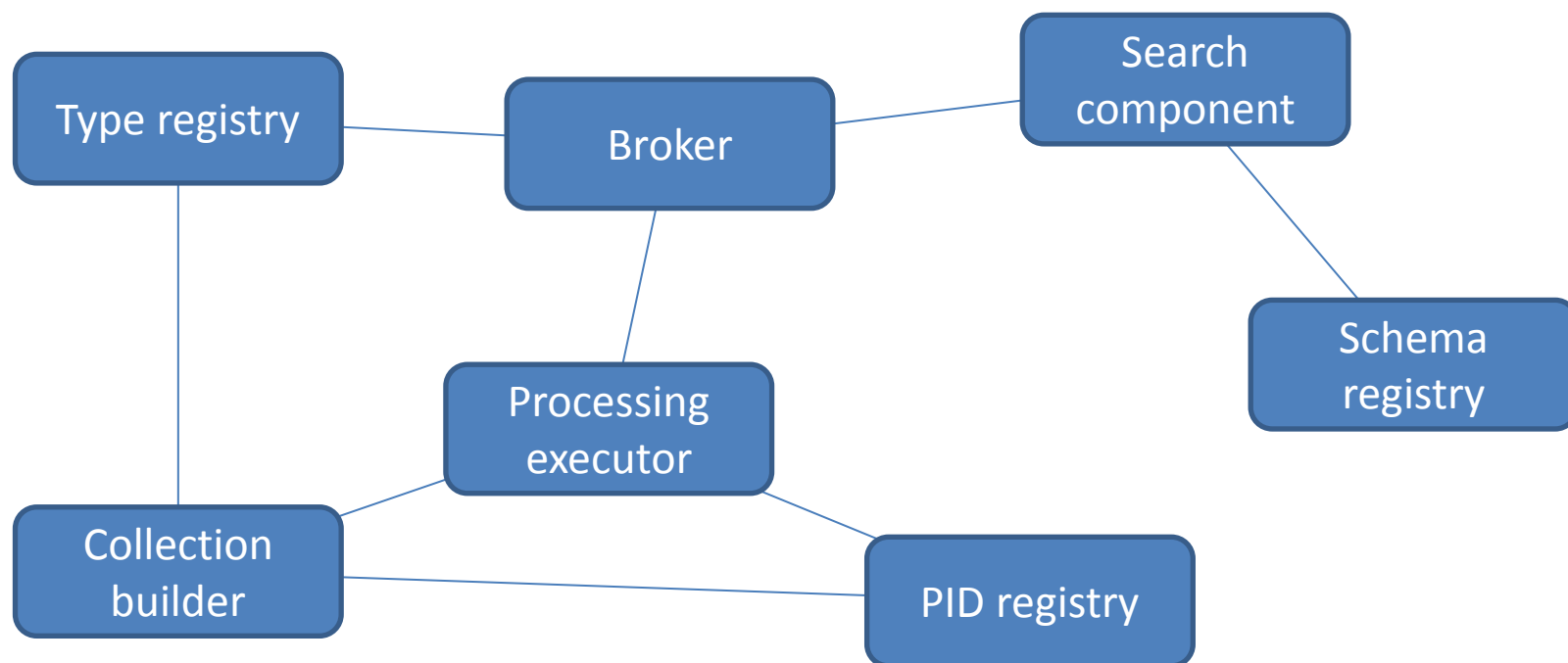
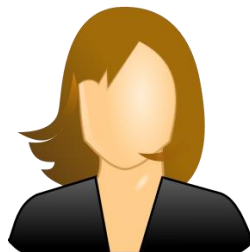
# Global Digital Object Cloud (GDOC)



L. Lannom / DFIG



# GDOC and reusable data service components



Thank you for your attention.

[weigel@dkrz.de](mailto:weigel@dkrz.de)