



Tobias Weigel, Katharina Berger, Stephan Kindermann, Michael Lautenschlager
German Climate Computing Center (DKRZ)

Introduction

Prototype
impression

Data preparation

Versioning for CMIP6 in the Earth System Grid Federation

Initial registration

... and PIDs!

Version updates

EUDAT2
perspective

End-user tools

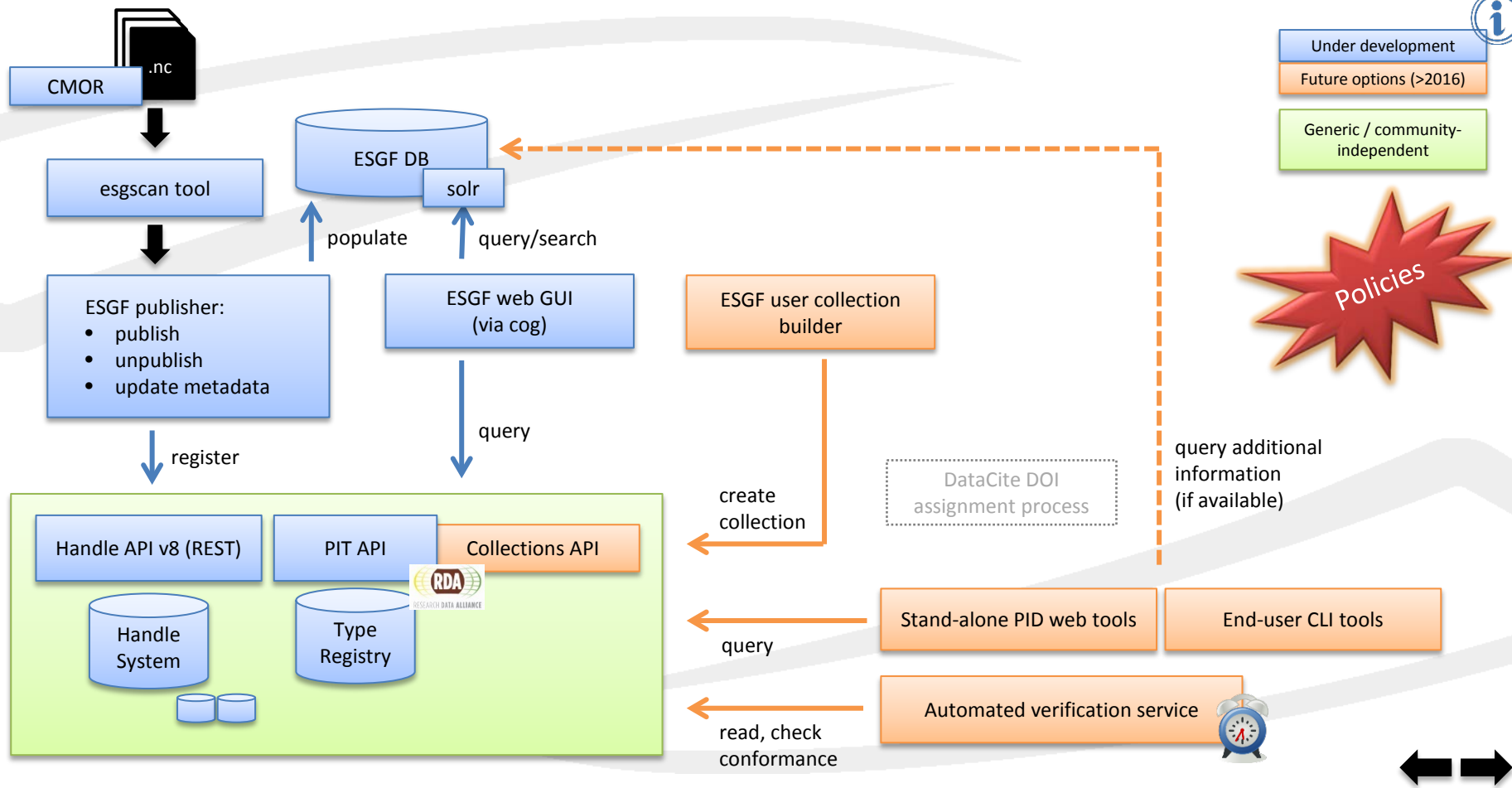
References



Motivation

- No common ESGF approach to versioning, unclear processes
- Demonstrate usefulness of wide-scale low-level PID usage within operational e-infrastructure
- Controlled versioning at this scale will be new for CMIP6





What is required?

- Technical development (esgf publisher)
- Agreement on pioneering nodes
- Definition of **policies** to be enforced
- DKRZ Handle service and future coordination

Until end of 2015...



Home

Essential versioning policies

- Versioning can only be trustworthy if everyone adheres to the policies.
- Enforce use of ESGF tools as opposed to unmonitored changes in the file system
- Unified version numbers: YYYYMMDDxx
 - recommended for all future projects using ESGF
 - mandatory if automated version management is to be used



Prototype impression

[Search Help](#)
[Search Controlled Vocabulary](#)

Home Search Tools Login Help

Current Selections

[remove all](#)
[\(x\) project.obs4MIPs](#)
[\(x\) data_node.bmbf-ipcc-ar5.dkrz.de](#)

Search Categories

- Project
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model

Examples: *temperature, "surface temperature", climate AND project:CMIP5 AND variable:hus*.
 To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

Search All Sites Show All Replicas Show All Versions

< 1 > displaying 1 to 1 of 1 search results

Display datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Results **Data Cart**

Show all Filter over text

[Globus Online All Selected](#) [WGET All Selected](#) [Remove All](#)

cordex.output.EUR-11i.MPI-CSC.MPI-M-MPI-ESM-LR.historical.r1i1p1.REMO2009.v1.day.clt.v20140220**carbon.dkrz.de** [Show Files](#) | [WGET](#) | [Globus Online](#) | [Remove](#)
 (Total Number of Files for All Variables: 12)

obs4MIPs.FUB-DWD.SSMI-MERIS.mon.v20140616**bmbf-ipcc-ar5.dkrz.de** [Hide Files](#) | [WGET](#) | [Globus Online](#) | [Remove](#)
 (Total Number of Files for All Variables: 4)

obs4MIPs.FUB-DWD.SSMI-MERIS.mon.v20140616.prwErr_SSMI-MERIS_L3_v1-00_200301-200812.nc|bmbf-ipcc-ar5.dkrz.de [HTTP](#) | [Globus Online](#) | [OPENDAP](#)
tracking_id: a9b1bfc-4b73-4295-8ed6-6b586bf1be02
checksum: 149ee38e24e819b5d04c34f6ed7b375 (MD5)
pid: http://hdl.handle.net/10876/ESGF/a9b1bfc-4b73-4295-8ed6-6b586bf1be02

File PID

PID := prefix+tracking_id

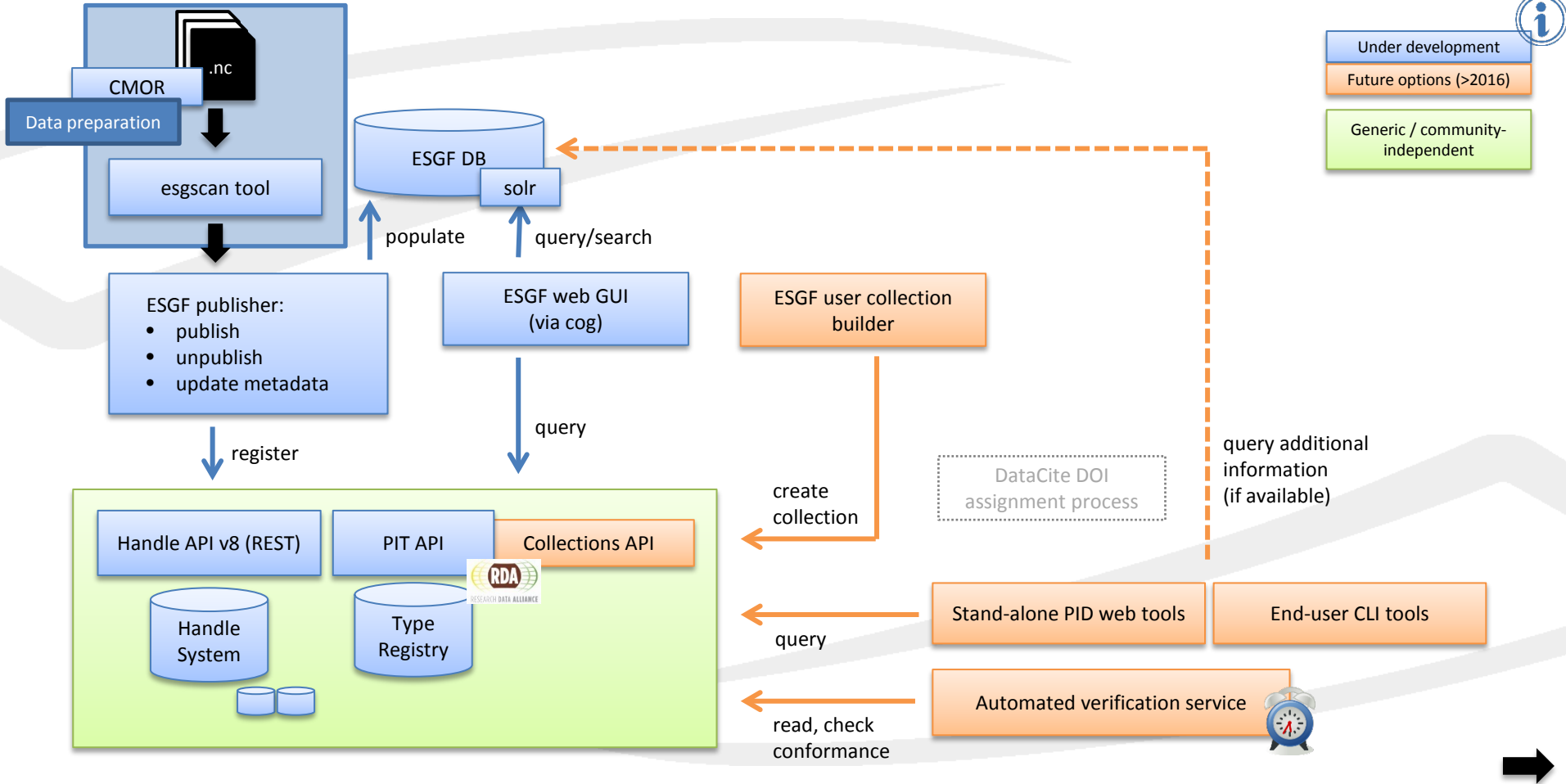
Results **Data Cart**

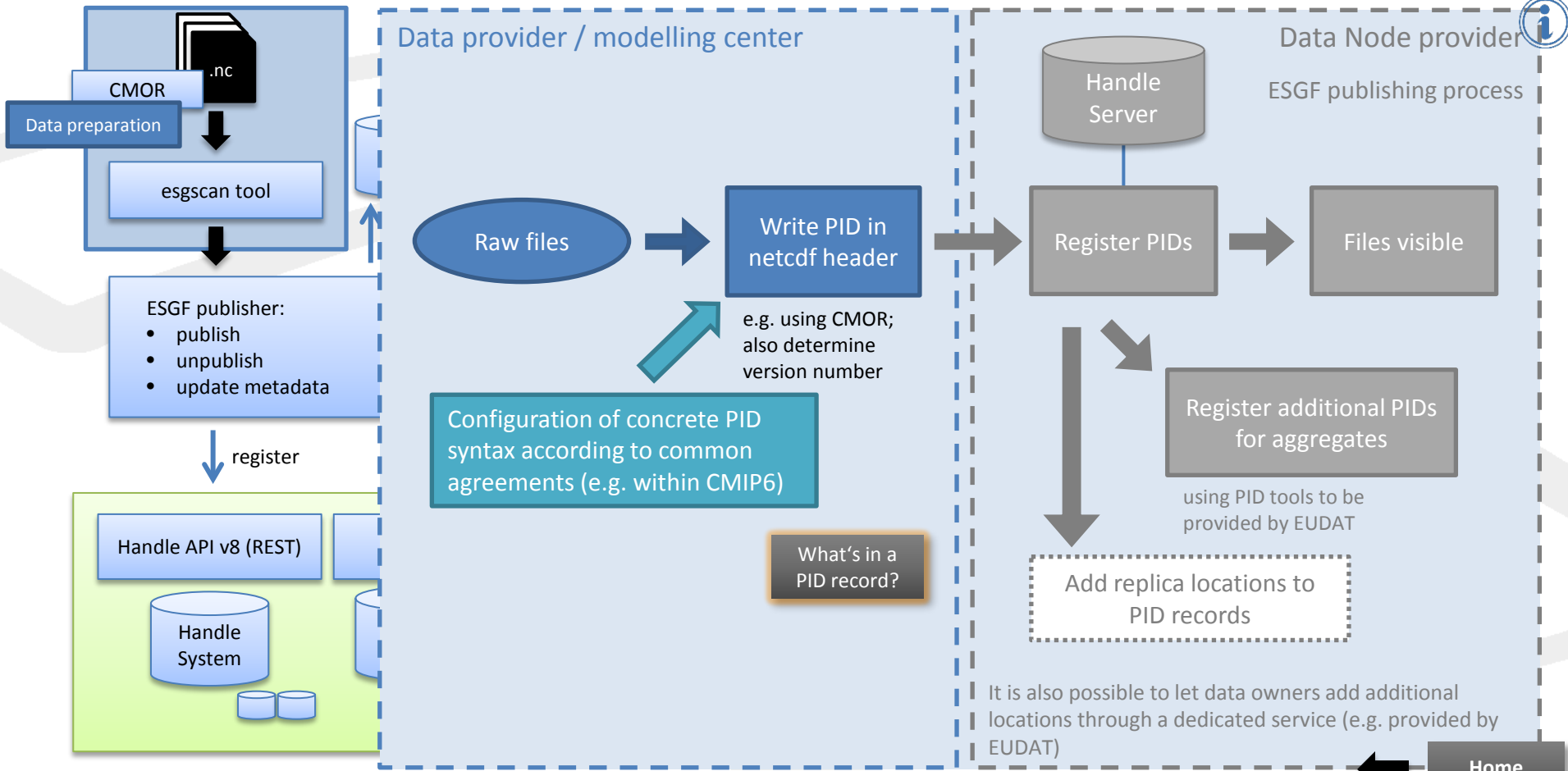
project=obs4MIPs_institute=FUB-DWD_instrument=SSMI-MERIS_time_frequency=mon_variable=prw
 Data Node: bmbf-ipcc-ar5.dkrz.de
Version: 20140616
 Description: obsVapour - Total Column Water Vapour monthly mean from SSMI-MERIS
PID: http://hdl.handle.net/10876/ESGF/4ee9d37b-6454-44bf-b3ef-e738b2ecedb4
 Further options: Add To Cart Technical Note

Dataset PID

What happens when clicking on a PID?

[Home](#)





Home ←



Example PID records (DWD obs4MIPs prototype)

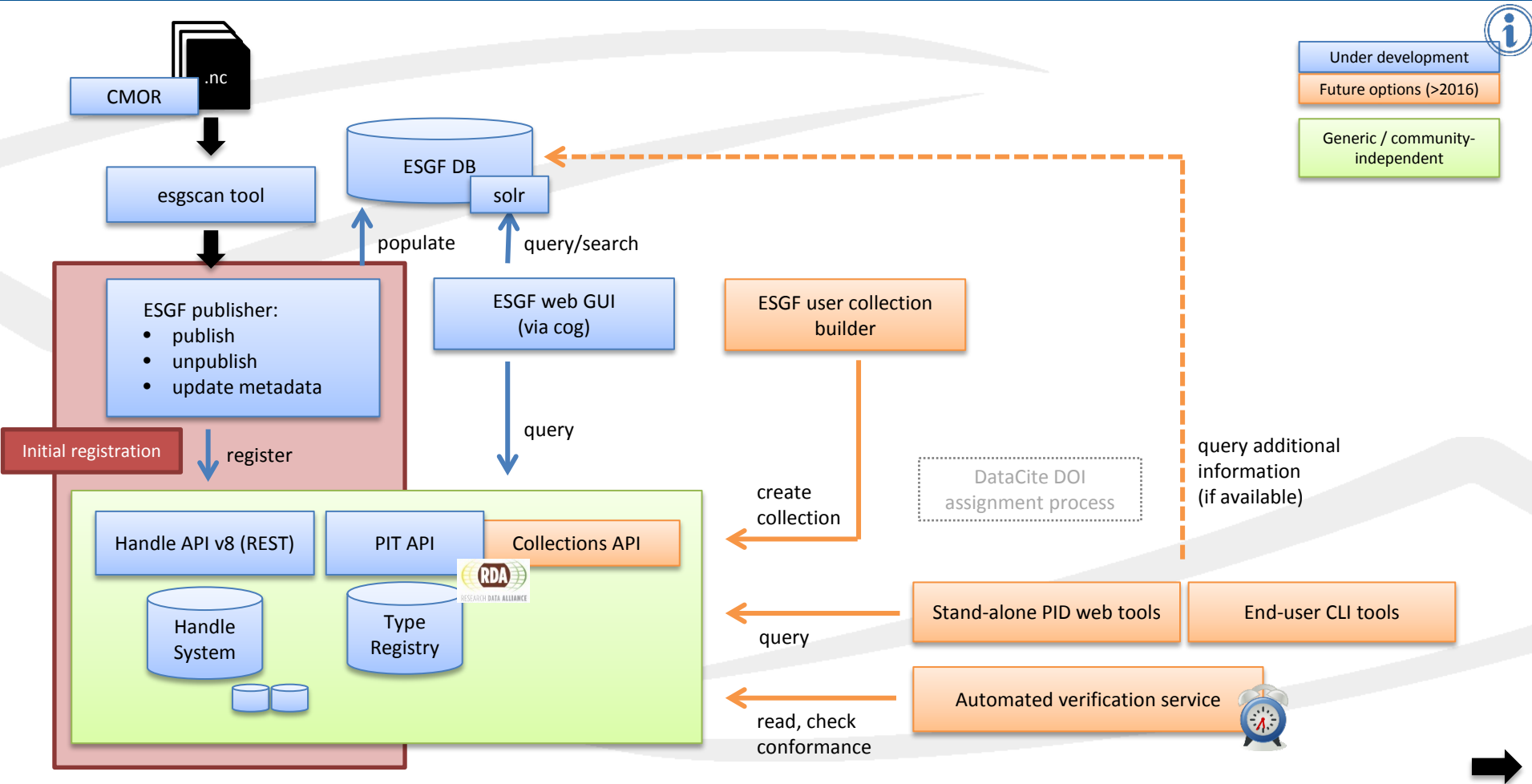
Dataset: 10876/ESGF/4ee9d37b-6454-44bf-b3ef-e738b2ecedb4

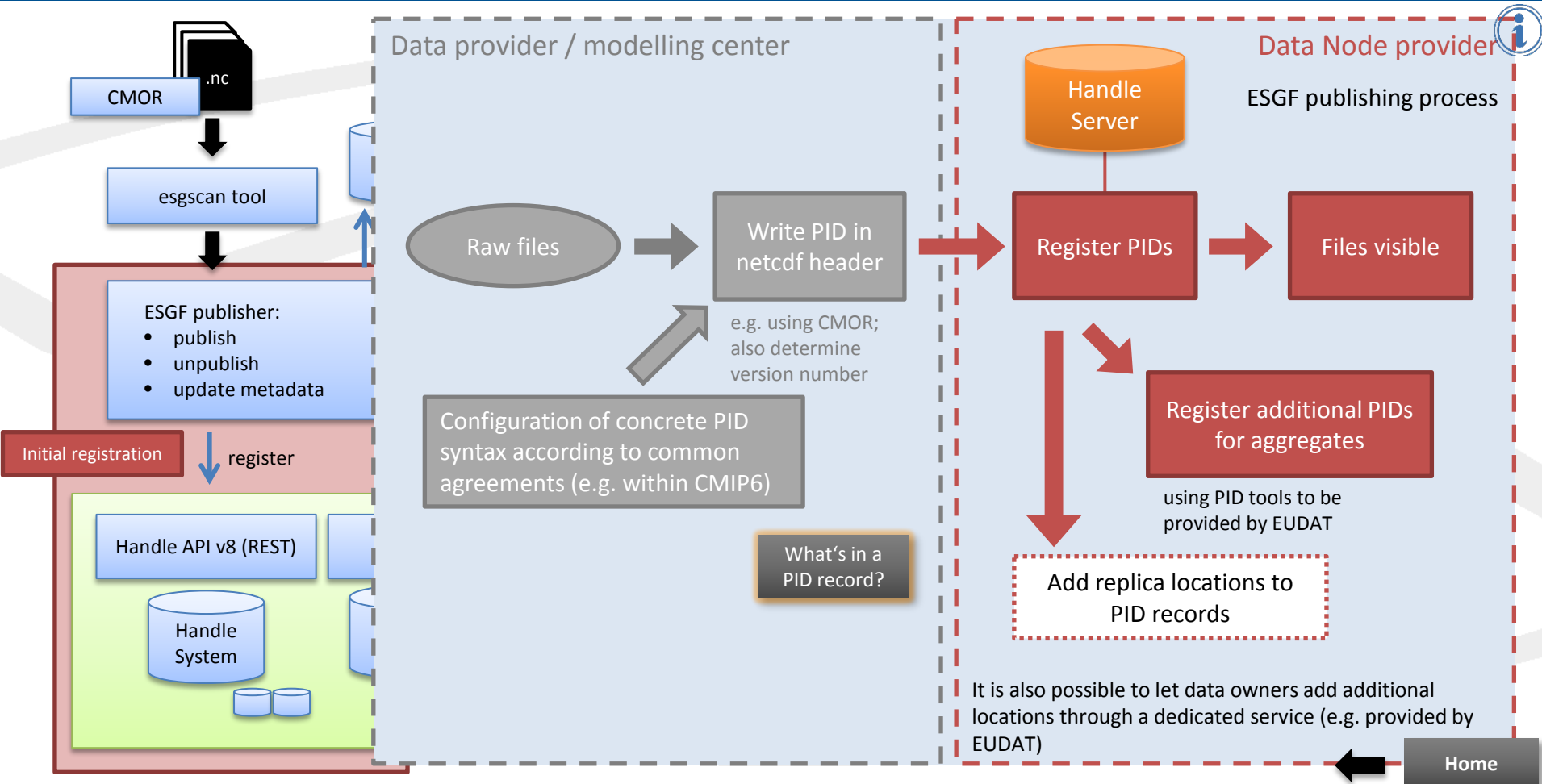
Key	Value
URL	http://bmbf-ipcc-ar5.dkrz.de/thredds/esgcat/3/obs4MIPs.FUB-DWD.SSMI-MERIS.mon.v20140616.html
DRS name	obs4MIPs/observations/FUB-DWD/Obs-SSMI-MERIS/obs/mon/atmos/prw
Publication date	2014-06-16
Version number	20140616
Children	["10876/ESGF/a9b1bfbc-4b73-4295-8ed6-6b586bf1be02", ...]

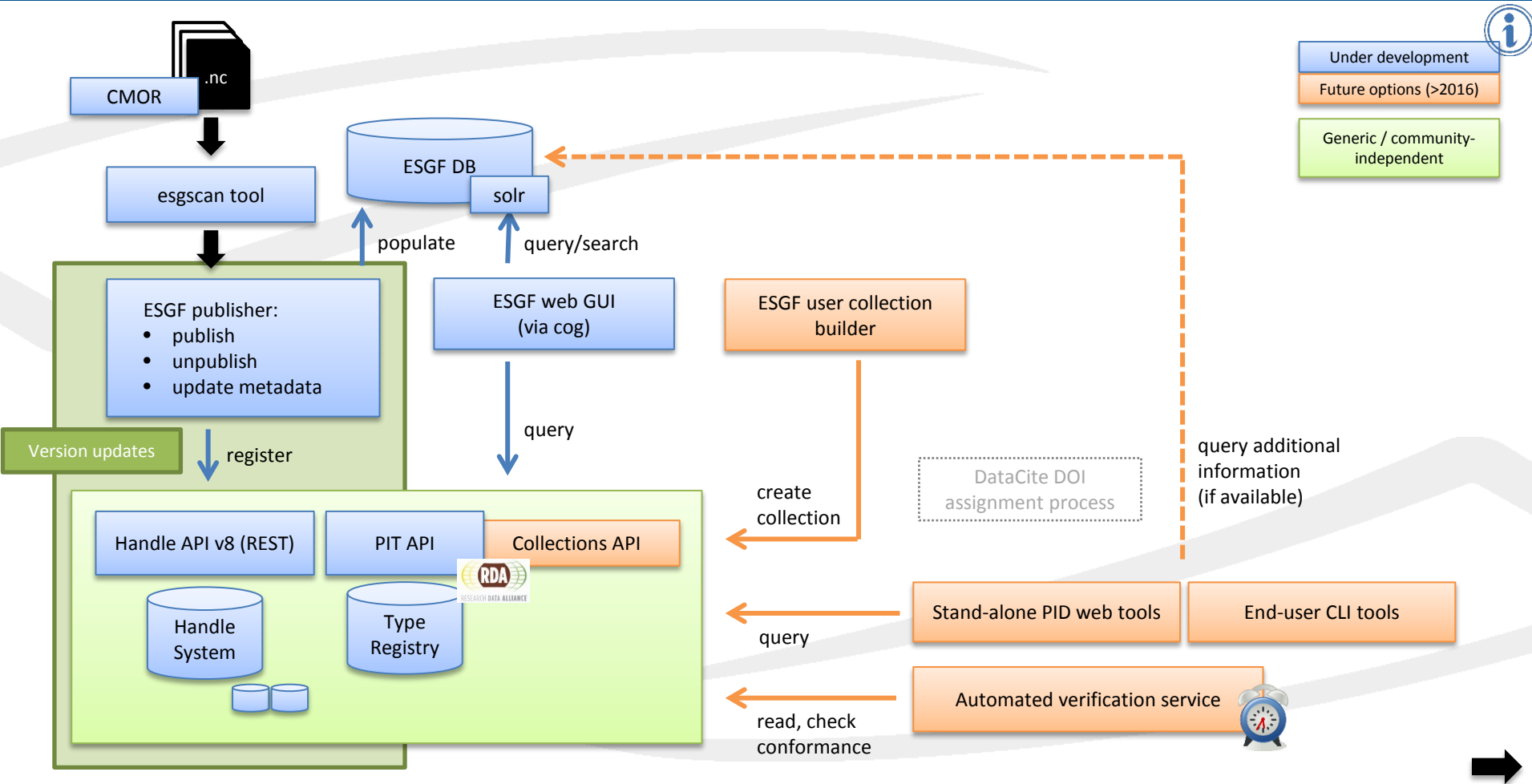
File: 10876/ESGF/a9b1bfbc-4b73-4295-8ed6-6b586bf1be02

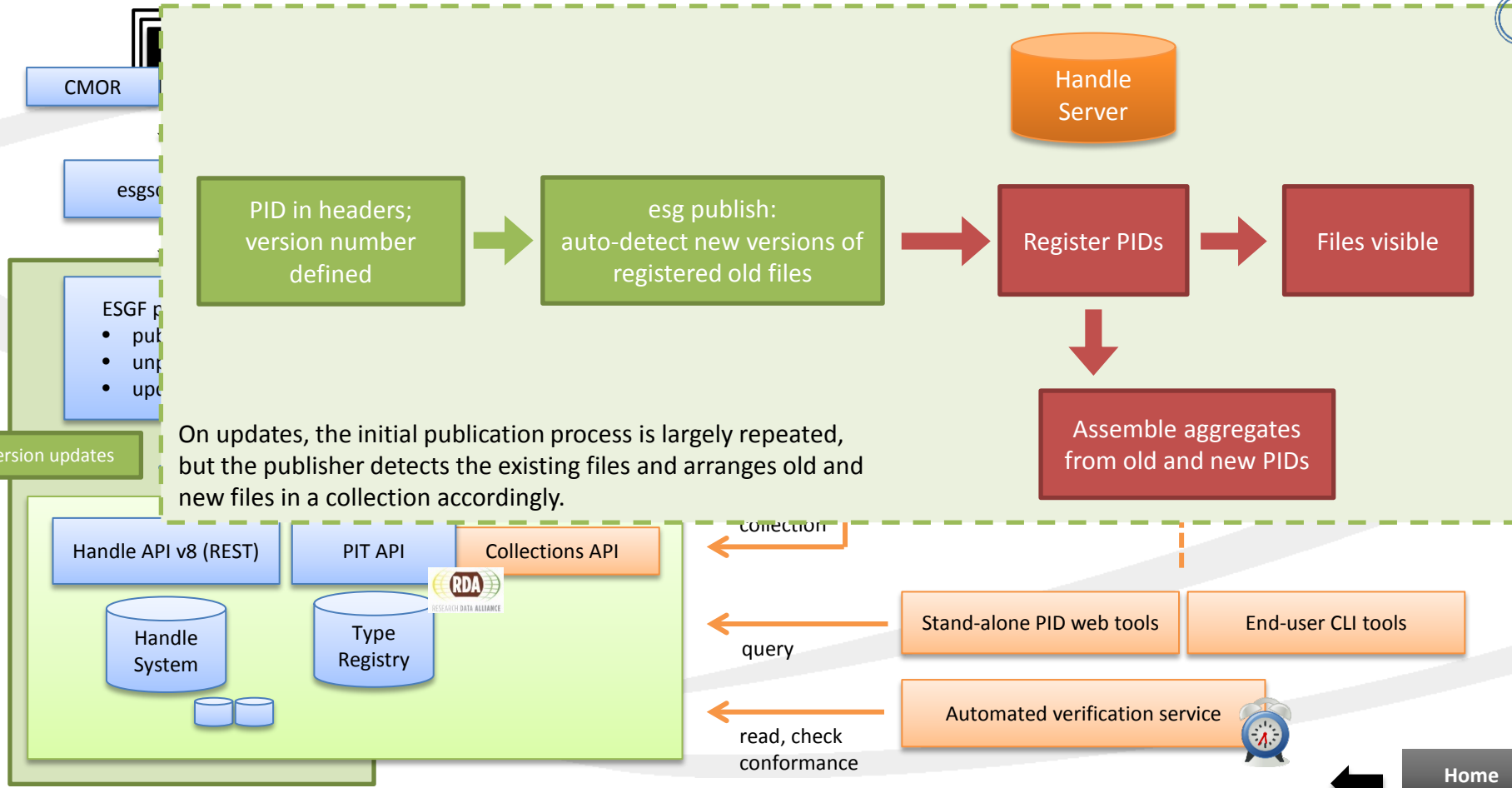
Key	Value
URL	http://bmbf-ipcc-ar5.dkrz.de/thredds/fileServer/obs4MIPs/observations/FUB-DWD/Obs-SSMI-MERIS/obs/mon/atmos/prw/prwErr_SSMI-MERIS_L3_v1-00_200301-200812.nc
DRS name	prwErr_SSMI-MERIS_L3_v1-00_200301-200812.nc
Publication date	2014-06-16
Checksum (MD5)	F49ee38e24e819b5d04c534f6ed7b375
Size	50989760
Parent	10876/ESGF/4ee9d37b-6454-44bf-b3ef-e738b2ecedb4

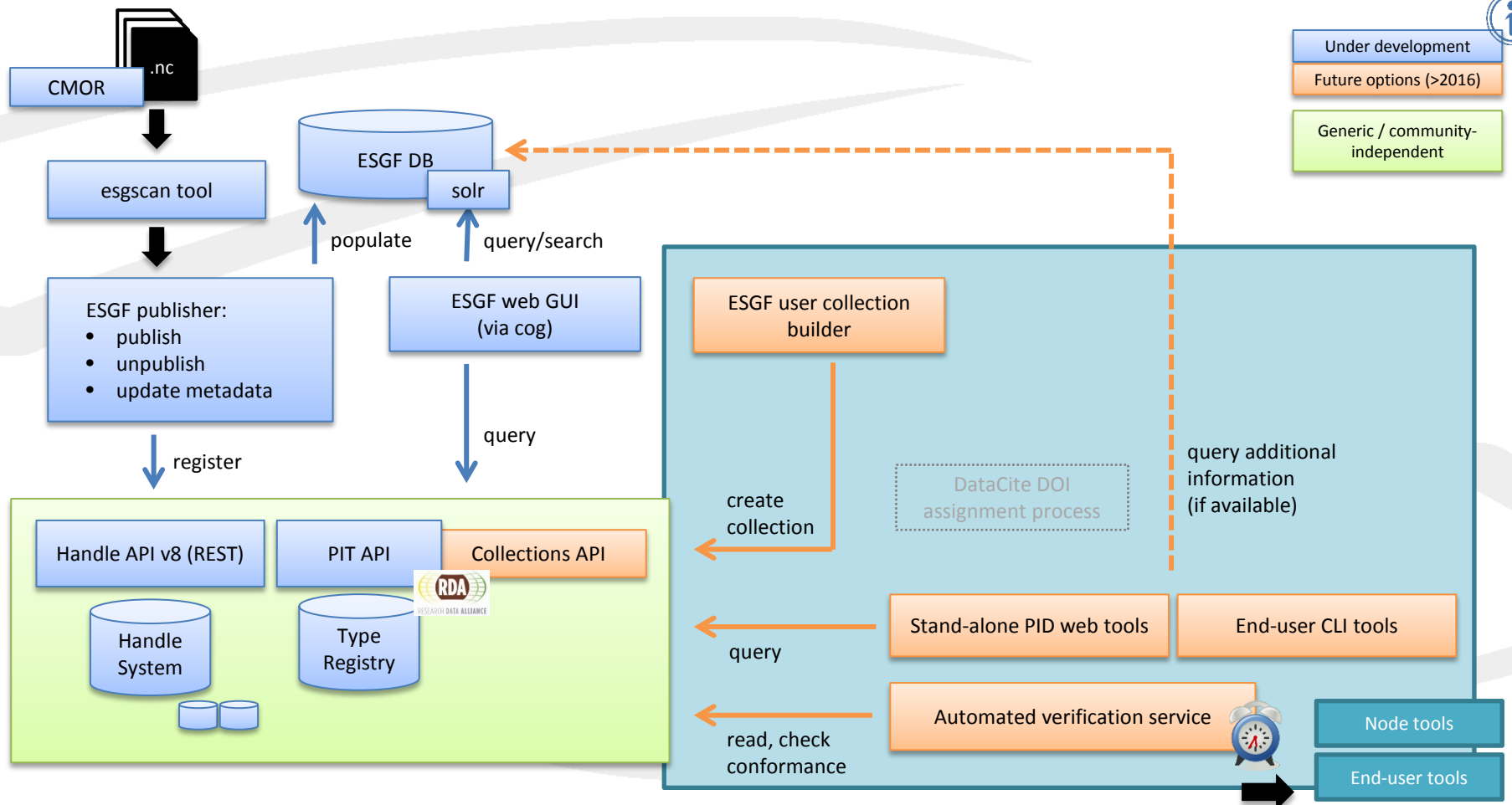
Back













PID quality management



Automated PID verification service



Issue manager

Determine action

Mark PID as tombstone; provide tombstone record info

Possibly include reference to new version/replacement

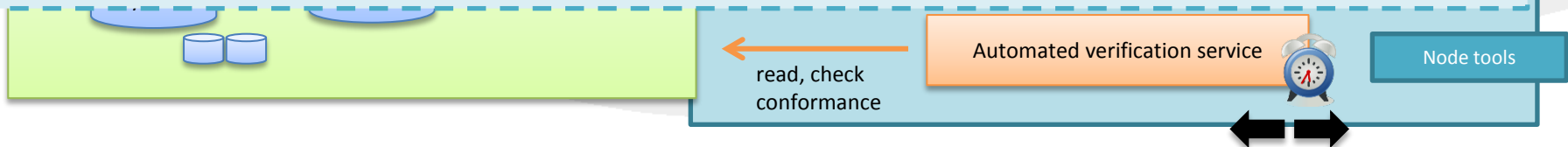
Handle Server

Update PID record with new location

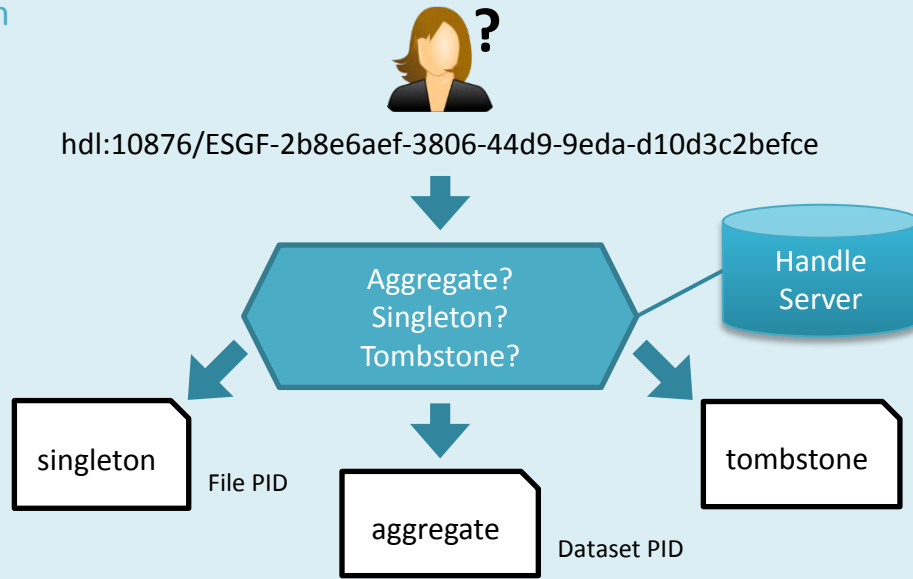
... based on additional knowledge to be acquired

Other external trigger factors

Parts of this process should be supported by EUDAT/ESGF tools to make it more scalable and reduce current manual effort



Basic PID resolution



Individual information page services provided by ESGF based on EUDAT tools

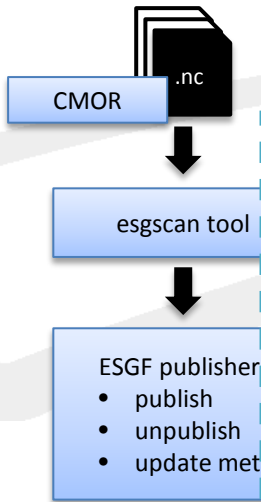
Web landing pages could offer: data download, versioning information, replication information, ...

- Under development
- Future options (>2016)
- Generic / community-independent

ional

er CLI tools

End-user tools



Possible command line tools:

- wget for PID'ed data with smooth authentication
- info tool
- get latest version
- ... all Python-based!

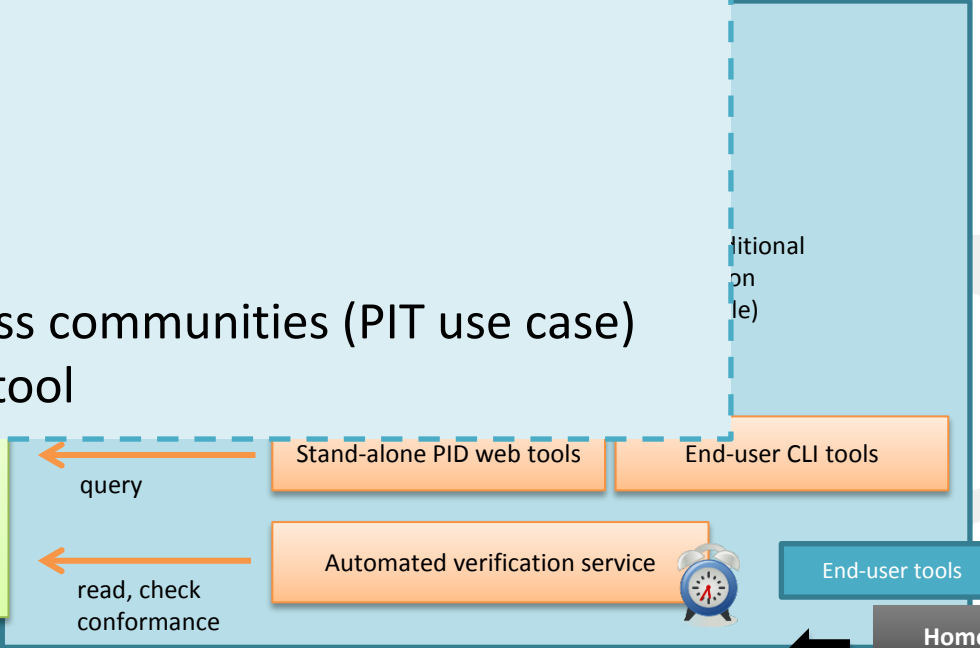
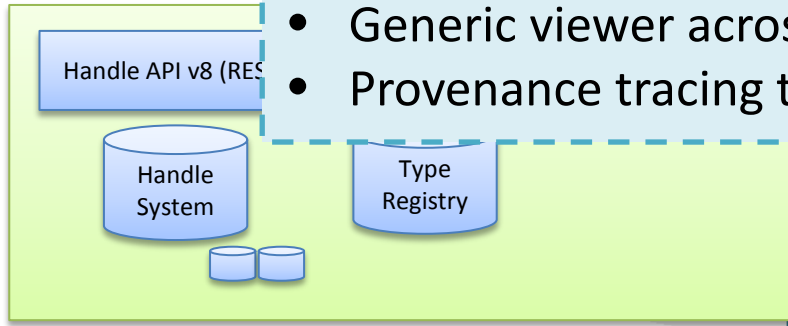
Possible web tools:

- Generic viewer across communities (PIT use case)
- Provenance tracing tool

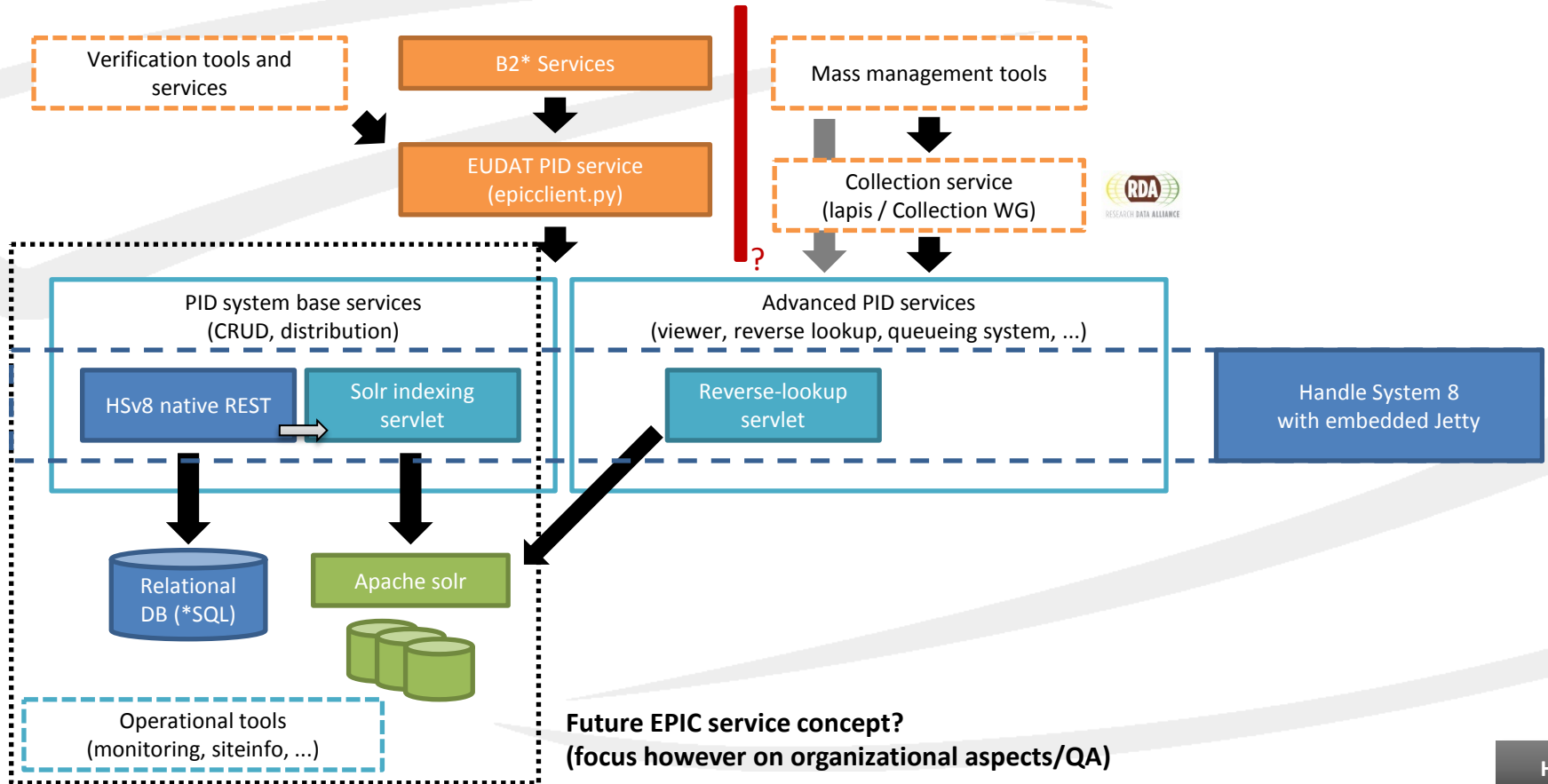
Under development

Future options (>2016)

Generic / community-independent



Envisioned EUDAT2 PID services architecture



Index

- ➔ ■ Home
 - ➔ ■ Motivation
 - ➔ ■ Architecture overview
 - ➔ ■ Requirements
 - ➔ ■ Policies
- ➔ ■ Prototype impression
- ➔ ■ Data preparation
 - ➔ ■ Modelling center perspective
 - ➔ ■ Example PID records
- ➔ ■ Initial registration
 - ➔ ■ Data node perspective
- ➔ ■ Version updates
 - ➔ ■ Version update process
- ➔ ■ End-user tools
 - ➔ ■ PID quality management
 - ➔ ■ Basic PID resolution
 - ➔ ■ Possible CLI and web tools
- ➔ ■ EUDAT2 perspective
- ➔ ■ References



References

- Meehl, Moss, Taylor, Eyring, Stouffer, Bony, Stevens (2014): Climate Model Intercomparisons: Preparing for the Next Phase. EOS Trans. AGU, Vol. 9, No. 9. doi:10.1002/2014eo090001
- Weigel, Lautenschlager, Toussaint, Kindermann (2013): A Framework for Extended Persistent Identification of Scientific Assets. Data Science Journal, Vol. 12. doi:10.2481/dsj.12-036
- Weigel, Kindermann, Lautenschlager (2013): Actionable Persistent Identifier Collections. Data Science Journal, Vol. 12. doi:10.2481/dsj.12-058
- Weigel, DiLauro, Zastrow: RDA Recommendation: PID Information Types. Under review.

- <http://www.handle.net>
- <http://cnri.reston.va.us>