

c2camp - ENES

RDA P13 SIDE MEETING

ENES: CMIP6 data use case

CMIP6: Climate Model Intercomparison Project, phase 6

Data infrastructure support: Earth System Grid Federation (ESGF)

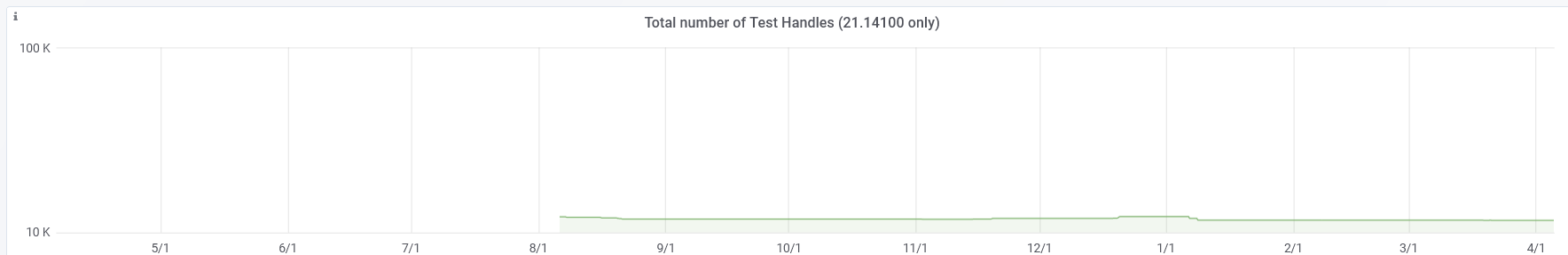
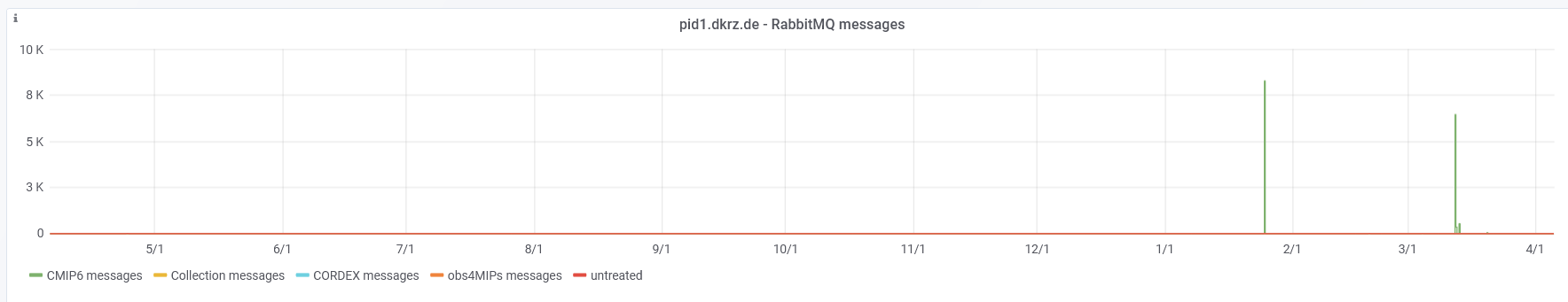
All files and datasets receive Handles as part of e-infrastructure workflow

- 100 mio+ objects – scalability concerns

Technical setup:

- 20+ publishing nodes -> RabbitMQ -> Handle server(s)
- Motivation: Reliability, scalability
- Upcoming change: Handle assignment will be put on critical path





Search the menus (Alt+/)

100% £ % .0 100 123 Arial 10 B I S A

fx

Short URL: <https://goo.gl/EtxDmo>

	A	B	C	D	E	F	G	H	I	
1	Short URL: https://goo.gl/EtxDmo	Model ready	Spin up started	Deck simulation started	Post processing workflow ready	Data ready for ESGF	Data on ESGF			
2	IPSL-CM6A-LR	Yes	Yes	Yes	yes (dr2xml->online)	Yes	Yes			
3	IPSL-CM6A-MR	Testing	NO	NO	Almost	NO	NO			
4	CCCma	Testing	Testing	No	Almost	No	No			
5	CCCR-IITM	Almost	Yes	NO	Almost	NO	NO			
6	E3SM-project (DOE)	Yes	Yes	Yes	Yes	Yes	17 variable subset			
7	NASA GISS	Yes (GISS-E2.1, late 2017)	Yes	Yes	Yes	Yes	Yes			
8	CNRM-CM6-1	Yes	Yes	Yes (march 2018)	yes (dr2xml->online)	Yes	Yes			
9	CNRM-ESM2-1	Yes	Yes	Yes (march 2018)	yes (dr2xml->online)	yes	Yes			
10	CNRM-CM6-1-HR	Yes	Yes	NO	yes (dr2xml->online)	NO	NO			
11	MOHC (HadGEM3-GC31-LL)	Yes	Yes	Yes	Almost	NO	NO			
12	MOHC (HadGEM3-GC31-MM)	Yes	Yes	Yes	Almost	NO	NO			
13	MOHC (UKESM1-0-LL)	Yes	Yes	Yes	Almost	NO	NO			
14	NOAA GFDL (CM4)	Yes	Yes	Yes	Yes	Yes	500+ datasets			
15	NOAA GFDL (ESM4)	Almost	NO	NO	Almost	NO	NO			
16	MPIESM1 (LR)	Almost	Yes	NO	Almost	NO	NO			
17	MPIESM1 (HR)	Yes	Yes	Yes	Almost	NO	NO			
18	MPIESM2	NO	NO	NO	NO	NO	NO			
19	DLR	Yes	Almost (August 2017)	Almost (August 2017)	Almost	NO	NO			
20	NorESM	Almost	NO	NO	Almost	NO	NO			
21	AWI-CM	Yes (since August 2017)	Yes (since August 2017)	NO	NO (August 2017)	NO	NO			
22	NCAR-CESM	Yes	Yes	Yes	Yes	Yes	Starting			
23	CMCC-ESM2	Almost	NO (fall 2017)	NO	Almost	NO	NO			
24	CMCC-CM2-HR4	Yes	Yes	NO	Almost	NO	NO			
25	CMCC-CM2-HR5	NO	NO	NO	NO	NO	NO			
26	CMCC-CM2-SR5	Yes	Yes	NO	Almost	NO	NO			
27	EC-EARTH	NO	NO (late 2017)	NO	NO	NO	NO			
28	CAS FGOALS-F3	Yes	Yes	Yes	Almost	NO	NO			
29	CAS FGOALS-G3	Almost	NO	NO	Almost	NO	NO			
30	CAS ESM	Almost	Almost (since August 2017)	Almost (August 2017)	NO	NO	NO			
31										
32										

+ Sheet1

Explore

c2camp for CMIP6

Current benefits:

- Established better practice for dealing with versions, replicas, and enable status tracking via PID Kernel Information

Desired future benefits:

- Tracking data usage – CMIP6 referencing tool planned
- Automating data replication – white paper drafted

RDA Recommendation on PID Kernel Information

7 Guiding Principles for PID Kernel Information

- Independent of specific infrastructure or technologies
- Geared towards minimizing human interaction, long-term stability of processes relying on Kernel Information

Draft Kernel Information profile

Exemplary high-level architecture

Use cases and community adoption

<https://doi.org/10.15497/rda00031>

Migration of records

For adoption, migration of existing and continuously created records is required. Timeline currently unclear.

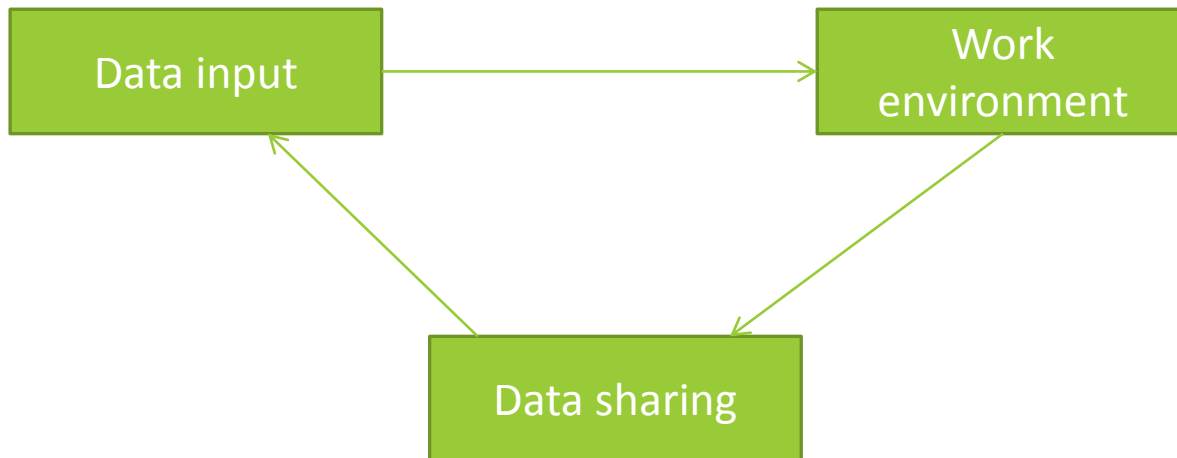
Current structure	PID KI profile	Comments
	KernellInformationProfile	
	digitalObjectType	
URL_ORIGINAL_DATA	digitalObjectLocation	
FIXED_CONTENT	digitalObjectPolicy	Special policy record format as in Recommendation
CHECKSUM	etag	
	dateModified, dateCreated	Can be requested from solr index
VERSION_NUMBER	version	
REPLACES	wasRevisionOf	

Data analytics: Short loop

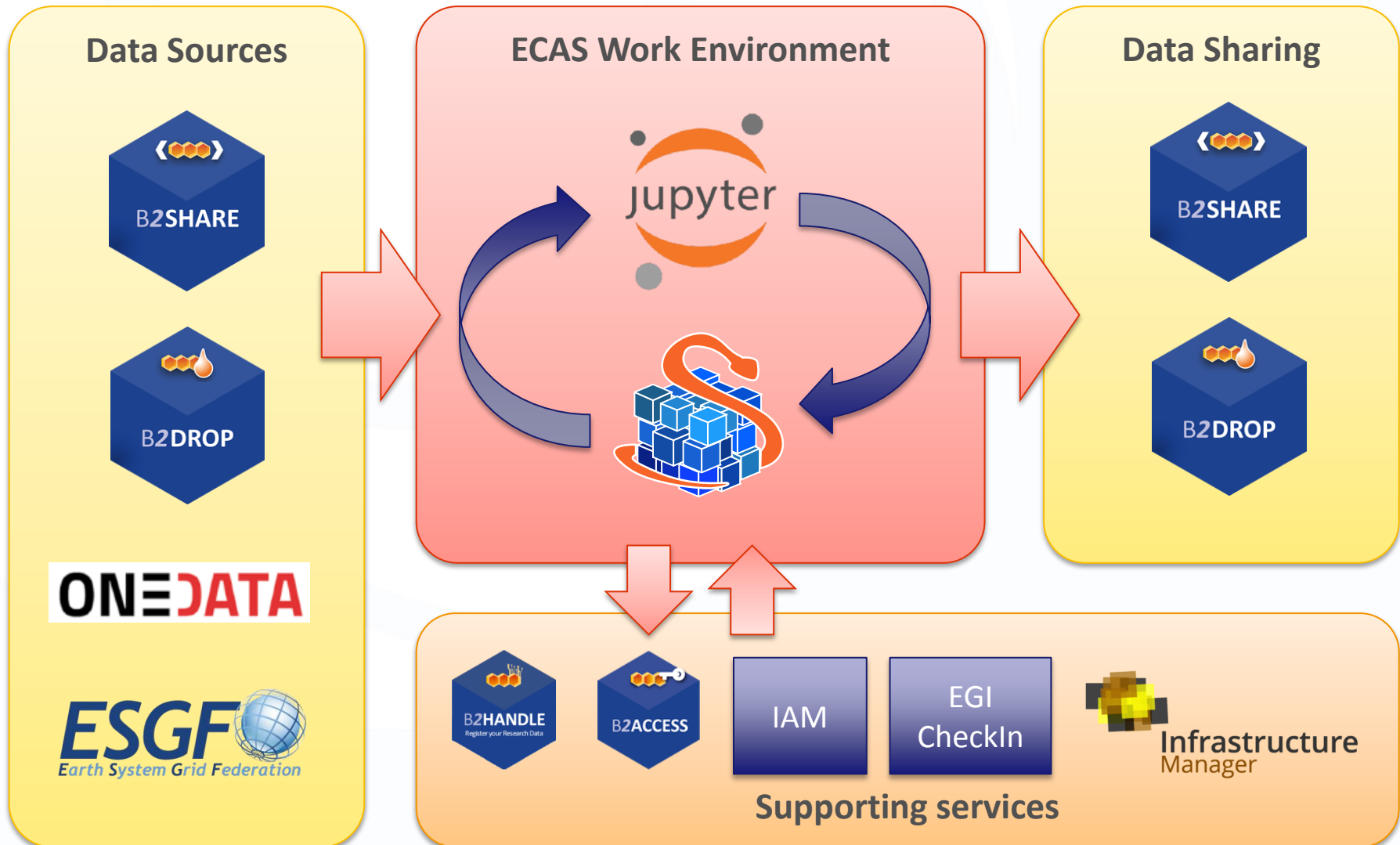
Opportunity: practical demonstration of key c2camp ideas in limited area

- DOIP, Kernel Information, collection management, process automation

Also brings ML topics into c2camp?



EOSC-hub Service architecture and interfaces



Future perspectives for RDA work from P13

Support for PID Kernel Information by non-Handles

- Possible session with PID IG at P14?

PID KI and DTR adoption by services in practice, focusing on automation aspects

- Focus of Data Fabric IG at P14?

New group work:

- Type and KI profile management and governance, KI + Collections
- PID policy registries