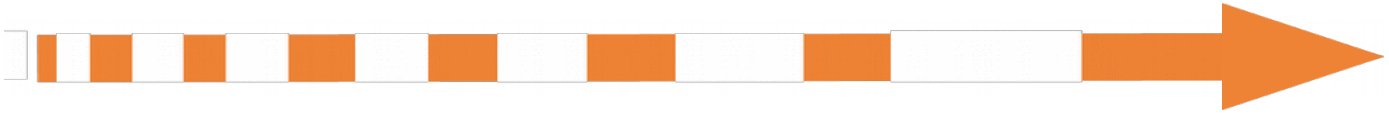
Data Management in XDC



Paul Millar paul.millar@desy.de

… **with slides stolen from:** Daniele Cesini, Fernando Aguilar, Patrick Fuhrmann, Michael Schuh

My thanks go to them!

eXtreme DataCloud

Data Management for extreme scale computing

eXtreme DataCloud is co-funded by the Horizon2020 Framework Program – Grant Agreement 777367

XDC Objectives

- The eXtreme DataCloud is a software development and integration project

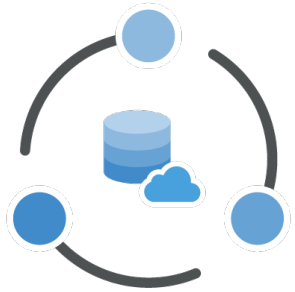
- We are developing scalable technologies for federating storage resources and managing data in highly distributed computing environments
  - Focus efficient, policy driven and Quality of Service based DM

- The targeted platforms are the current and next generation e-Infrastructures deployed in Europe
  - European Open Science Cloud (EOSC)
  - The e-infrastructures used by the represented communities

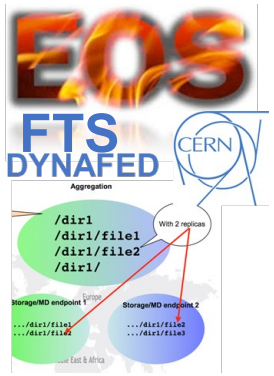
XDC Foundations

- **XDC is based on**
  - the INDIGO-DataCloud data management activity
  - the experience of the project partners on data-management
- **Improve existing, production quality, federated data management services, by…**
  - adding missing functionality requested by research communities,
  - harmonising coherently across European e-Infrastructures.

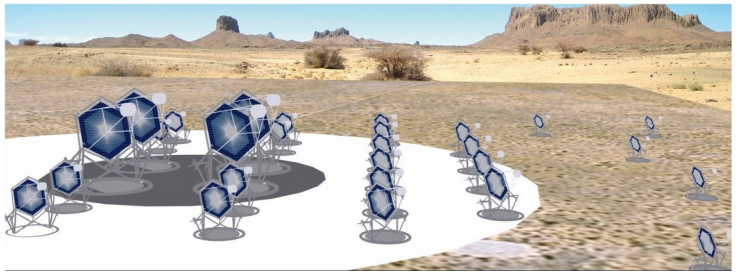
INDIGO PaaS Orchestrator

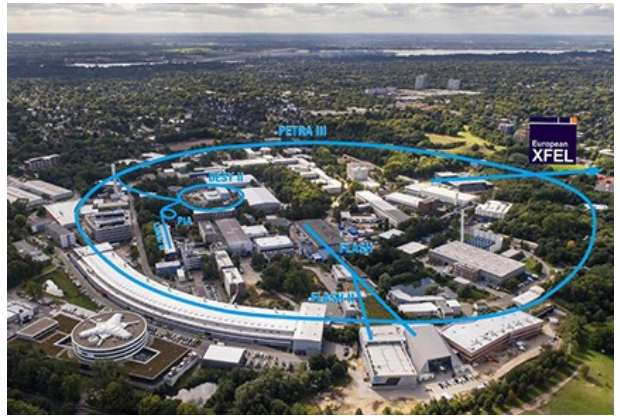


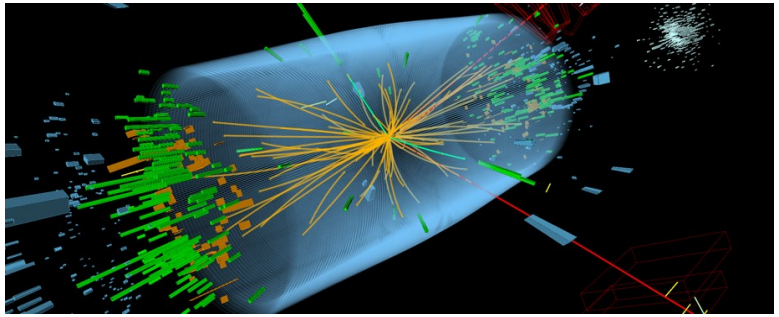
INDIGO CDMI Server

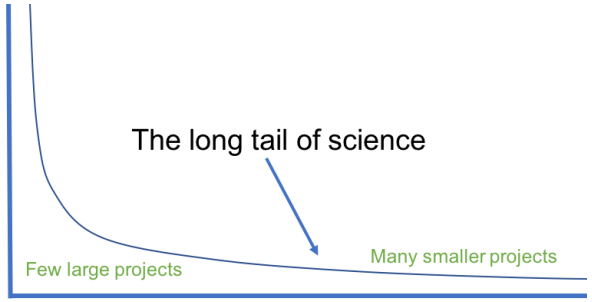


Represented research communities









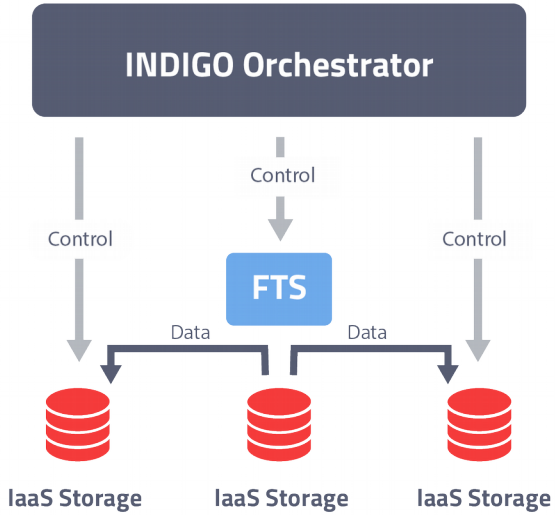
The New Functionalities

- Intelligent & Automated Dataset Distribution
  - Orchestration to realise a policy-driven data management
  - Data distribution policies based on Quality of Service (i.e. disks vs tape vs SSD) supporting geographical distributed resources (cross-sites)
  - Software lifecycle management
- Processing, automatically triggered when data is ingested.
- Data management based on access patterns; e.g.,
  - move unused data to slower, cheaper storage,
  - move “hot” data to faster, more expensive storage.
- Smart data caching
- Metadata management
- Sensitive data handling: secure storage and encryption

Policy driven Data Management

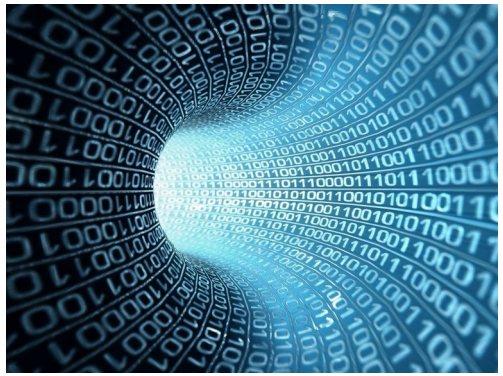
Intelligent & Automated Dataset Distribution

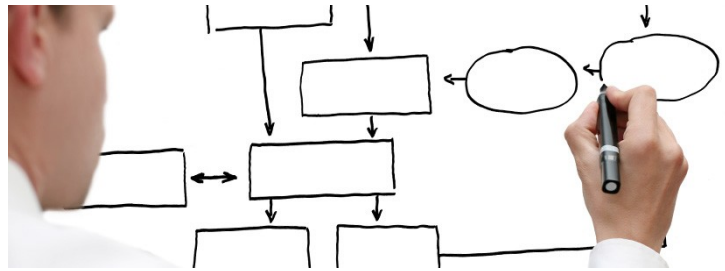
- A typical workflow
  - Initially the data is stored on low latency devices for fast access
  - To ensure data safety, the data will be replicated to a second storage device and will be migrated to custodial systems (tape or S3 appliances)
  - Eligible users will get permission to restore archived data as necessary
  - After a grace period, Access Control will be changed from “private” to “open access”
- Data management based on access pattern

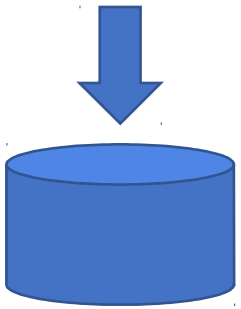


Data pre-processing during ingestion

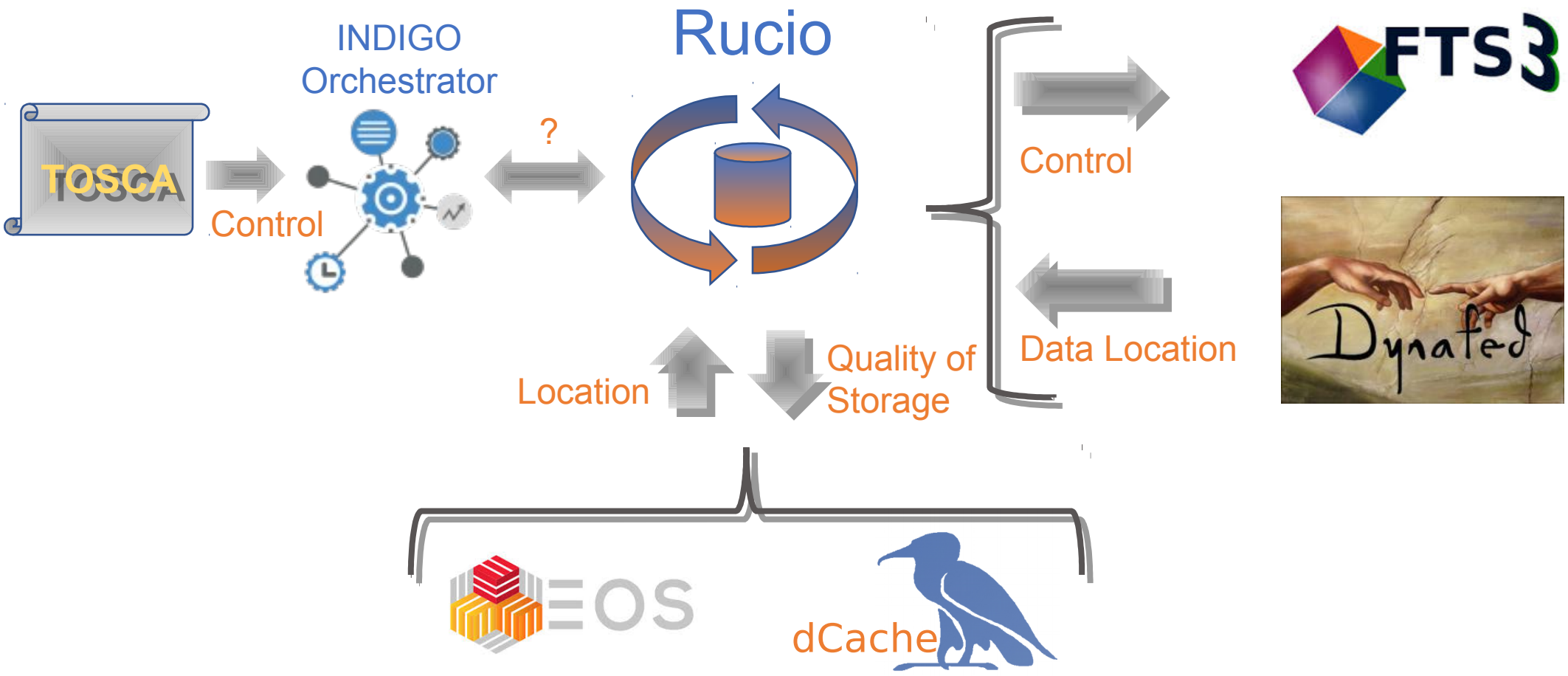
- Automatically run user-defined applications and workflows when data is uploaded
  - e.g., skimming, indexing, metadata extraction, consistency checks
- The solution will discover new data at specific locations.
- INDIGO Orchestrator is used to execute user-defined application and workflows on scalable resources.
- Support writing the results (derived data).





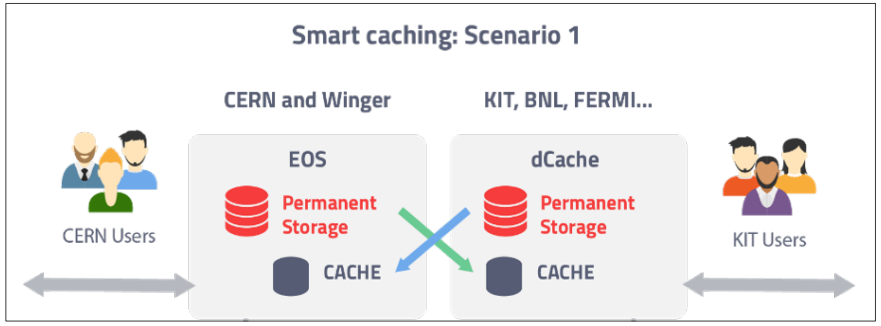


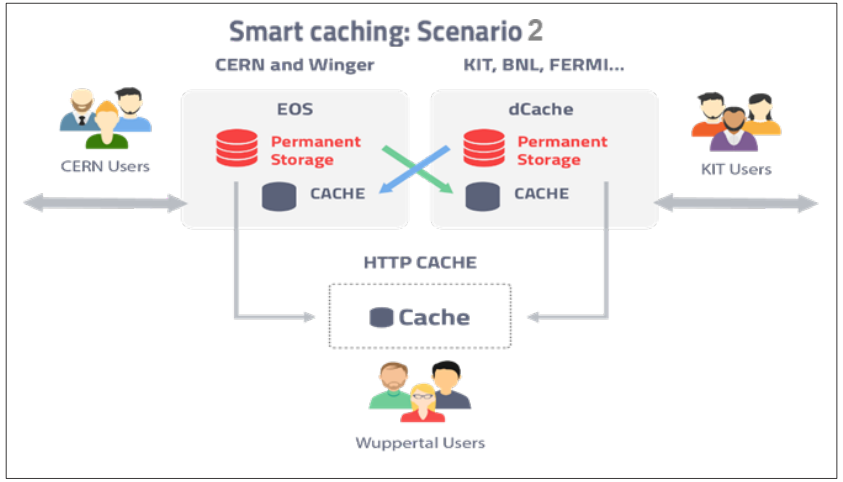
Orchestrating data flow

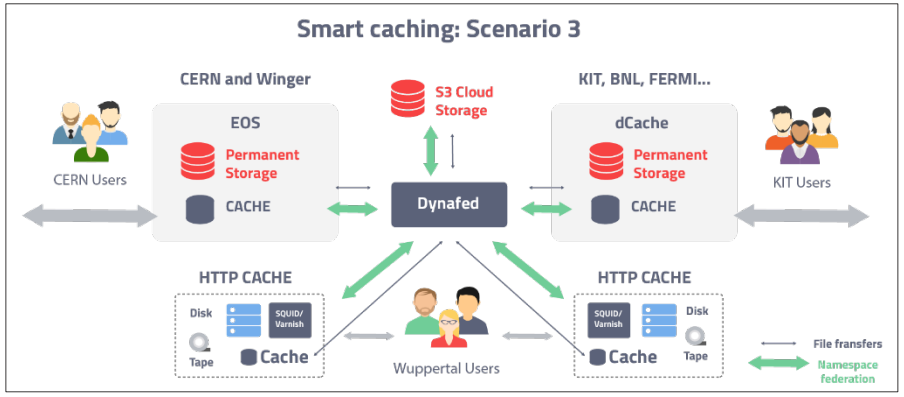


Smart caching

- Develop a global caching infrastructure for supporting the following building blocks:
  - Dynamic integration of satellite sites by existing data centres
  - Creation of stand-alone caches modelled on existing web solutions
  - Federatio of the above to create a large scale caching infrastructure

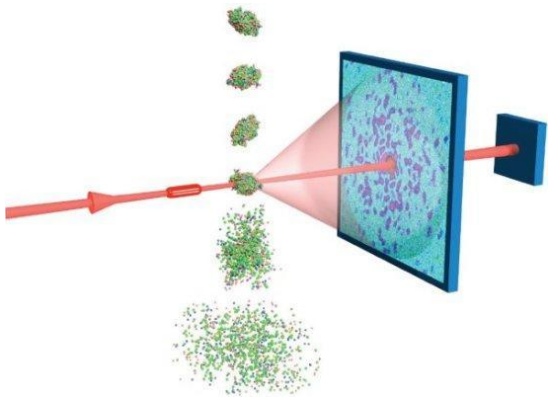


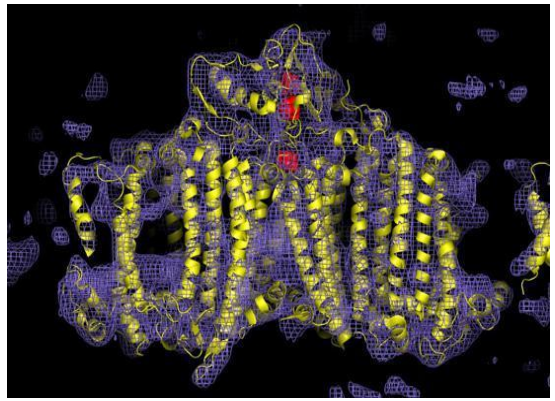


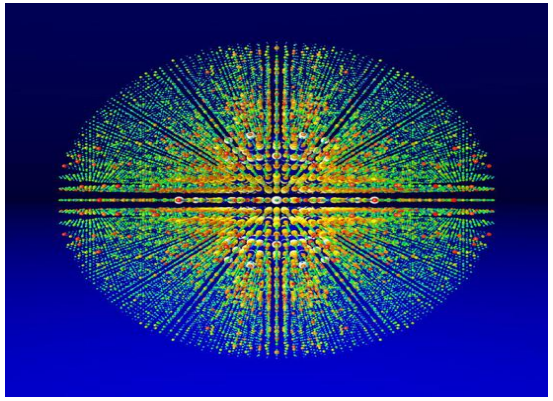


Photon Science @ (or close to) DESY

- Photon Sources
  - European XFEL
  - FLASH
  - PETRA III
- Science
  - Health and Biology
  - Crystallography
  - Energy
  - Catalytic Processes
  - Information Technology

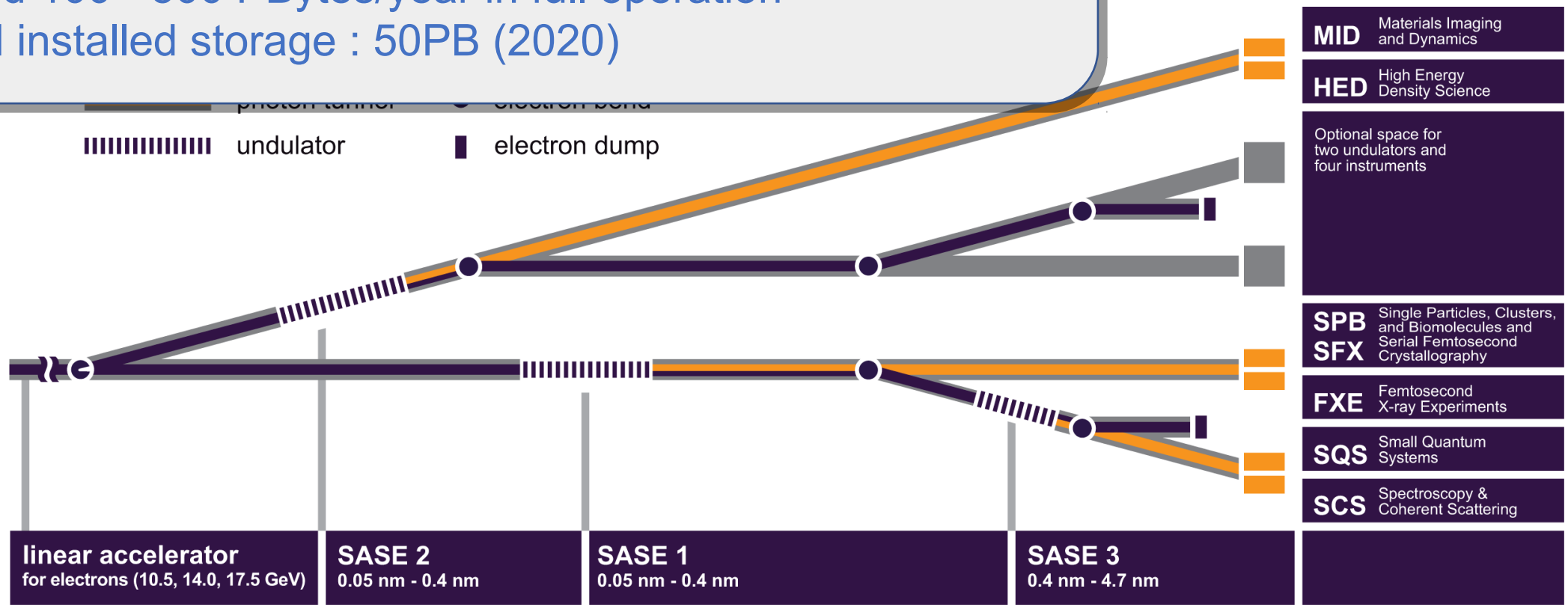




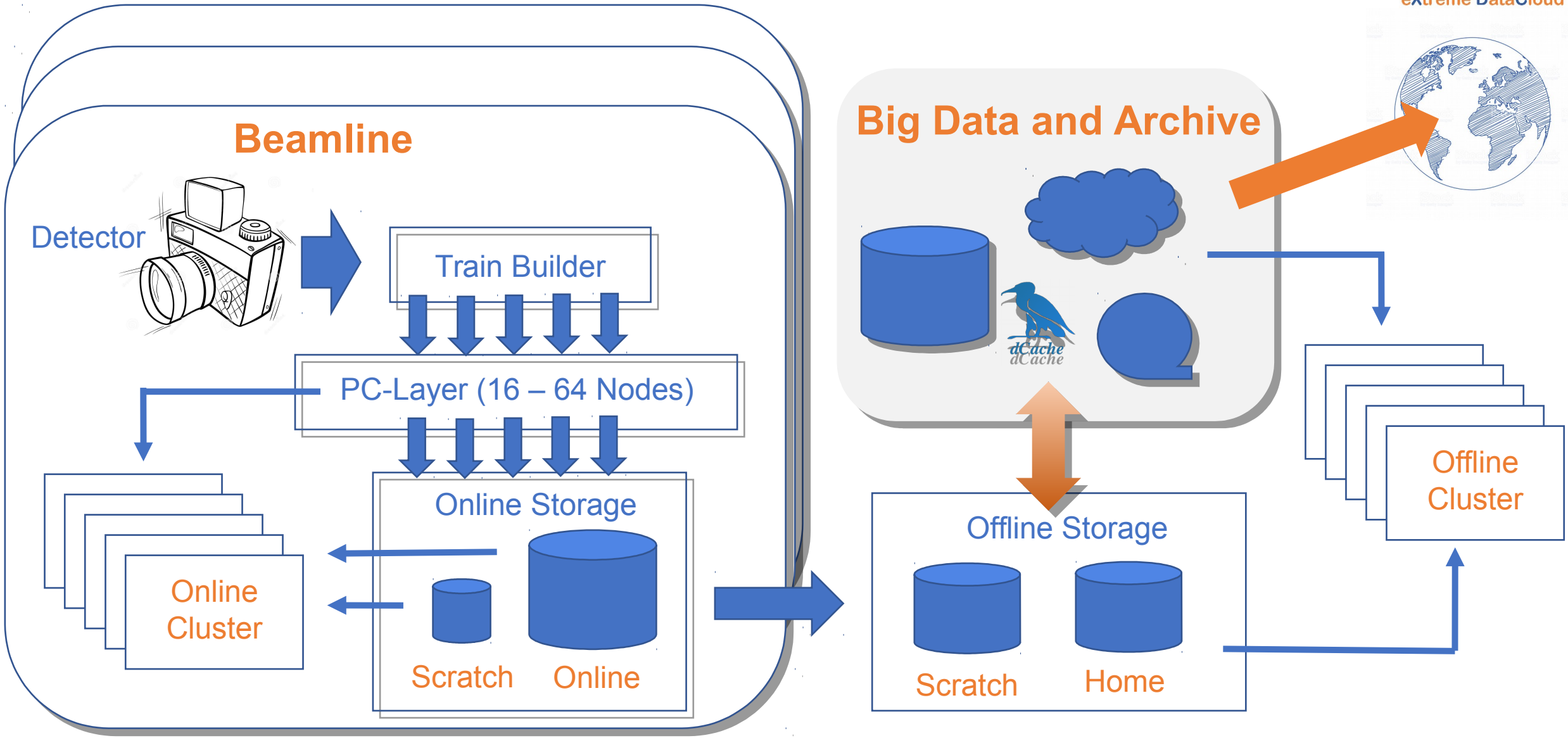


XFEL Cheat Sheet

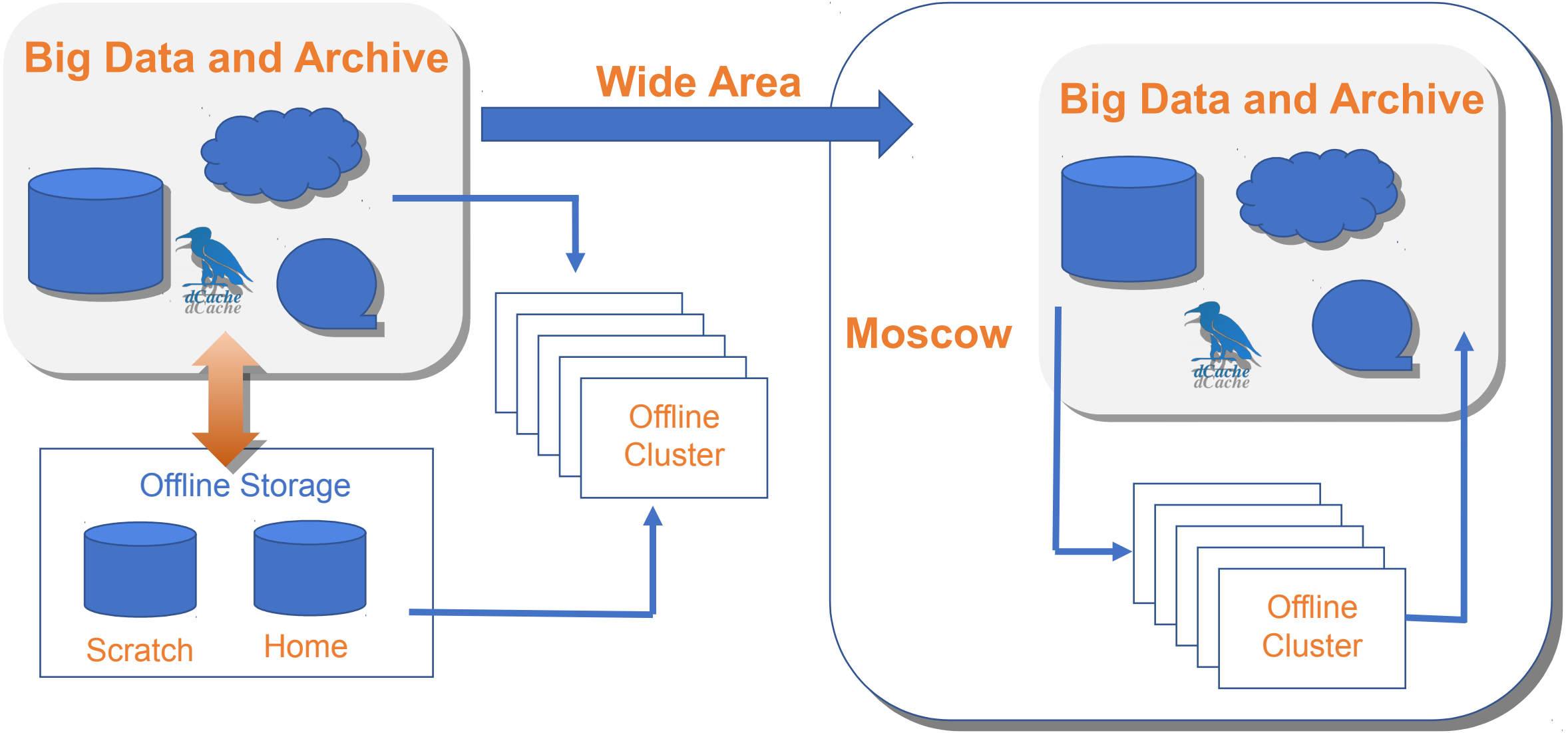
4 M Pixel Detector : 30 GBytes / sec about 1 ExaByte / year
Now : 50 MBytes/sec - 5 TBytes / Day
First 3 month , ½ PetaByte
Expected 100 - 500 PBytes/year in full operation
Planned installed storage : 50PB (2020)



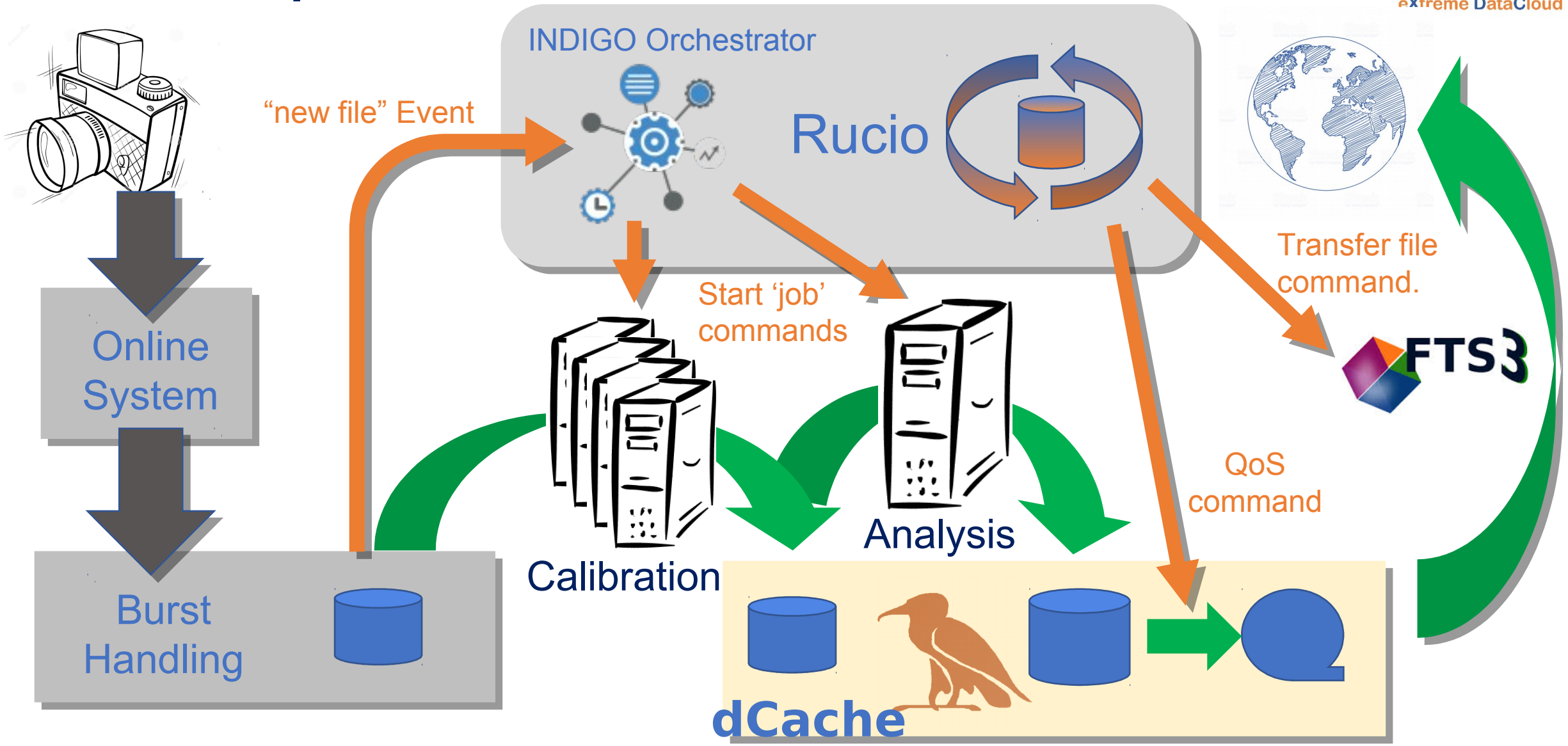
European XFEL high level data flow



European XFEL high level data flow



“European XFEL”-like Use Case



Conclusion

- XDC has an ambitious development plan for data management services
  - We want to support very diverse use cases and requirements
  - We need really a modular and flexible approach in building our platform
- We will support as much as possible standards protocols to make the solutions as general as possible
- Sustainability of the products:
  - Provide upstream to the original project all the changes developed by XDC
  - Involving the user communities in exploiting the XDC outputs in their production environments
  - Pushing XDC developments in the EOSC Service Catalogue

Thanks for listening!

The Einfra-21-2017 Call

- (a) Support to Public Procurement of innovative HPC systems, PPI
- (b) Research and Innovation Actions for e-Infrastructure prototypes
  - 1 - Universal discoverability of data objects and provenance
  - **2 – Computing e-infrastructure with extreme large datasets**
- Service prototypes should follow common interfaces to access and analyse underlying data *collected/stored in different platforms, formats, locations and e-infrastructures […] tested against requirements of* ***very large or highly heterogeneous research data sets***.
- Funds development of **service prototypes** at **TRL6+**
  - **Bring to TRL8 and include in a unified service catalogue in 2018+**
- **Budget per proposal**: **2.5-3M€**

XDC Consortium

ID	Partner	Country	Represented Community	Tools and system
1	INFN (Lead)	IT	HEP/WLCG	INDIGO-Orchestrator, INDIGO-CDMI(*)
2	DESY	DE	Research with Photons (XFEL)	dCache
3	CERN	CH	HEP/WLCG	EOS, DYNAFED, FTS
4	AGH	PL		ONEDATA
5	ECRIN	[ERIC]	Medical data	
6	UC	ES	Lifewatch	
7	CNRS	FR	Astro [CTA and LSST]	
8	EGI.eu	NL	EGI communities	

- 8 partners, 7 countries
- 7 research communities represented + EGI
- XDC Total Budget: 3.07Meuros
- XDC started on Nov 1st – will run for 27 months

From the EU reviewers of the proposal: ***"Consortium as a whole is well formed, bringing together the world-class expertise, experience and infrastructure to enable the successful achievement of the proposal's objectives."***

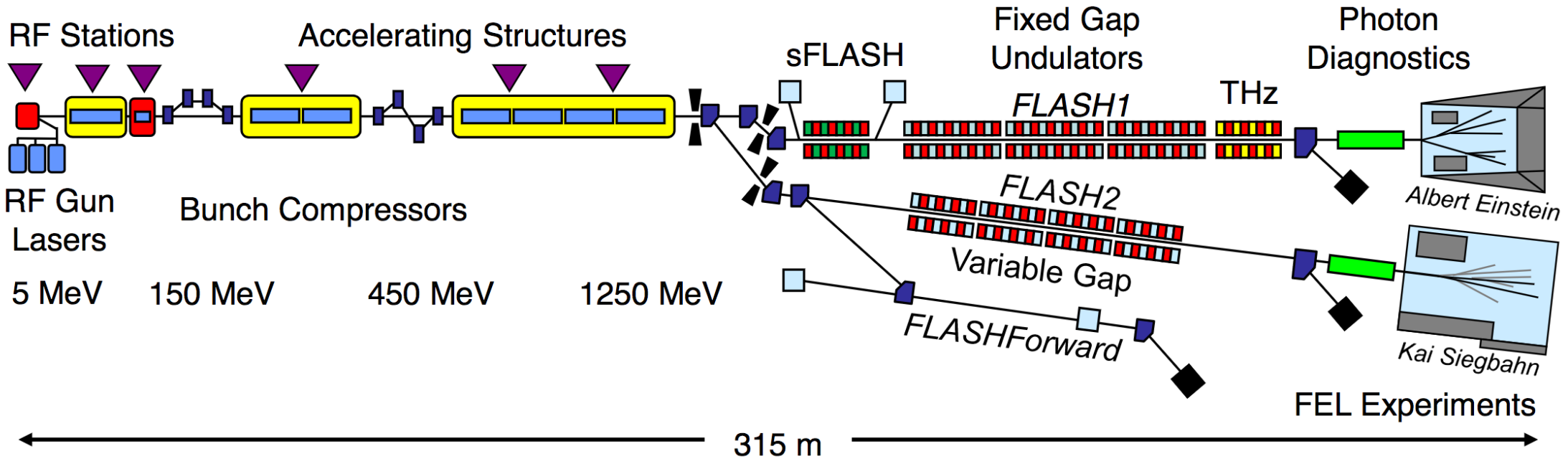
The plan for the next couple of years

Main Milestones

- Research communities requirements for new functionalities collected — PM3
- Research communities requirements analysis performed — PM6
- Project architecture detailed — PM6
- Development schedule defined — PM6
- Event with User Communities — PM9 - Joint with DEEP in Santander
- XDC reference releases – 1 — PM12
- XDC reference releases – 2 — PM24
- Functionalities and scalability demonstrated — PM27

Basics of a free electron laser

- SASE: Self-Amplified Spontaneous Emission
- Example : FLASH



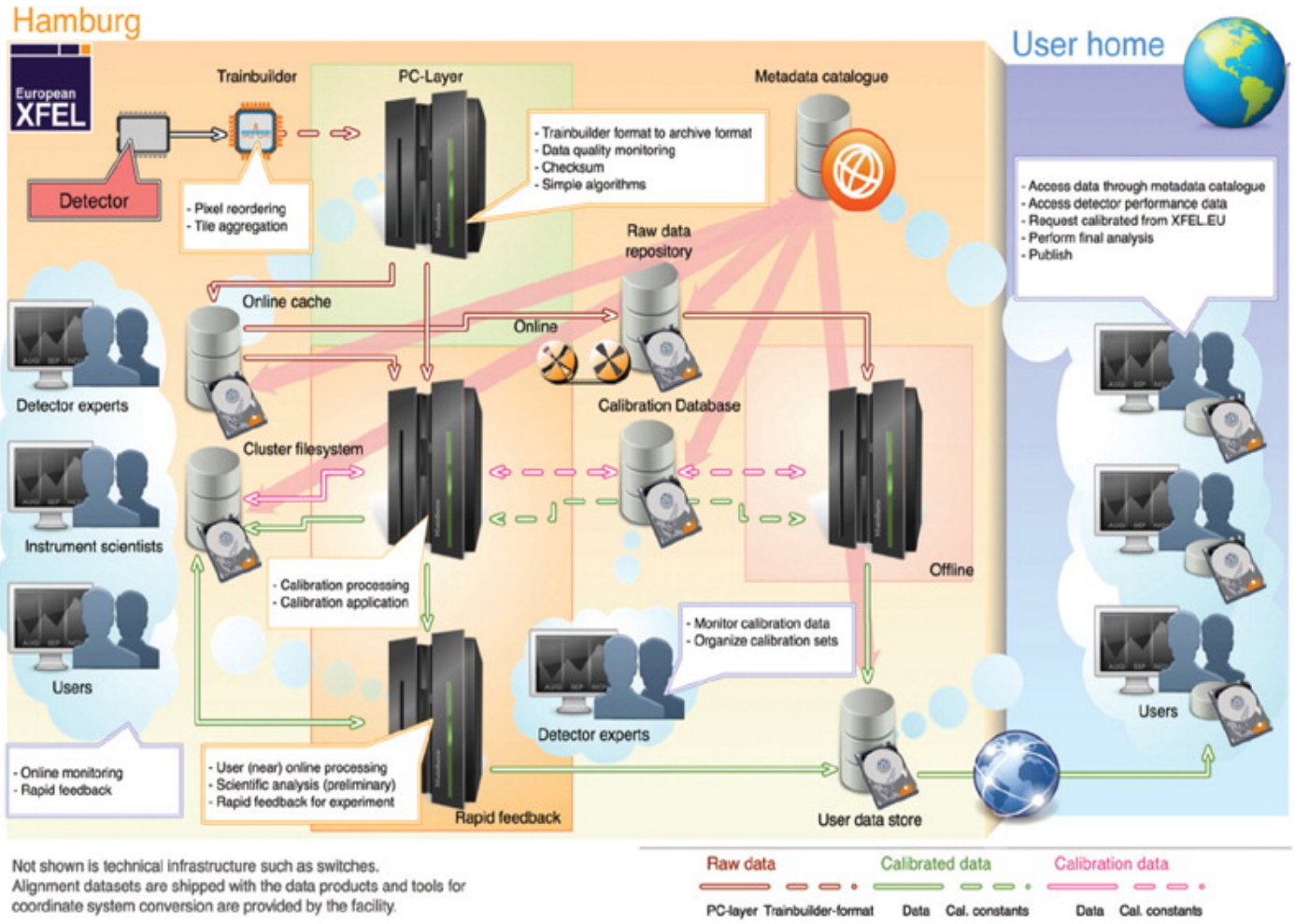
Use Case Diagram

Online:

- Exclusive access to cluster during beam time, only from experiment rooms (Karabo control system).
- On-the-fly peak detection, ROI selection and user defined software stack can read ZeroMQ stream.

Offline:

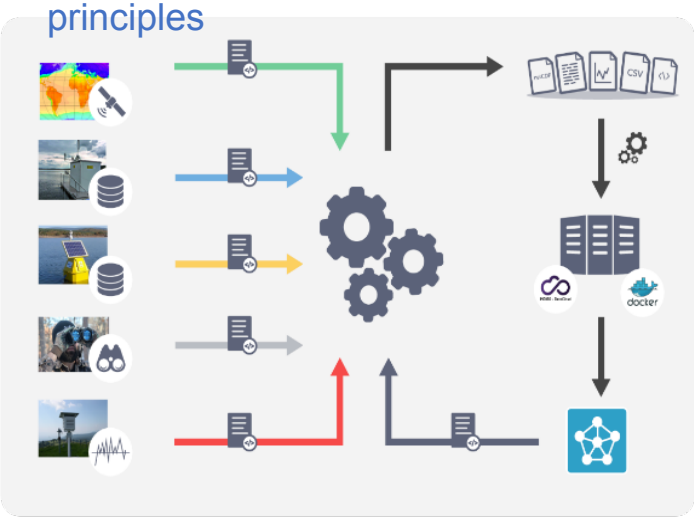
- Calibrated data provision for analysis on the HPC Cluster and in the OpenStack cloud environment.



Metadata handling use cases

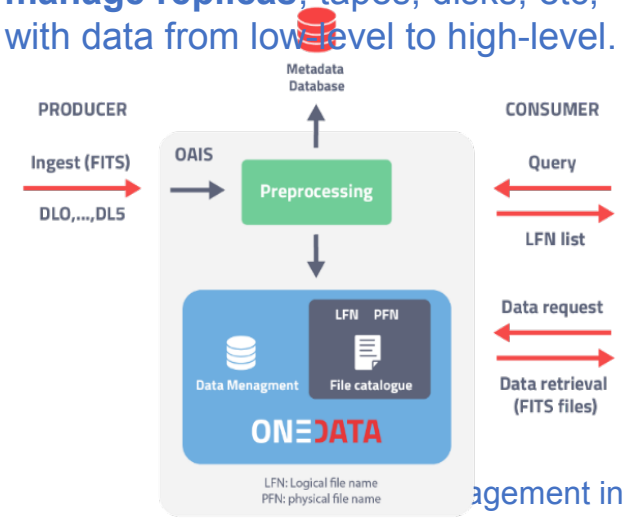
LIFEWATCH

- Metadata management to handle heterogeneous and large datasets
  - Different data types, formats, source and ways to access
  - e.g. Copernicus data: ~16PB per year
- Used as input for water quality forecasting systems
- Use of standards like EML (Ecological Metadata Language) and adopting best practices like FAIR+R principles



CTA

- The CTA distributed archive lies on the « Open Archival Information System » (OAIS) ISO standard. Event data are in files (FITS format) containing all metadata.
- Metadata are extracted from the ingested files, with an automatic filling of the metadata database.
- Metadata will be used for the further query of archive.
- The system should be able to **manage replicas**, tapes, disks, etc, with data from low-level to high-level.



ECRIN

- Clinical trial data objects available for sharing with others
  - a variety of access mechanisms
  - wide variety of different locations
    - growing number of general and specialised data repositories
    - trial registries
    - Publications
    - the original researchers' institutions
  - 'discoverability' will become much worse in the future as more and more materials is made available for sharing

XDC Use Case – Physics with Photons

Brief Description (European XFEL, PETRA III, FLASH)

- European XFEL, the world's largest free electron laser, will vastly increase progress in molecular imaging and enable scientists to observe atomic details that were previously inaccessible. At the DESY campus, we also operate the most brilliant synchrotron radiation source world wide PETRA-III and the Free Electron Lasers FLASH and FLASH-II. DESY is providing services throughout all stages of data processing, analysis, storage and management.

Community Size

- 11 European countries are participating in the XFEL project, which is fully operational since September 2017. Beam time is free of charge and open to scientists world wide. PETRA-III and FLASH have more than 5000 active users.

Champions

- Contact in XDC: Patrick Fuhrmann, Tigran Mkrtchyan
- Science Contract: Michael Schuh, Frank Schlünzen, Prof. Hans Fangohr

Use Case Description

General Status

- European XFEL produces high intensity X-ray laser pulses at new levels of intensities. With resolution of atomic length scale and on a femto second time scale, scientists can assemble movies of chemical reactions in unprecedented detail. XFEL provides 10 experimental stations to serve scientific and industrial users from a wide range of domains in physics, chemistry, material science, biology and nano-technology. Data processed and stored by DESY will be produced at up to 30GB/s.

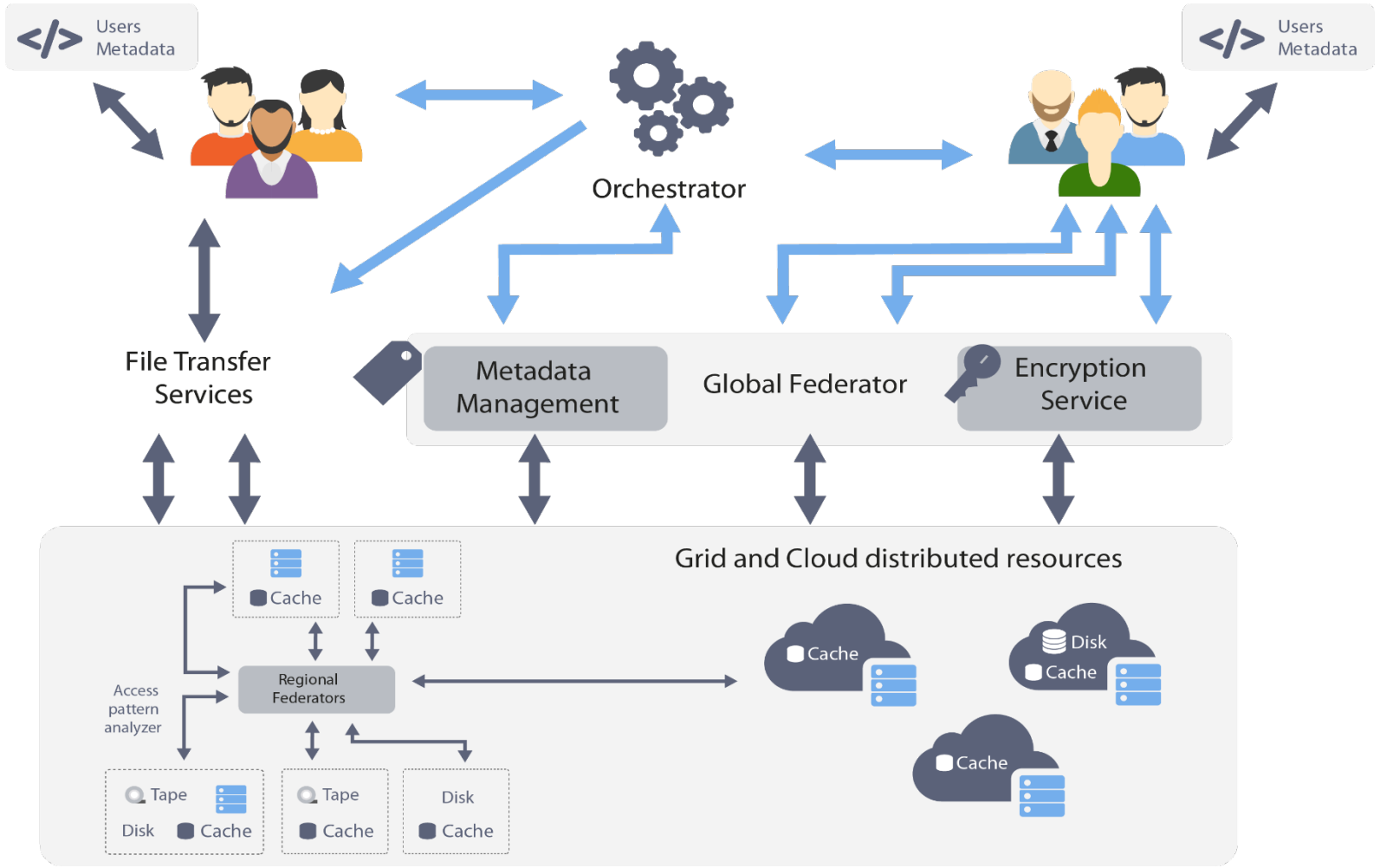
Roles

- Researcher/User: Control experiment, produce raw data diffraction images, first online data analysis. Reconstruct images in offline analysis, apply domain specific analysis.
- Instrument scientist: System and configuration management, beam time high availability. Online and offline analysis software stacks.
- Cloud/HPC DevOps: Operate and develop HPC and cloud platform

Current Status vs. Expectation

- Deployment on HPC Clusters to be extended to the cloud.
- Offline analysis clusters to be scaled for a growing, highly distributed user community.

XDC high level architecture



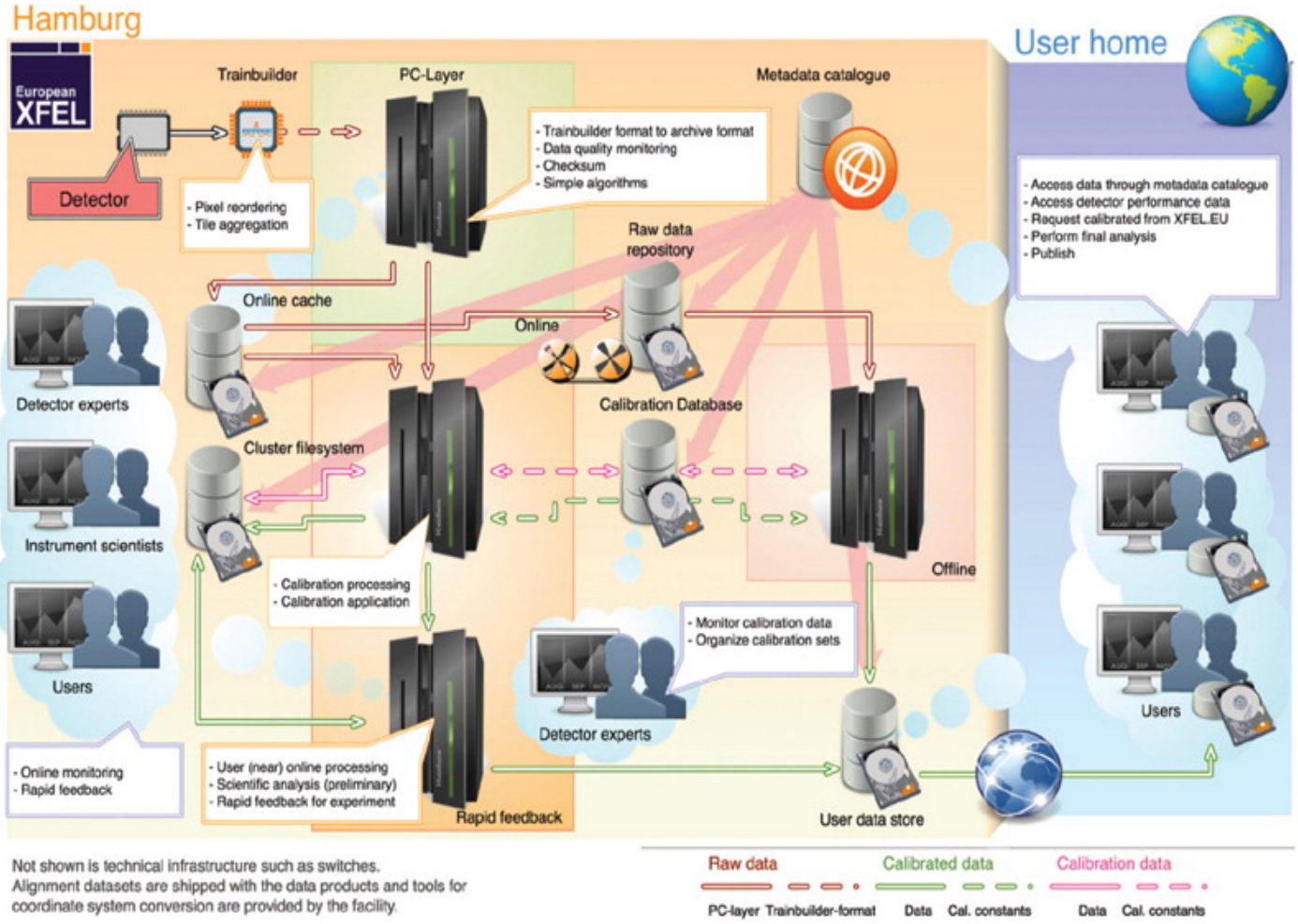
Use Case Diagram

Online:

- Exclusive access to cluster during beam time, only from experiment rooms (Karabo control system).
- On-the-fly peak detection, ROI selection and user defined software stack can read ZeroMQ stream.

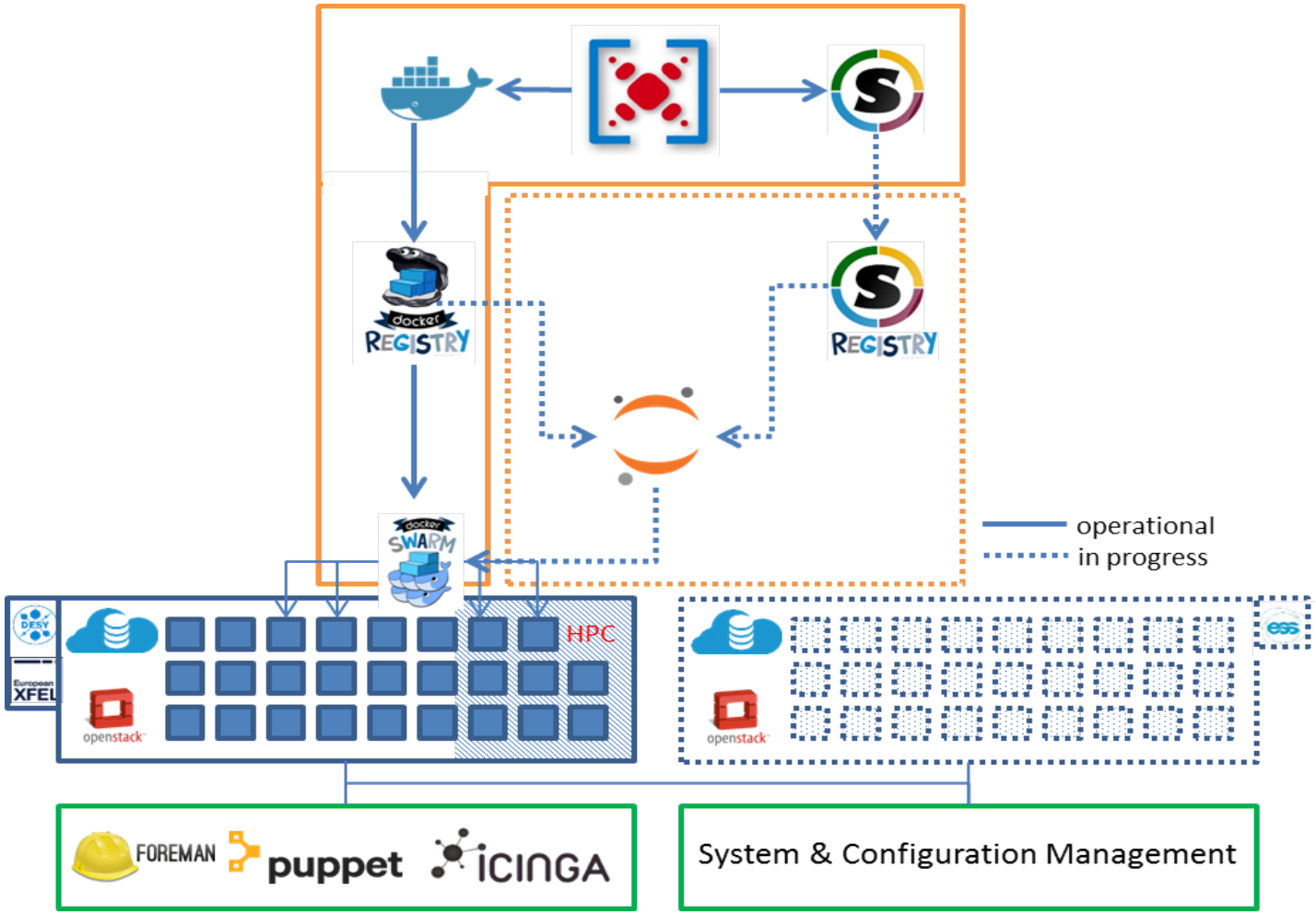
Offline:

- Calibrated data provision for analysis on the HPC Cluster and in the OpenStack cloud environment.



Use Case Diagram (Computing)

- Container: Docker available on online and offline clusters, providing OpenMPI and Infiniband drivers. Analysis tools delivered and demonstrated using Jupyter notebooks on the HPC cluster, connected to users Web-Browser on desktop/laptop.
- Cloud: OpenStack environment growing from 150 to 500 Cores.



About the Data ...

- Formats, sizes, legal issues:
  - HDF5 files in XFEL data schema specification.
  - Raw data volumes of 15GB per second and detector.
  - Usage of the experiment is free of charge,
  - but publishing is mandatory.
  - Data management plans of the Science Users apply. (not from the facility)
- Metadata:
  - User portal to the European XFEL (UPEX) gives access to metadata catalogue, storage and compute resources.
- Special needs:
  - Separate flows for raw data, calibrated data and data taken for detector.
  - Calibration and geometries applied by instrument scientists.
  - License aware registries for docker, singularity.

About the Data ...

- DMP : There is a draft DMP for the European-XFEL in general, which is not signed yet. I can use the content but I can't make the content available.
- PETRA III(IV) and the European XFEL themselves don't produce the actual data. This is done by XFEL customers (beam line users). Therefor the European XFEL doesn't really own the data.
- Persistent identifiers not yet in discussion.

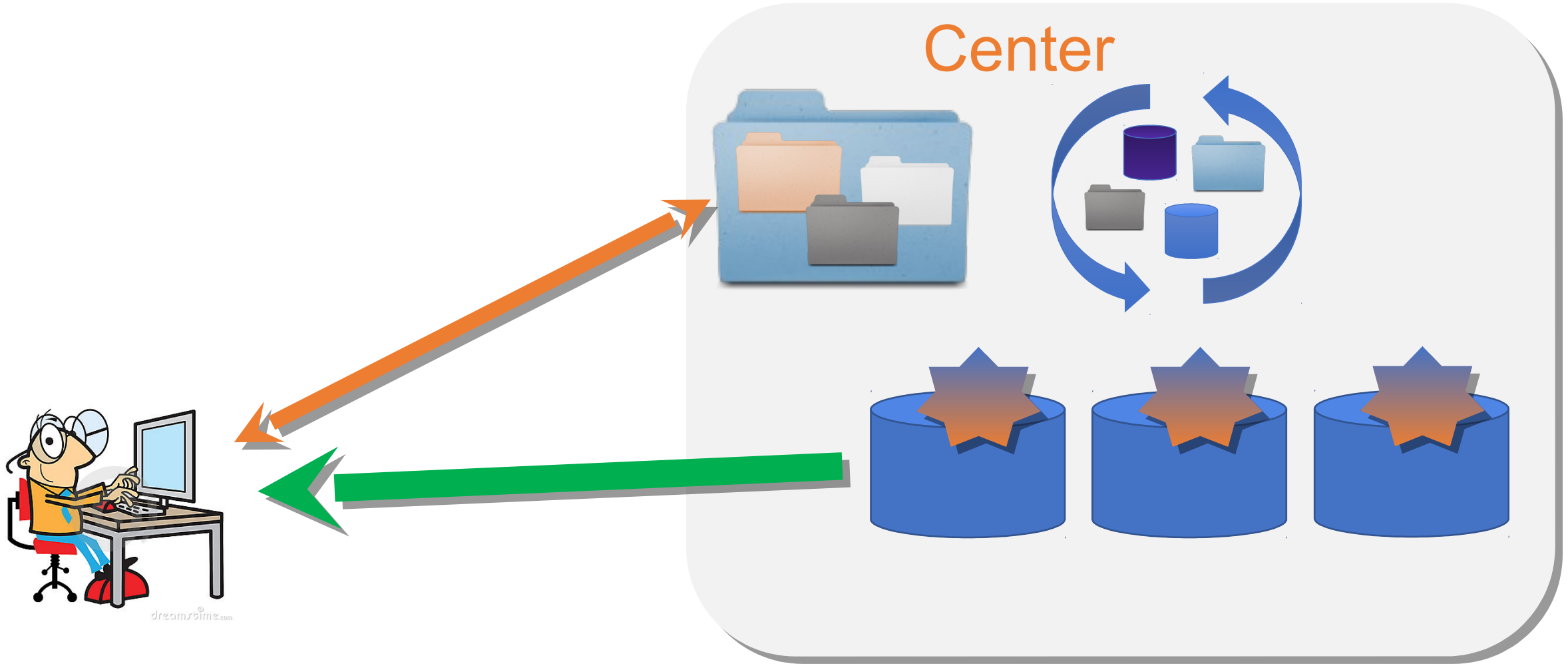
Now on the caching part

Prerequisite

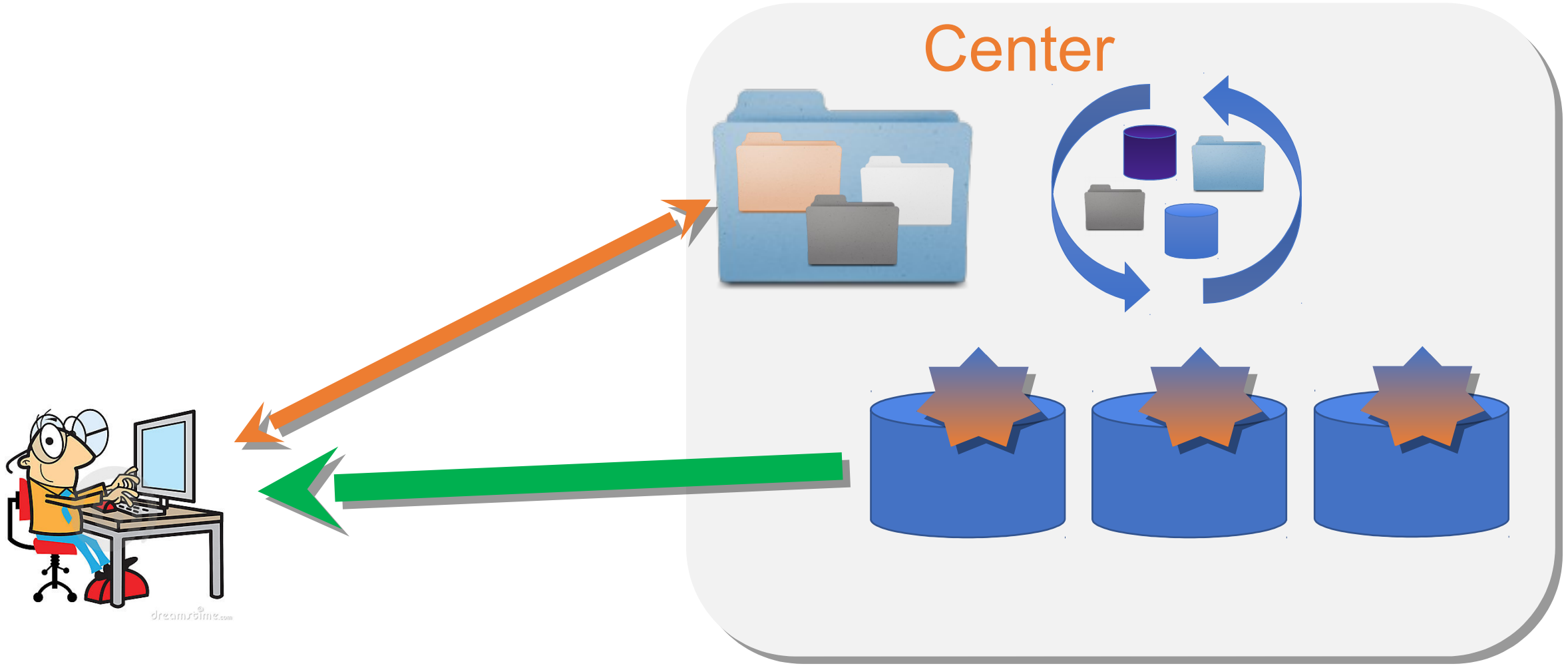
,

:=

Access: control and data flow



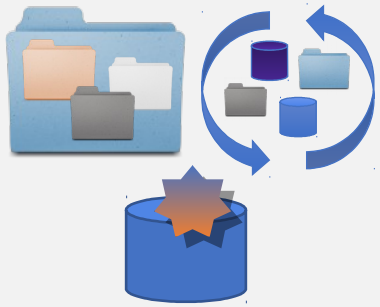
Access: control and data flow



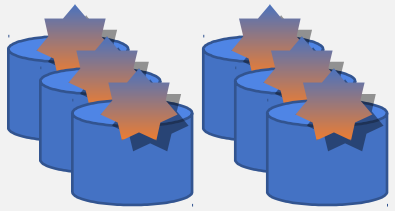
Examples

dCache

Denmark

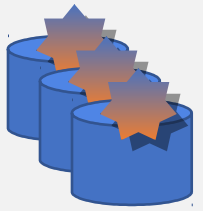


Sweden



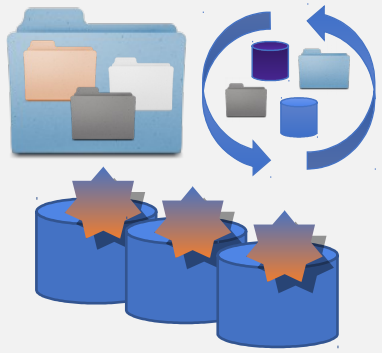
Norway

Finland



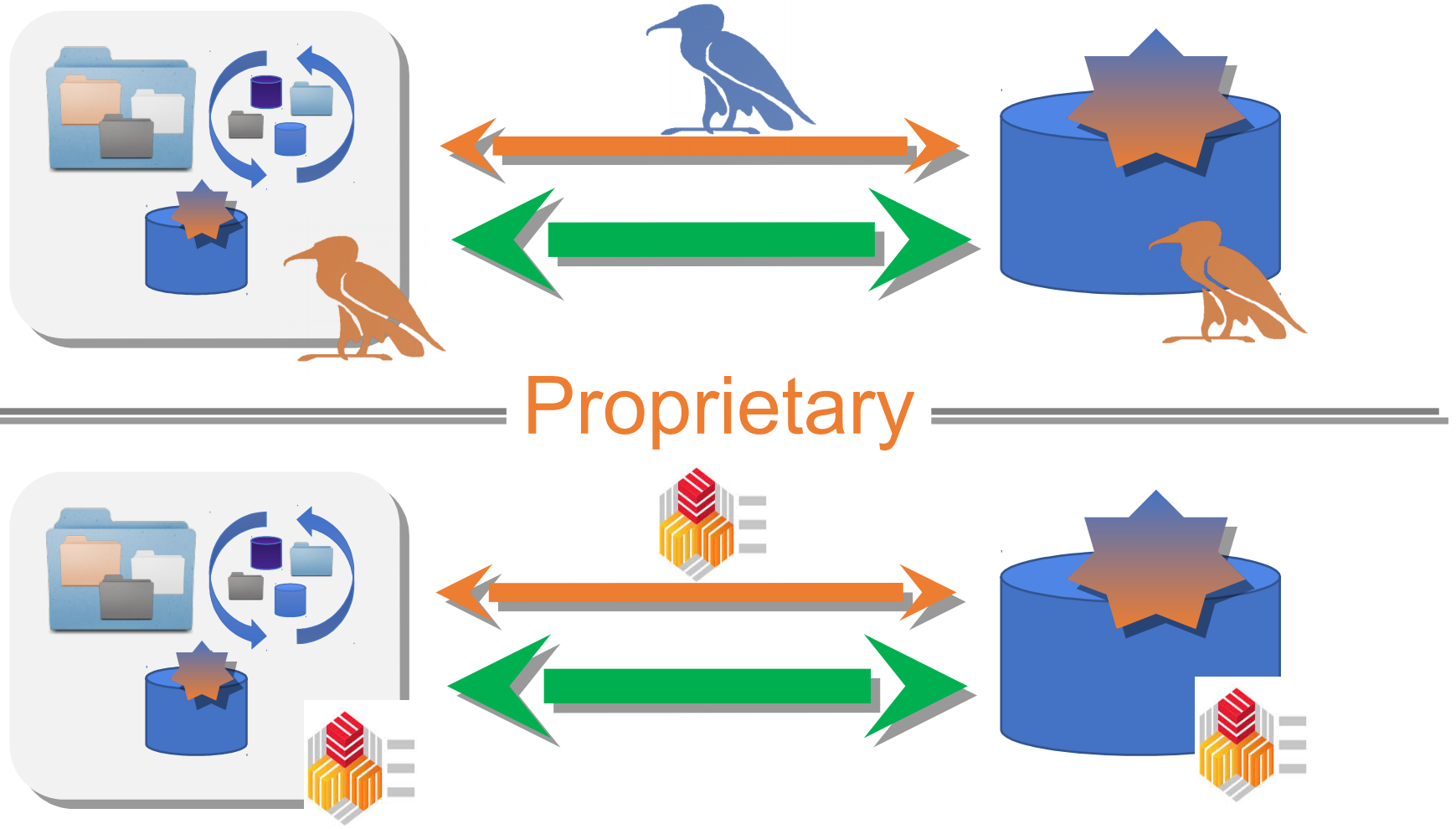
EOS

CERN

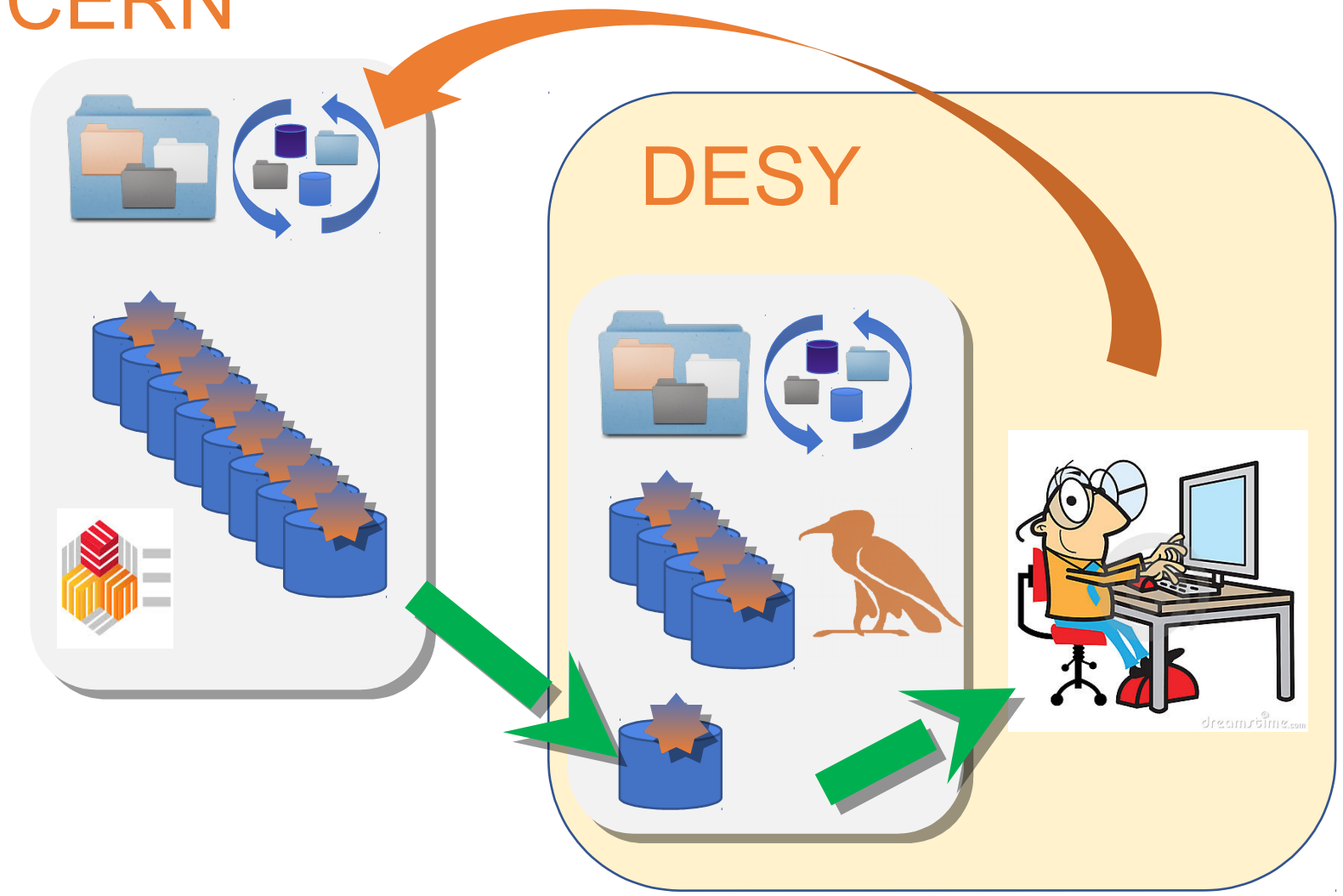


Budapest (Wigner)

Current situation (naturally)



That would be nice to have



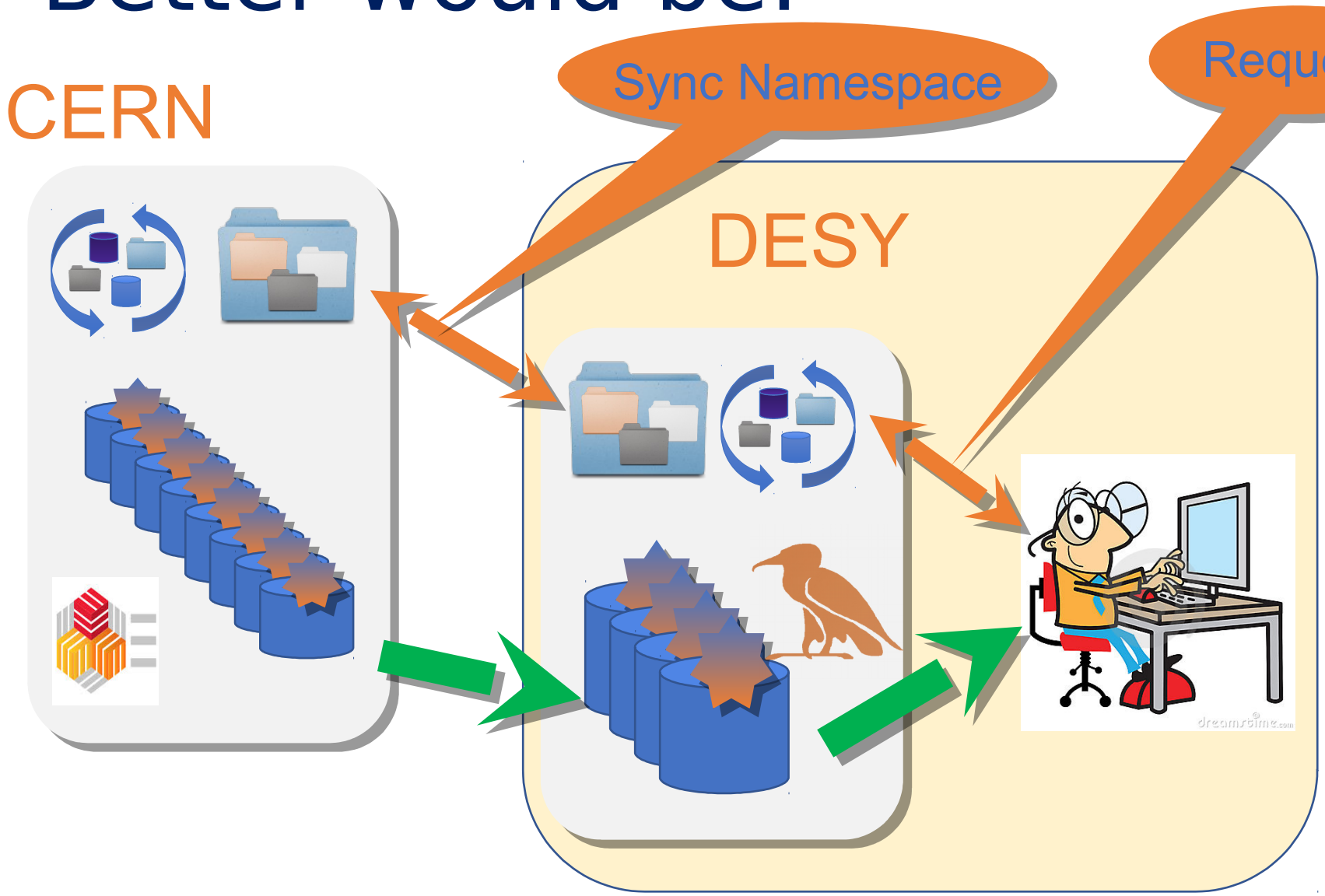
Advantages

- no additional software stack needed at sites.

Still disadvantage

- Local data not accessible in case data link is down or central service not available

Better would be:



Advantages

- Same software stack as we currently have at the sites.
- After the data has been transferred to the local storage system, a name space entry has been created locally and the data is available at the local site independently of the remote network link and the availability of the central service.