The eXtreme-DataCloud project

eXtreme DataCloud

Data Management for extreme scale computing

Daniele Cesini

info<at>extreme-datacloud.eu

www.extreme-datacloud.eu

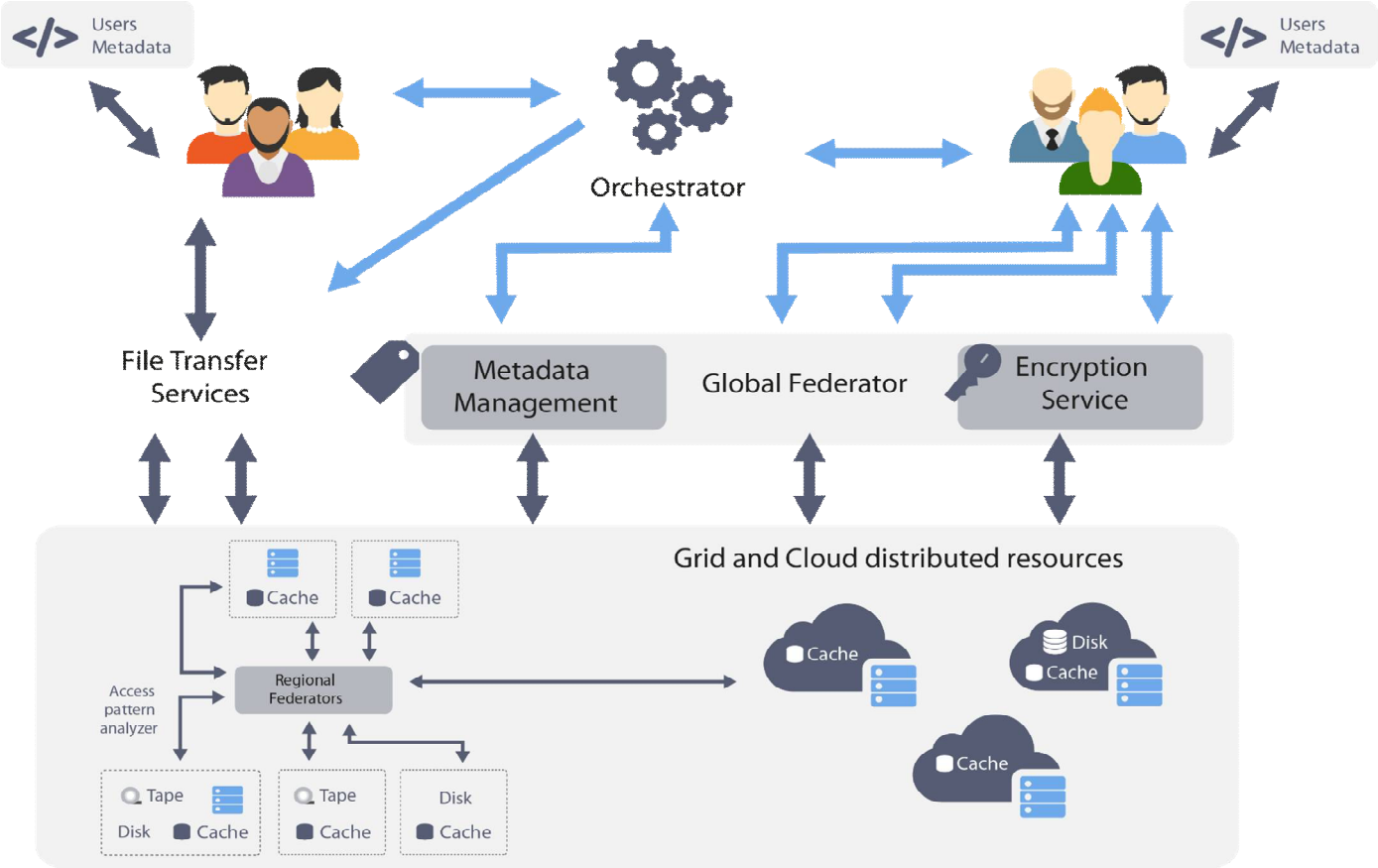
eXtreme DataCloud is co-funded by the Horizon2020 Framework Program – Grant Agreement 777367

Copyright © Members of the XDC Collaboration, 2017-2020

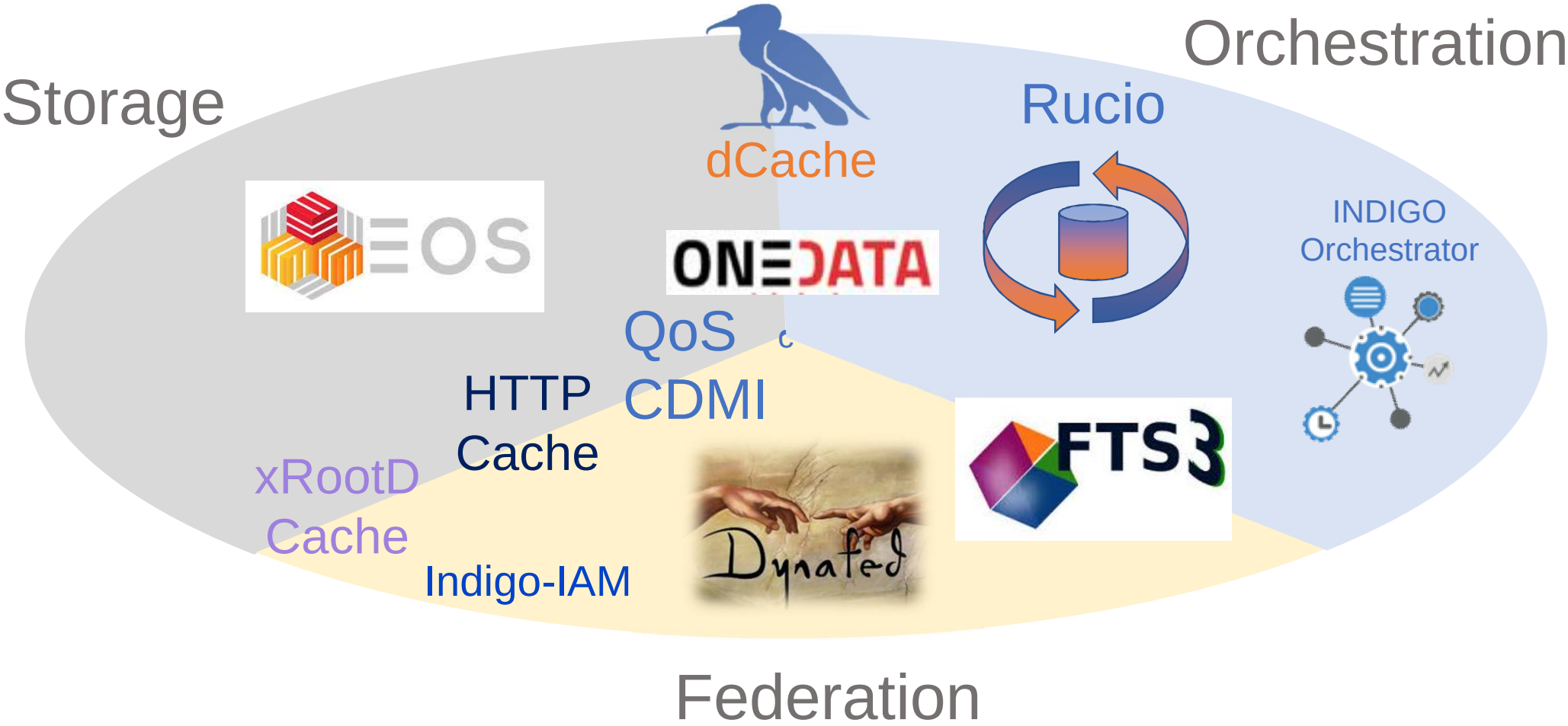
XDC Technical Topics

- Intelligent & Automated Dataset Distribution
  - Orchestration to realize a policy-driven data management
  - Data distribution policies based on Quality of Service supporting geographical distributed resources (cross-sites)
    - QoS beyond disk vs tape: availability, reliability, durability, latency
  - Data lifecycle management
- Data (pre)processing during ingestion
  - Storage Notifications based on events
- Smart caching
  - Transparent access to remote data without the need of a-priori copy
- Metadata management
- Data management based on access patterns
  - Move to ‘glacier-like’ storage unused data, move to fast storage “hot” data
    - at infrastructure level
- Sensitive data handling
  - secure storage and encryption

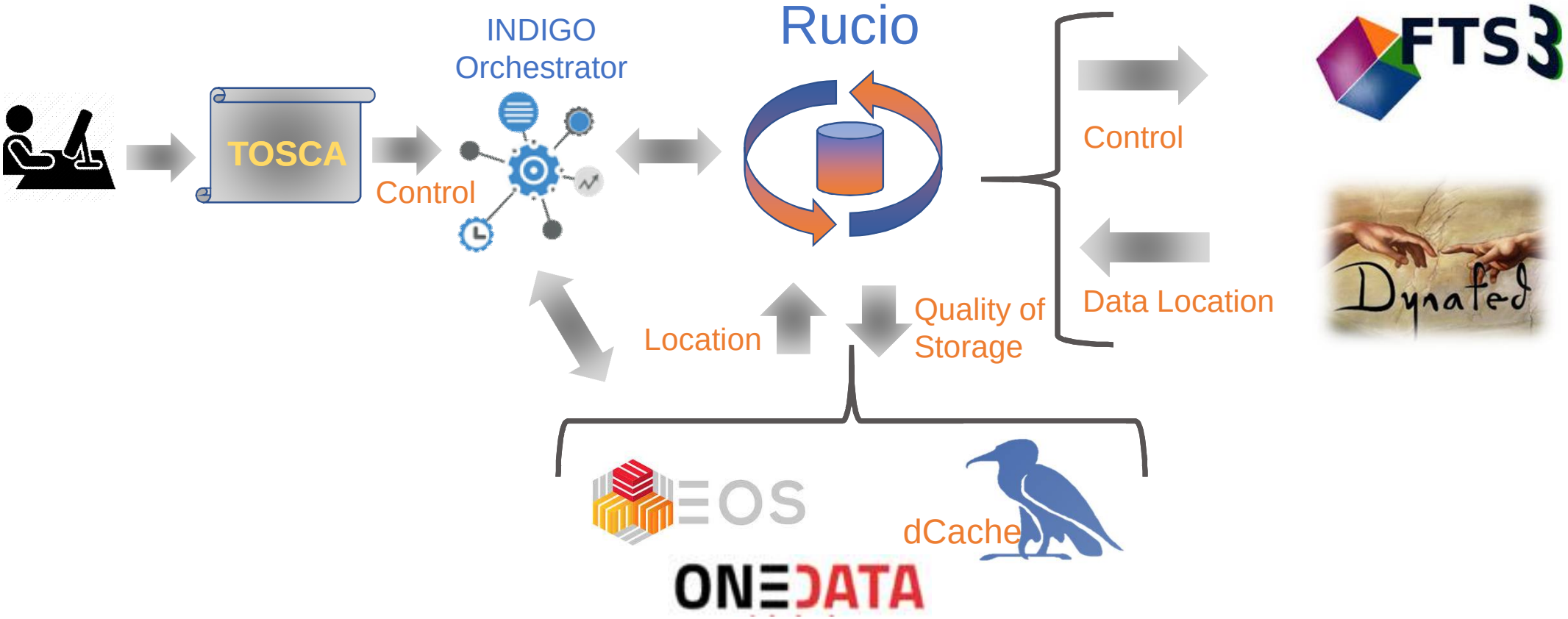
XDC high level architecture



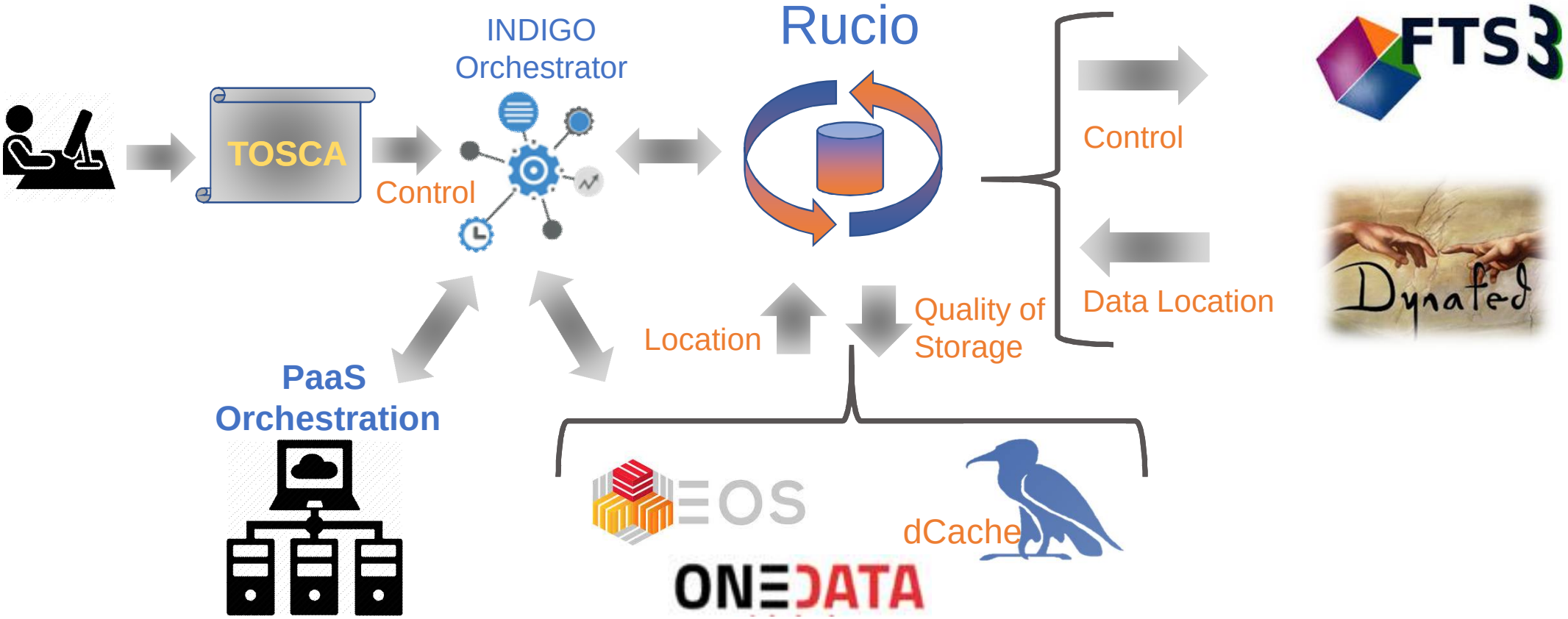
XDC Toolbox


Orchestration
Storage
Rucio
dCache
EOS
ONEDATA
INDIGO
Orchestrator
QoS
CDMI
HTTP
Cache
xRootD
Cache
FTS3
Indigo-IAM
Dynafed
Federation

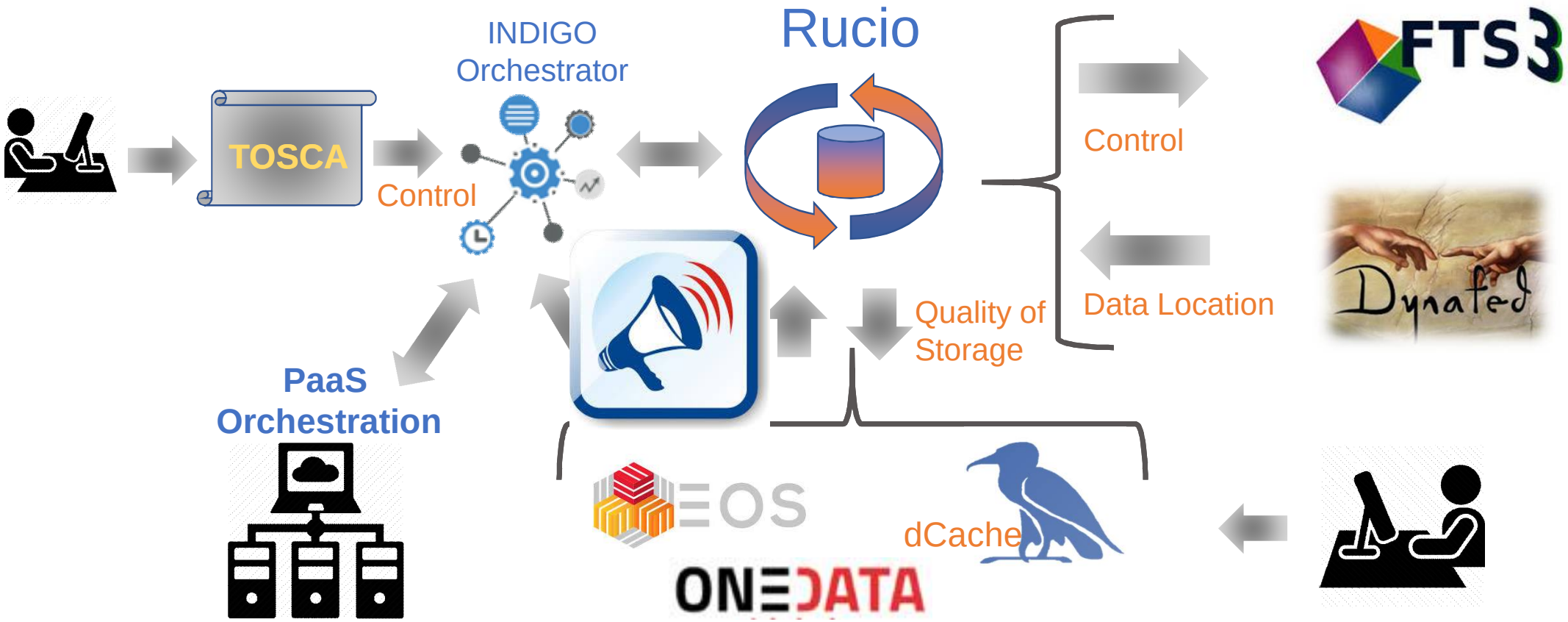
Expected Control Flow for Orchestration and Policy Driven Data Management



Expected Control Flow for Orchestration and Policy Driven Data Management

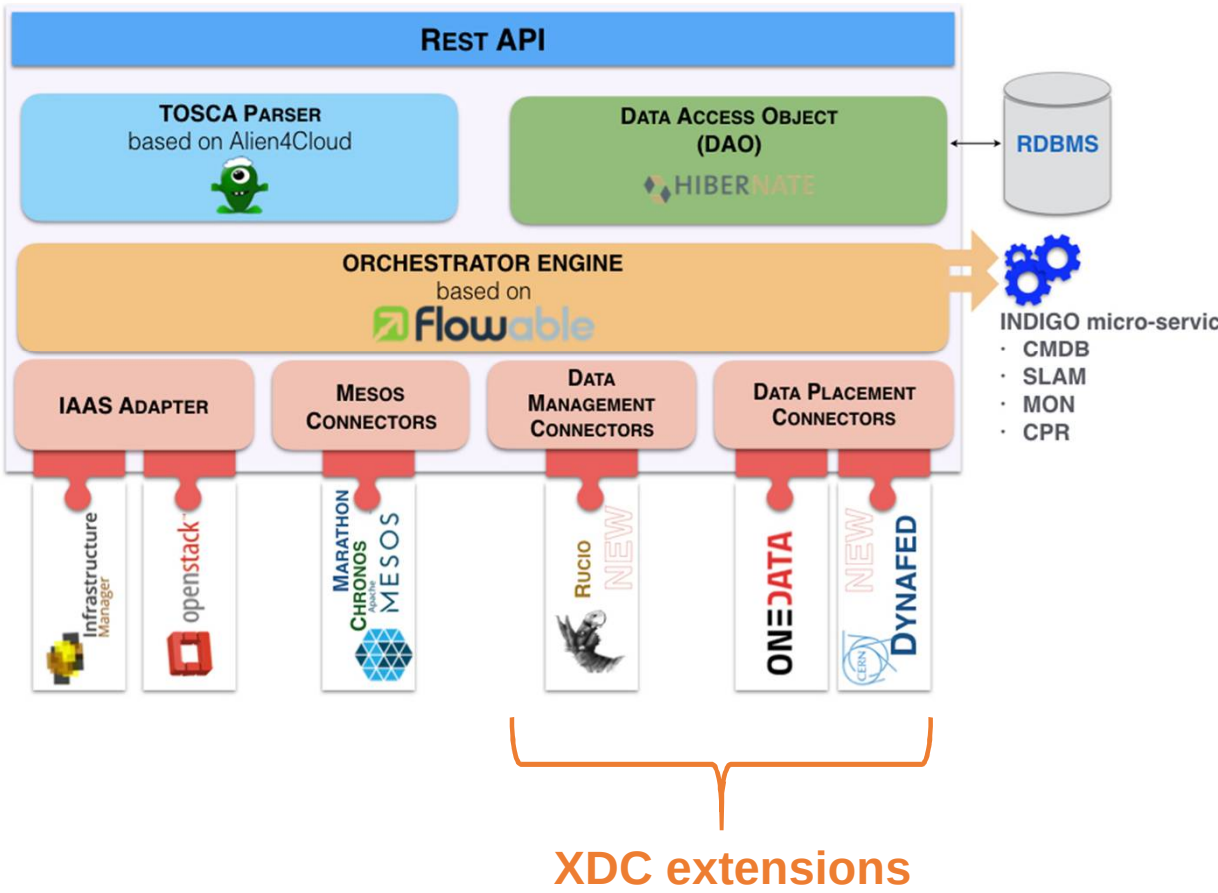


Expected Control Flow for Orchestration and Policy Driven Data Management



INDIGO Orchestrator Overview

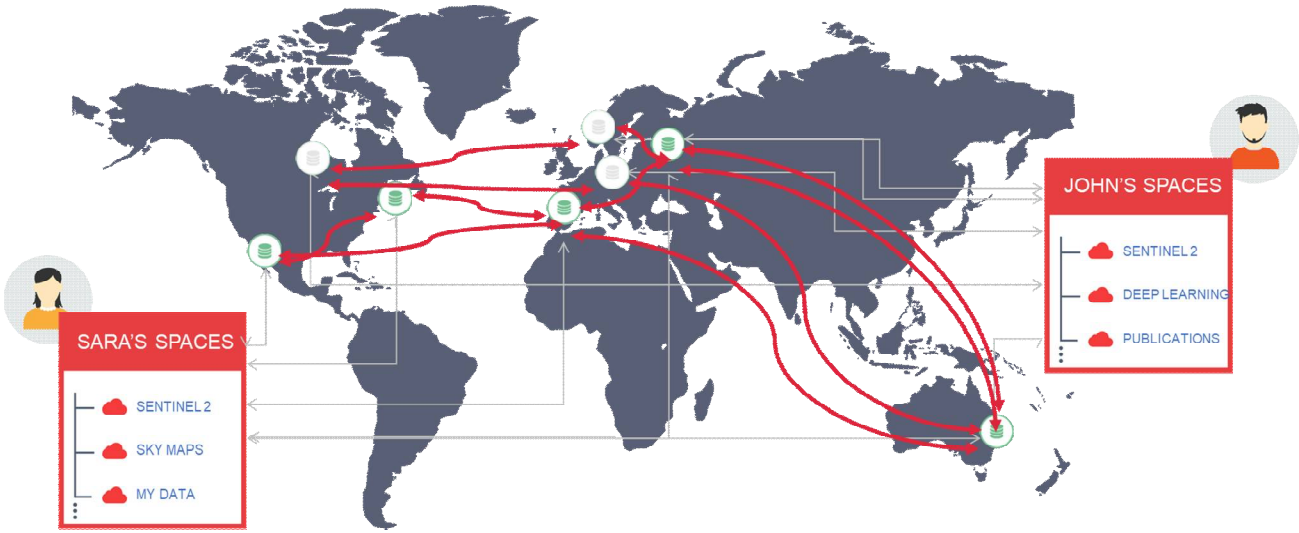
- The Orchestrator accepts TOSCA requests
- It can run three main workflows in the PaaS:
  - Deploy
  - Undeploy
  - Update
- The Orchestration layer is being extended in order to address the new requirements:
  - move data between distributed storages
  - specify different QoS for replicas
  - launch and monitor user defined processing jobs at ingestion time



Data in Hybrid Cloud Environments: Onedata

- ONEDATA is a storage federator that allows to use resources backed by providers worldwide
  - Providers deploy *Oneprovider* services near physical storage resources
  - Users use *Onezone* web interfaces
    - APIs available
    - Local mounting on users machines available
  - Storage is organized into **Zones**
    - federations of providers
    - enable the creation of closed or interconnected communities

https://onedata.org



XDC – Release 1

Release date 31/10/2018

Component	XDC functionalities
EOS	- Caching with xcache for geographic deployment: Xcache deployed at a remote centre to accelerate its local CPU - External storage adoption (Through an S3 or a WebDAV interface) - External data adoption (Data already present on a system described above can be incorporated into EOS)
dCache	- New QoS types integration, aggregated QoS for storage federations - OpenIDConnect support in dcache_view - dcache storage events (SSE notifications): Allow non-dCache agent to get notified that something of interest happen inside dCache
FTS+gfal	- QoS support: can now accept a QoS job - OpenIDConnect support - QoS in gfal (gfal with basic cdmi client) – python bindings available
Dynafed	- Integration of OIDC authentication
Caching infrastructure based on xCache or http	- deployment receipts for geographically distributed caches (via xcache) - deployment receipts for scalable local caches (via xcache and http)
PaaS Orchestrator	- Implementation of Dynafed plugin – interaction via INDIGO IAM oauth2 token - Enhancement of ONEDATA plugin
ONEDATA	- Performance and stability improvements - new RADOS driver

XDC Contacts

- POSTERS:
  - 4: Water Quality monitoring and forecasting with eXtreme-DataCloud
  - 5: CTA use case for XDC solution
  - 21: The adoption of the HTTP/XRootD protocols for a new data caching architecture in WLCG experiments
  - 32: dCache events driving data management at the European XFEL: Evaluating event-driven cloud computing and storage management
- Website: www.extreme-datacloud.eu
- **@XtremeDataCloud on Twitter**
- **Mailing list: info<at>extreme-datacloud.eu**