

# Post-CWP perspectives for national R&D initiatives - CMS Italy

Tommaso Boccali - INFN Pisa

# Small and incomplete list of efforts I know about

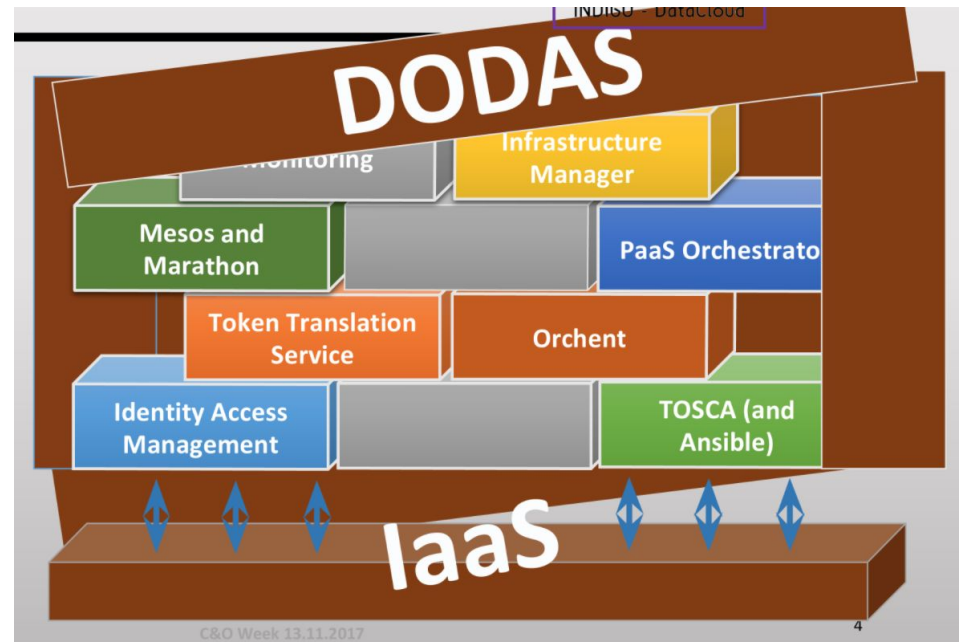
- H2020 related efforts
- INFN ongoing efforts
- Other ...

# H2020 projects

- Building on the success of INDIGO-Datacloud, three projects were submitted + approved in the first half on 2017 - starting 2018; all with INFN as first or second biggest contributor
- **EOSC-Hub (PI: EGI.eu-NL)**
  - Main “connector” project for European Open Science Cloud
  - DODAS thematic service (CMS is proponent and main target)
- **XDC: The eXtreme DataCloud (PI: INFN-IT)**
  - Develop scalable technologies for federating storage resources and managing data in highly distributed computing environments
- **Deep-HyperCloud (PI: CSIC-Spain)**
  - Work on intensive computing techniques on the analysis of very large datasets
  - Also on specialized hardware components, like GPUs, low-latency interconnects

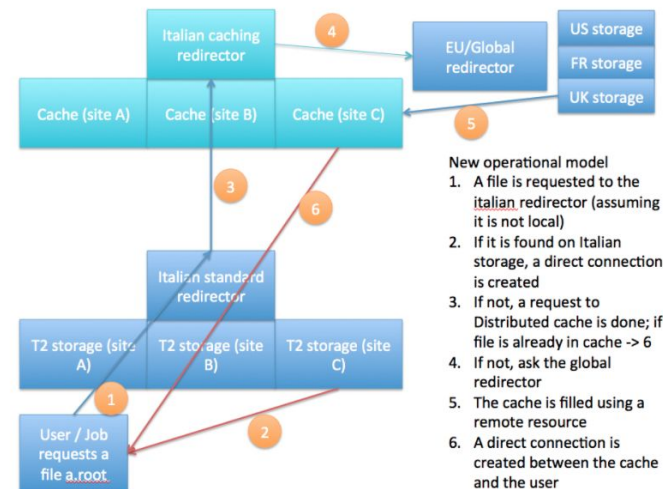
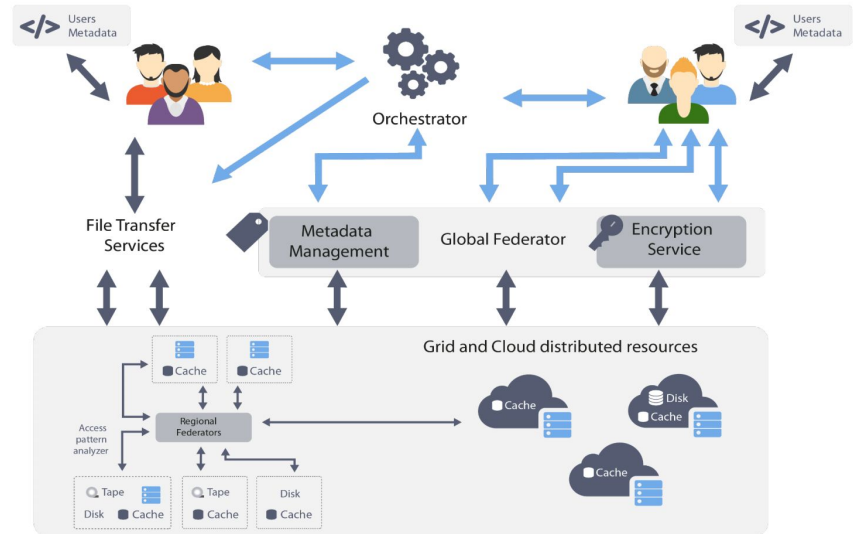
# More in details - DODAS

- DODAS: ideas are:
  - Describe via TOSCA templates full data centers, with pets and cows, with all the interdependencies
  - TOSCA translates seamlessly to all the major open/proprietary cloud implementations
  - Already tests on Azure, Openstack.
  - Implement full stat of authorization / authentication via IAM and TTS
  - Next step: extend to storage, including virtual site proxies, caches, ...
  - Sinergies with XDC



# XDC

- XDC: focus on data. Starting from existing software, integrate with added functionalities
- In Italy, most interest in developing a global caching infrastructure supporting the following building blocks:
  - Dynamic integration of satellite sites by existing data centres
  - Creation of standalone caches modelled on existing web solutions
  - Federation of the above to create a large scale caching infrastructure
- In a sentence, distributed smart caching ...



- New operational model**
1. A file is requested to the Italian redirector (assuming it is not local)
  2. If it is found on Italian storage, a direct connection is created
  3. If not, a request to Distributed cache is done; if file is already in cache -> 6
  4. If not, ask the global redirector
  5. The cache is filled using a remote resource
  6. A direct connection is created between the cache and the user

# ESCAPE (INFRAEOSC-04)

- Being prepared in these days;
- INFN entering as principally as HL-LHC
- Main interest is “DataLake prototyping”
  - CNAF participating as Computing center
  - Qa
  - Interest in AAI infrastructure

- THE MAJORITY OF THE PARTICIPATING PROJECTS ARE ESFRI OR ESFRI LANDMARKS:

CTA	ESFRI	HL-LHC	ESFRI-Landmark
SKA	ESFRI-Landmark	FAIR	ESFRI-Landmark
KM3Net	ESFRI	EGO-Virgo	
EST	ESFRI	JIVE	ERIC
ESO/E-ELT	ESFRI-Landmark	(LSST)	“observer”

# (new)manpower

- 12 INFN positions for LHC SW/CMP R&D in early 2017 - contracts starting early 2018
- CMS:
  - 2 positions (PI, BA) on Machine Learning development for analysis, DQM, ...
  - 1 position (MIB) for Tracking with GPU
  - 1 position (PG) in data management, caches, DODAS
- We would like these persons (usually senior postdocs) to participate actively / take leading role in the efforts
  
- If successful, such contracts could become the norm in next years

# More on the Infrastructure side

- As you know, CNAF is off since Nov 9th, due to a flood, till ~ mid Feb.
- It was already planned moving to a separate site “in steps”:
  - **2018:** most of the CPU moved to PRACE Tier-0 CINECA (10 km away) - storage at CNAF
    - Interesting test for aggregation of remote physical sites, dedicated network at 500 GBit/s (to be upgraded to 1200) available
    - We already have the first jobs running in this setup
  - **2019+:** move the facility to the nearby ECWMF center, being built
    - Just got 15MEur of specific funding
- Review of the existing 30+ INFN computing centers ongoing
  - Expect a reduction to mostly the 10 in WLCG hierarchy





# More standard activities (CMS mostly)

- Deep Learning / Machine Learning for Physics
  - Physics (PI, BA, BO, ...) (PI), Btag (PI), Tracking/seeding (BA, PI), Trigger (BO)
  - GPU tracking and analysis (BA, CNAF, MIB)
- Data Management tests
  - Rucio tests (PG)
  
- ATLAS is following roughly the same path
  - Trying to form links, not easy at the moment

# Conclusions

- More to come:
  - Next round of national level funding for specific projects in the next months
  - We see a large focus on ML/DL (even too much, sometimes just an attempt of using the buzzword?)
- European Open Science Cloud (EOSC) going to get the largest share of EU funding in IT in the next years
  - Not yet a well definite concept, but a clear focus for EU
- EU also trying to push for HPC@EU
  - 1 BEur on the table
  - Not yet defined through calls and grants
  - INFN involved in previous projects (EuroExa, ExaNest)