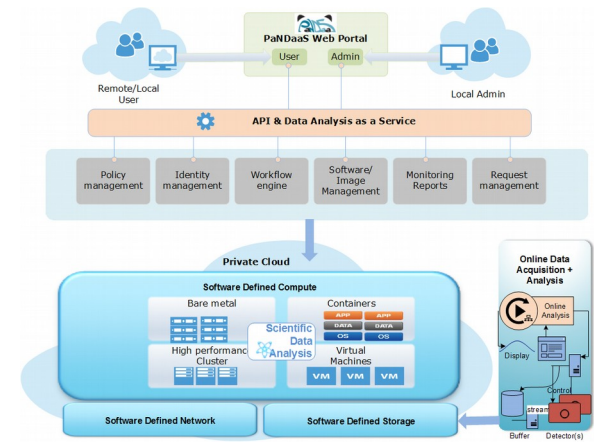




| The European Synchrotron

## VISION

***Provide users with services which give them remote access to their data and to pre-installed data analysis software packages to browse, reduce and analyse their data via remote login and/or a web interface***



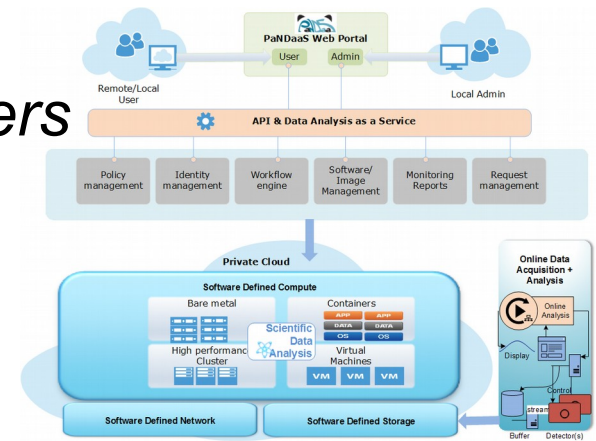
## GOALS for CALYPSO+

- ***Deploy a prototype remote services***
- ***Test typical use cases with real users***
- ***Prepare a blueprint + position paper for Photons***
- ***Collaborate on technical + scientific + with Neutron***



## IT TECHNOLOGY

*Debian Linux* for operating system  
*GPFS + NFS/SMB* for central file servers  
*LDAP* for on-site identity management  
*OpenStack* for private cloud  
*HNSciCloud* for public cloud  
*Python/C++* for data analysis  
*Docker* for packaging



## PANDATA products

- *ICAT* for metadata catalogue
- *Umbrella* for identity management



## HUMAN

**Benoit** – *sys admin*

**Elodie** – *sys admin*

**Alejandro** – *data manager*

**Armando** – *data analyser*

**Olof** – *data analyser*

**Andy** – *project manager*

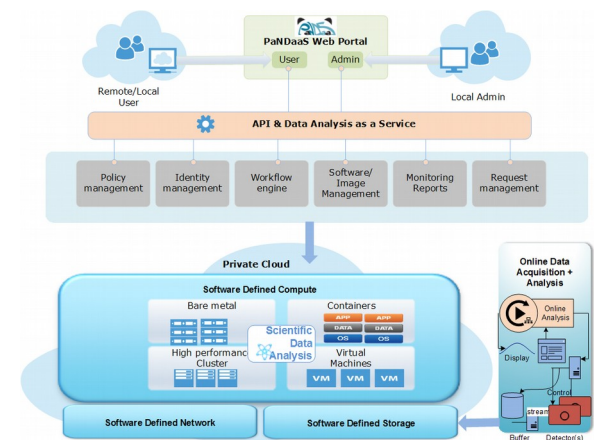
**Rudolf** – *director*

**Scientists + Users**



## COMPUTING

- *Hardware for **small cluster***
- **Silx software** for data analysis
- *Remote access to **central file server***



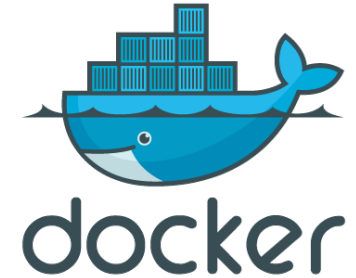
Hardware used for the pilot:

- 3 R630 with SLC SSD.
- 1 R630 with rotating HDD
- 3 R730xd
- 2 SAN equalogic.
- Local 40G network

Goal of the pilot:

- Learn openstack
- Test several architectures





**Kolla** has been chosen to **deploy openstack**:

-All **openstack services** are packages and deployed as Containers (compute, storage, network and others) to simplify DevOps.

-This gives a lot of flexibility, as this is independent from the system. Setup and configuration are also clearly separated from operating containers that are constant.

-**Cost** is pretty **high** at the beginning to **learn** the **tool**, but we do think there will be a **great benefit** when deploying production nodes, or when we will have to upgrade.



-**GlusterFS** vs. **Ceph**: GlusterFS is easier and seems more performant. But as **Ceph** is included by **default** with **Kolla**, we are giving it a try. So far so good, we are also **pretty happy** about it for now.

-We are **not** going to use **SAN** arrays. Too complex, and seems more interesting to use **local drives** through Ceph.

-We have an **operational platform**, but there are still openstack modules we need to explore:

- **Swift**: for object storage
- Part of **Neutron** (network)
- Ceph administration.

- ✓ • Project Coordination + workshops (technical coordination)
- ✓ • Science Use Case definition and collation (spectroscopy, tomography, diffraction, MX, industrial use cases)
- ✓ • Packaging for applications (Docker, Debian)
- ✓ • Cloud setup and deployment (OpenStack)
  - Configuration of site and test sites
- ✓ • Port and package applications and examples (Silx)
- ✓ • Umbrella authentication (IdP, eduGain)
- ✓ • User portal development (web development)
  - Definition, development and distribution of mini demonstrator platform to test sites
- ✓ • Report on results and links to HNSciCloud + EOSC (participant of HNSciCloud + EOSC WG)



- Browse and display data remotely via web
- Run standard analysis programs via web / remote login
- Store workflows for data analysis processing
- Follow standard workflows for data processing
- Data analysis tutorials and documentation online
- Get scientific results instead raw data
- Export data to public cloud for further analysis
- Access to more hardware resources than standard laptop

# BIG PICTURE

