JRA2

Demonstrator of a Photon Science Analysis Service

This talk ...

Data Analysis = Major Bottleneck @ Light sources

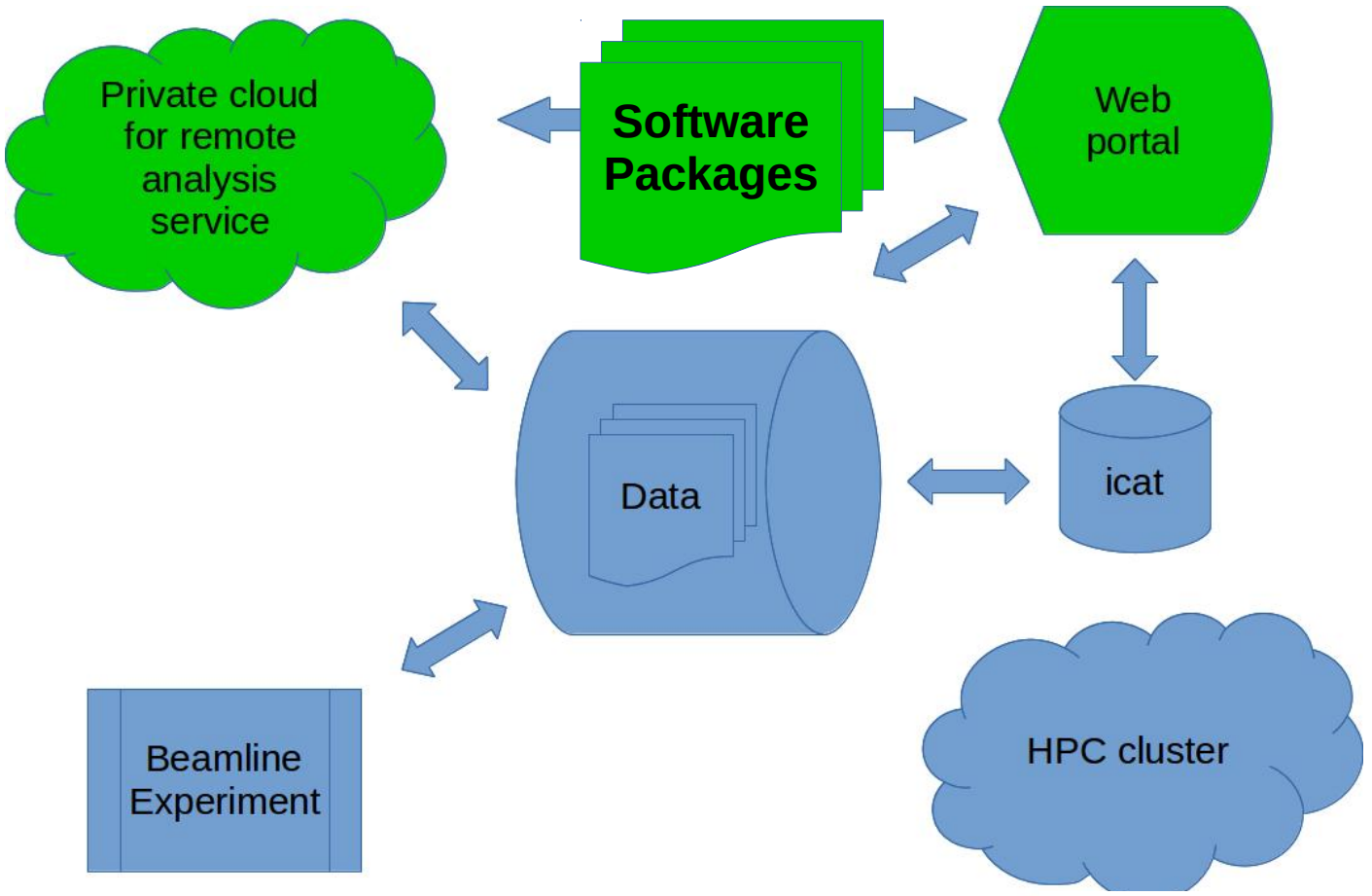
- **Most money** is invested in accelerator, beamline equipment, detectors, computing hardware and infrastructure
- **Little or no money** is invested in making professional data analysis software – *increasing data makes problem worse every year!*

JRA2 budget

H2020-INFRAIA01-2016-2017

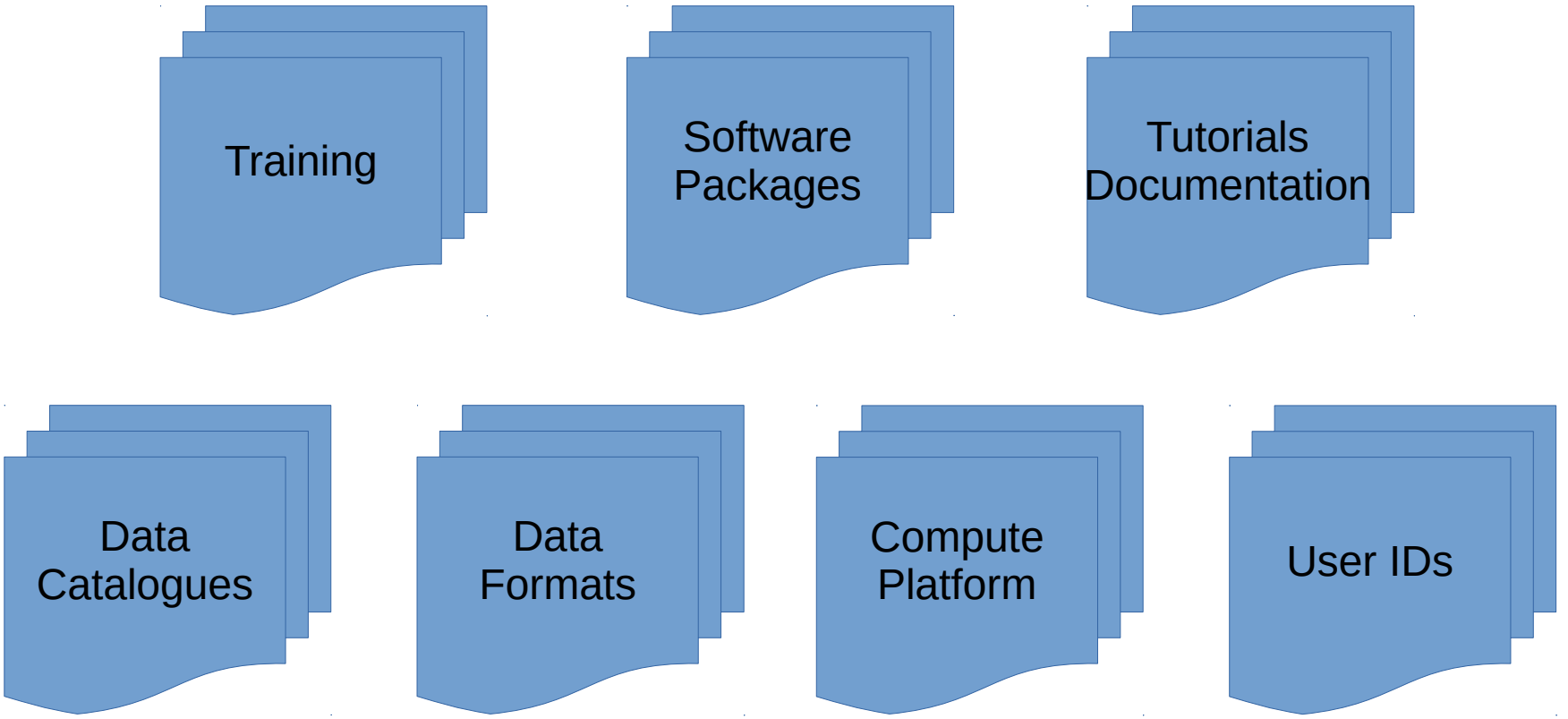
Work package number	24	**Lead beneficiary**				PSI/ ESRF	
Work package title	**JRA2: Demonstrator of a Photon Science Analysis Service (DaaS)**						
Participant number	15	12	4	6	7	8	9
Short name of participant	**PSI**	**ESRF**	CELLS	DESY	DIAMOND	ELETTRA	XFEL
Person/months per participant:	30	28	12	15	6	12	0
Participant number	18						
Short name of participant	SOLEIL						
Person/months per participant:	6						
Start month	1			**End month**	48		

A Platform for Remote Data Analysis as a Service



CALIPSOplus

We need many building blocks for data analysis !



Build on existing solutions !

HERCULES European School

Workshops

Training

Software Packages

Tutorials Documentation

jupyterhub

https://software.pan-data.eu

Data Catalogue(s)

Data Format(s)

Compute Platform

User ID(s)

ICAT

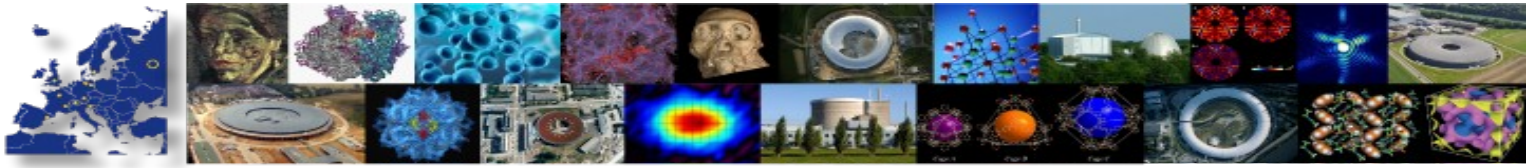
NeXus HDF

HPC / VMs / CLOUD

umbrella

pandata

http://pan-data.eu/



PaNData Software Catalogue

PaNdata Softwa... ×

https://software.pan-data.eu 150% Search

Getting Started AWS PANDAAS Working Gro... TANGO tango sympa pan-data.eu Time and Absences TweetDeck TLaGit PyFAI Tango Tango Controls · Webu Meditation Timer

Browse Software Institutes sign up log in

search for software... Add new software

PaNdata Software Catalogue

PaNdata software catalogue is a database of software used mainly for data analysis of neutron and photon experiments.

This database can be freely consulted. It gives an overview of software available for neutron and photon experiments and their use with respect to instruments at experimental facilities.

By registering and logging-in new software can be entered and it will appear in the database after moderation.

Browse software

Recently added software

ANKAphase

ANKAphase processes X-ray inline phase-contrast radiographs by reconstructing the projected thickness of the object(s) imaged. The tool uses a single-...

BONSU

Bonsu is an interactive phase retrieval suite, designed for phase retrieval with real-time visualisation in both two and three dimensions. It includes...

Recently updated software

- iFit
- ANKAphase
- WIEN2k

FEL Software Catalogue

Appl Cryst JAC Journal of Applied Crystallography

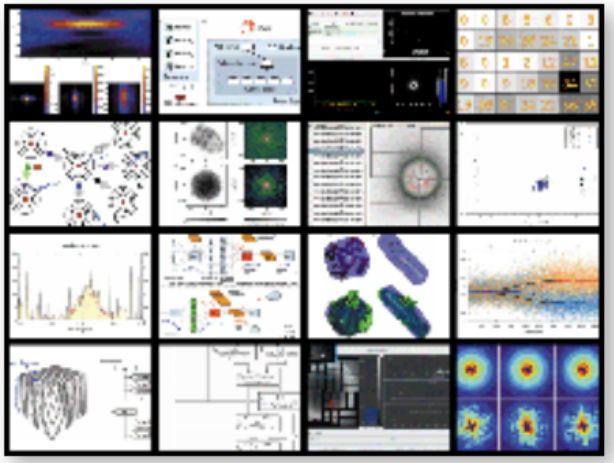
Appl Cryst JAC JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN: 1600-5767

CCP-FEL: a collection of computer programs for FEL research (August 2016)

Guest editors: Filipe R. N. C. Maia, Thomas A. White, N. Duane Loh and Janos Hajdu

This virtual special issue of *Journal of Applied Crystallography* brings together a series of specially commissioned articles describing software for free-electron laser research. These articles were published in the journal between April and August 2016.



Cover illustration: CCP-FEL: a collection of computer programs for free-electron laser research.

Task 1 – Design a Platform for Remote Data Analysis as a Service

- Task leader - **ESRF**
- Effort - 18m
- Build on existing solutions
- Technologies – HPC, VMs, Docker, OpenStack, Umbrella
- Challenge – each site has different IT setup
- Result – a common architecture

Deliverables

Deliverable Number	D24.2
Deliverable Title	Blueprint on implementing a DAAS platform
Lead Beneficiary	ESRF
Type	Report
Dissemination Level	Public (including the Commission Services)
Due date of delivery	Month 18 (originally month 12 but extended by 6 months)

Deliverable report

D24.2

Blueprint on implementing a DAAS platform *(month 12)*

WP24

Demonstrator of a Photon Science Analysis Service (DaaS)

deliverable report

PROJECT DETAILS

PROJECT ACRONYM
CALIPSOplus

GRANT AGREEMENT NO:
xxxxxxxx

START DATE
01/05/2017

PROJECT TITLE
Demonstrator of a Photon Science Analysis Service (DaaS)

CALL Horizon 2020-H2020-INFRAIA-2016-2017
INFRAIA-01-2016-2017: Convenient Access to Light Sources Open to Innovation, Science and to the World (CALIPSOplus)

DELIVERABLE DETAILS

WORK PACKAGE ID
WP24

WORK PACKAGE TITLE
WP24

Task 1 –

Design a Platform for Remote Data Analysis as a Service

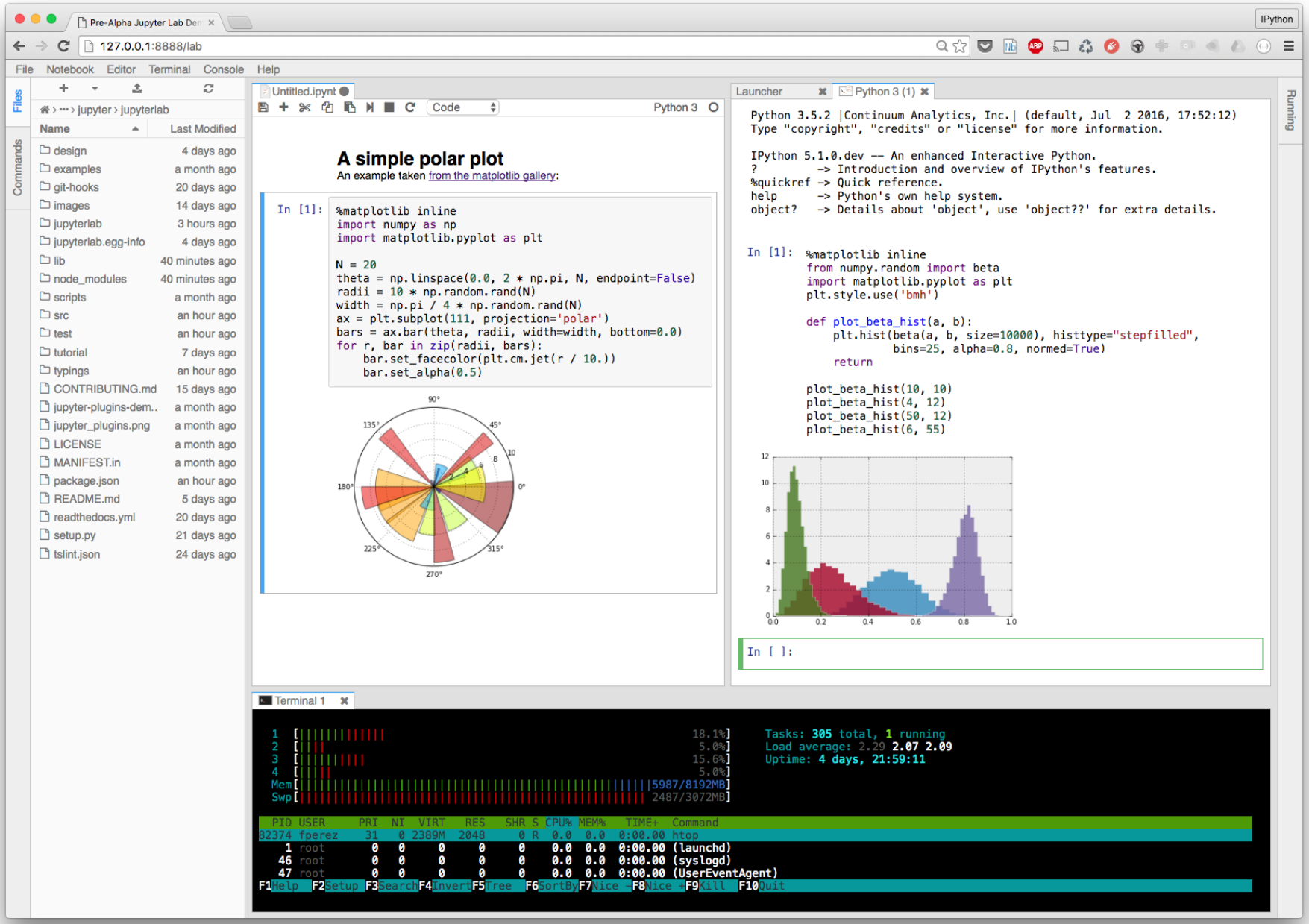
Task 2 – Collect and Compare Offline Data Analysis Software

- Task leader - **PSI**
- Effort - 8m
- Use cases – SAXS, Tomography, Diffraction
- Software – To be defined
- Users – Industrial users are a good candidate
- Result – list of common generic and specific data analysis tasks

Task 3 – Implement DaaS platform on each site

- Task leader - **ESRF**
- Effort - 23m
- Sites – ESRF, PSI, ALBA, DLS, SOLEIL, ELETTRA
- Software – Architecture defined in Task 1
- Result – a common generic service for Data Analysis

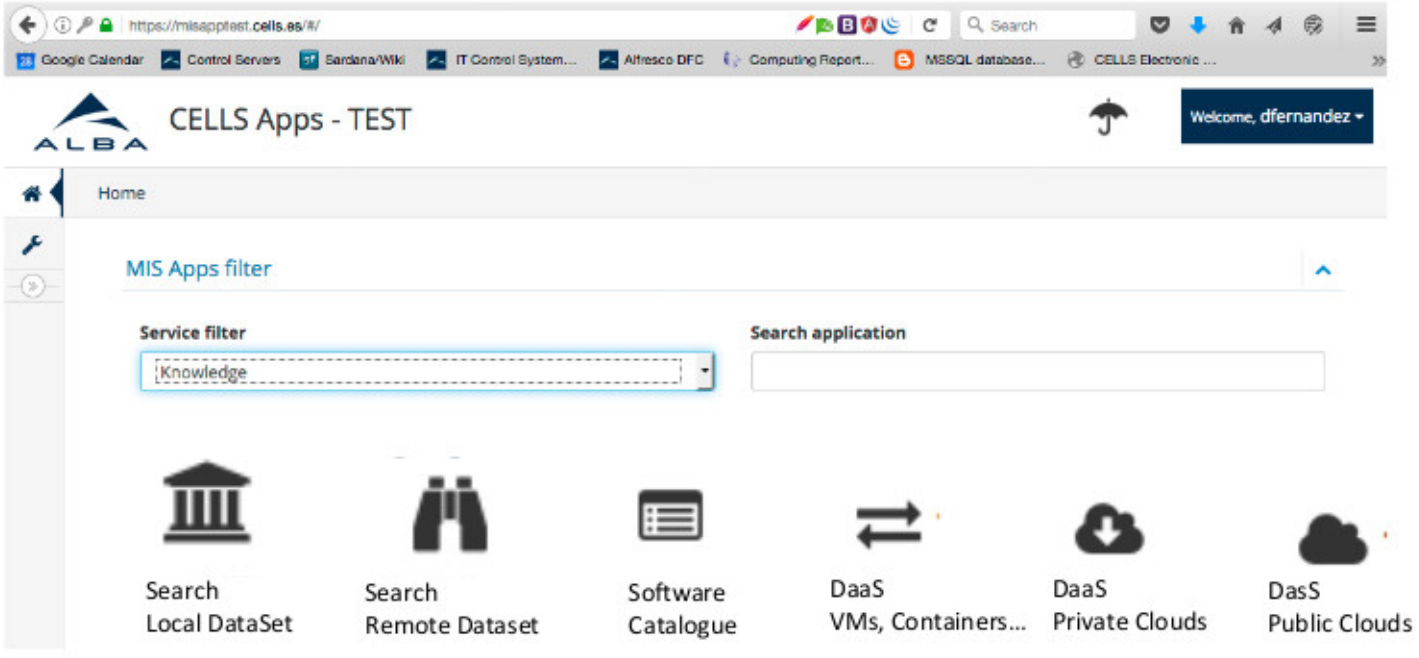
Task 4 – Platform to enable generic notebook service based on JupyterHub



Task 4 – Design and Implement DaaS platform portal

- Task leader - **ALBA**
- Effort - 16m
- Sites – ESRF, PSI, ALBA, DLS, SOLEIL, ELETTRA
- Software – Re-use existing solution
- Users – Friendly + Industrial users
- Result – a common portal for accessing DAAS

Task 4 – Design and Implement DaaS platform portal



Task 5 – Deploy and Package at least 2 Data Analysis software packages

- Task leader - **DESY**
- Effort - 20m
- Software – Identified by Task 2
- Users – Friendly + Industrial users
- Result – data analysis software packages

Task 6 – Extend and Deploy Umbrella as standard authentication service

- Task leader - **PSI**
- Effort - 18m
- Sub-tasks – Consolidate existing solution, integrate with eduGain, Jisc, icat, solve sso security issues
- Result – Umbrella integration and support

Task 7 – Test Use Cases with real Users

- Task leader - **ELETTRA**
- Effort - 6m
- Users – Friendly + Industrial Users
- Result – feedback from real users

Deliverables

Deliverable Number	D24.5
Deliverable Title	Report on test and deployment of mini demonstrator on at least six sites
Lead Beneficiary	ELETTRA
Authors	George Kourousias (ELETTRA) Ivan Adrian (ELETTRA) Daniel Salvat (ALBA) Alex Camps (ALBA) Aidan Campbell (ESRF) Andy Götz (ESRF) Gregory Viguier (SOLEIL) Tom Schoonjans (DIAMOND) Kai Kaminski (PSI) Johannes Reppin (DESY)
Type	Report
Dissemination Level	Public
Due date of delivery	Month 24

Report on test and deployment of mini demonstrator on at least six sites

Test Survey structure and summary

The system as a whole is based on blueprint presented in D24.2 *Blueprint on implementing a DAAS platform,* and is made of multiple interconnected software components and services, many of which are still under development. Even if the project has common objectives, each facility may have slightly different needs from the portal and some have more advanced deployments than other. Note that the final deployment will be reported during the D25.7 Workshop. This document reports the testing of the system as of M24 by presenting an extensive survey and its summary. This survey has a formal section of 9 question (section A) and a general one where useful information regarding the testing and deployment is reported (section B; 7 questions). Section A is the examination of the system through *Functional* and *Non-functional Testing*. The outline of this examination is based on the following structure:

Task 7 –

Test Use Cases with real Users

Calipsoplus JRA2 Deliverables

D#	Deliverable name	Task	Planned delivery date
D24.1	Report on kick-off meeting workshop for the CALIPSOplus partners to present their needs for remote data analysis		M3
D24.2	Blueprint on implementing a platform and manuals for the implementation at the different sites		M18
D24.3	Cross site use case requirements report including comparison of existing solutions		M12
D24.4	Software packages for the selected experiment use cases ready to install and run		M18
D24.5	Report on test and deployment of mini demonstrator on at least six sites		M24
D24.6	White paper on sustainability of HHScienceCloud and European Open Science Cloud for synchrotron and FEL applications		M36
D24.7	Organisation of a workshop to present the results of the DaaS demonstrator and obtain feedback from users on how this approach fits the current needs. The output of the workshop will be a white paper on how the remote data analysis needs of the photon community can be best served using modern computing paradigms		M36

http://www.calipsoplus.eu/jra2-daas/#Del_JRA2

Batch schedulers survey

Site	Batch Scheduler	Satisfied
HZDR	Slurm	yes
CELLS	Slurm	yes
DLS	UGE	yes
DESY	Slurm	yes
ILL	Torque+Maui	yes
PSI	Slurm	moderately
ELETTRA	Slurm	good enough
SOLEIL	Slurm	extremely
MAXIV	Slurm	yes

Learning from the PAST ...

"Those who don't study history are doomed to repeat it. Yet those who *do* study history are doomed to stand by helplessly while everyone else repeats it."

Conclusion

- JRA2 was a small step in the right direction (preparing for COVID-19)
- It fostered collaboration and exchange on **data analysis as a service**
- The **demonstrator** helped some users with generic and some specific data analysis tasks
- The **long term goal** was to prepare for bigger projects in the future to solve the data analysis problem

- **PaNOSC** and **ExPaNDS ...**

CALIPSOplus

JRA2 Deliverables

(**D1**) **Kickoff** workshop (**M3**) - **27 June 2017 @ ESRF**

(**D2**) **Blueprint** on implementing a platform (**M12**).

(**D3**) Cross site **use case** requirement report (**M12**)

(**D4**) **Software packages** for the selected experiment use cases (**M18**)

JRA2 Deliverables

(**D5**) **Report** on test and deployment of demonstrator on six sites (**M24**)

(**D6**) **White paper** on suitability of HNSciCloud and European Open Science Cloud (EOSC) for synchrotron and FEL applications (**M36**)

(**D7**) **Workshop** to present the results of the DaaS demonstrator and obtain feedback from users (**M36**).