PaNOSC + ExPaNDS tools for scientists + science clusters

Andy Götz (ESRF, PaNOSC coordinator)
Talk outline

1. Science clusters tools and platforms
2. PaN Tools for visualization, processing, remote analysis
3. PaN Training platforms
4. PaN DMPs
<table>
<thead>
<tr>
<th>Category</th>
<th>EOSC-Life</th>
<th>ENVRI-FAIR</th>
<th>ESCAPE</th>
<th>PaNOSC</th>
<th>SSHOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAI</td>
<td>LS-AAI</td>
<td>IAM &amp; existing</td>
<td>Umbrellald</td>
<td>AARC BPA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X509 infrastructure</td>
<td></td>
<td>implementations</td>
<td></td>
</tr>
<tr>
<td>Data Repositories</td>
<td>Covid portal FAIRsharing catalog</td>
<td>Certified data centers at RI level ENVRI-Hub</td>
<td>HEP Open Data Portal Virtual observatory Software catalogues</td>
<td>ICAT + Scicat Federated search, Human Organ Atlas Software catalogue,</td>
<td>&gt; 40 certified data centres (CTS)</td>
</tr>
<tr>
<td>Metadata standards</td>
<td>Fairsharing.org Bioschemas.org</td>
<td>DCAT-AP</td>
<td>FITS</td>
<td>Nexus/HDF5</td>
<td>CMDI, DDI</td>
</tr>
<tr>
<td>Data transfer</td>
<td>Globus, http, rsync</td>
<td>https</td>
<td>FTS (http, gridftp), EUDAT solutions</td>
<td>Globus, http, rsync</td>
<td>http, smtp</td>
</tr>
<tr>
<td>Training platforms</td>
<td>TESS</td>
<td>ENVRI Training Catalogue</td>
<td>Summer schools, training material in OSSR</td>
<td>Moodle + TESS</td>
<td>SSH Training Toolkit</td>
</tr>
<tr>
<td>Tools used / developed</td>
<td>DS-Wizard Galaxy WorkflowHub.eu</td>
<td>FIP, KB, ENVRI-Hub, Jupyter, VRE</td>
<td>Data Lake, OSSR, VO tools, VRE, Jupyter, etc. Citizen Science platforms</td>
<td>DS-Wizard, Jupyter, H5Web, VISA, VINYL,</td>
<td>SSH Open Marketplace; Virtual Collection Registry; Jupyter</td>
</tr>
</tbody>
</table>

**Science Clusters are providers + consumers of Tools + Platforms**
Active Data Management Plans

1. ExPaNDS and PaNOSC have adopted active DMPs
2. Active DMPs are updated at different phases of the project
3. ESS and ESRF have chosen to use DS Wizard developed by Elixir
4. Example of implementation @ ESRF
   1. Automatically generates a DMP automatically for every proposal
   2. 50 out of 82 questions are automatically filled in from DP/User/Data Portals
   3. DMPs offer a structured way to communicate information
   4. Users can use the DMP for satisfy funders requirements
   5. Next step is to use the DMPs to ensure users can manage their data
DMPs need to be more useful to convince users to use them.
H5Web Visualization Ecosystem

Generic HDF5 file viewer

- Integrated into **ESRF data portal**, for viewing files generated during experiments
- Available as **JupyterLab** and **VS Code extensions**, and soon as part of stand-alone web service, **myHDF5**, for viewing local and hosted HDF5 files

Used in various web applications at ESRF including:
- **Braggy**, diffraction image viewer (screenshot above)
- **Daiquiri**, beamline control and data acquisition software

https://github.com/silx-kit/h5web

PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.
H5Web Visualization Ecosystem

• Visualisation in the web is a common requirement for many tools

• H5Web provides a modular solution for plotting in ReactJS applications

• H5Web has been welcomed by many communities e.g. photon + neutron science, neuroscience, astronomy, space, microscopy, materials science, environment, commercial companies ...
Support for Jupyter notebooks

- **Jupyter service** now available at all PaNOSC and most ExPaNDS sites
- **Jupyter on Slurm** service developed: [https://github.com/silx-kit/jupyterhub_moss/](https://github.com/silx-kit/jupyterhub_moss/)
- **H5Web** Jupyterlab plugin developed
- **VISA** provides Jupyter service
- **PaN e-learning** platform provides Jupyter as a service
- **PaNOSC summer school** trained participants to program in Python using Jupyter
- **EGI** provided Jupyter and Binder as a service

Ten computer codes that transformed science
doi: [https://doi.org/10.1038/d41586-021-00075-2](https://doi.org/10.1038/d41586-021-00075-2)
Example Jupyter service @ https://jupyter-slurm.esrf.fr

- Enables users to run Jupyter Notebook on ESRF SLURM cluster

Unique users@ESRF: 156 (monthly average), 276 (total) over 4 months

PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.
Open Science with Jupyter notebooks

- Notebooks are a combination of code, output and annotation in one document.
- If used appropriately, makes publications reproducible.
- For example: one notebook per figure in publication.
- Notebooks from reproducible publications make the work re-usable.
- Currently, lots of time is used by researchers to repeat the work of others, before they can advance science.

EOSC could provide training on making reproducible publications for FAIR data.

VISA - Remote Data Processing/analyses

New compute instance
Please fill in the details below to create a new compute instance.

Experiments
Select the experiments you wish to associate with your dataset.

- instance not associated to any specific experiments

Computing Environment
Choose an environment

- Desktop staging
- Desktop
- Bliss

Choose hardware requirements

- 4 Cores 4GB memory
- 8 Cores 16GB memory
- 16 Cores 32GB memory
- 32 Cores 128GB memory

Infrastructure for remote data processing / analysis

Users dedicated VM

Access to data

Access to Provisioning of scientific SW using CVMFS and Containers

Access to the GPUs, HPC cluster

Infrastructure based on OpenStack

Development led by ILL in the scope of the PaNOSC project
Open Training - courses on Pan-learning.org

Achieving 100% Open Educational Resources:

1. Publish training material on pan-training.org
2. Develop learning material on pan-learning.org

Welcome to the e-Learning platform

This e-Learning platform hosts free education and training for scientists and students. Below you will find courses on both the theory of photon and neutron scattering and how to use python code or software for data reduction and modelling.

https://e-learning.pan-training.eu/moodle/
https://pan-training.eu/

PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.
Conclusion

1. Science clusters are providers + consumers of tools and platforms which need to be supported by the EOSC

2. Software tools make a big difference for scientists

3. Science clusters will actively collaborate around tools in the future