Global collaboration platform for Data Science

Jakub T. Mościcki, IT Storage, CERN
Gavin Kennedy, Cloud Services, AARNet
Integrated Service for Data Science

Users *collaborate on data* using an increasing number of applications.

Data available on **all devices**: mobile, laptops, desktops

Data easily *sharable* with individuals and groups

**Concurrent editing**

**Web-based Analysis**

Ready-to-go environment “one click away”

Integrated with **entire data repository**
Future Federated Analysis Platform

Advancing state of the art

One-click to create user groups, share projects and data

Domestic and remote users in the same collaborative workflow.

Full metadata awareness in the research workflow.

WLCG Grid

Integrate distributed computing capabilities with data sharing systems.

Application & data workflow.
The Project: CS³ Mesh

- 12 partners to create initial infrastructure
- connect existing, sustainable services
- 200K+ existing users
- Deliver a Global CS3 Collaboration Service for researchers, educators, data curators, analysts, …
- Provide an interoperable platform to easily share and deploy applications and software components to extend functionality of the service.
# Partners

## Coordinator

<table>
<thead>
<tr>
<th>Organization</th>
<th>Address</th>
<th>Activity type</th>
<th>Country</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH</td>
<td>Route De Meyrin Cern, 1211 Geneva 23</td>
<td>Research Organisations</td>
<td>Switzerland</td>
<td>WLCG T0</td>
</tr>
<tr>
<td>DANMARKS TEKNISKE UNIVERSITET</td>
<td></td>
<td></td>
<td>Denmark</td>
<td>UNI</td>
</tr>
<tr>
<td>SURFSARA BV</td>
<td></td>
<td></td>
<td>Netherlands</td>
<td>WLCG T1</td>
</tr>
<tr>
<td>INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK</td>
<td></td>
<td></td>
<td>Poland</td>
<td>HPC</td>
</tr>
<tr>
<td>CESNET ZAJMOVE SDRUZENI PRAVNICKYCH OSOB</td>
<td></td>
<td></td>
<td>Czechia</td>
<td>NREN</td>
</tr>
<tr>
<td>AARNET PTY LTD</td>
<td></td>
<td></td>
<td>Australia</td>
<td>NREN</td>
</tr>
<tr>
<td>SWITCH</td>
<td></td>
<td></td>
<td>Switzerland</td>
<td>NREN</td>
</tr>
<tr>
<td>WESTFAELISCHE WILHELMS-UNIVERSITAET MUENSTER</td>
<td></td>
<td></td>
<td>Germany</td>
<td>UNI</td>
</tr>
<tr>
<td>AILLERON SA</td>
<td></td>
<td></td>
<td>Poland</td>
<td>TECH ENTERPRISE</td>
</tr>
<tr>
<td>CUBBIT SRL</td>
<td></td>
<td></td>
<td>Italy</td>
<td>TECH STARTUP</td>
</tr>
<tr>
<td>JRC-JOINT RESEARCH CENTRE- EUROPEAN COMMISSION</td>
<td></td>
<td></td>
<td>Belgium</td>
<td>EC</td>
</tr>
<tr>
<td>FUNDACION ESADE</td>
<td></td>
<td></td>
<td>Spain</td>
<td>BUSINESS SCHOOL</td>
</tr>
</tbody>
</table>
## Work packages

<table>
<thead>
<tr>
<th>WP</th>
<th>Title</th>
<th>Leader</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Project Management</td>
<td>CERN</td>
</tr>
<tr>
<td>2</td>
<td>Federated Infrastructure</td>
<td>SURFSARA</td>
</tr>
<tr>
<td>3</td>
<td>Technology &amp; Foundation</td>
<td>CERN</td>
</tr>
<tr>
<td>4</td>
<td>Users &amp; Applications</td>
<td>PSNC</td>
</tr>
<tr>
<td>5</td>
<td>Dissemination, exploitation &amp; outreach</td>
<td>DTU</td>
</tr>
</tbody>
</table>
Collaborative Workflows

Integrate existing experience and technology

- Share, access, synchronize
- Metadata & tagging, Open Data (OpenAIRE, Zenodo, …)
- Data Science: Jupyter Notebooks (SWAN, …)
- Concurrent editing
- On-demand data transfers (Direct, FTS, DTN, Rucio, …)
Interoperability

- Add thin layer on top of existing services
- Use existing fabric
- Use existing standards
  - Introduce new APIs only if needed
- ScienceBox
- Close collaboration with industry
- Integrate into upstream products
Community Service

cs3community.org
2019—Worldwide Community

- 147 participants
- 70 institutions
- 25 countries
CLOUDSTOR IS A RESEARCH DATA AND APPLICATION INTEGRATOR
AARNETS ROLE IN THE CS3 MESH

• Non-funded partner in the project contributing to:
  • Standards for metadata handling
    • WP4: Users and Applications – Open Data, Metadata and Data Transfer
  • Facilitate integration with member and partner infrastructure, e.g. Pawsey, NCI, UQ, etc.
    • WP3: Federated Infrastructure – Storage interoperability
  • Use cases from research domains, e.g. HASS, bio, astronomy, etc
    • WP5: Outreach – New CS3 User Communities
BENEFITS TO AARNET COMMUNITY

• Benefits
  • Seamless international collaboration for users
  • Contribute to CS3 APIs that will extend CloudStor’s functionality
  • Create an environment for deployment of tools and applications to international collaborators.
    • Integrate to other platforms
    • Work collaboratively on the platforms across separate cloud platforms
    • Think of co-editing documents between international collaborators.
  • Access new tools and services created internationally.
BENEFITS TO AARNET COMMUNITY

• Benefits
  • Seamless international collaboration for users
  • Contribute to CS3 APIs that will extend CloudStor’s functionality
  • Create an environment for deployment of tools and applications to international collaborators.
    • Integrate to other platforms
    • Work collaboratively on the platforms across separate cloud platforms
    • Think of co-editing documents between international collaborators.
  • Access new tools and services created internationally.
Workshop on **Cloud Services** for **Synchronisation** and **Sharing**

27-29 January 2020
Copenhagen, Denmark

Project presentation and discussion

[cs3.deic.dk](http://cs3.deic.dk)
Win a Google Home Mini

How to enter:

1. Attend a session with an AARNet speaker. The more sessions you attend the greater your chance is to win.
2. Pick up your token.
3. Take your token to the AARNet booth to enter into the draw. Ten winners will be drawn.
Demo Schedule

Want to know more about AARNet Services? Visit the AARNet booth to participate in any of the following demonstrations.

**Tuesday**
- **11.00am – 11.15am**
  Science DMZ
- **1.15pm – 1.30pm**
  High speed uploading to CloudStor
- **1.40pm – 1.55pm**
  CloudStor Basics
- **3.50pm – 4.05pm**
  CloudStor SWAN (Service for Web-based Analysis)

**Wednesday**
- **8.45am – 9.00am**
  CloudStor Basics
- **11.00am – 11.15am**
  High speed uploading to CloudStor
- **1.15pm – 1.30pm**
  Science DMZ
- **1.40pm – 1.55pm**
  CloudStor SWAN (Service for Web-based Analysis)
- **3.50pm – 4.05pm**
  CloudStor Basics

**Thursday**
- **8.45am – 9.00am**
  High speed uploading to CloudStor
- **11.00am – 11.15am**
  Science DMZ
- **1.15pm – 1.30pm**
  CloudStor Basics
- **1.40pm – 1.55pm**
  CloudStor SWAN (Service for Web-based Analysis)