

Report from WP3

Gareth Murphy



panosc

**photon and neutron
open science cloud**

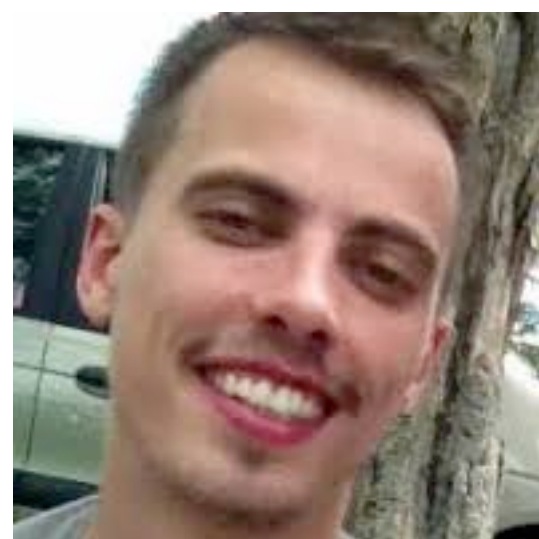
@garethcmurphy

@essneutron

Goals

- “The work package will create an API that existing catalog solutions can adopt to allow for ***seamless integration into EOSC via OpenAIRE..***”
- The API will allow for FAIR exposure of the data at the individual institutions through a catalogue service
- The API will allow federation, and exposure of metadata relevant for the area
- Existing APIs (e.g OAI-PMH) and communities (e.g. openarchives.org, Dublin Core Metadata Initiative (DCMI), OpenAIRE) will be taken into account
- The API will enable domain specific search extensions aware of the metadata definitions and usage at photon and neutron facilities

People



Jamie



Emiliano



Tobias



Lajos



Luis



Alex



Stuart



Alessandro



Gareth



Fredrik

We all have different catalogues

Table 1

Partner	CERIC	ESS	ELI	ESRF	ILL	XFEL
Catalogue	VUO (online storage NOT a catalog) -> ICAT	SciCat	TBD	ICAT	ILL Own	MyMdC
URL	https://vuo.elettra.trieste.it	https://scicat.esss.se	---	https://datahub.esrf.fr	https://data.ill.eu	https://in.xfel.eu/metadata
Login required	Yes	Yes	---	Yes	Yes	Yes
File formats	NeXus, HDF5, ASCII and many others	NeXus	---	EDF, SPEC, MCA, CBF, CCD, MCCD, HDF5, NeXus	NeXus and ILL Ascii	HDF5
Database	Oracle	MongoDB	---	Oracle and MongoDB	Oracle	MySQL and PostgreSQL
Language	Plsql, Python	Javascript	---	JAVA and Javascript	PHP	App: Ruby(onRails), Client: Python
Main technologies	WebDAV, Guacamole	Angular	---	React, NodeJS, EJB, JPA	Symfony, JQuery	Rails
Number of public datasets/files	0/0	We need a common interface -> PaNOSC WP3				
Using OAI-PMH	No					
Minting DOIs	Yes	Yes	---	Yes	Yes	Yes
Data/embargo policy	Not defined	Embargoed for 3 years	---	Embargoed for 3 years, ESRF Data Policy	Embargoed for 3 to 5 years, ILL Data Policy	Embargoed for 3 with possible extension to 5 years, XFEL Data Policy
Number of instruments connected to data catalogue	None	None	---	17	54	16

Use Cases	
Actor	Story
Scientist	I would like to know the states of Machine (i.e Energy) during an experiment
PTM	I would like to search for all metadata produced from the experiment (note that this are application specific)
Scientist	I would like to know the type of the software used for the analysis of data
Scientist	I want to be able to search for all data on water at the PanOSC facilities
Scientist	I would like to search for experiments/datasets by a particular metadata entity [eg sample, formula, title, scientific technique, sample type]
Scientist	I would like to search for detector type
Scientist	I would like to know the instruments setup
Scientist	I would like to retrieve the logbook of an experiment
Scientist	I would like to find all experimental data associated with a proposalId or experimental number
Scientist	I would like to find all experimental data matching abstract/proposal text
Scientist	I would like to see the proposal details associated with experimental data (abstract, title, experimental dates)
Scientist	I would like to search for data from a particular institute / instrument
Scientist	I would like to search for data acquired during a particular period
PTM	I would like to find my own data (experimental data from my proposals)

Actors	
Actor	Description
Proposal Team Member (PTM)	Person involved in a proposal
Open Access User (OAU)	Public member wanting to analyse open data
Scientist	Regroups Proposal Team Member and Open Access User
Admin	IT Staff
Facility Management (FM)	Directors, PaNOSC stakeholders (requiring metrics, usage, KPI...)
WP4 (Data Analysis Services)	Uses WP3 API to obtain metadata and to obtain access to data

WP4 Use Cases

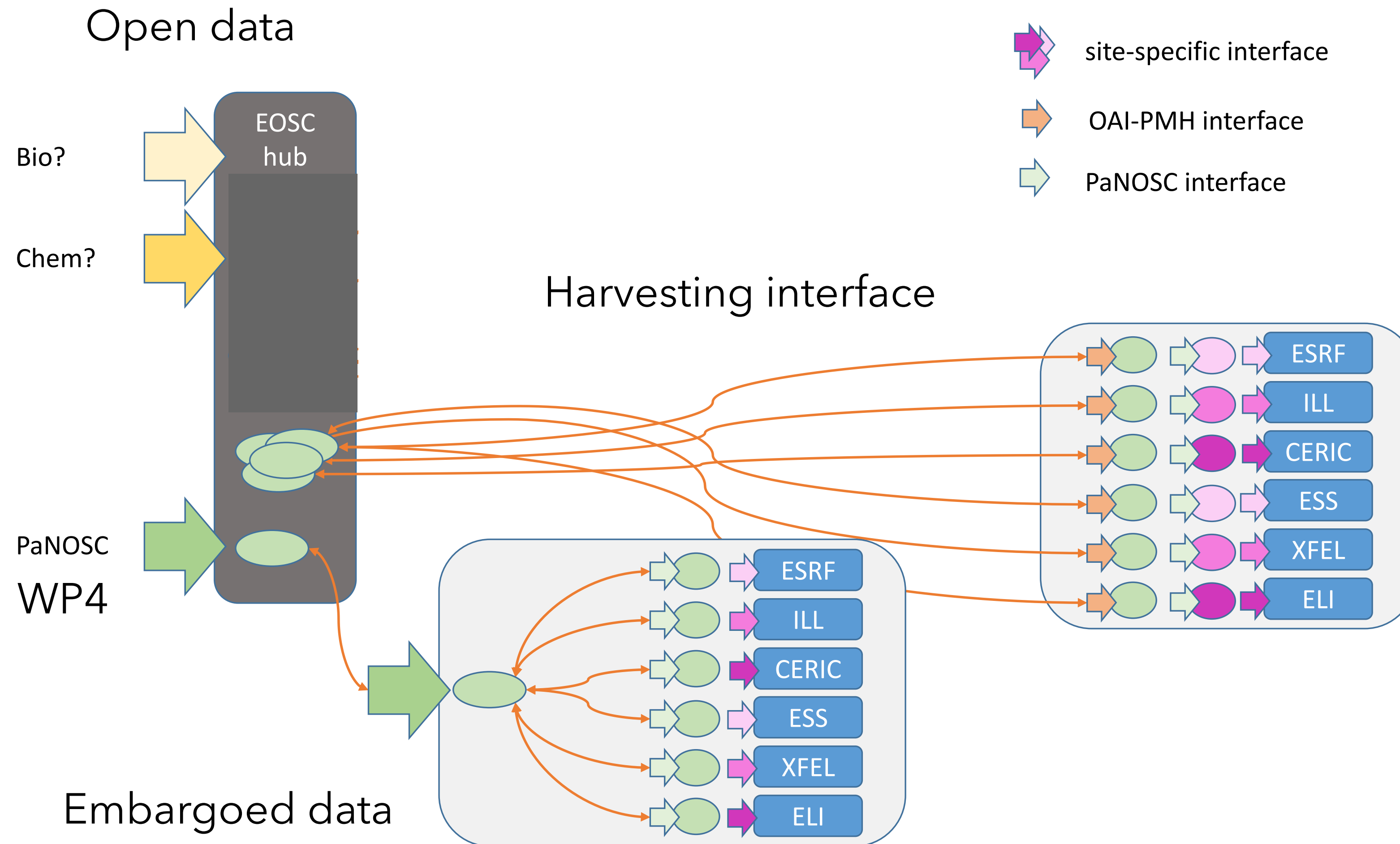
Actor	Story	
WP4	For the authenticated user, I would like to obtain a list (full metadata) of their associated proposals or data from different institutes	
WP4	I would like to obtain a list of open data (restricted metadata) from different institutes using specific search criteria	
WP4	Having identified a proposal or datafile/set of interest I would like to obtain the data in order to to analyse it	TI ol m er
WP4	I would like to upload processed data to the institute's catalogue	M

Harvesting vs authenticated search

- For EOSC, we need to provide all our metadata to be “harvested”. No login, anyone can access, truly open data
- For analysis users, can query for their own (embargoed) metadata. Requires login, authentication, securing data and metadata



Routes for open data and embargoed data



Metadata - core vs domain specific

- Name
- Abstract
- Date

"Dublin Core Metadata"

- Proposal
- URL to Binder/Jupyter Hub
- Technique (absorption, spectrometry)
- Wavelength (units? nm, Å, SI, keV?)
- Sample Thickness
- Beam CenterX, Y
- SampletoCCDDistance

"Domain Specific Metadata"



```
354 dataset:
355   required:
356     - parameters
357     - proposal
358     - id
359     - name
360     - startDate
361     - endDate
362     - location
363   properties:
364     parameters:
365       type: array
366       items:
367         type: object
368         properties:
369           name:
370             type: string
371           value:
372             type: string
373     proposal:
374       required:
375         - name
376         - id
377       properties:
378         name:
379           type: string
380         id:
381           type: number
382       type: object
383     id:
384       type: number
385     name:
386       type: string
387     startDate:
388       type: string
389     endDate:
390       type: string
391     location:
392       type: string
393     technique:
394       type: string
395     sampleName:
396       type: string
397     chemicalFormula:
398       type: string
```

PANOSC API Draft 0.1 OAS3

PANOSC API Draft for data catalog WP3

MIT

default ▾

GET	/proposals	
GET	/proposals/{proposalId}	Gets a specific proposal with a proposal identifier
GET	/proposals/{proposalId}/schedules	Gets schedules (session or visit) for a given proposal
GET	/proposals/{proposalId}/schedules/{scheduleId}/datasets	Gets datasets for a specific schedule of a proposal
GET	/proposals/{proposalId}/schedules/{scheduleId}/datasets/{datasetId}/files	Gets a list of files associated to a dataset
GET	/datasets	Returns a set of datasets given a PID
GET	/datasets/{datasetId}/status	
GET	/instruments	Returns all instruments for the facility

Schemas ▾

Swagger Editor

https://editor.swagger.io

AppsPaNOSC WP4 Me...V20 Test Beamlin...PANOSC WP3 kic...SciCatScicat @ ESSESS Public Datase...doisciCatProject/cat...Other Bookmarks

Swagger EditorSupported by SMARTBEARFileEditInsertGenerate ServerGenerate Client

```
1 openapi: 3.0.0
2 info:
3   title: PANOSC API Draft
4   description: PANOSC API Draft for data catalog WP3
5   version: "0.1"
6   license:
7     name: MIT
8     url: 'https://opensource.org/licenses/MIT'
9 paths:
10  '/proposals':
11    get:
12      responses:
13        '200':
14          description: Returns all proposals
15          content:
16            application/json:
17              schema:
18                $ref: '#/components/schemas/proposalsResponse'
19  '/proposals/{proposalId}':
20    get:
21      summary: Gets a specific proposal with a proposal identifier
22      parameters:
23        - in: path
24          description: Proposal's identifier
25          name: proposalId
26          schema:
27            type: integer
28            required: true
29      responses:
30        '200':
31          description: Returns a specific proposal for a given identifier
32          content:
33            application/json:
34              schema:
35                $ref: '#/components/schemas/proposalResponse'
36        '400':
37          $ref: '#/components/responses/error400'
```

schedule >

instrument >

member >

datafile >

dataset {

parameters* > [...]

proposal* > {...}

id* number

name* string

startDate* string

endDate* string

location* string

technique string

sampleName string

chemicalFormula string

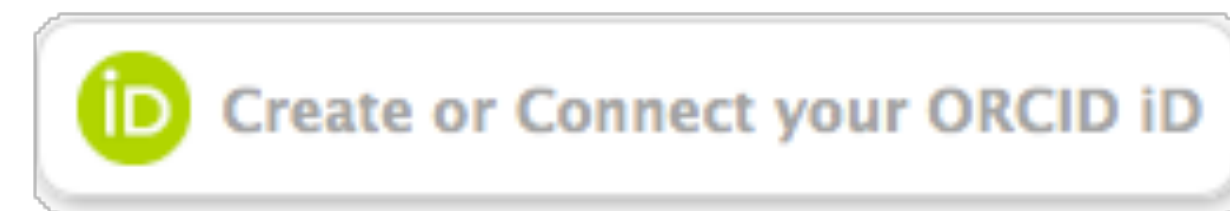
wavelength number

}

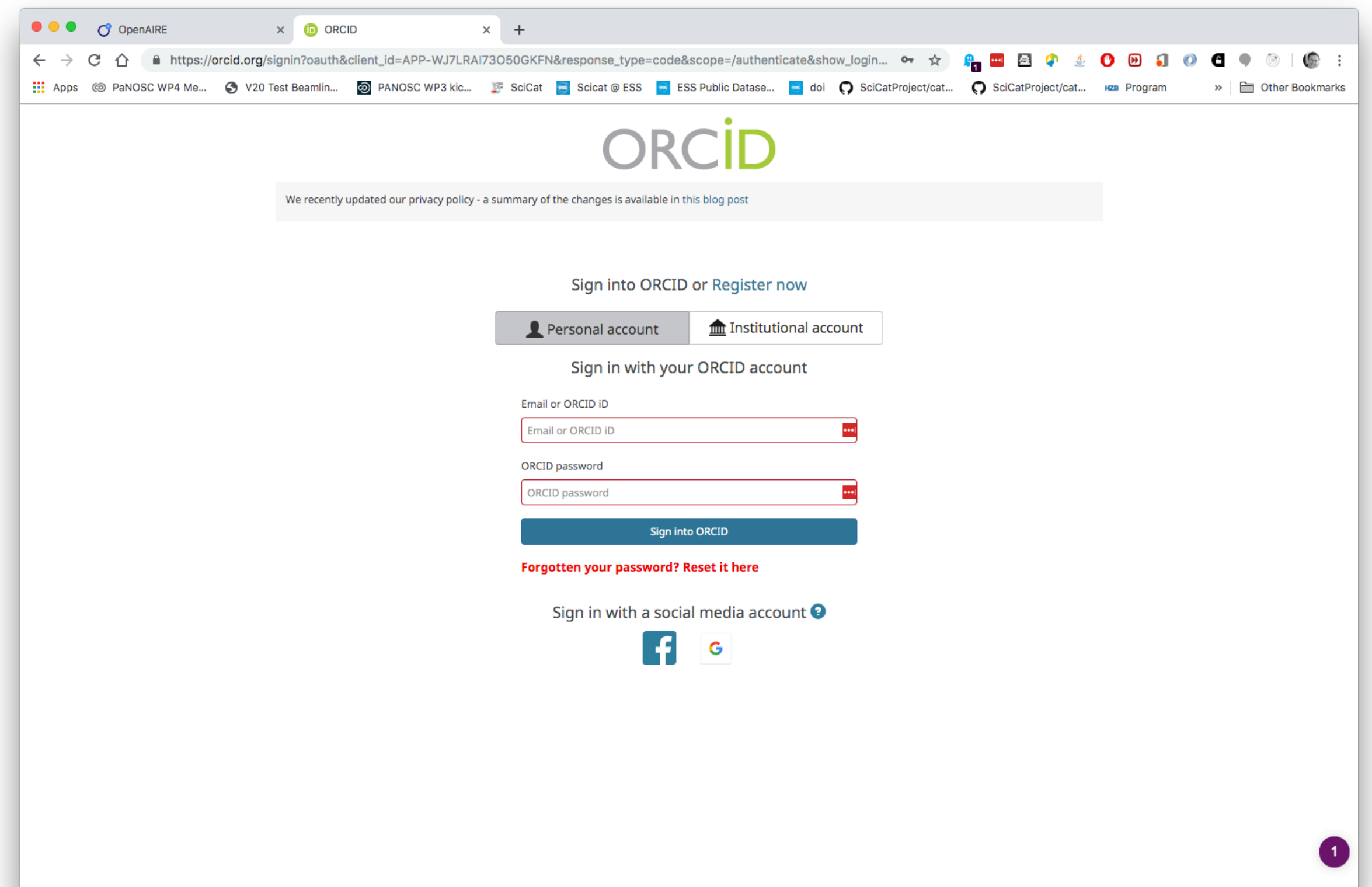
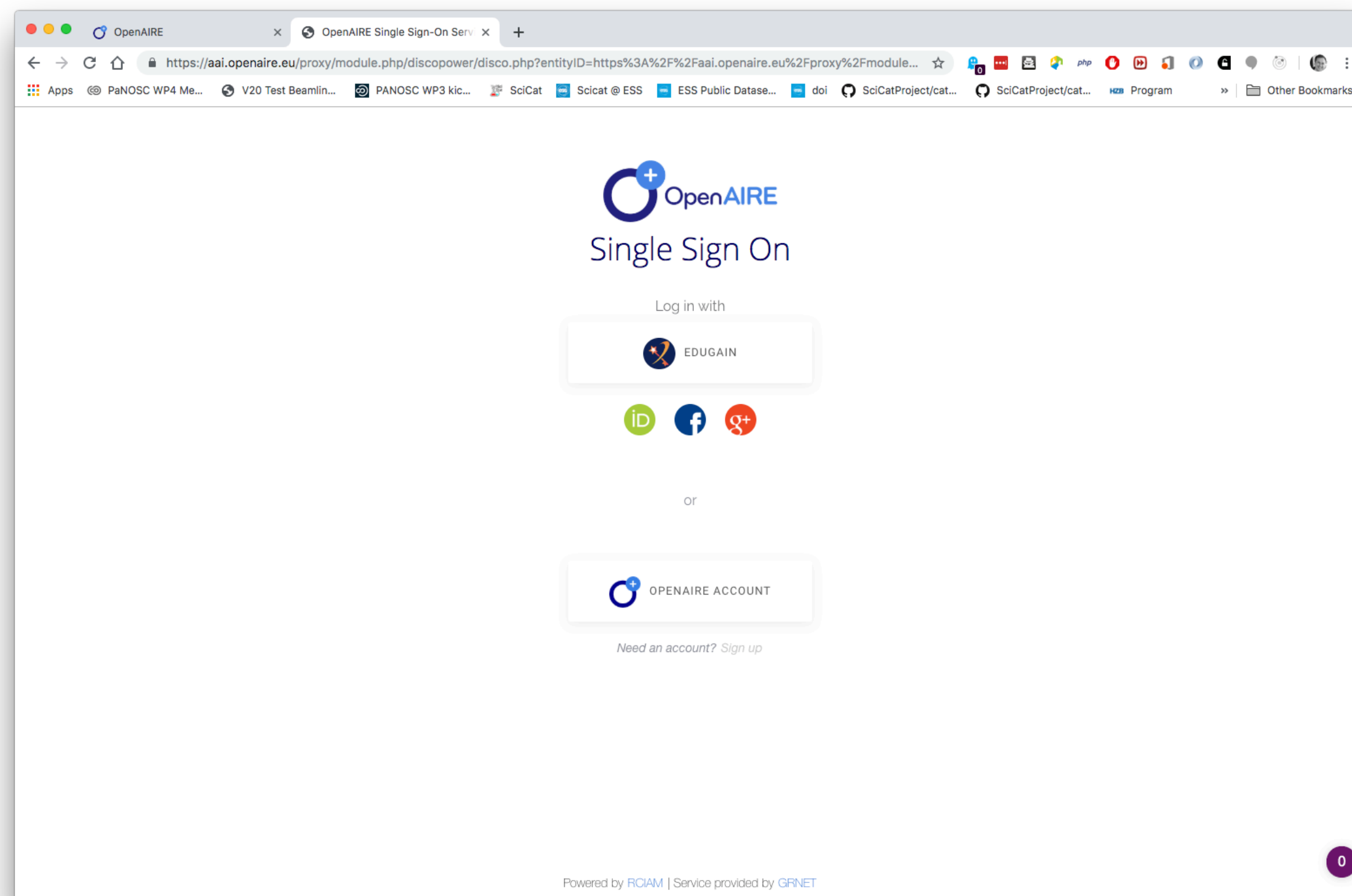
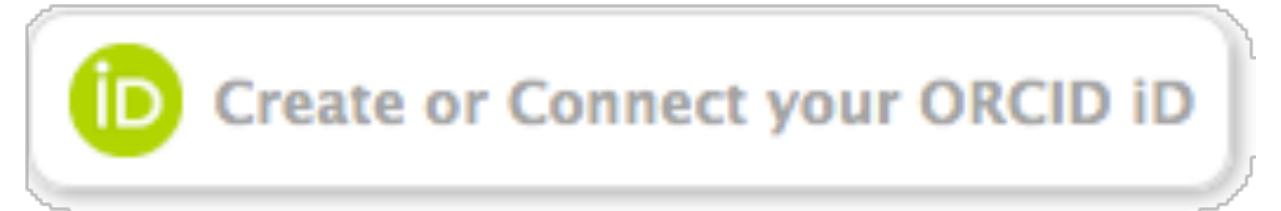
status >

How to authenticate for embargoed data

- WP6
- Edugain
- UmbrellaID
- ORCID signin
- WP2



- Each PanOSC institute implements Sign-In With ORCiD
- <https://members.orcid.org/api/integrate/orcid-sign-in>
- Users can then login with ORCiD at PaNOSC demonstrator
- If they have an accounts e.g. ESS with the will get back embargoed data
- If they have no account they get back only public data
- Can see demo at OpenAIRE



<https://workshops.ill.fr/event/218/>

PaNOSC WP3 Workshop (17-18 x

https://workshops.ill.fr/event/218/

Apps PaNOSC WP4 Me... SWAP board - Agil... V20 Test Beamlin... PANOSC WP3 kic... SciCat Scicat @ ESS Other Bookmarks

Europe/Paris English Login

PaNOSC WP3 Workshop

17-18 September 2019
ILL
Europe/Paris timezone

- Overview
- Timetable
- Accommodation
- Travel to EPN Campus
- Registration

The ILL will host a PaNOSC WP3 Workshop from **Tuesday 17th September to 18th September 2019** with the principal objective of advancing the specifications of the FAIR data API as defined in Task 3.1.

The programme will be updated over the following weeks, but suggested discussions and presentations are:

Day 1:

1. API endpoints
 1. Determine endpoints that satisfy all facilities
2. Search and API Query formulation
 1. NeXus: keywords, quantities, units
 2. Formulating complex searches (analyse use of a DSL?)
3. Returned data structure
 1. Mandatory and optional data
4. How to aggregate facility results (for use in a common portal)
5. API Validators

Day 2:

1. Harvesting by EOSC repos, OAI-PMH, extended schemas
2. Initial thoughts on Federated Search Demonstrator (Task 3.2)
3. Ties to optional EOSC services (authentication, file transfer...)
4. Internal milestones, deliverables and risks

The timetable will be finalised closer to the event.