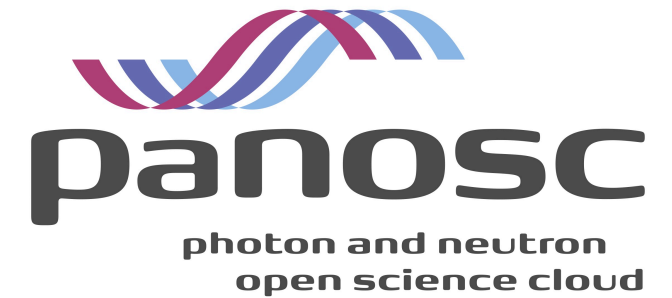# Introduction to myMdC
the European XFEL metadata catalogue

Luís Maia,
On behalf of European XFEL

11th February 2020

Agenda

■ **Introduction**

  ■ Context

  ■ Purpose

  ■ Landscape & Architecture

■ myMdC application

  ■ Roles and permissions

  ■ Key features

  ► Proposal authorisation, team administration and access to real data

  ► Proposal Run types and Samples

  ► Proposal workflow and notifications

  ► Run management activities (quality assessment, calibration request, describe runs)

  ► DatCite (DOI) management

  ► DAQ integration and

  ► Data Source Groups management

  ► (…)

■ Governance & Resources

■ Next steps (& EOSC)

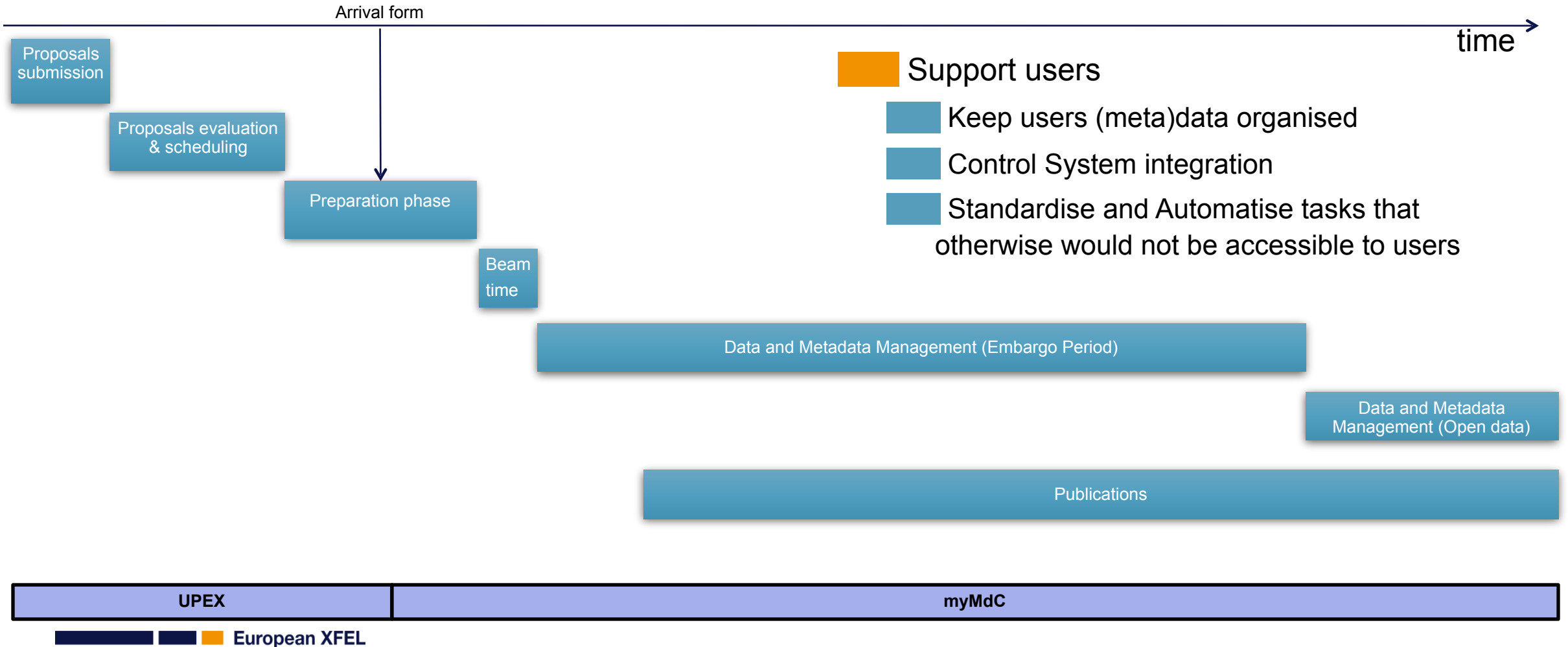**European XFEL**

Context (I)
Scientific Data Management

Organize and manage data in a coherent way

Aggregate and record large volumes of data

Provide possibility of inspecting data during experiment

Enable users to efficiently analyse experiment data when back at home institute

Prepare for open access era

https://www.xfel.eu/users/experiment_support/policies/scientific_data_policy/index_eng.html

Scientific data policy   Policy for samples and substances   User-supplied instrumentation/equipment   Electrical work

Users > Experiment support > Policies > Scientific data policy

Scientific data policy

Scientific data policy
As approved by the European XFEL Council on 30 June 2017

DOWNLOAD

Data retention policy

DOWNLOAD

General principles

Open access

Long term data curation

Embargo period with restricted access

European XFEL

Context (II)
Users support (and Experiments lifecycle)

Context (III)
Data Management Activities

■ Catalog and document the data generated at European XFEL

■ Data migration services

■ Keep track of data location

■ Allow data publishing, exportation and authorisation management

■ Manage permissions and accesses to the data in an uniform manner and guarantee that only authorised people have access to the (meta)data

Storage	Quota	Safety	Lifetime	Comment
dcache.raw	None	Tape Archive	6 months (t.b.c)	Raw data
raw	None	None	2 months (t.b.c)	Fast accessible raw data, lifetime not guaranteed
usr	5TB	Snapshots + Tape Backup	24 months (t.b.c)	user data, results
proc	None	None	6 months (t.b.c)	processed data e.g. calibrated
scratch	None	None	6 months	Temporary data (lifetime not guaranteed)
dcache.cal	None	Tape Archive	10 years	Calibration constants
cal	None	Tape archive	6 months	
user home	20GB	Snapshots + Tape Backup	Lifetime of the account	
archive.raw	None	-	Long-term	Long term means 5 years and XFEL will strive for 10 years
archive.cal	None	-	10 years	

■ Data Retention Policies control and execution

■ Data Quality Assessment

■ Data Calibration

■ Data Archiving

■ Data Cleanup

European XFEL

myMdC Purpose (I)

Provide a means of storing, retrieving and query raw and run based data in an organised way

Organize and manage data and metadata in a coherent way

Assist on the execution of the Data Management policies

Manage data and metadata authorisation and responsibilities (especially during the Embargo period)

Manage and restrict the GLOBAL unique entities used during data taking

   Proposal number

   Proposal path

   Sample

   Run type

European XFEL

myMdC Purpose (II)

Glue together different systems and services

UPEX
Users and internal proposals
Proposals Experiment team & Samples sheet
Check and sync users accepted data policies
Trigger proposal team snapshot

Registry
Users accounts
Staff organisational group
Instrument support groups*

LDAP
Users Proposal groups*

GPFS
Repository location
Universal folders management
Proposal folders creation
Runs data migration
Folders removal*
Data Audit and reconciliation

Elog
Elog logbook config file generation
Integration with Elog replacement tool

SLURM
Priority queues reservation for beamtime*

Karabo - DAQ
Universal Run Management
Run tagging and basic details
Number of files and size
Data Source Groups definition

Calibration Pipeline
Trigger Processed data generation*

calCat
Document used CCV on data calibration

Grid-FTP / Globus / Off-shore data exportation
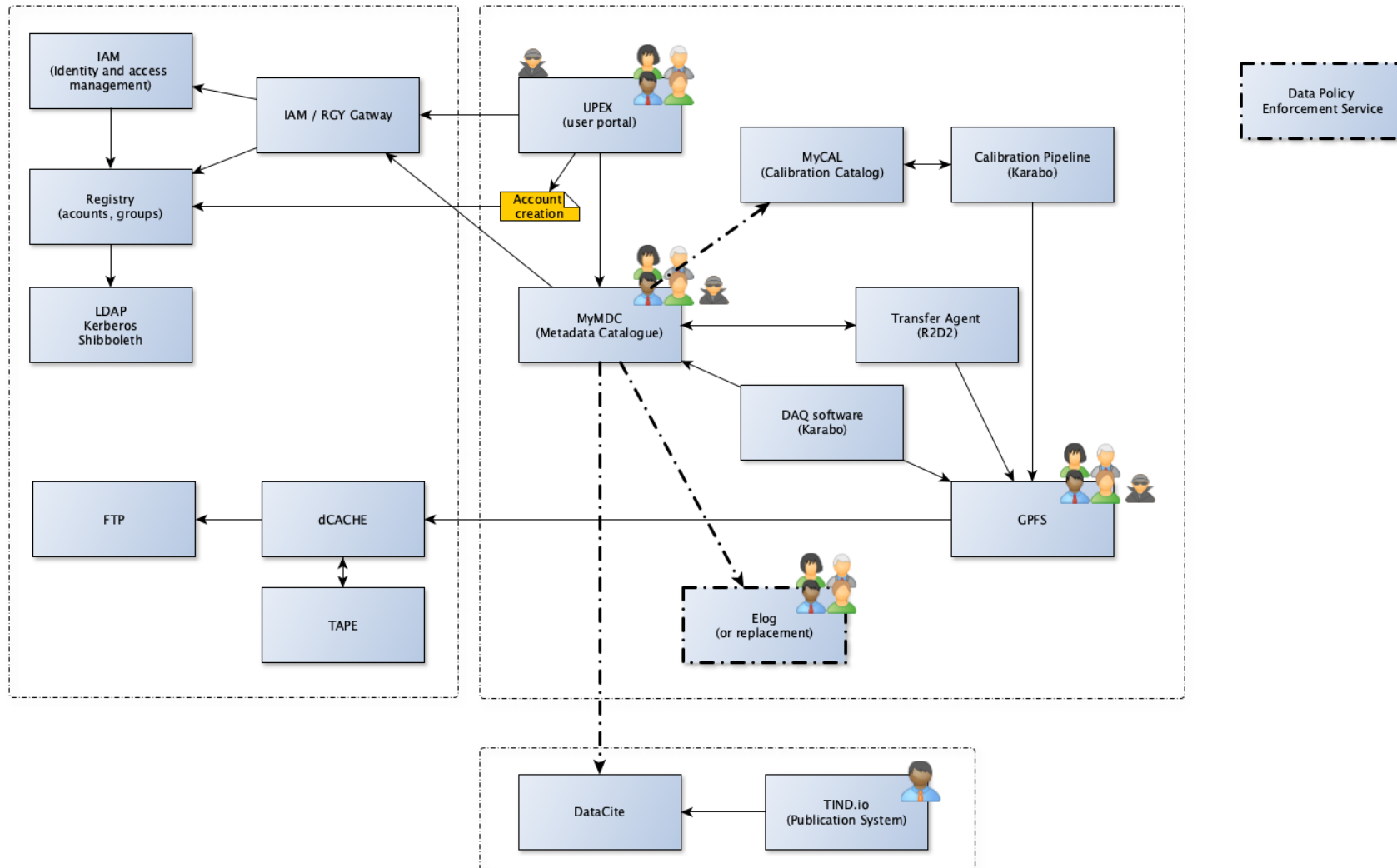Transfer Agent Off-shore data migration

DataCite / Tind.io
DOI generation and publish*
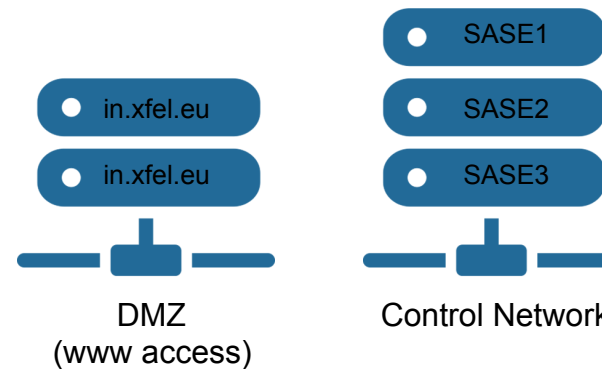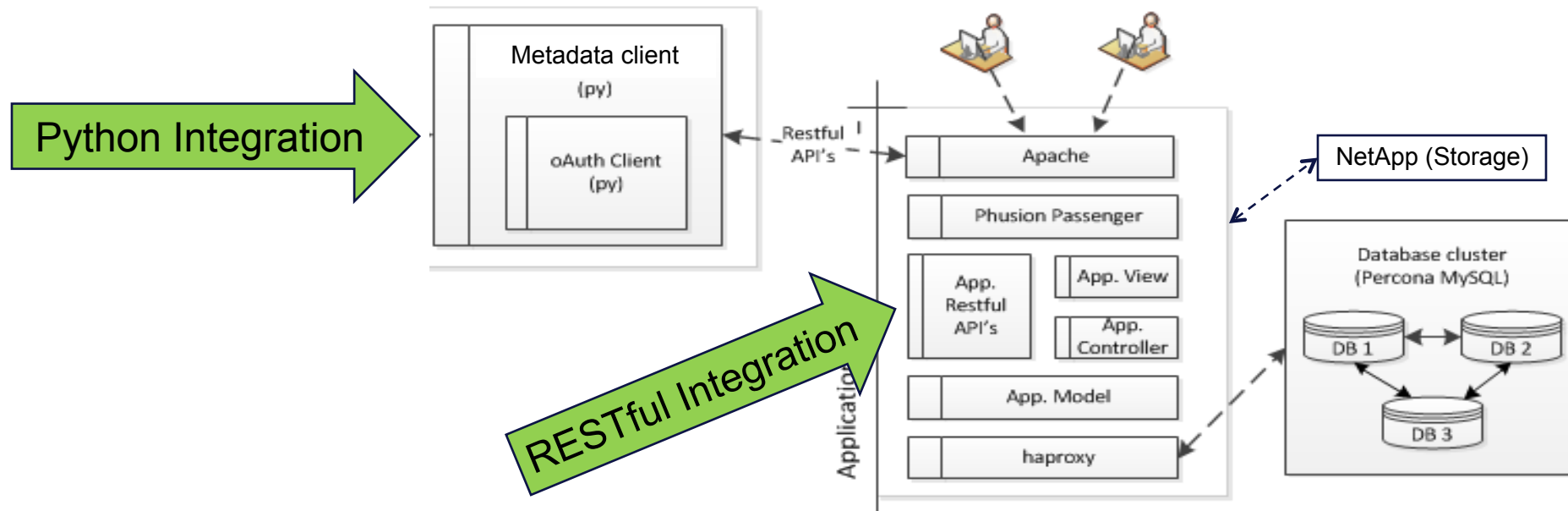Public DOI end point

Data Policy enforcement Service

Notifications*

Statistics

European XFEL

Landscape

Architecture

Agenda

🟧 Introduction

   🟦 Context

   🟦 Purpose

   🟦 Landscape & Architecture

🟧 myMdC application

   🟦 **Roles and permissions**

   🟦 Key features

      ► Proposal authorisation, team administration and access to real data

      ► Proposal Run types and Samples

      ► Proposal workflow and notifications

      ► Run management activities (quality assessment, calibration request, describe runs)

      ► DatCite (DOI) management

      ► DAQ integration and

      ► Data Source Groups management

      ► (…)

🟧 Governance & Resources

🟧 Next steps (& EOSC)

European XFEL

Roles and permissions (I)
myMdC roles

Administration

admin is the role assigned to myMdC administrators and ITDM OCD personnel

Users & Staff

read_only is the default role for everybody accessing myMdC

user is the role granted to users with UPEX account and the latest Data Policy accepted

Systems

Specific roles to be assign to system integration local users

List of Roles

Dashboard / Roles

≡ List

| | Filter | | ⟳ Refresh | ✕ |

	Id	▼	Name	Flg default
☐	1		admin	0
☐	2		trusted	0
☐	3		user	0
☐	4		read_only	1
☐	6		KARABO_SYS	0
☐	7		UPEX_SYS	0
☐	8		GPFS_SYS	0
☐	9		CAL_SYS	0

8 roles

European XFEL

Roles and permissions (II)
Responsibilities in the context of instrument and proposal

LS: Instrument's Leading Scientist

> Any user in myMdC that is associated as the leading scientist to the Instrument where the given proposal is taking place

BE: Instrument's Beamline Expert

> Any user in myMdC that is associated as an expert user to the Instrument where the given proposal is taking place

SG: Instrument's Support Group

> Any user in myMdC that is a member of the support group associated with the instrument where the given proposal is taking place (e.g. SPBDATA)

European XFEL

PI: Principal Investigator

> Any user in myMdC

MP: Main Proposer

> Any user in myMdC

LC: Local Contact

> Any user in myMdC that is a Beamline Expert

T_Part: Execution team Participant

> Any user in myMdC

T_DM: Execution team Data Manager

> Any user in myMdC

Others: Non users or non members

> Any user in myMdC not having any responsibility on the proposal and not being Administrator

Roles and permissions (III)
Proposal permission Grid

What / Who	PI	MP	T_Part	T_DM	LC	BE	LS	SG	Others
Index/search Proposals	Y	Y	Y	Y	Y	Y	Y	Y	N
Show Proposal (General, PI, Runs, Team)	Y	Y	Y	Y	Y	Y	Y	Y	N**
Edit Proposal	Y	Y	N	N^1	Y	Y^2	Y	N	N
New Proposal	N	N	N	N	N	N	N	N	N
Delete Proposal	N	N	N	N	N	N	N	N	N
History Page	Y	Y	Y	Y	Y	Y	Y	Y	N
Public Proposal DOI page	Y	Y	Y	Y	Y	Y	Y	Y	Y
Proposal Runs List	Y	Y	Y	Y	Y	Y	Y	Y	N**
Proposal Show Run Type	Y	Y	Y	Y	Y	Y	Y	Y	N**
Proposal Edit Run Type	Y	Y	N	N	N	Y^3	Y'	N	N
Proposal Create Run Type	Y	Y	N	N	N	Y^3	Y'	N	N
Proposal Delete Run Type	N	N	N	N	N	N	N	N	N
Proposal Show Sample	Y	Y	Y	Y	Y	Y	Y	Y	N**
Proposal Edit Sample	Y^4	Y^4	Y^4	Y^4	N	Y^3	Y'	N	N
Proposal Create Sample	N	N	N	N	N	Y^3		N	N
Proposal delete Sample	N	N	N	N	N	N	N	N	N
Proposal Index Runs	Y	Y	Y	Y	Y	Y	Y	Y	N**
Proposal Runs List Update	Y	Y	N	Y	N	Y^3	Y'	N	N
Proposal Create Runs	N	N	N	N	N	N	N	N	N
Proposal Delete Runs	N	N	N	N	N	N	N	N	N

Grant access to data

Notes:
(*): includes the ability to assess the quality of the run
('): instrument's leading scientist is also always a Beamline expert
(1): data managers can change Preferred Data output field
(2): for proposals in states: R, A, I
(3): for proposals in states: T, N, WLC, DLC, WOFC, DOFC, WONC, DONC, ONNA, R, A, I, F
(4): for proposals in states: N, WLC, DLC, WOFC, DOFC, WONC, DONC, ONNA, R, A, I
(**): Proposal is open data

European XFEL

Agenda

Introduction

- Context
- Purpose
- Landscape & Architecture

myMdC application
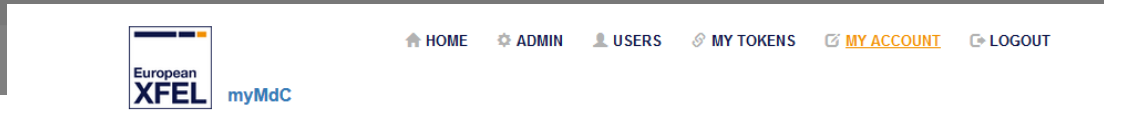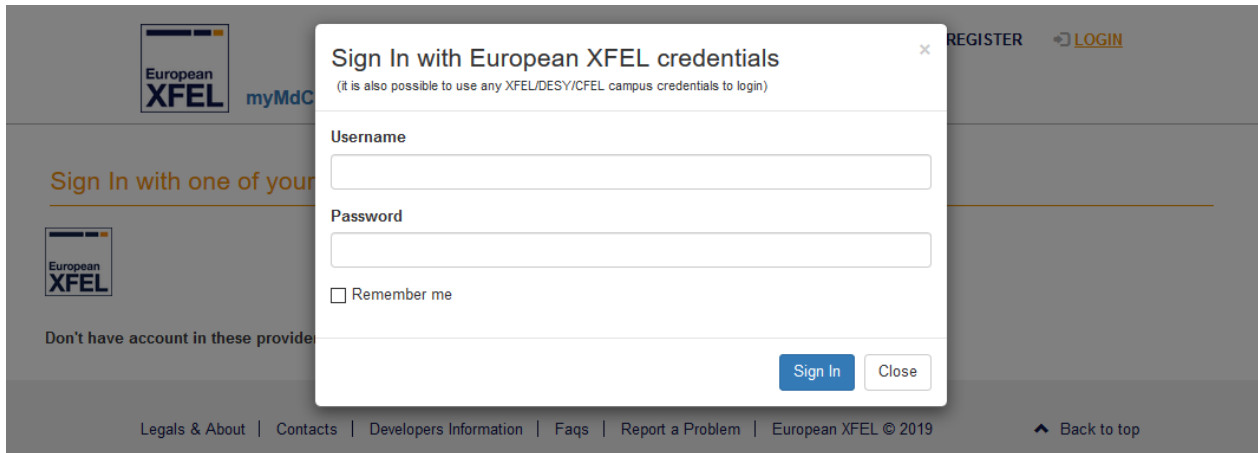
- Roles and permissions
- **Key features**
  - ► Proposal authorisation, team administration and access to real data
  - ► Proposal Run types and Samples
  - ► Proposal workflow and notifications
  - ► Run management activities (quality assessment, calibration request, describe runs)
  - ► DatCite (DOI) management
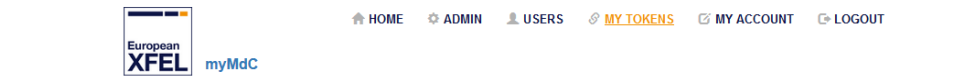 - ► DAQ integration and
  - ► Data Source Groups management
  - ► (…)

Governance & Resources

Next steps (& EOSC)

European XFEL

Key Features
Authentication
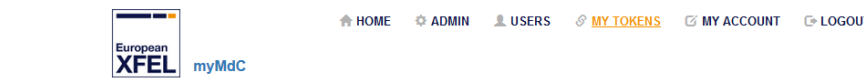
Key Features
Application Layout & Responsiveness

Administrators Only!

Key Features
Proposal details

Key Features
Proposal workflow and notifications

myMdC Proposals' Beamline status workflow

N: New
WLC: Waiting LDAP Configuration
DLC: LDAP Configuration Done
WOFC: Waiting GPFS Offline Configuration
DOFC: GPFS Offline Configuration Done
WONC: Waiting GPFS Online Configuration
DONC: GPFS Online Configuration Done
ONNA: GPFS Online Configuration Not Applicable
R: Ready
A: Active
I: Inactive
F: Finished (Beamtime)
FR: Finished and Reviewed
WCLO: Waiting (GPFS) Online Folder Removal
CLO: Closed
C: Cancelled
T: Test

Key Features
Proposal DOIs integration

Key Features
Proposal runs (data) management

Key Features
Proposal team and History

Proposal no. 900054

Back | Edit | Clone | + Add another | Runs | Delete | Beamtime status ▾

General | Public Information | Runs | Team | History

Experiment Team

Principal Investigator:	Djelloul Boukhelef - (boukhele) as Data Manager & Participant
Main Proposer:	Nasser Al-Qudami - (alqudami) as Data Manager & Participant
Team members 1:	Luis Maia - (maial) as Participant
Team members 2:	Jolanta Sztuk-Dambietz - (jsztuk) as Data Manager & Participant
Team members 3:	Maurizio Manetti - (manettim) as Participant
Team members 4:	Krzysztof Wrona - (wrona) as Data Manager & Participant
Team members 5:	Peter Zalden - (zaldenp) as Data Manager & Participant
Team members 6:	Janusz Malka - (jmalka) as Data Manager & Participant
Team members 7:	Juergen Hannappel - (hannappj) as Participant

Download: CSV | JSON | Elog format

Back | Edit | Clone | + Add another | Runs | Delete | Beamtime status ▾

Team members

* User	Data access	
Luis Maia	Participant	✕
Djelloul Boukhelef	Data Manager &	✕
Jolanta Sztuk-Dambietz	Data Manager &	✕
Maurizio Manetti	Participant	✕
Krzysztof Wrona	Data Manager &	
Peter Zalden	Data Manager &	
Nasser Al-Qudami	Data Manager &	
Janusz Malka	Data Manager &	
Juergen Hannappel	Participant	

⊕ add member

Audit fields

Created by	Janusz Malka

Proposal no. 900054

Back | Edit | Clone | + Add another | Runs | Delete | Beamtime status ▾

General | Public Information | Runs | Team | History

History

2019-02-14 15:27:59 +0100 – ## System ##

Preferred data output	Processed	⇒	Raw

2019-01-29 18:23:55 +0100 – ## System ##

Repository	---	⇒	XFEL_GPFS_ONLINE_SASE_2

2019-01-29 03:41:47 +0100 – Luis Maia

Beamtime status	Active	⇒	Finished (Beamtime)

2019-01-23 16:24:10 +0100 – Janusz Malka

Team changes	---	⇒	Juergen Hannappel (Participant)

2018-10-19 16:39:06 +0200 – Janusz Malka

Beamtime status	Ready	⇒	Active

2018-09-27 16:38:46 +0200 – Janusz Malka

Team changes	---	⇒	Janusz Malka (Data Manager & Participant)

2018-09-25 17:30:09 +0200 – Luis Maia

Beamtime status	GPFS Online Configuration Done	⇒	Ready

2018-09-25 17:28:04 +0200 – Luis Maia

Beamtime status	GPFS Offline Configuration Done	⇒	GPFS Online Configuration Done

European XFEL

Key Features
Integration with data acquisition

Key Features
Data Source Groups management

European XFEL

Key Features
RESTful API's

```
exflqr17747:metadata_client maial$ python3 maial_py.py
{'success': True, 'info': 'Got proposal successfully', 'app_info': {}, 'data': {'id': 85,
'number': 900054, 'title': 'TST SASE2 DAQ tests', 'abstract': '', 'proposal_folder': 'p900
054', 'def_proposal_path': '/gpfs/exfel/exp/TSTSA2/201831/p900054/raw', 'doi': '10.22003/X
FEL.EU-DATA-900054-00', 'url': '', 'instrument_id': 110, 'instrument_cycle_id': 31, 'princ
ipal_investigator_id': 51, 'main_proposer_id': 411, 'local_contact_id': None, 'orig_reposi
tory_id': 2, 'last_run': 82, 'begin_at': '2018-09-25T00:00:01.000+02:00', 'end_at': '2018-
12-31T23:59:58.000+01:00', 'release_at': '2021-12-31T23:59:59.000+01:00', 'num_delivered_s
hifts': None, 'flg_public': False, 'flg_proposal_system': None, 'flg_preferred_data_output
': 'R', 'flg_available': True, 'flg_beamtime_status': 'F', 'beamtime_start_at': None, 'bea
mtime_end_at': None, 'description': '', 'beamtimes': [{'id': 73, 'begin_at': '2018-09-25T0
0:00:01.000+02:00', 'end_at': '2018-12-31T23:59:58.000+01:00', 'flg_available': True, 'des
cription': None}], 'team': [{'id': 1674, 'user_id': 51, 'data_access_role': 1}, {'id': 167
5, 'user_id': 170, 'data_access_role': 1}, {'id': 1677, 'user_id': 391, 'data_access_role'
: 1}, {'id': 1678, 'user_id': 399, 'data_access_role': 1}, {'id': 1679, 'user_id': 411, 'd
ata_access_role': 1}, {'id': 1697, 'user_id': 665, 'data_access_role': 1}, {'id': 1673, 'u
ser_id': 1, 'data_access_role': 2}, {'id': 1676, 'user_id': 231, 'data_access_role': 2}, {
'id': 1959, 'user_id': 1131, 'data_access_role': 2}], 'users_ids': [[51, 'boukhele'], [170
, 'jsztuk'], [391, 'wrona'], [399, 'zaldenp'], [411, 'alqudami'], [665, 'jmalka'], [1, 'ma
ial'], [231, 'manettim'], [1131, 'hannappj']]}}
[exflqr17747:metadata_client maial$
```

```
#!/usr/bin/python3
from metadata_client.metadata_client import MetadataClient

# Necessary configuration variables to establish a connection
user_id = '1d_____ab'
user_secret = '84b_____25'
#
user_email = 'luis.maia@xfel.eu'
#
metadata_web_app_url = 'https://in.xfel.eu/metadata'
token_url = 'https://in.xfel.eu/metadata/oauth/token'
refresh_url = 'https://in.xfel.eu/metadata/oauth/token'
auth_url = 'https://in.xfel.eu/metadata/oauth/authorize'
scope = ''
base_api_url = 'https://in.xfel.eu/metadata/api/'

proposal_number = 900054

# Generate the connection
client_conn = MetadataClient(client_id=user_id,
                             client_secret=user_secret,
                             user_email=user_email,
                             token_url=token_url,
                             refresh_url=refresh_url,
                             auth_url=auth_url,
                             scope=scope,
                             base_api_url=base_api_url)

all_proposal_runs = MetadataClient.get_proposal_info(client_conn, proposal_number)
#all_proposal_runs = MetadataClient.get_proposal_runs(client_conn, proposal_number)
print(all_proposal_runs)
```

Python library: https://in.xfel.eu/gitlab/ITDM/metadata_client

European XFEL

Key Features
Statistics

	Mar. 2019	Jan. 2020
Users	1296	1886
Groups	346	662

	Mar. 2019	Jan. 2020
Instruments	15	20
Instrument Cycles	59	115
Repositories	11	12

	Mar. 2019	Jan. 2020
Porposals	144	287
Proposals with published DOI	3	4
Samples in all proposals	540	1238
Run Types in all proposals	488	1136
Users in all proposals	2833	6039
Runs in all proposals	29736	64596
Runs migrated to Maxwell	18304	43341
Runs Calibrated (via myMdC)	463	8018

Runs Assessment
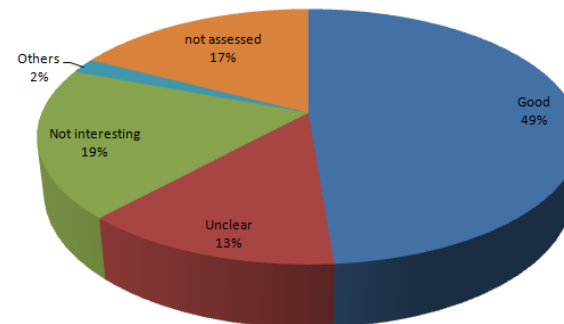
Runs Assessment (Jan 2020)

RUNS ASSESSMENT (JAN. 2020)	
Good	38.428
Unclear	4.913
Not interesting	14.120
Not assessed	6.620
Others	515
TOTAL RUNS	**64.596**
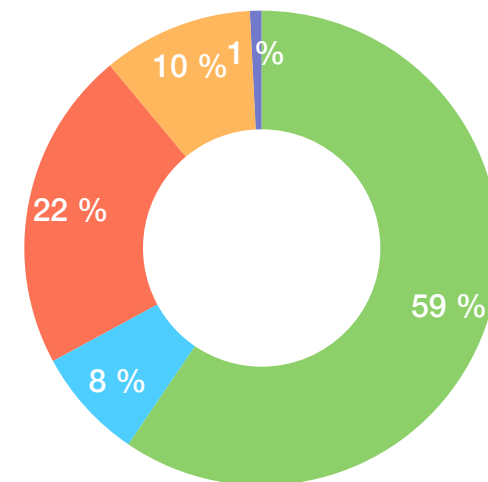
European XFEL

Agenda

Introduction

Context

Purpose

Landscape & Architecture

myMdC application

Roles and permissions

Key features

► Proposal authorisation, team administration and access to real data

► Proposal Run types and Samples

► Proposal workflow and notifications

► Run management activities (quality assessment, calibration request, describe runs)

► DatCite (DOI) management

► DAQ integration and

► Data Source Groups management

► (…)

Governance & Resources

Next steps (& EOSC)

European XFEL

Governance & Resources

Product Owner

Data Department (Operational/Support groups)

Architecture Owner

Web DevOps

Team Members

3 FTEs(<50%)

(…)

Stakeholders

Data Management experts

Detector experts

Instrumentation groups

Instruments experts

Users

Administration
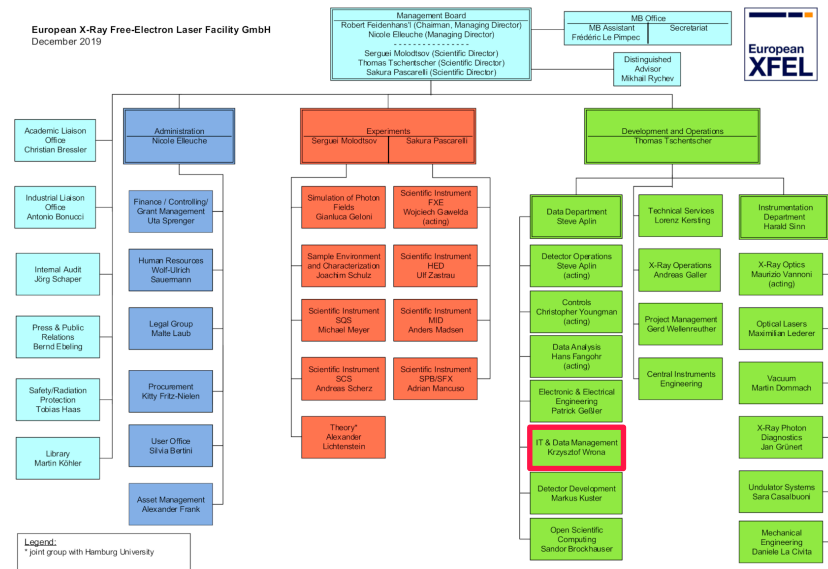
European XFEL

Agenda

■ Introduction

  ■ Context

  ■ Purpose

  ■ Landscape & Architecture

■ myMdC application

  ■ Roles and permissions

  ■ Key features

    ► Proposal authorisation, team administration and access to real data

    ► Proposal Run types and Samples

    ► Proposal workflow and notifications

    ► Run management activities (quality assessment, calibration request, describe runs)

    ► DatCite (DOI) management

    ► DAQ integration and

    ► Data Source Groups management

    ► (…)

■ Governance & Resources

■ **Next steps (& EOSC)**

European XFEL

Next Steps

Areas of improvement

- Proposals search functionalities and information usability
- Metadata Client and Metadata Tools python libraries
- Data Audit and Reconciliation
- Data Processing pipelines integration
- Data Source Groups definition and integration within Karabo Control System
- Repositories, data migration and data calibration services integration with generic "Transfer Agent" distributed task Executer
- Data reduction R&D project (Event lists / multiple data assessment stages / etc.)
- Integration with the future electronic logbook – definition of interfaces
- Data Policy Enforcement Service
- (…)

Areas of improvement - EOSC related

- Datacite membership
- Automatised DOI generation by Local Contact and Leading Scientists
- Job Scheduler
- Data migration from XFEL/DESY campus to NCBJ (via Transfer Agent)
- data.xfel.eu
- Globus "integration"
- OAI-PMH integration
- Integration with cloud initiatives within PaNOSC (e.g. OpenAIRE, B2FIND, etc.)
- (…)

European XFEL

Links

myMdC: Metadata Catalog Web Server

  https://in.xfel.eu/metadata/ (website, intro, FAQs)

  https://in.xfel.eu/gitlab/ITDM/metadata_catalog (code)

  https://in.xfel.eu/redmine/projects/metadata-catalogue (questions, suggestions, issue tracking & project planning)

  Email it-support@xfel.eu (for urgent support)

Metadata_client: Metadata Client python RESTful APIs client

  https://in.xfel.eu/gitlab/ITDM/metadata_client (code)

  https://in.xfel.eu/redmine/projects/metadata-catalogue (questions, suggestions, issue tracking & project planning)

  Email it-support@xfel.eu (for urgent support)

European XFEL

Thank you!

Special thank you to all the people that contributed to the project

European XFEL

Data Retention Policy – <u>under review</u>

Storage	Quota	Safety	Lifetime	Comment
dcache.raw	None	Tape Archive	6 months (t.b.c)	Raw data
raw	None	None	2 months (t.b.c)	Fast accessible raw data, lifetime not guaranteed
usr	5TB	Snapshots + Tape Backup	24 months (t.b.c)	user data, results
proc	None	None	6 months (t.b.c)	processed data e.g. calibrated
scratch	None	None	6 months	Temporary data (lifetime not guaranteed)
dcache.cal	None	Tape Archive	10 years	Calibration constants
cal	None	Tape archive	6 months	
user home	20GB	Snapshots + Tape Backup	Lifetime of the account	
archive.raw	None	-	Long-term	Long term means 5 years and XFEL will strive for 10 years
archive.cal	None	-	10 years	

European XFEL

myMdC – DB model