

PAUL SCHERRER INSTITUT



**WIR SCHAFFEN WISSEN - HEUTE  
FÜR MORGEN**

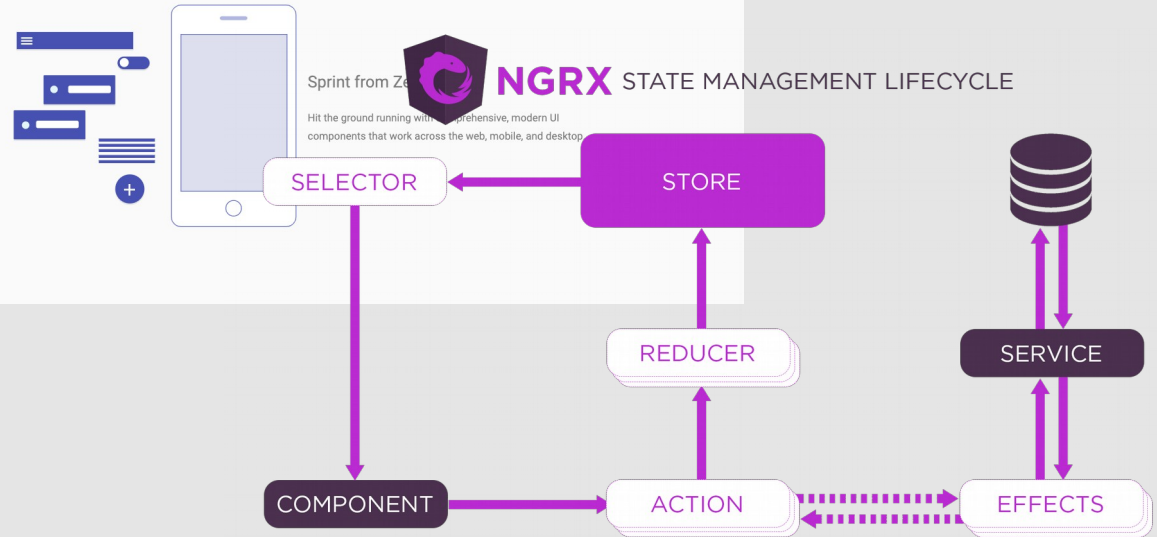
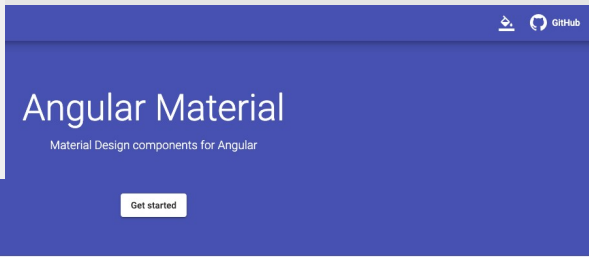
**Luke Gorman - Software Engineer (Paul Scherrer Institut)**

- Data catalog management system for scientific data
- Aims to support the management of the whole data lifecycle
- Open source (<https://github.com/scicatproject>)
- Modular design to cope with evolving requirements from scientists
- Technical architecture based on Microservices with the latest web technologies

# The Stack - Client



## ANGULAR



- Built using Angular and NgRx
- Created with Typescript
- Component based architecture to reuse throughout the application
- Based on Angular Material Widget Set
- Responsive, Standards compliant

# The Stack – Back End



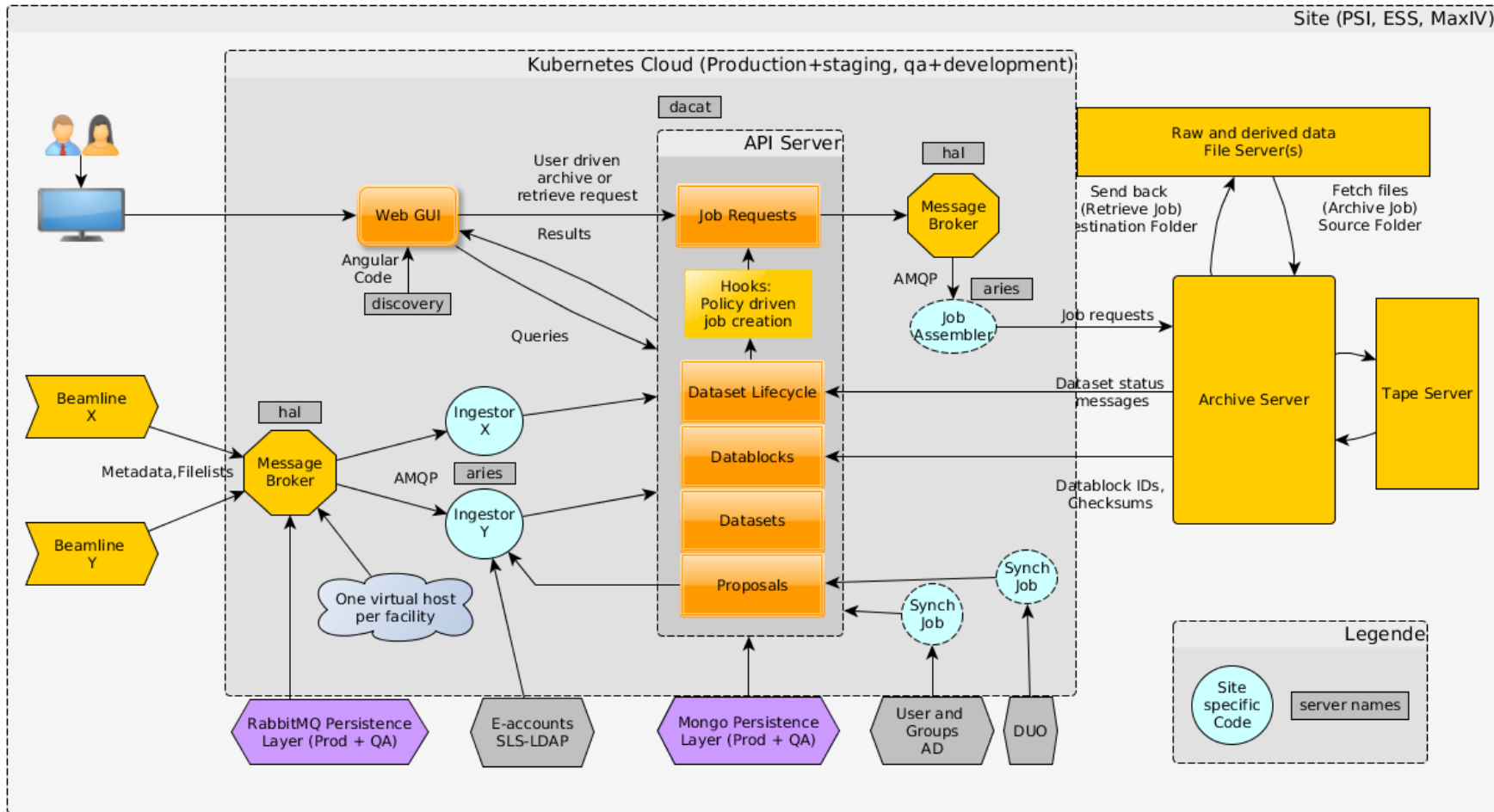
## LoopBack

### Express

Fast, unopinionated, minimalist  
web framework for Node.js



# Overall Data Catalog Architecture



# Why use it?

- Allows to add metadata to your files to give them more meaning
- Allows you to organize data and make the data findable
- Allows you to publish your data and make it citable via a digital object identifier (DOI)
- Can be used both for raw (experiment) data and derived (analysis results) data



# 1. Define Datasets

- Simply list all the file names that logically belong to a measurement (e.g. a data taking run). In the simplest case this is just the name of a folder: all files within are taken as the dataset files
- Define data **ownership** by assigning a p-group via Digital user Office DUO. Only members of the pgroup get access to the data. Membership defined by beamline manager BM or principal investigator PI
- Define minimal administrative metadata (more is possible)

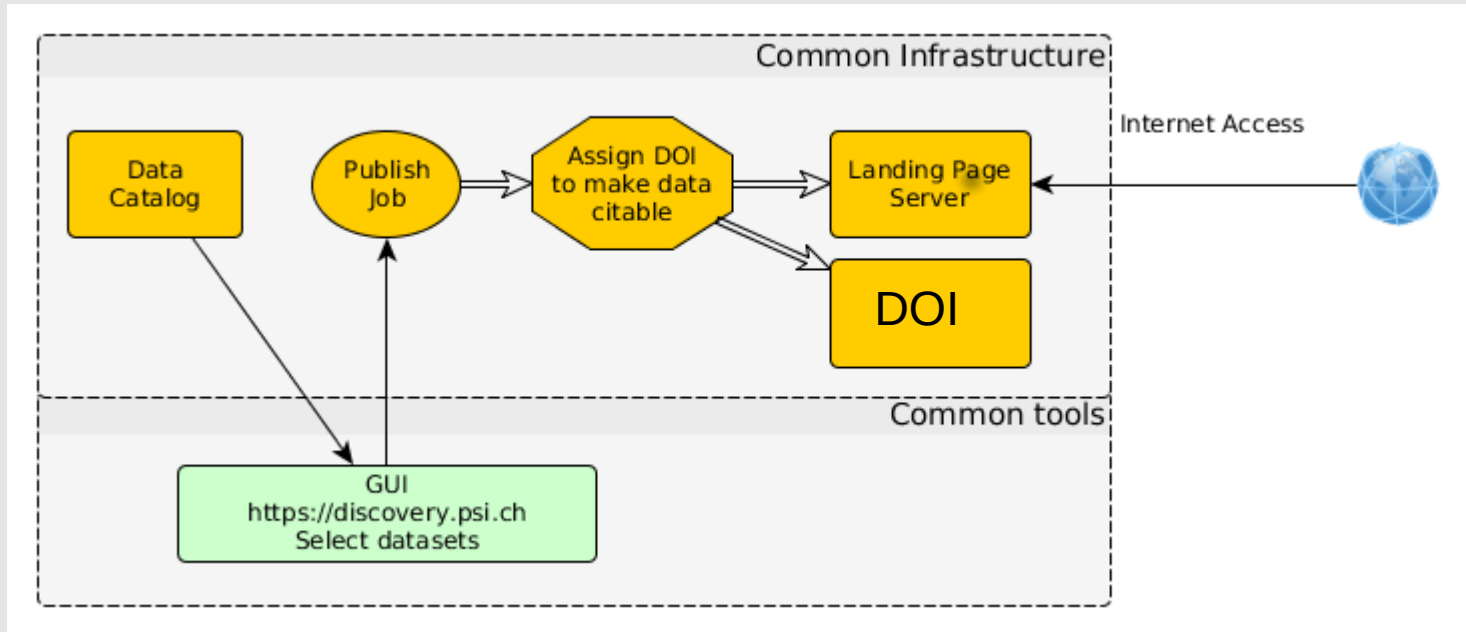
```
{  
  "principalInvestigator": "federica.marone@psi.ch",  
  "creationLocation": "/PSI/SLS/TOMCAT",  
  "sourceFolder": "/data/experiment/run/17",  
  "type": "raw",  
  "ownerGroup": "p16623"  
}
```

## 2. Define Scientific Metadata

- The definition of scientific meta data is completely up to the scientific discipline.
- Ideally follow a standard if it exists, e.g. Nexus based HDF5 files
- Scientific metadata can also be added later
- Just an example:

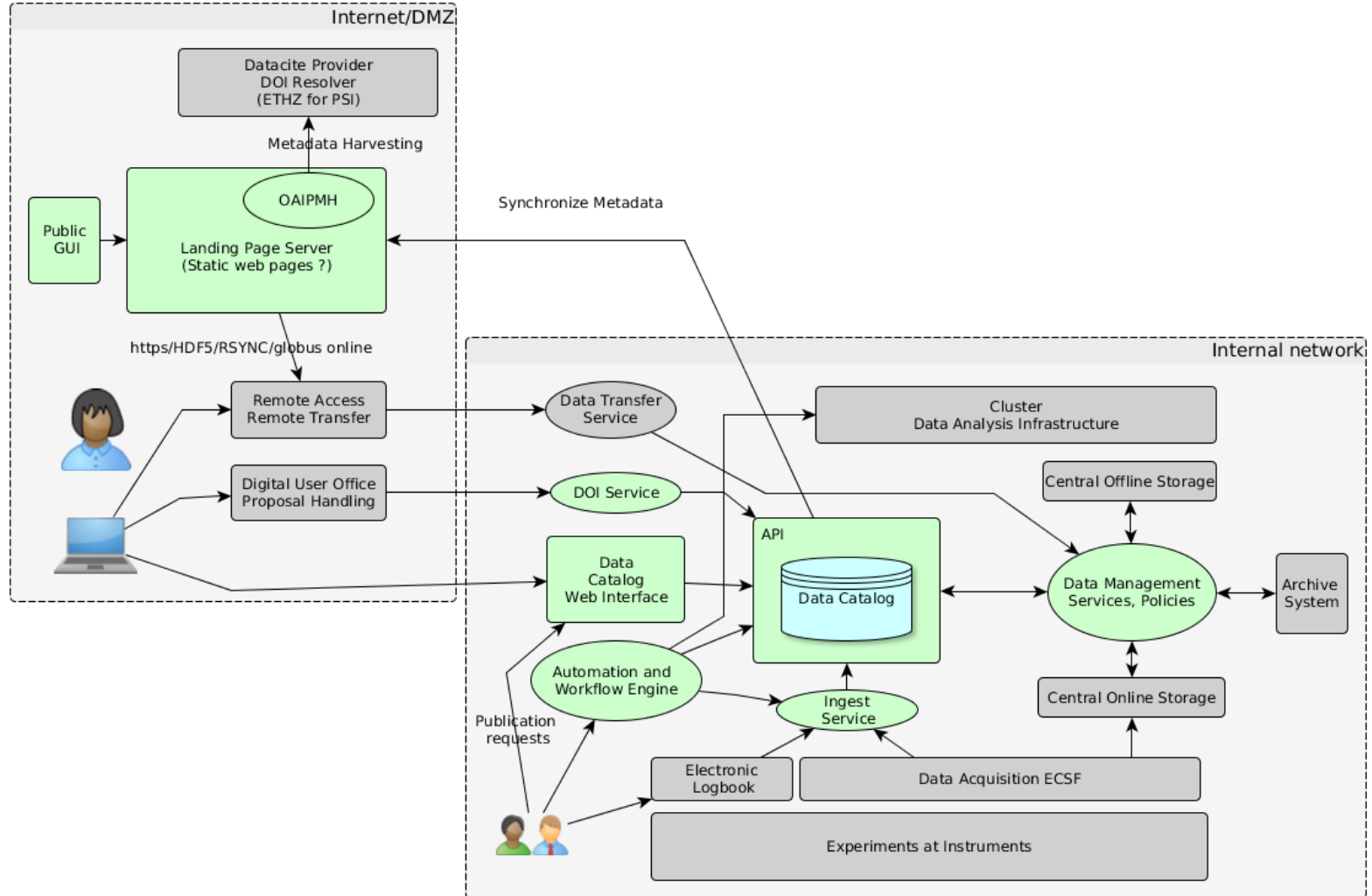
```
"beamlineParameters": {  
  "Monostripe": "Ru/C",  
  "Ring current": {  
    "v": 0.402246,  
    "u": "A"  
  },  
  "Beam energy": {  
    "v": 22595,  
    "u": "eV"  
  }  
},  
"detectorParameters": {  
  "Objective": 20,  
  "Scintillator": "LAG 20um",  
  "Exposure time": {  
    "v": 0.4,  
    "u": "s"  
  }  
}  
}...
```

# How it works: publishing data

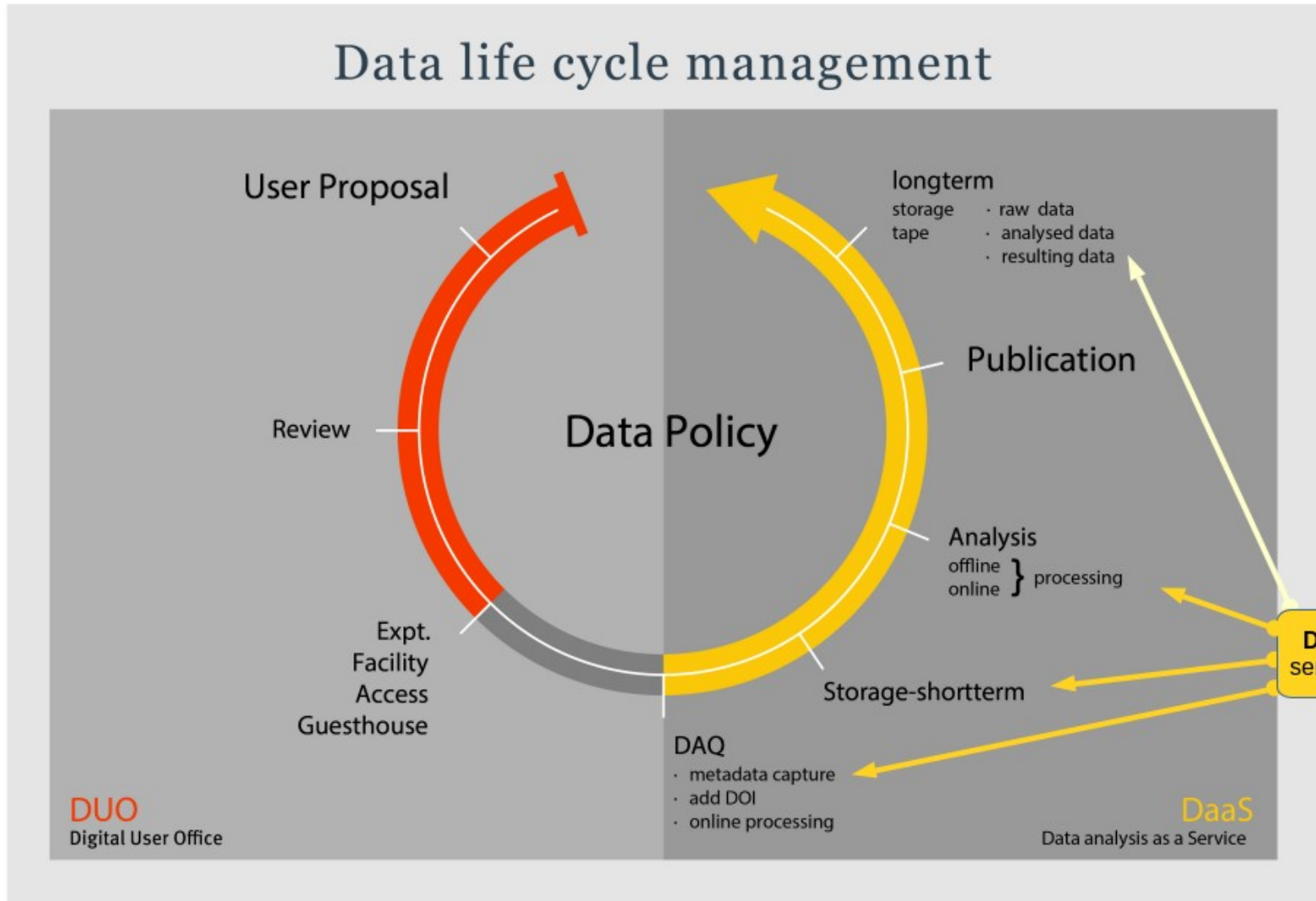


Driven by researchers who want to publish in a journal

# Architecture Update for Publication Workflow

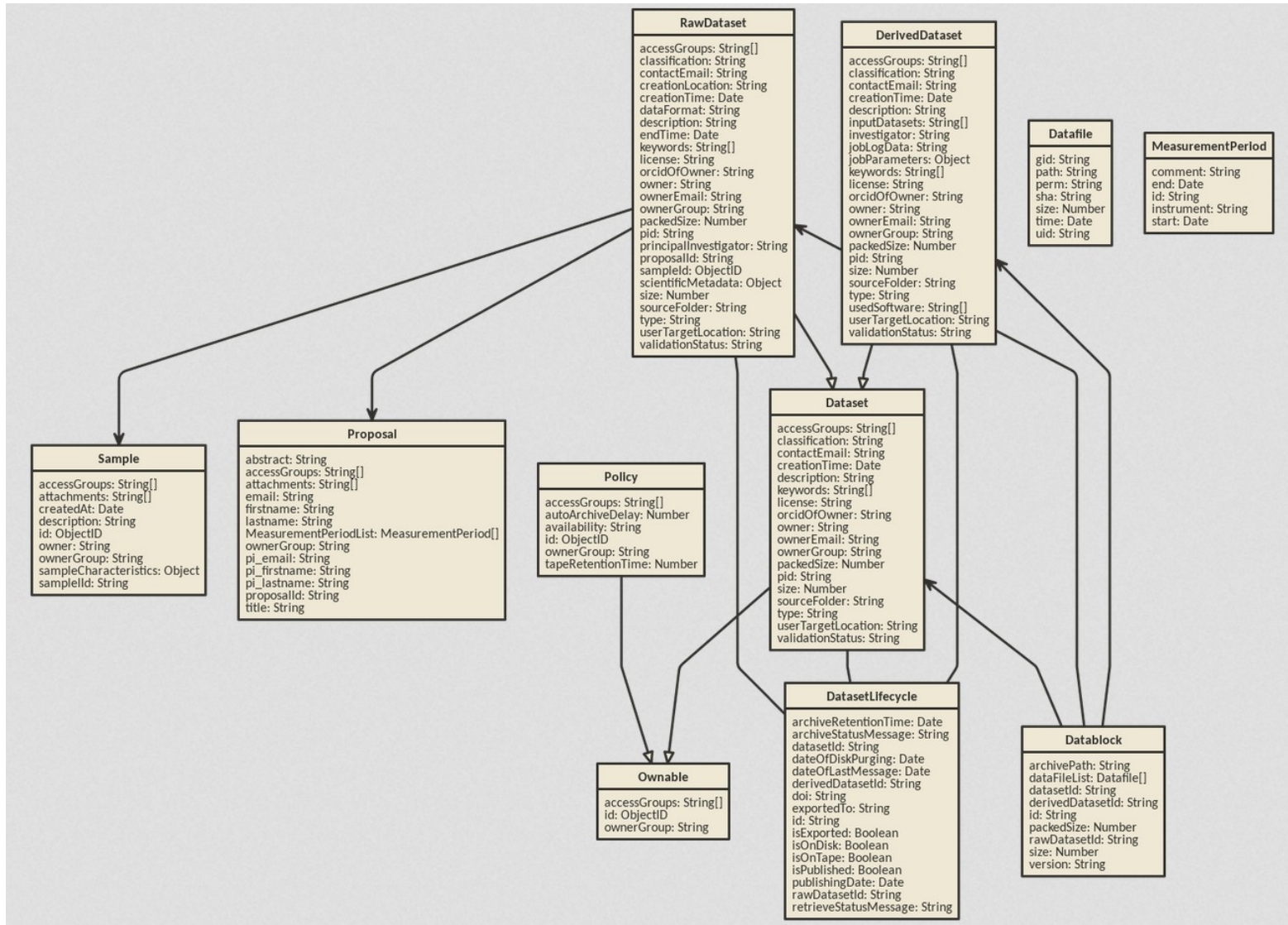


# Bigger Picture: Data Lifecycle



- MongoDB backend
- NoSQL - Document based storage, collections, "schemaless"
- Map/Reduce/Facet Aggregation queries
- Powerful indexing and support for file storage
- Fault tolerant and drivers for most languages

# Data Model: Raw/Derived Datasets, Proposals...



Jupyter Hub

About the data

	<b>Name</b>	disk2/test_grass_21keV_20us_/tif
	<b>Description</b>	test_grass_21keV_20us
	<b>Owner</b>	Susan Braovac
	<b>PID</b>	20.500.11935/8576c6e1-fe80-4807-a1ce-5f1af998d6d2
	<b>Source Folder</b>	/sls/X02DA/Data10/e17081/disk2/test_grass_21keV_20us_/tif

Structural information

	<b>Type</b>	raw
	<b>Version</b>	2.6.3
	<b>Proposal</b>	<a href="#">Xray tomographic microscopy of retreated wood from the Oseberg finds</a>
	<b>Size</b>	11 GB
	<b>Orcid</b>	unknown

Administrative information

	<b>Creation Time</b>	2018-04-26 13:44
	<b>Principal Investigator</b>	<a href="mailto:susan.braovac@khm.uio.no">susan.braovac@khm.uio.no</a>
	<b>Creation Location</b>	/PSI/SLS/TOMCAT
	<b>Owner Group</b>	p17081
	<b>Access Groups</b>	slstomcat
	<b>Contact Email</b>	<a href="mailto:susan.braovac@khm.uio.no">susan.braovac@khm.uio.no</a>

Scientific Metadata

View Edit

Name	Value	Unit
Beamlineparameters	<pre>{   "Ring current": {     "v": 401.431,     "u": "mA"   },   "energy": {     "v": 14.999,     "u": "keV"   },   "Monostripe": "R",   "Filter": "No Filter 100%",   "OP-Filter1": "1C" }</pre>	



# Screenshot of API Explorer

StrongLoop API Explorer
Token Not Set  [Set Access Token](#)

---

**Datafile** Show/Hide | List Operations | Expand Operations

---

**Dataset** Show/Hide | List Operations | Expand Operations

GET /Datasets Find all instances of the model matched by filter from the data source.

PUT /Datasets Update an existing model instance or insert a new one into the data source.

POST /Datasets Create a new instance of the model and persist it into the data source.

**Response Class (Status 200)**

Model | Model Schema

```

"string"
],
"RepositoryOfRecord": "string",
"MutabilityFlag": true,
"Version": "string",
"TypeOfScienceMetaData": "string",
"ArchiveLocation": "string",
"DisposalTime": "2016-08-24",
"Exported": true,
"ExportedTo": "string",
"Published": true,

```

Response Content Type

**Parameters**

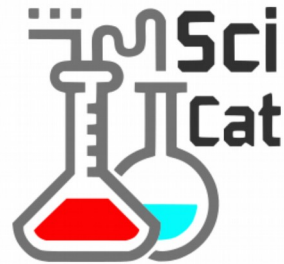
Parameter	Value	Description	Parameter Type	Data Type
data	<div style="border: 1px solid #ccc; height: 60px; width: 100%;"></div>	Model instance data	body	Model   Model Schema

Parameter content type:

```

{
  "DOI": "string",
  "InitialFolderLocation": "string",
  "Size": 0,
  "Checksum": "string",
  "CreationTime": "2016-08-24",
  "CreationLocation": "string",
  "Description": "string"
}

```



Username

ESS account username

Password

ESS account password

[SciCat Privacy Notice](#)

Log in

Remember me

Datasets

Search Clear

Text Search

Location  
/PSI/SL5/TOMCAT

Group  
p17081

Type  
raw

Keywords

Select a date range  
2/5/2018 - 2/6/2020

+ Add Condition

+ Create Dataset

- My Data
- Public Data
- All
- Archivable
- Retrievable
- Work In Progress
- System Error
- User Error

Retrieve Add to Cart

Items per page: 25 1 - 25 of 48

<input type="checkbox"/>	Name	Source Folder	Size	Start Time	Type	Image
<input type="checkbox"/>	disk2/test_grass_15keV_40us_/tif	..._40us_/tif	9 GB	2018-04-26 Thu 13:47	raw	
<input type="checkbox"/>	disk2/test_grass_21keV_20us_/tif	..._20us_/tif	11 GB	2018-04-26 Thu 13:44	raw	
<input type="checkbox"/>	disk2/test_grass_21keV_50us_/tif	..._50us_/tif	11 GB	2018-04-26 Thu 13:40	raw	
<input type="checkbox"/>	disk2/test_grass_21keV_/tif	...21keV_/tif	11 GB	2018-04-26 Thu 13:33	raw	
<input type="checkbox"/>	disk2/test_Lena_10x_21keV_4mm_6001_/tif	..._6001_/tif	34 GB	2018-04-26 Thu 13:02	raw	
<input type="checkbox"/>	disk2/test_Lena_10x_21keV_4mm_/tif	...V_4mm_/tif	23 GB	2018-04-26 Thu 12:24	raw	
<input type="checkbox"/>	disk1/057_PSI_26.5_20x_20keV_/tif	...20keV_/tif	19 GB	2018-04-26 Thu 12:04	raw	
<input type="checkbox"/>	disk1/058_PSI_26_20x_20keV_/tif	...20keV_/tif	19 GB	2018-04-26 Thu 11:52	raw	
<input type="checkbox"/>	disk1/057_PSI_26.5_20x_20keV_/tif	...20keV_/tif	19 GB	2018-04-26 Thu 11:40	raw	
<input type="checkbox"/>	disk1/055_PSI_26.5_20x_/tif	...5_20x_/tif	19 GB	2018-04-26 Thu 11:21	raw	
<input type="checkbox"/>	disk1/054_PSI_24_20x_/tif	...4_20x_/tif	19 GB	2018-04-26 Thu 10:53	raw	
<input type="checkbox"/>	disk1/052_PSI_25_20x_/tif	...5_20x_/tif	19 GB	2018-04-26 Thu 10:39	raw	
<input type="checkbox"/>	disk1/053_PSI_21_20x_/tif	...1_20x_/tif	19 GB	2018-04-26 Thu 10:27	raw	
<input type="checkbox"/>	disk1/052_PSI_23_20x_/tif	...3_20x_/tif	19 GB	2018-04-26 Thu 10:14	raw	

Datasets 20.500.11935/8576c6e1-fe80-4807-a1ce-9f1af998d6d2

- Details
- Datafiles
- Reduce
- Attachments
- Logbook
- Lifecycle
- Admin

Jupyter Hub

**About the data**

- Name** disk2/test\_grass\_21keV\_20us\_1.tif
- Description** test\_grass\_21keV\_20us
- Owner** Susan Braovac
- PID** 20.500.11935/8576c6e1-fe80-4807-a1ce-9f1af998d6d2
- Source Folder** /sis/X02DA/Data10/e17081/disk2/test\_grass\_21keV\_20us\_1.tif

**Structural information**

- Type** raw
- Version** 2.6.3
- Proposal** Xray tomographic microscopy of retreated wood from the Oseberg finds
- Size** 11 GB
- Orcid** unknown

**Administrative information**

- Creation Time** 2018-04-26 13:44
- Principal Investigator** susan.braovac@khm.uio.no
- Creation Location** /PSI/SLS/TOMCAT
- Owner Group** p17081
- Access Groups** slstomcat
- Contact Email** susan.braovac@khm.uio.no

**Scientific Metadata**

View Edit

Name	Value	Unit
Beamlineparameters	["Ring current":{"v":401.431,"u":"mA"},	
	energy{"v":14.999,"u":"keV"},Monostripe":{"R-Filter":{"No Filter":100%,"DP-Filter1":{"IC Al":{"DP-Filter Z":{"No Filter":{"DP-Filter 2":{"Filter"/>	
Detectorparameters	["Camera":{"PCO.Edge	
	5.5;"Microscope":{"Opt.Peter MB	
	op;"Objective":10;"Scintillator":{"LuAG-C	
	20 um;"Exposure time":{"v":20,"u":"ms"},"Delay time":{"v":0,"u":"ms"},"Millisecond	
Scanparameters	shutter":{"not used":{"X-ROI Start":1,"X-ROI End":2560},"ROI Start":1,"ROI --	
	End":2160,"Actual pixel size":{"v":0.65,"u":"um"},"Microscope x	
	position":{"v":160.34,"u":"mm"},"Microscope y	
	position":{"v":135.68,"u":"mm"},"Microscope z	
Scanparameters	position":{"v":93,"u":"mm"},"	
	["Sample folder":{"v":"/sis/X02DA/data/e1	
	/Data10/disk2/test_grass_21keV_20	
	/Raw Source":{"v":"/sis/X02DA/Data10	
Scanparameters	/e17081/disk2/test_grass_21keV_20	
	/tif"File	
	Path":{"v":"/test_grass_21keV_20us_1.tif	
	of projections":1001,"Number of	
Scanparameters	darks":10,"Number of flats":30,"Number	
	inter-flats":0,"Flat frequency":0,"Rot Y r	
	position":{"v":0,"u":"deg"},"Rot Y max	
	position":180,"Rotation axis	
Scanparameters	position":{"Standard":"Angular step":	
	["v":0.18,"u":"deg"},"Sample	
	In":{"v":0,"u":"um"},"Sample Out":{"v":	
	10000,"u":"um"}]	

[Show Metadata](#)



index.jpeg

### Edit Settings

1 selected groups: ▼

Auto Archive: **True** ▼

Auto Archive Delay: **1** ▼

Tape Redundancy: **high** ▼

Archive Email Notification: **True** ▼

Archive Emails to be Notified: \_\_\_\_\_

Retrieve Email Notification: **False** ▼

Retrieve Emails to be Notified: \_\_\_\_\_

	Retrieve Email Notification	Retrieve Emails To Be Notified
p100	true	bertram.astor@grumble.com
stri	true	string
p174	true	egon.meier@psi.ch
p170	false	g
a-35	true	edgar.barabas@psi.ch
nn	true	string
a-35	true	edgar.barabas@psi.ch
p163	true	filip-karol.leonarski@psi.ch
p175	true	
p175	true	
p176	true	
p2018	true	
p20181741	true	
p13388	true	bjarne.almqvist@erdw.ethz.ch
p17079	true	dfgh
p12345	true	bertram.astor@grumble.com
p34123	true	egon.meier@web.de
p20181957	true	
p20190424	true	
p20181791	true	
p20190808	true	
p20190880	true	
p20182045	true	
AATEST	true	
n17078	false	

# Management and deployment

- Different deployment options, with and without container technology are available
- Preferred method is via a Kubernetes cluster
- Each microservice is built into a docker file and saved in a registry
- Kubernetes and Helm package manager deploys the containers and handles:
  - Routing
  - Scaling
  - Server failures
  - Updates



# Kubernetes Management Dashboard

+ CREATE

☰ Workloads > Deployments

**Cluster**

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

**Namespace**

All namespaces ▾

**Workloads**

- Daemon Sets
- Deployments
- Jobs
- Pods
- Replica Sets
- Replication Controllers
- Stateful Sets

**Discovery and Load Balancing**

- Ingresses
- Services

**Config and Storage**

- Config Maps
- Persistent Volume Claims
- Secrets

## Deployments ☰

	Name	Namespace	Labels	Pods	Age	Images	
✓	<a href="#">tomcat-ingestor</a>	production	name: tomcat...	1 / 1	8 hours	nodered/node-r...	⋮
✓	<a href="#">tomcat-ingestor</a>	qa	name: tomcat...	1 / 1	8 hours	nodered/node-r...	⋮
✓	<a href="#">dacat-api</a>	production	name: dacat...	1 / 1	8 hours	registry.psi.ch:5...	⋮
✓	<a href="#">dacat-api</a>	qa	name: dacat...	1 / 1	8 hours	registry.psi.ch:5...	⋮
✓	<a href="#">melanie-product...</a>	production	app: melanie-... chart: rabbit... heritage: Tiller release: mela...	1 / 1	8 hours	bitnami/rabbitm...	⋮
✓	<a href="#">melanie-qa-rabb...</a>	qa	app: melanie-... chart: rabbit... heritage: Tiller release: mela...	1 / 1	8 hours	bitnami/rabbitm...	⋮
✓	<a href="#">kubernetes-das...</a>	kube-system	k8s-app: kube...	1 / 1	9 hours	gcr.io/google_c...	⋮
✓	<a href="#">nginx-ingress-co...</a>	kube-system	k8s-app: ngin...	1 / 1	6 days	gcr.io/google_c...	⋮
✓	<a href="#">default-http-bac...</a>	kube-system	k8s-app: defa...	1 / 1	a month	gcr.io/google_c...	⋮
✓	<a href="#">tiller-deploy</a>	kube-system	app: helm name: tiller	1 / 1	a month	gcr.io/kubernet...	⋮

1 - 10 of 12    < >

lukegorman

Explore Filter

PEOPLE

- henrikjohansson
- stephanegli
- Gareth and henrikjohansson
- Empty room

ROOMS

- scicat
- Gareth

LOW PRIORITY

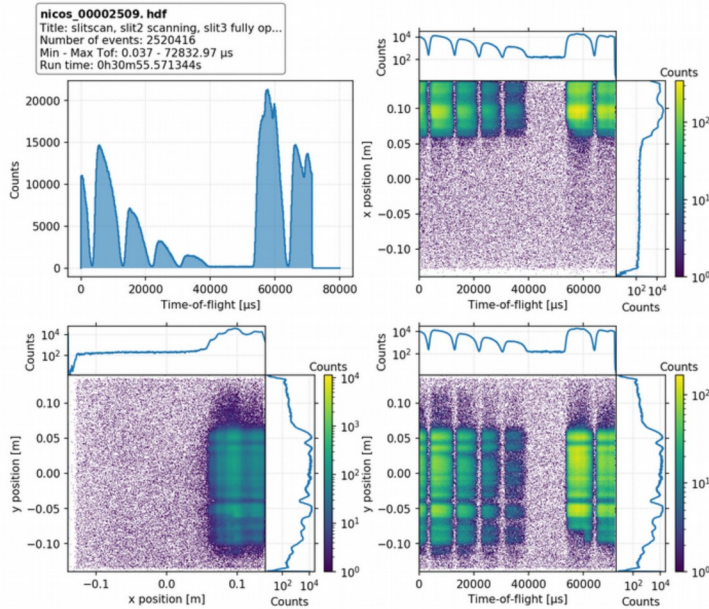
YC7SZ5

YC7SZ5 V20 <2019-11-27>

scicatbot

The file /data/kafka-to-nexus/nicos\_00002509.hdf was created. See [https://scicat.esss.se/datasets/20.500.12269%2F2509nicos\\_00002509.hdf](https://scicat.esss.se/datasets/20.500.12269%2F2509nicos_00002509.hdf) for details

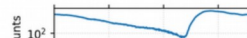
The file /data/kafka-to-nexus/nicos\_00002509.hdf was created. See [https://scicat.esss.se/datasets/20.500.12269%2F2509nicos\\_00002509.hdf](https://scicat.esss.se/datasets/20.500.12269%2F2509nicos_00002509.hdf) for details



Download plot of the data

The file /data/kafka-to-nexus/nicos\_00002510.hdf was created. See [https://scicat.esss.se/datasets/20.500.12269%2F2510nicos\\_00002510.hdf](https://scicat.esss.se/datasets/20.500.12269%2F2510nicos_00002510.hdf) for details

**nicos\_00002510.hdf**  
 Title: WFM Low Res messing with chopper 1 and 2...  
 Number of events: 1687555  
 Min - Max ToF: 0.009 - 345451.469  $\mu$ s



Send a message (unencrypted)...



Invite to this room

- tobiasrichter
- henrikjohansson
- fredrikbolmsten
- stephanegli
- michelebrambilla
- Afonso
- Matthew Jones
- @Jonas
- matt.clarke
- lukegorman
- Gareth
- anderspettersson
- scicatbot
- Oliver
- judithhouston
- andrewjackson
- danielnixon
- neilvaytet

Filter room members