

**DAMNIT** 

# The Data And Metadata iNspection Interactive Thing

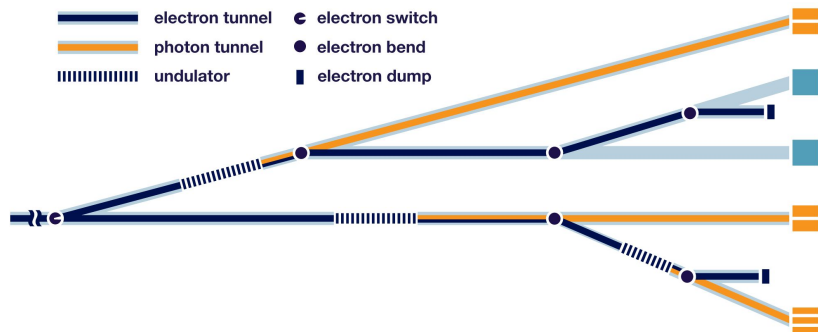


Thomas Michelat

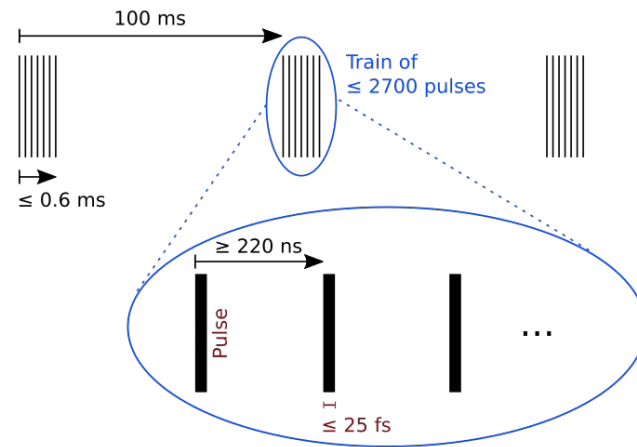
Data Analysis, European XFEL

NOBUGS 2024, Grenoble

## Facility overview

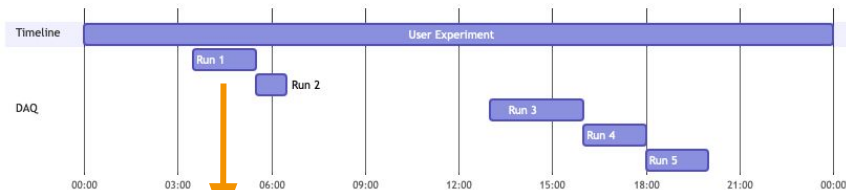
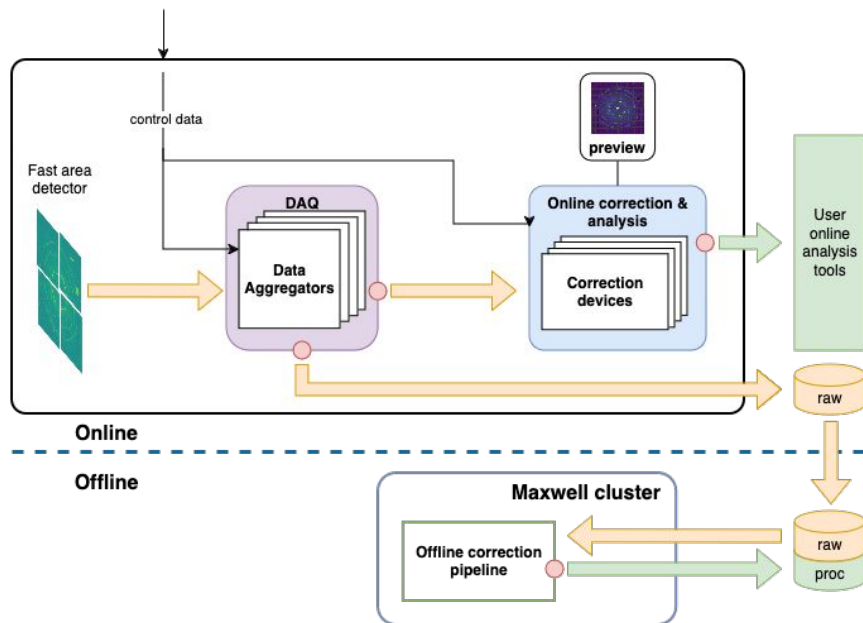


- 3 beamlines covering soft & hard X-rays
- 2-3 instruments per beamline, 7 in total
- Multiple endstations per instrument
- Pulses split across beamlines, all three operating at the same time



- Burst mode similar to FLASH
- 10 Hz trains with  $\leq 2700$  pulses each split across all beamlines
- Typically each instrument receives hundreds of pulses

# Data Workflow at European XFEL



```
r0001
[ 19G] RAW-R0001-AGIPD00-S00000.h5
[ 21G] RAW-R0001-AGIPD00-S00001.h5
[ 19G] RAW-R0001-AGIPD01-S00000.h5
[ 20G] RAW-R0001-AGIPD01-S00001.h5
[ 19G] RAW-R0001-AGIPD02-S00000.h5
[ 20G] RAW-R0001-AGIPD02-S00001.h5
[ 19G] RAW-R0001-AGIPD03-S00000.h5
[ 21G] RAW-R0001-AGIPD03-S00001.h5
[ 19G] RAW-R0001-AGIPD04-S00000.h5
[ 21G] RAW-R0001-AGIPD04-S00001.h5
[ 19G] RAW-R0001-AGIPD05-S00000.h5
[ 21G] RAW-R0001-AGIPD05-S00001.h5
[ 20G] RAW-R0001-AGIPD06-S00000.h5
[ 22G] RAW-R0001-AGIPD06-S00001.h5
[ 19G] RAW-R0001-AGIPD07-S00000.h5
[ 21G] RAW-R0001-AGIPD07-S00001.h5
[ 18M] RAW-R0001-AGIPD500K2G00-S00000.h5
[ 18M] RAW-R0001-AGIPD500K2G00-S00001.h5
[ 11M] RAW-R0001-AGIPD500K2G00-S00002.h5
[ 15M] RAW-R0001-DA01-S00000.h5
[ 15M] RAW-R0001-DA01-S00001.h5
[ 11M] RAW-R0001-DA01-S00002.h5
[ 16M] RAW-R0001-DA02-S00000.h5
[ 16M] RAW-R0001-DA02-S00001.h5
[ 5.0M] RAW-R0001-DA02-S00002.h5
[ 1.5G] RAW-R0001-DA06-S00000.h5
[ 1.5G] RAW-R0001-DA06-S00001.h5
[ 386M] RAW-R0001-DA06-S00002.h5
[ 2.4G] RAW-R0001-DA07-S00000.h5
[ 2.4G] RAW-R0001-DA07-S00001.h5
[ 496M] RAW-R0001-DA07-S00002.h5
[ 826M] RAW-R0001-DIGI01-S00000.h5
[ 826M] RAW-R0001-DIGI01-S00001.h5
```

## What's DAMNIT?

- Users create **spreadsheets** with **run data**

p2024 UM Beamtime Log .XLSX

Fichier Édition Affichage Insertion Format Données Outils Aide

Menus 100% 123 Par défaut 10

	A	B	C	D	F	G	H	I	J	K	L	M	N
	Date	Time	Sample	Photon Energy	Detector distance	Detector Gain	Pulses per Train	Injection type	Num. frames	Num. hits	Hit rate (%)	Extended Comments	
1													
2													
3	1	23-06-2024	19:04:43	Dark									
4	2	23-06-2024	19:49:42	Test DAQ	1 keV								
5	3	23-06-2024	20:00:25	Test DAQ	1 keV				1020				
6	4	23-06-2024	20:02:23	Dark			4			600			
7													
8	5	24-06-2024	8:25:18	Dark									
9	6	24-06-2024	12:16:33	Dark		85 mm	1						
10	7	24-06-2024	12:41:37	Test run	1.0 keV	85 mm	1						
11	8	24-06-2024	12:43:05	Test run	1.0 keV	85 mm	1	10	ITV1	4500			
12	9	24-06-2024	13:25:48	Test run	1.2 keV	85 mm	1	1	ITV1	3200			
13	10	24-06-2024	15:15:23	Sample A	1.2 keV	85 mm	1	1	ITV1	5750			
14	11	24-06-2024	15:25:52	Sample A	1.2 keV	85 mm	1	1	ITV1	1900			
15	12	24-06-2024	16:01:20	Sample B	1.2 keV	120 mm	1	1	ITV1	5555	230	4,14%	
16	13	24-06-2024	16:11:44	Sample B	1.2 keV	120 mm	1	1	ITV1	5800	243	4,19%	
17	14	24-06-2024	16:22:59	Sample B	1.2 keV	120 mm	1	1	ITV1	5750	253	4,40%	
18	15	24-06-2024	16:33:08	Sample B	1.2 keV	120 mm	1	1	ITV1	6150	255	4,15%	
19	16	24-06-2024	16:44:29	Sample B	1.2 keV	120 mm	1	1	ITV1	6000	298	4,97%	
20	17	24-06-2024	16:54:40	Sample B	1.2 keV	120 mm	1	1	ITV1	4800	140	2,92%	
21	18	24-06-2024	17:04:46	Sample B	1.2 keV	120 mm	1	1	ITV1	6100	212	3,48%	
22	19	24-06-2024	17:29:36	Sample B	1.2 keV	120 mm	1	1	ITV1	5900	182	3,08%	

- Original motivation

Create run tables automatically

# What's DAMNIT?

- Original motivation
  - Create run tables automatically
- Users create spreadsheets with run data
- Lots of flexibility and facility integration
- Useful during an experiment
  - steer and interpret data
- And after an experiment
  - refine parameter, post processing

p2024 UM Beamtime Log

Fichier Édition Affichage Insertion Format Données Outils Aide

Q Menu

	A	B	C	D	F	G	H	I	J	K	L	M	N
	Date	Time	Sample	Photon Energy	Detector distance	Detector Gain	Pulses per Train	Injection type	Num. frames	Num. hits	Hit rate (%)	Extended Comments	
1													
2													
3	1	23-06-2024	19:04:43	Dark									
4	2	23-06-2024	19:49:42	Test DAQ	1 keV								
5	3	23-06-2024	20:00:25	Test DAQ	1 keV				1020				
6	4	23-06-2024	20:02:23	Dark			4		600				
7													
8	5	24-06-2024	8:25:18	Dark		85 mm	1						
9	6	24-06-2024	12:16:33	Dark		85 mm	1						
10	7	24-06-2024	12:41:37	Test run	1.0 keV	85 mm	1						
11	8	24-06-2024	12:43:05	Test run	1.0 keV	85 mm	1	10	ITV1	4500			
12	9	24-06-2024	13:25:48	Test run	1.2 keV	85 mm	1	1	ITV1	3200			
13	10	24-06-2024	15:15:23	Sample A	1.2 keV	85 mm	1	1	ITV1	5750			
14	11	24-06-2024	15:25:52	Sample A	1.2 keV	85 mm	1	1	ITV1	1900			
15	12	24-06-2024	16:01:20	Sample B	1.2 keV	120 mm	1	1	ITV1	5555	230	4.14%	
16	13	24-06-2024	16:11:44	Sample B	1.2 keV	120 mm	1	1	ITV1	5800	243	4.19%	
17	14	24-06-2024	16:22:59	Sample B	1.2 keV	120 mm	1	1	ITV1	5750	253	4.40%	
18	15	24-06-2024	16:33:08	Sample B	1.2 keV	120 mm	1	1	ITV1	6150	255	4.15%	
19	16	24-06-2024	16:44:29	Sample B	1.2 keV	120 mm	1	1	ITV1	6000	298	4.97%	
20	17	24-06-2024	16:54:40	Sample B	1.2 keV	120 mm	1	1	ITV1	4800	140	2.92%	
21	18	24-06-2024	17:04:46	Sample B	1.2 keV	120 mm	1	1	ITV1	6100	212	3.48%	
22	19	24-06-2024	17:29:36	Sample B	1.2 keV	120 mm	1	1	ITV1	5900	182	3.08%	



Data And Metadata Inspector

Table

Run table	Context file
164	azimuthal integration
165	float32: (13266, 28, 700)
166	float32: (1280, 28, 700)
167	float32: (13311, 28, 700)
168	float32: (1596, 28, 700)
169	float32: (1313, 28, 700)
170	float32: (13317, 28, 700)
171	float32: (1317, 28, 700)
172	float32: (1289, 28, 700)
173	float32: (19104, 28, 700)
174	float32: (2399, 28, 700)
175	float32: (4389, 28, 700)
176	float32: (4146, 28, 700)
177	float32: (1239, 28, 700)
178	float32: (5752, 28, 700)
179	float32: (3915, 28, 700)
180	float32: (6730, 28, 700)
181	float32: (1925, 28, 700)
182	float32: (5687, 28, 700)
183	float32: (3948, 28, 700)
184	float32: (1352, 28, 700)
185	float32: (3302, 28, 700)
186	float32: (4460, 28, 700)
187	float32: (5326, 28, 700)
188	float32: (3647, 28, 700)
189	float32: (2859, 28, 700)

Micro-plot

azimuthal overview vs. Event (run 168)

Mean (log) over all trains and pulses

lpd\_lq\_ref\_diff vs. Event (run 168)

Std Normalized difference (reference averaged over all trains)

Plotting controls

Plot summary for all runs

Plot for selected runs

Autoscale

Display hover annotations

Dynamic aspect ratio

Autoscale









Display hover annotations

Dynamic aspect ratio

Event (run 168)

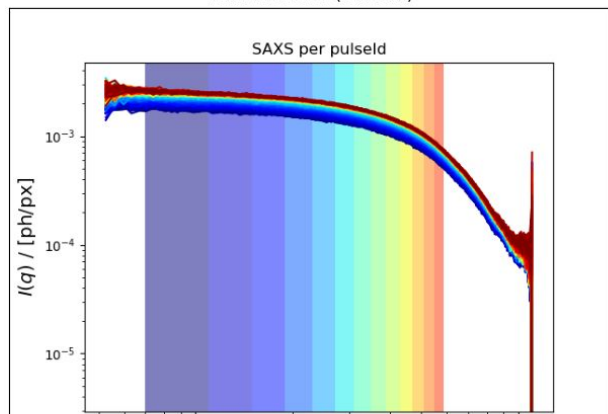
# Overview

MHz XPCS at MID (p005397 / Christian Gutt)

Run	Comment	3M intensity [t]	Run type	Velocity X	Photon energy [keV]	Sample X	fastXPCS	Scan type	get_beam_px	g2	XTD1 trans.	Sample Y	2d	Sample	Trains
375	375	<b>1376.6029</b>	XPCS	0.4000	10.0000	<b>-1.6719</b>	Done	dmesh 1.0s: HW_MID_EXP_SAM_MOTOR_SSHEX_Y (0.035 -> -0.035, 1 steps)		0.0207	<b>18.7209</b>		1.1 Ferr_10dex_10wt	2644	
376	376	<b>1370.2007</b>	XPCS	0.4000	10.0000	<b>-1.7057</b>	Done	dmesh 1.0s: HW_MID_EXP_SAM_MOTOR_SSHEX_Y (0.035 -> -0.035, 1 steps)		0.0207	<b>16.0810</b>		1.1 Ferr_10dex_10wt	2646	
377	377	<b>1373.9078</b>	XPCS	0.4000	10.0000	<b>-1.6734</b>	Done	dmesh 1.0s: HW_MID_EXP_SAM_MOTOR_SSHEX_Y (0.035 -> -0.035, 1 steps)		0.0207	<b>15.9408</b>		1.1 Ferr_10dex_10wt	2644	
<b>378</b>	<b>378</b>	<b>1368.2666</b>	XPCS	0.4000	10.0000	<b>-1.6953</b>	Done	dmesh 1.0s: HW_MID_EXP_SAM_MOTOR_SSHEX_Y (0.035 -> -0.035, 1 steps)		0.0207	<b>15.8009</b>		1.1 Ferr_10dex_10wt	2645	

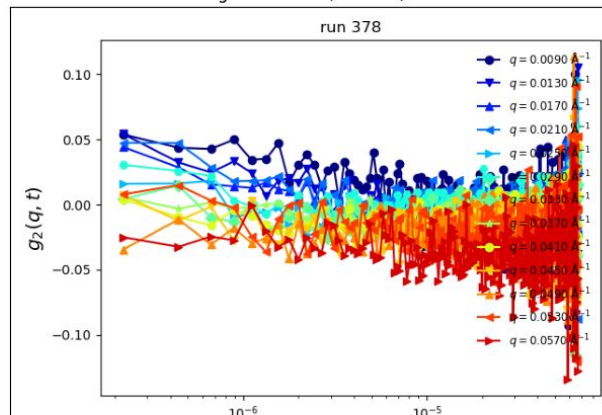
amore-proto (on max-exf461.desy.de)

saxs vs. Event (run 378)



amore-proto (on max-exf461.desy.de)

g2 vs. Event (run 378)

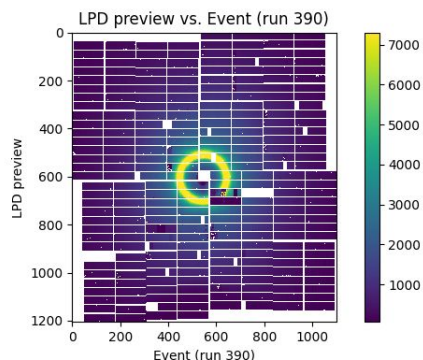


# Overview

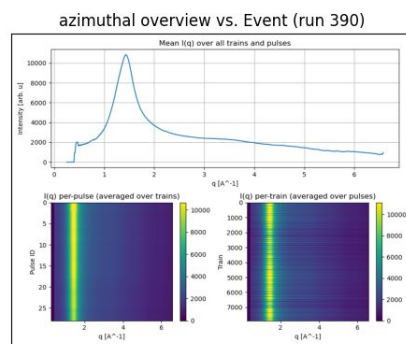
HERFD-XAS and WAXS at FXE (p004507 / Christopher Milne)

Mov Y	Mov X	APD digi (2_C)	APD digi (2_A)	LPD preview	Raw size [TB]	muthal overvi	Laser Atten	Trains	Mov Z	Laser HWP800	M intensity [L	On-Off (-refere	AM_S delay [fs w	digitizer (2_	gitizer_intensi	S(q) On-Off	Pulses
380.0000	395.6000	2.101e+04	-4.659e+04		0.3116		360.0005	3229	1267.4920	20.0005	708.8757		-1010.0620	1786.7396		28	
381.0000	395.6000	2.047e+04	-4.538e+04		0.5658		360.0005	5867	1267.4920	20.0005	666.0659		-1003.8207	1786.7178		28	
382.0000	395.6000	2.057e+04	-4.550e+04		0.2971		360.0005	3079	1267.4920	20.0005	713.9459		-996.3392	1786.6836		28	
383.0000	395.6000	1.991e+04	-4.410e+04		0.2769		360.0005	2864	1267.4920	20.0005	644.1906		-993.9528	1786.6718		28	
384.0000	395.6000	1.938e+04	-4.276e+04		0.3169		360.0005	3289	1267.4920	20.0005	705.5331		-991.0693	1786.7739		28	
385.0000	395.6000	1.962e+04	-4.324e+04		0.3416		360.0005	3549	1267.4920	20.0005	710.3880		-989.0382	1786.6315		28	
386.0000	395.6000	1.948e+04	-4.303e+04		0.3266		360.0005	3390	1267.4920	20.0005	702.3126		-987.3088	1786.7166		28	
387.0000	395.6000	1.956e+04	-4.334e+04		1.2218		360.0005	1.270e+04	1267.4920	20.0005	670.8028		-980.7382	1786.8088		28	
388.0000	395.6000	1.852e+04	-4.020e+04		0.0169		360.0005	158	1267.4920	20.0005	655.6574		-977.9391	1786.9373		28	
389.0000	395.6000	1.927e+04	-4.188e+04		0.3258		360.0005	3376	1267.4920	20.0005	677.9234		-977.9320	1786.8607		28	
390.0000	395.6000	1.870e+04	-4.049e+04		0.7597		360.0005	7892	1267.4920	20.0005	633.4589		-980.1849	1786.8159		28	
391.0000	395.6000	1.975e+04	-4.254e+04		0.1534		360.0005	1589	1267.4920	20.0005	691.5504		-982.0406	1786.8762		28	

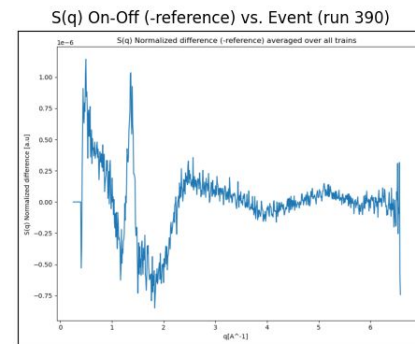
amore-proto (on max-exfl461.desy.de)



amore-proto (on max-exfl461.desy.de)

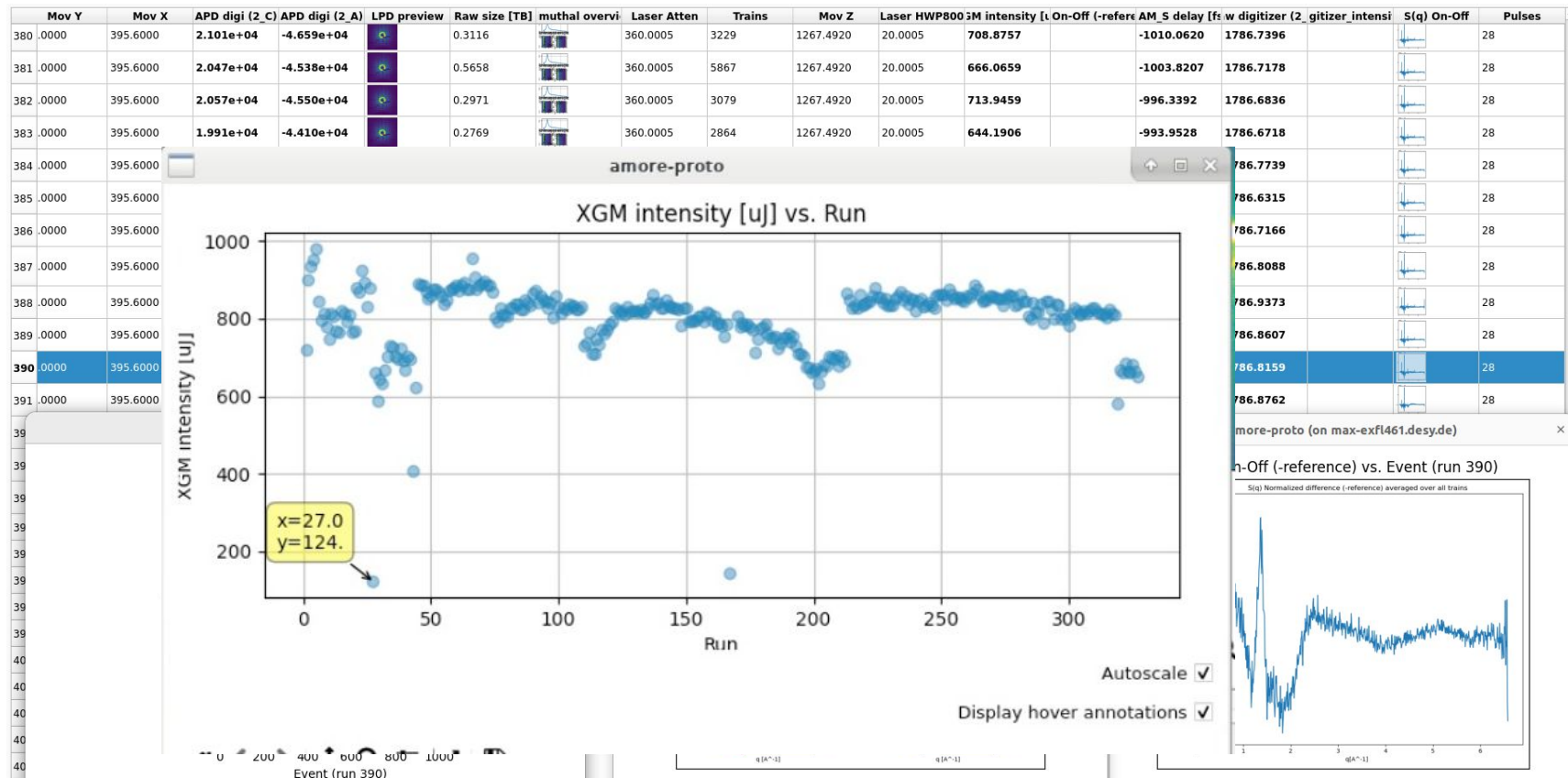


amore-proto (on max-exfl461.desy.de)



# Overview

HERFD-XAS and WAXS at FXE (p004507 / Christopher Milne)

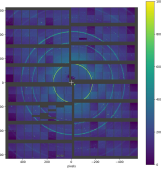




## Context file

```
@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

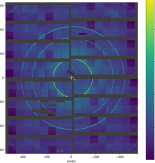
@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])
```

Run	Detector preview	Integration
7		<b>0.2655</b>

## Context file

```
@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])
```

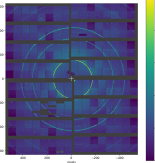
Run	Detector preview	Integration
7		<b>0.2655</b>

- A **Variable** generates a **column** in the table

## Context file

```
@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])
```

Run	Detector preview	Integration
7		<b>0.2655</b>

- A **Variable** generates a **column** in the table
- Attributes provide **customization**

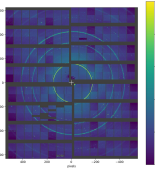
## Context file



```
RAW-R0001-AGIPD00-S00000.h5
RAW-R0001-AGIPD00-S00001.h5
RAW-R0001-AGIPD01-S00000.h5
RAW-R0001-AGIPD01-S00001.h5
RAW-R0001-AGIPD02-S00000.h5
RAW-R0001-AGIPD02-S00001.h5
```

```
@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])
```

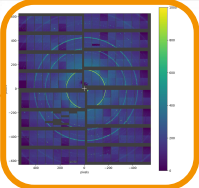
Run	Detector preview	Integration
7		<b>0.2655</b>

- A **Variable** generates a **column** in the table
- Attributes provide **customization**
- **Run data** accessible through **EXtra-data**

## Context file

```
@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])
```

Run	Detector preview	Integration
7		<b>0.2655</b>

- A **Variable** generates a **column** in the table
- Attributes provide **customization**
- **Run data** accessible through **EXtra-data**
- **Preview** returned data in the cell

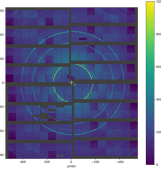
## Context file

```

@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])

```

Run	Detector preview	Integration
7		<b>0.2655</b>

- A **Variable** generates a **column** in the table
- Attributes provide **customization**
- **Run data** accessible through **EXtra-data**
- **Preview** returned data in the cell
- Interconnected function, structured as **DAG**

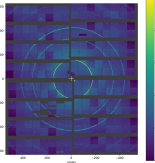
## Context file

```

@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])

```

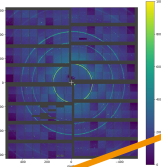
Run	Detector preview	Integration
7		<b>0.2655</b>

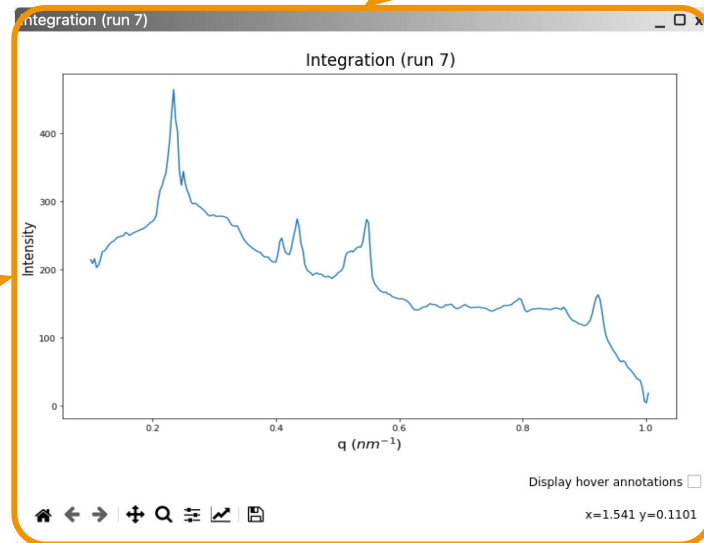
- A **Variable** generates a **column** in the table
- Attributes provide **customization**
- **Run data** accessible through **EXtra-data**
- **Preview** returned data in the cell
- Interconnected function, structured as **DAG**
- Defined **summary** data to be displayed

## Context file

```
@Variable(title="Detector preview")
def detector(run):
    agipd = AGIPD1M(run)
    image = _some_processing(agipd)
    return image

@Variable(title="Integration", summary="max")
def integration(run, image: "var#detector"):
    I, q = integrate1D(image)
    return xr.DataArray(I, dims=["q"], coords=[q])
```

Run	Detector preview	Integration
7		<b>0.2655</b>

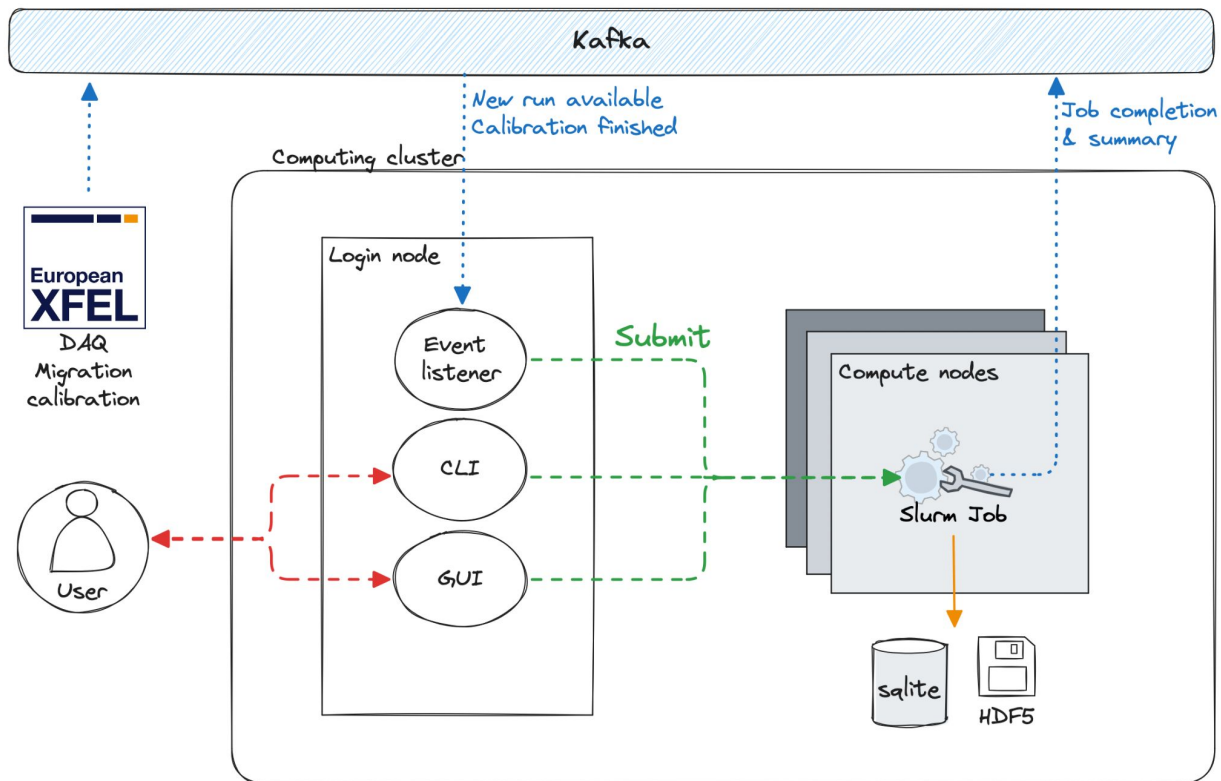




## Architecture - Processing workflow

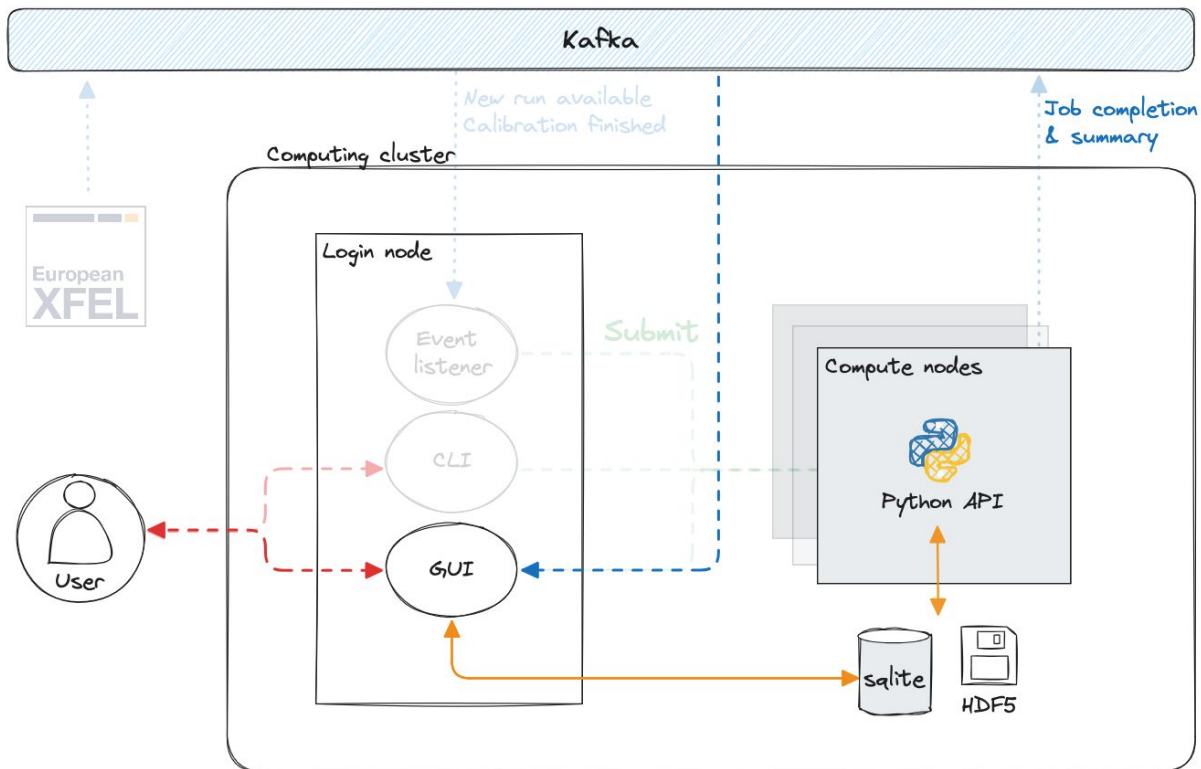
- **Automated** - Kafka
- **Distributed** - Slurm
- Variable **results** are **saved**

- Summary: **SQLite**
- Data: **HDF5/NetCDF**



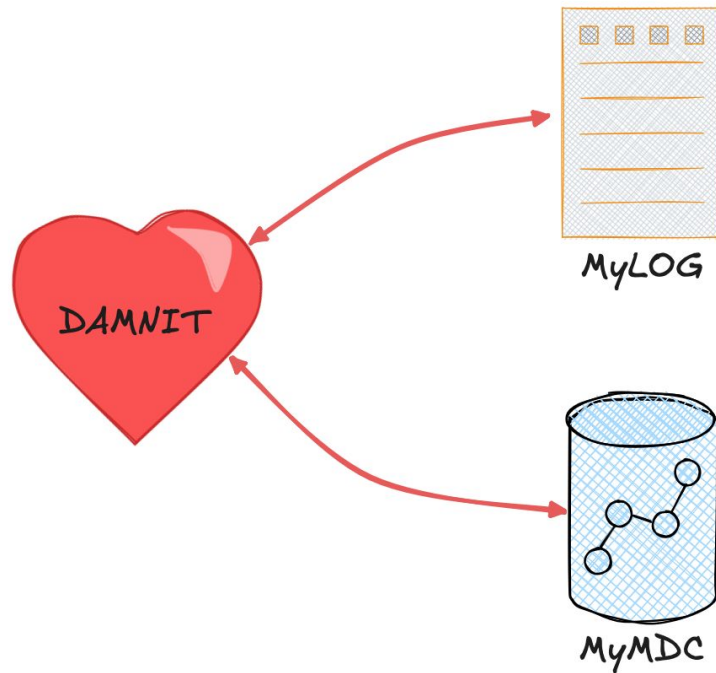
## Architecture - User interaction

- **Automated** - Kafka
- **Distributed** - Slurm
- Variable **results** are **saved**
  - Summary: **SQLite**
  - Data: **HDF5/NetCDF**
- **Qt GUI** live update - Kafka
- **Python API**



## Architecture - Services integration

- **Automated** - Kafka
- **Distributed** - Slurm
- Variable **results** are **saved**
  - Summary: **SQLite**
  - Data: **HDF5/NetCDF**
- **Qt GUI** live update - Kafka
- **Python API**
- Electronic **Logbook** publication - MyLog
- **Metadata** Catalog - MyMDC
  - Retrieve information
  - (future) export defined set of metadata



## Future - Web frontend

- **Frontend:** React, Redux, Apollo GraphQL
- **Backend:** FastAPI, Strawberry GraphQL

MID p6996 - Christian Gutt

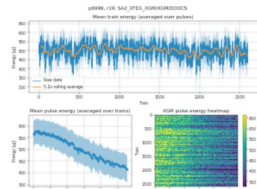
LTP - MHz XPCS enabled studies of dynamics, interactions and aggregation phenomena in protein solutions (derived from 3348)

gth	Raw size [TB]	Proc size [TB]	XGM intensity	Pulses	Run type	Sample	Scan type	XGM overview	XTD1 trans.	XTD1
	1.99	0.0507	368.3	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.113	494.2	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0973	481.9	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0978	498.6	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	2.00	0.0886	485.1	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0888	484.3	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0886	499.7	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0886	500.8	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0883	494.4	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	2.00	0.0884	481.6	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0882	459.7	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.97	0.0872	476.3	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	2.00	0.0883	498.8	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0878	483.4	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0874	493.1	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	2.00	0.0853	489.5	280	XPCS	Ferritin	dmesh 1.0s: fs		0.0169	
	1.99	0.0769	495.4	280	XPCS	Buffer_Michel	dmesh 1.0s: fs		0.0169	
	2.00	0.0769	482.3	280	XPCS	Buffer_Michel	dmesh 1.0s: fs		0.0169	

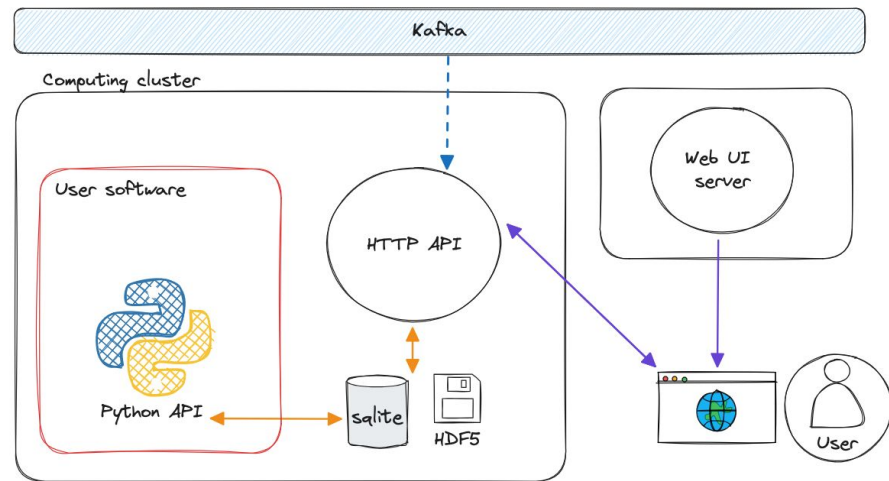
Run: 19

Timestamp 14:00:16  
 Trains 2678  
 Run length 0:04:26  
 Raw size [TB] 1.993746523866  
 Proc size [TB] 0.088570979324  
 XGM intensity [uJ] 500.76483154296875  
 Pulses 280  
 Run type XPCS  
 Sample Ferritin  
 Scan type dmesh 1.0s: fssy (0.035 -> -0.035, 1 steps), fssx (0.0 -> 52.0, 1 steps)

XGM overview

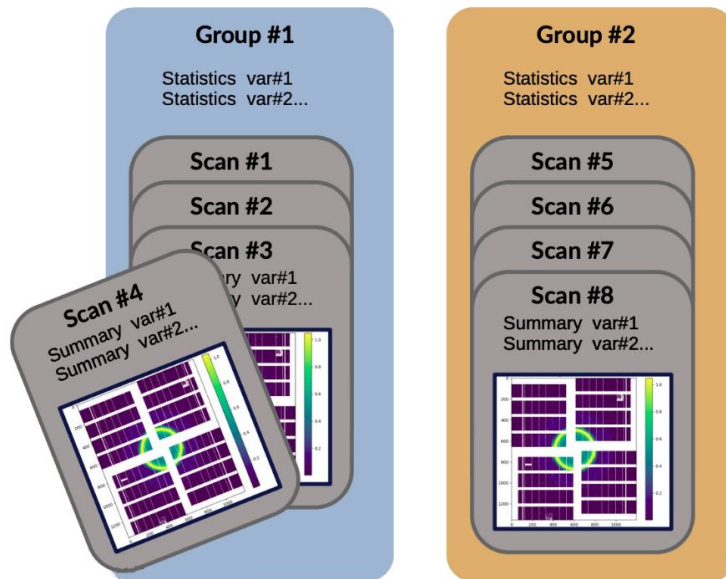


XTD1 trans. 0.016867853454485965



## Future

- **Web frontend**
  - Accessible from anywhere and any device
- **Grouping**
- **More integration**
  - Sample database
  - Metadata collection
- **Reproducibility**
  - E.g. storing code
  - History of variables



## Summary

- **DAMNIT** currently offers:
  - A **processing** framework
  - That executes **user-defined** code aiming for maximum **flexibility**
  - Built on top of our analysis software ecosystem
- A **Qt** application **frontend**
  - That shows a **run table**
  - That lets the user explore Variable's
  - Soon to be replaced by a web application
- A **Python API** to access processed data
- Usage across **all instruments** for everything from **metadata collection** to full-blown **analysis**.
- Open source: <https://github.com/European-XFEL/DAMNIT>
- Documentation: <https://damnit.rtd.io>
- contact: [da@xfel.eu](mailto:da@xfel.eu)