

STREAMLINE



WP4 Build Capacity

D4.2

Report on workflows for online data reduction and analysis for four techniques used at ESRF beamlines

Expected date 14/10/2022



STREAMLINE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870313.

PROJECT DETAILS

PROJECT ACRONYM	PROJECT TITLE
STREAMLINE GRANT AGREEMENT NO:	Sustainable research at micro and nano X-ray beamlines THEME
870313	H2020-INFRADEV-2018-2020 Development and long-term sustainability of new pan-European research infrastructures
START DATE	
15/11/2019	
DELIVERABLE DETAILS	
WORK PACKAGE ID	EXPECTED DATE
WP4	14/11/2020
WORK PACKAGE TITLE	DELIVERABLE TITLE
Build Capacity	Report on workflows for online data reduction and analysis for four techniques used at ESRF beamlines
WORK PACKAGE LEADER	DELIVERABLE DESCRIPTION
Andy Götz	Report on workflows for online data reduction and
DELIVERABLE ID	analysis for four techniques used at ESRF beamlines
D4.2	- Task 4.4
	PERSON RESPONSIBLE FOR THE DELIVERABLE
	Andy Götz

ATIRF

⊠ R- Report

□ P - Prototype

D - Demonstrator

□ O - Other

DISSEMINATION LEVEL

⊠ P - Public

□ PP- Restricted to other programme participants & EC:

□ RE – Restricted to a group

□ CO – Confidential, only for members of the consortium

REPORT DETAILS

VERSION	DATE	NUMBER OF PAGES
1.0 DELIVERABLE REPORT AUTHOR(S)	14/11/202215FOR MORE INFO PLEASE CONTACT	
Wout de Nolf, Olof Svensson, GiannisKoumotsos, Henri Payno, Andy GötzSTATUS	Andy Götz	
□ Template		Draft
🛛 Final		☑ Released to the EC

Contents

Introduction	4
ESRF Workflow System (ewoks)	4
Use case #1 – Structural biology	6
Use case #2 – Powder diffraction	10
Use case #3 – Spectroscopy	11
Use case #4 - Tomography	12
Use case #5 - Dark-field microscopy	15





Introduction

When implementing online data processing and beamline automation, the specifics of the calculations or automation steps need to be implemented case by case, but the following features are common for all use cases:

- **Reproducibility, reusability and data provenance**: the process should be fully described in a *recipe* that can be stored and transferred in different ways. This allows experts to create the recipe and anyone can execute it:
 - scientists that are not experts in X-ray technique
 - the acquisition control system
 - reviewers or readers of scientific publications
- **Collaboration**: the process described by a recipe should be *modular* so that experts in different fields can combine their knowledge that create the recipe
- **Traceability**: process execution needs to be *monitored* and the *history* of all executions needs to be recorded
- Different time and place: the same recipe can be used for online and offline data processing, executed *locally* or *remote* (e.g. computer cluster)

The **workflow** approach fits these needs where the recipe describes a workflow of tasks with data transfer between the tasks and links that can be conditional (i.e. links that are followed when predefined conditions are met). The tasks are most often written in python. For beamline automation, workflows with loops and conditional links are common.

The aim of this task was to develop four (or more) workflows within Streamline. The Ewoks workflow system, new workflows and the adaption of existing workflows (use cases #1, #4 and #5) described below were developed within Streamline.

ESRF Workflow System (ewoks)

Many *workflow management systems* exist (<u>https://s.apache.org/existing-workflow-systems</u>) each providing different features and having different strengths and weaknesses. With *ewoks* (<u>https://doi.org/10.5281/zenodo.6075054</u>) we created a meta workflow system which allows developers to *implement tasks* and scientists to *create workflows*, independent of the workflow management system. This means workflow creation is independent of the execution environment (locally, parallelized, distributed) and the way the user interacts with and visualises workflows. (command line, desktop GUI or web GUI).

Currently ewoks supports the following third-party workflow management systems

- dask: execution engine which supports parallel and distributed execution (i.e. cluster) (<u>https://www.dask.org/</u>)
- orange3: provides a desktop GUI to create and execute workflows (<u>https://orangedatamining.com/</u>)
- pypushflow: execution engine which supports cyclic workflows and conditional links (<u>https://pypushflow.readthedocs.io/en/latest/</u>)
- celery: trigger workflows remotely (job scheduling) (<u>https://docs.celeryq.dev/en/stable/</u>)

Ewoks itself provides

- a common definition to describe workflows (based on *networkx*, supporting different formats such as json and yaml)
- a common way to describe and implement tasks for ewoks workflows
- sequential execution of directed cyclic graphs (i.e. no loops or conditional links)
- web support to create and execute workflows (React frontend with a REST backend)
- start workflows on a SLURM cluster using the SLURM REST API
- event mechanism to monitor and visualise the execution of a workflow, regardless of how and where it is executed

Ewoks provides the following interfaces to create and execute workflows

- Python API: execute workflows (locally or remote), dynamic in-memory workflow creation (useful for data-parallel workflows where the side of the workflow depends on the side of the data)
- Command Line API: execute workflows locally or remote
- REST API: store workflows, execute workflows in the backend or remote, websocket for workflow execution monitoring
- Desktop GUI: create and execute workflows
- Web GUI: create, execute and monitor workflows

The web GUI was developed in order to provide access and fully exploit the ewoks infrastructure. Its main purpose is to give the user the ability to visualise and edit workflows as graphs on a canvas. The GUI can handle complicated graphs with subgraphs and conditional links. The user can easily inspect, build and parameterize ewoks graphs in a feature-rich graphical environment. A printscreen of the workflow editing interface is presented in Figure 1.

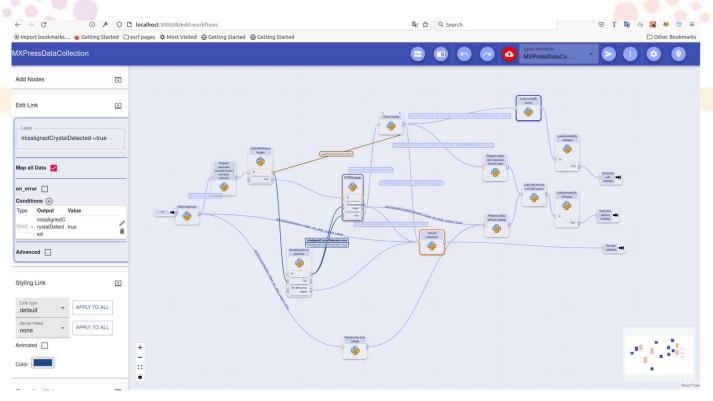


Figure 1: Ewoks workflow editing interface.

Alongside building workflows, the GUI provides the ability to launch the execution of workflows through a REST API. To achieve this an environment is under development for monitoring workflows that are being executed or having finished execution for investigating results and errors. Figure 2 presents the interface for monitoring executed workflows.

· 11 5916686.626-465.2666.939/460:493 13.4111 Thu Sep 01 2022 13.43.81 Thu Sep 01 2022 35211 Koumouts 10 · 11 6976686.626-465.2666.939/450:4936 145615 Wed Oct 05 2022 145615 Wed Oct 05 2022 5678 Koumouts 10 Execution Events progress outputs inputs was the control to the contro to the control to the control to the control to the control to t	← → C ③ ✓ O D localhost: 3000/#/mor ④ Import bookmarks ④ Getting Started □ esrf pages ♦ Most Visit		g Started	🖣 යු 🔍 Search	h	0		📓 🐸 💮 ≡) Other Bookmark:
· 11 5106686-6465-6056-569746-269 154414 Tu Sep 01 202 154.38 Tu Sep 01 202 5521 Kounout 1 · 11 670606-2502-4466-6456-5058368 145615 Wed Oct 05 202 145615 Wed Oct 05 202 1457.0 Kounout 1 Control 11 670606-2502-4466-6456-5058368 1598-10202 145615 Wed Oct 05 202 1457.0 Kounout 1 2022-10-05714-56.15.658402-0200 - <t< th=""><th>Categories • Open Workflow • Stat</th><th></th><th>mm/dd/yyyy FILTER</th><th>] =</th><th></th><th></th><th></th><th></th></t<>	Categories • Open Workflow • Stat		mm/dd/yyyy FILTER] =				
n 11 ee7dc08c-c502 448c-a59b-137454abad03 14 56:15 Wed Oct 05 202 14 57:14 Wed Oct 10 202 25 59:66 Koumouts 11 2022:10:03714:56:15.810571+02:00 11 31059442:64:06:4598-64:15 Ha97eck 14 57:20 Wed Oct 10 2022 16:57.41 Wed Oct 10 2022 16:47.41 Wed Oct 10 2022 25 396 Koumouts 11 2022:10:03714:56:15.810571+02:020 11 31059442:64:06:4598-6151Ba97eck 14:57.20 Wed Oct 10 2022 16:57.41 Wed Oct 10 2022 25 396 Koumouts 11	details	workflow_i	id ↑ job_id	Started	Ended	process_id	user_name	host_name
Time progres outputs hpus error error_message node_id task_id	□	11	591c66f8-6c3e-465c-b608-f3e74c9c34	d 13:44:14 Thu Sep	01 2022 13:44:36 Thu Sep 01 2022	35821	koumouts	Ikoumoustos
2022-10-05T1456:15.656975-02.00 start 2022-10-05T1456:15.6683402-02.00 ed true Traceback graphs cannot execute cyclic graphs 2022-10-05T1456:15.6683402-02.00 ed true Traceback graphs cannot execute cyclic graphs cannot execute cyclic graphs 2022-10-05T1456:15.810571+02.00 11 31b9442-8d0-454e-ad6b-77e40ae2d0t 14.57.20 Wed Oct 05 2022 14.57.41 Wed Oct 05 2022 54678 koumouts III 0 - 11 31b9442-8d0-454e-ad6b-77e40ae2d0t 14.57.20 Wed Oct 05 2022 16.41.49 Wed Oct 12 2022 2596 koumouts III 0 - 11 049e6338-17c6-4ad18-46f4eae80b-77e40ae2d0t 14.57.20 Wed Oct 12 2022 16.54.149 Wed Oct 12 2022 2596 koumouts III 0 - 11 049e6338-17c6-4ad18-46f4eae80b-77e40ae2d0te 16.41.20 Wed Oct 12 2022 16.54.07 Hu Oct 27 2022 16.25.07 Hu Oct 27 2022 16.25.05 Hu Oct 27 2022 16.25.05 Hu Oct 27 2022 16.25.05 Hu	□ ^	11	ee7dc08c-c502-448c-a59b-13745dab9c	6 14:56:15 Wed Oct	05 2022 14:56:15 Wed Oct 05 2022	54678	koumouts	Ikoumoustos
2022-10-05T1456:15.656975-02.00 start 2022-10-05T1456:15.6683402-02.00 ed true Traceback graphs cannot execute cyclic graphs 2022-10-05T1456:15.6683402-02.00 ed true Traceback graphs cannot execute cyclic graphs cannot execute cyclic graphs 2022-10-05T1456:15.810571+02.00 11 31b9442-8d0-454e-ad6b-77e40ae2d0t 14.57.20 Wed Oct 05 2022 14.57.41 Wed Oct 05 2022 54678 koumouts III 0 - 11 31b9442-8d0-454e-ad6b-77e40ae2d0t 14.57.20 Wed Oct 05 2022 16.41.49 Wed Oct 12 2022 2596 koumouts III 0 - 11 049e6338-17c6-4ad18-46f4eae80b-77e40ae2d0t 14.57.20 Wed Oct 12 2022 16.54.149 Wed Oct 12 2022 2596 koumouts III 0 - 11 049e6338-17c6-4ad18-46f4eae80b-77e40ae2d0te 16.41.20 Wed Oct 12 2022 16.54.07 Hu Oct 27 2022 16.25.07 Hu Oct 27 2022 16.25.05 Hu Oct 27 2022 16.25.05 Hu Oct 27 2022 16.25.05 Hu	Execution Events							
Bit Statistics 15 588369-02:20 stati 2022-10-0571456:15 588369-02:20 end Traceback rus Cannot cell is:: Cannot graphs Cannotis Cannot graphs <	Time		progress outputs inputs	type error e	error_traceback error_message	node_id task_id	task_uri	
2022-10-05T14:56:15.663402:02:00 end true Traceback (most recent call la eannot execute cyclic graphs eannot execute cyclic graphs 2022-10-05T14:56:15.663402:02:00 11 31b59442:8db0.454e-ad6b.77e40ae2d0fc 14:57:20 Wed Oct 105 2022 14:57:41 Wed Oct 05 2022 54678 kournouts 11 0 11 31b59442:8db0.454e-ad6b.77e40ae2d0fc 14:57:20 Wed Oct 105 2022 16:41:49 Wed Oct 12 2022 25396 kournouts 11 0 11 049e6338-17c64-9d1:edf1:a683205Med86 17:38:14 Tue Oct 18 2022 16:41:49 Wed Oct 12 2022 76250 kournouts 11 0 11 049e6338-17c64-9d1:edf1:a683205Med86 17:38:14 Tue Oct 18 2022 16:36:05 Thu Oct 27 2022 51271 kournouts 11 0 11 049e638-17c64-9d1:edf1:a683205Med86 16:35:39 Thu Oct 27 2022 16:36:05 Thu Oct 27 2022 51271 kournouts 11 0 11 98dace8-9e1a-4315-a816-2f8de5ab2c3 16:50:02 Thu Oct 27 2022 15:20:35 Thu Oct 27 2022 51271 kournouts 11 0 11 98dace8-9e1a-4315-a816-2f8de5ab2c3 16:50:34 Thu Oct 27 2022 51271 kournouts 11 0 11 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>								
2022-10-05T14:56:15.663402:02:00 end true frue frue <td>2022-10-05T14:56:15.588369+02:00</td> <td></td> <td></td> <td>start</td> <td></td> <td></td> <td></td> <td></td>	2022-10-05T14:56:15.588369+02:00			start				
2022-10-05T14:56:15.810571+02:00 end true true <td>2022-10-05T14:56:15.663402+02:00</td> <td></td> <td></td> <td>end true</td> <td>(most recent execute cyclic</td> <td></td> <td></td> <td></td>	2022-10-05T14:56:15.663402+02:00			end true	(most recent execute cyclic			
Image: Constraint of the constraint	2022-10-05T14:56:15.810571+02:00			end true	(most recent execute cyclic			
Image: Constraint of the constraint	□	11	31b59442-8db0-454e-ad6b-77e40ae2d	fc 14:57:20 Wed Oct	05 2022 14:57:41 Wed Oct 05 2022	54678	koumouts	Ikoumoustos
- 11 42b245ac-ddc4-4e8-9b23-1dal0fece8bb 16:35:39 Thu Oct 27 2022 16:36:00 Thu Oct 27 2022 51271 kournouts II - 11 363dace8-9e1a-43f5-a81c-2f8te5bab2c3 16:50:02 Thu Oct 27 2022 16:50:23 Thu Oct 27 2022 51271 kournouts II - 11 9648025e-6a61-4e86-bl42:7172b86ad618 16:50:34 Thu Oct 27 2022 16:50:55 Thu Oct 27 2022 51271 kournouts II - 11 18ab80b8 e17d-4f11.838c-756dc382d672 16:52:24 Thu Oct 27 2022 16:52:46 Thu Oct 27 2022 51271 kournouts II		11	bb8a81b2-24c8-4fce-bd39-8c151ba97e	c 16:41:20 Wed Oct	12 2022 16:41:49 Wed Oct 12 2022	25396	koumouts	Ikoumoustos
Image: Constraint of the constraint	□	11	049e6338-17c6-49df-8dff-a683205b49	6 17:38:14 Tue Oct	18 2022 17:38:35 Tue Oct 18 2022	76250	koumouts	lkoumoustos
Image: Constraint of the constraint	□	11	42b245ac-ddcd-4e48-9b23-1daf0fece8	b 16:35:39 Thu Oct :	27 2022 16:36:00 Thu Oct 27 2022	51271	koumouts	Ikoumoustos
□ ✓ 11 18ab80b8-e17d-4f11-838c-756dc382d672 16:52:44 Thu Oct 27 2022 51271 kournouts Ⅱ		11	363dace8-9efa-43f5-a81c-2f8fe5bab2	3 16:50:02 Thu Oct	27 2022 16:50:23 Thu Oct 27 2022	51271	koumouts	Ikoumoustos
	□ ~	11	9648025e-6a6f-4e86-bf42-7172b86ad6	8 16:50:34 Thu Oct :	27 2022 16:50:55 Thu Oct 27 2022	51271	koumouts	Ikoumoustos
	□	11	18ab80b8-e17d-4f11-838c-756dc382d6	2 16:52:24 Thu Oct :	27 2022 16:52:46 Thu Oct 27 2022	51271	koumouts	Ikoumoustos
11 36b17615-a0d0-499c-bcbc-903dd4e52811 10:42:30 Mon Nov 07 2022 10:42:30 Mon Nov 07 2022 11819 koumouts		11	36b17615-a0d0-499c-bcbc-903dd4e528	1 10:42:30 Mon Nov	07 2022 10:42:30 Mon Nov 07 2022	11819	koumouts	Ikoumoustos

Figure 2: Ewoks interface for monitoring executed workflows.

The main advantage of the web application is that it makes it a lot easier for scientists and engineers to build their own tasks and dynamically use them in their workflows.

Use case #1 – Structural biology

BES

Automation of MX (Macro-molecular crystallography) beamlines is becoming increasingly more important for synchrotron radiation facilities in general and for the ESRF in particular. The ESRF was the first facility to operate a MX beamline completely automatically in 2014 (MASSIF1, ID30A-1, <u>http://dx.doi.org/10.1107/S1600577515016604</u>). The operation of the other four MX beamlines (ID23-1, ID23-2, ID30A-3 and ID30B) is increasingly being automated, especially during nights and weekends, where automation takes over from the daytime experiment control piloted by experts.

The backbone of the MX beamline automation is a set of workflows called the Beamline Expert System (BES). The fully automatic workflows operating MASSIF 1 and other MX beamlines (<u>https://www.esrf.fr/MXPressWF</u>) are implemented using BES. These workflows not only allow fully automatic operation of the beamlines (<u>http://dx.doi.org/10.1107/S1399004715011918</u>) but also allow data mining from a large set of samples (<u>https://doi.org/10.1107/S2052252519008017</u>) thanks to the workflow database, which automatically records metadata during the experiment, and the MX beamline LIMS ISPyB. BES is also routinely used for interactive workflows including mesh scans, X-ray centring and another dozen workflows currently in production on the MX

beamlines. The MX beamlines typically execute several hundreds of fully automatic and interactive workflows every day of operation.

The BES workflows rely on the ESRF workflow system Ewoks. BES uses the pypushflow workflow engine through the ewoksppf API. Thanks to Ewoks, the workflows on the MX beamlines are now implemented in a similar way to other non-MX beamline workflows which has significantly improved the support and maintenance situation. Figure 3 presents an example MX workflow.

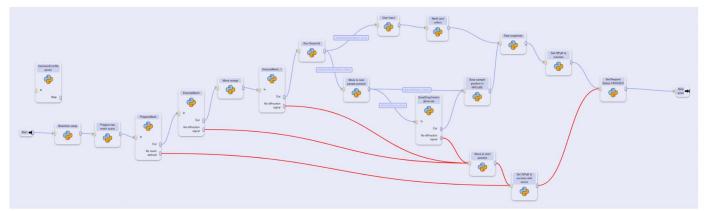


Figure 3: New workflow "Two Mesh Scans" developed using the Ewoks web interface.

Workflow(s)	MX
Beamlines	ID23-1, ID23-2, ID30A3, ID30-B
Users	Beamline scientists + users
When	Automatically or interactively
Input	2D diffraction images (CBF or H5)
Output	Scaled and merged intensities
Trigger	mxcube

Table 1: Summary of MX workflow(s).

CryoEM workflows

Single particle analysis (SPA) in cryo electron microscopy has become a routine structural technique to achieve sub 3 Å resolution. In spite of all the advances in this field, the data collection and image processing steps still suffer from low throughput. If cryo-EM SPA is to become a true complementary technique to X-ray crystallography, the following gaps need to be bridged.

1. Faster speed in data collection: Ongoing developments in hardware and software are resulting in more stable microscopes (less time spent in alignment procedures and equal results from lesser amount of data), better image recording devices (faster and higher DQE) and better data collection software. With these advances, a single SPA data set can be collected in less than 24 hours (currently implemented in only a few facilities worldwide).

2. On-the-fly feedback on the quality of the collected data: The faster the data is collected the more imperative it is to keep up with the image pre-processing which provides indications of the quality of the data that is being collected.

3. Final output in a desirable time frame: The ultimate goal is to achieve sub 3 Å resolution 3D structure of the macromolecule, but currently this is a very time-consuming process due to the overwhelming need for GPU and CPU resources and substantial expertise in cryo-EM image processing. A cryo-EM facility has to wait between 3 months to 3 years to witness the results of their users.

Since its commissioning in November 2017 as an international user facility for cryo-EM SPA, the ESRF cryo-electron microscope CM01 has been providing user service for SPA experiments with a routine data collection taking over 48 hours. The ESRF currently lacks the investment to address point 1 above. We have addressed point 2 by providing automatic pre-processing of data as well as automatic uploading of meta-data to the ISPyB LIMS and automatic archiving of meta-data and data via the ESRF data policy. The automatic pre-processing, upload of meta-data and archiving is implemented using the Scipion framework. Thanks to this Scipion workflow, users have been able to follow the experiment in real time using EXI/ISPyB and have at the end of the experiment access to motion-corrected files that are the starting point for further data processing (Figure 4).

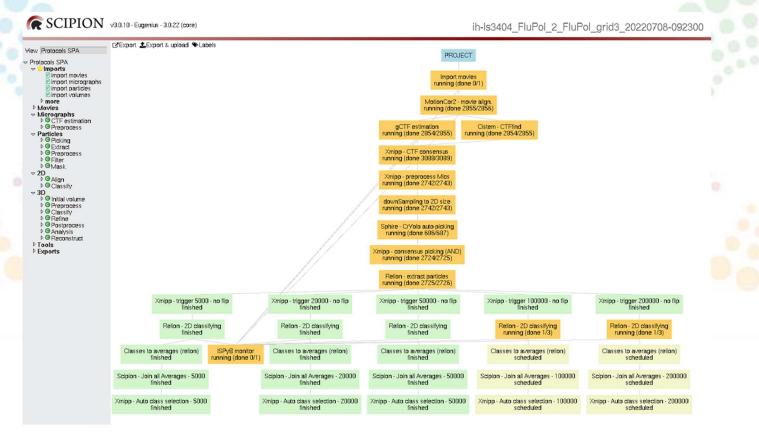


Figure 4: Scipion workflow for automatic 2D classification.

There are two main goals of this workflow:

• To help the CM01 scientists track each data collection collected and make sure the productivity is kept optimal. Although the pre-processing of the images gives an indication of the quality of the data collected, the real potential of the data to go to high resolution is indicated by the class averages. Moreover, with the speed of technological advances in the field, it is fair to anticipate faster acquisition modes being implemented for cryo-EM in the near future, in which case, the ESRF would be one step ahead of several facilities by having a pipeline to assess the

sample quality and choose whether the experiment should be continued or abandoned in favour of the next sample. This would make optimal use of beam time at the CryoEM instrument.

• To test the pipeline for 3D refinement, which is the ultimate aim. A data collection shown by the 2D workflow to give good results for initial class averages with a small data set would be continued to accumulate sufficient data to achieve optimal 2D class averages, which helps to prepare the input for 3D refinement steps.

Further developments of this workflow will include generation of the initial 3D model and 3D classification and auto refine. If this full pipeline could be implemented, either completely automated or with minimum user input, then a notable application would be for drug screening experiments. This application would be in addition to all the above listed advantages of use to academic users.

A new GUI for easy start and control of the automatic processing has been developed. Through it the user can launch on demand new tasks for the celery workers to execute as well as inspect the execution process. It provides dynamic forms with all needed parameters depending on the selected task. It was developed within the Daiquiri platform (<u>https://ui.gitlab-pages.esrf.fr/daiquiri/</u>) which is a web based framework for beamline control and data acquisition. A screenshot of the Cryocube interface as part of Daiquiri is given in Figure 5.

					<i>6</i>
Workers					+ New T
Host	Uptime 1 J	Status ↑↓	#Tasks	† ↓	
celery.opcm01@dgx01	238 hr 39 min	Running	1		
celery.svensson@cmproc3	335 hr 26 min	Ready	0		
	_				
5 -	1				
Tasks					
Tasks Id	Task	Recieved	Started	Finished	Status
	Task esrf.workflow.cm_process_worker.run_workflow	Recieved 13-10-2022 18:23:09	Started 13-10-2022 18:23:09	Finished 13-10-2022 18:23:12	
ld					Status PENDING PENDING
ld 98a83060-0f14-4a21-a074-7b5a2599a1d7	esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09	13-10-2022 18:23:09	13-10-2022 18:23:12	PENDIN
ld 98a83060-0f14-4a21-a074-7b5a2599ald7 babcd6e9-4ff1-47a3-94d7-7e551205e5ea	esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09 13-10-2022 18:20:13	13-10-2022 18:23:09 13-10-2022 18:20:13	13-10-2022 18:23:12 13-10-2022 18:20:16	PENDING
ld 98a83060-0fl4-4a21-a074-7b5a2599ald7 babcd8e9-4ff-47a3-94d7-7e551205e5ea be5069e7-4bba-466b-9fa8-ad74a37e1988	esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01	13-10-2022 18:23:12 13-10-2022 18:20:16 13-10-2022 18:19:04	PENDING PENDING PENDING PENDING
Id 98a83060-0fl4-4a21-a074-7b5a2599a1d7 babcd6e9-4ff-47a3-94d7-7e551205e5ea be5069e7-4bba-466b-9fa8-ad74c37e1988 b2b2d73b-c6d1-480c-b64d-24a9de4f878d	esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01 13-10-2022 18:17:37	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01 13-10-2022 18:17:37	13-10-2022 18:23:12 13-10-2022 18:20:16 13-10-2022 18:19:04 13-10-2022 18:17:40	PENDIN PENDIN PENDIN PENDIN PENDIN
Id 98a83060-0f14-4a21-a074-7b5a2599a1d7 babcd6e9-4f1-47a3-94d7-7e551205e5ea be5069e7-4bba-466b-9fa8-ad74a37e1988 b2b2d73b-c6d1-480c-b64d-24a9de4f878d 8583c679-ac4f-4edI-b2e3-2afa070a7cc2	esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01 13-10-2022 18:17:37 13-10-2022 18:17:37 13-10-2022 18:17:34	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01 13-10-2022 18:17:37 13-10-2022 18:16:34	13-10-2022 18:23:12 13-10-2022 18:20:16 13-10-2022 18:19:04 13-10-2022 18:17:40 13-10-2022 18:16:38	PENDING PENDING PENDING PENDING PENDING
Id 98a83060-0fl4-4a21-a074-7b5a2599ald7 babcd6e9-4ff1-47a3-94d7-7e551205e5ea be5069e7-4bbc-466b-9fa8-ad74a37e1988 be502d73b-c6d1-480c-b64d-24a9de4f878d 8583c679-ac4f-4ed1-b2e3-2afa070a7cc2 c43b8cca-83a4-4674-99b6-bf688d95c64d	esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:001 13-10-2022 18:17:37 13-10-2022 18:16:34 13-10-2022 18:15:30	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01 13-10-2022 18:19:73 13-10-2022 18:16:34 13-10-2022 18:16:34	13-10-2022 18:23:12 13-10-2022 18:20:16 13-10-2022 18:19:04 13-10-2022 18:17:40 13-10-2022 18:16:38 13-10-2022 18:16:38	PENDING
Id 98a83060-0fl4-4a21-a074-7b5a2599ald7 babcd6e9-4ff-47a3-94d7-7e551205e5ea be5069e7-4bbc-466b-9fa8-ad74a37e1988 b2b2d73b-c6dl-480c-b64d-24a9de4f878d 8583c679-ac4f-4edl-b2e3-2afa070a7cc2 c43b8cca-83a4-4674-99b6-bf688495c64d df553abe-50fd-48b4-a0e7-3a9dac8f3c5f	esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow esrf.workflow.cm_process_worker.run_workflow	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:001 13-10-2022 18:17:37 13-10-2022 18:16:34 13-10-2022 18:16:30 13-10-2022 18:15:30 13-10-2022 18:14:12	13-10-2022 18:23:09 13-10-2022 18:20:13 13-10-2022 18:19:01 13-10-2022 18:19:01 13-10-2022 18:17:37 13-10-2022 18:16:34 13-10-2022 18:15:30 13-10-2022 18:15:30	13-10-2022 18:23:12 13-10-2022 18:20:16 13-10-2022 18:19:04 13-10-2022 18:19:04 13-10-2022 18:17:40 13-10-2022 18:16:38 13-10-2022 18:15:33 13-10-2022 18:15:33	PENDING PENDING PENDING PENDING PENDING PENDING

Figure 5: CryoCube UI screenshot.

Workflow	cryoem
Beamlines	СМ01
Users	Beamline scientists + users
When	Manually after each acquisition
Input	Movies (TIFF)
Output	Motion corrected micrographs, CTF, picked particles, 2D classification
Trigger	daiquiri

Table 2: Summary of cryoem workflow.

Use case #2 – Powder diffraction

Powder diffraction experiments often require data reduction to obtain powder diffractograms that can be further analysed to obtain structural and quantitative information about the sample. Currently three such workflows are running

- ID31: a data reduction workflow is triggered automatically for scans that include a diffraction camera. The workflow includes normalisation and azimuthal integration with *pyfai* (<u>https://doi.org/10.5281/zenodo.5946928</u>).
- ID22: data conversion (HDF5 to ASCII) and data reduction workflows are triggered actively by the acquisition macros. The data reduction workflow includes rebinning with *multianalyzer* (<u>https://zenodo.org/record/7223895</u>) of data taken with multiple analyzer crystals and a large area detector.
- High-throughput powder diffraction (in use at ID31): a mobile system has been developed to fully automate the measurement of thousands of powders in a matter of hours. The acquisition includes the triggering of a data reduction workflow. The workflow includes calibration, normalisation and azimuthal integration with *pyfai* (<u>https://doi.org/10.5281/zenodo.5946928</u>).

File Edit View Widget Options Help	dows* – 🗆 🧶
CalibrateSingle CalibrateSingle Multitude CalibrateSingle Multitude CalibrateSingle Multitude CalibrateSingle Multitude CalibrateSingle Multitude Multitude CalibrateSingle Multitude Multitude CalibrateSingle Multitud	sample
Select a widget to show its data ption. See workflow sources, YouTube tornels on pen the welcome screet. It # T I I 2	Colbuschlage - 0 0 Togor Scould Retern (Re: score) Retern (Re:
The workflow is triggered by the acquisition control system for every sample	Bitarec (nin) 1000000000000 10000 Nun (pund) 10000 10000

Figure 6: High-throughput powder diffraction workflow.

Workflow	powder
Beamlines	ID22, ID31
Users	Beamline scientists + users
When	Automatically for each sample
Input	2D image (hdf5)
Output	1D spectrum (hdf5)
Trigger	bliss

Table 3: Summary of powder workflow.

Use case #3 – Spectroscopy

In the field of X-ray spectroscopy (e.g. EXAFS and XANES) one acquisition can take up to one hour. To judge the quality of the data being acquired, pre-processing is required. Considering the total time of an acquisition, this should be done during the measurement so that scientists can make decisions without losing valuable beamtime. The workflow (Figure 7) we developed includes normalisation, post-edge background subtraction and XAFS Fourier Transform with *xraylarch* (https://doi.org/10.5281/zenodo.6796308). It is currently operational at BM23.

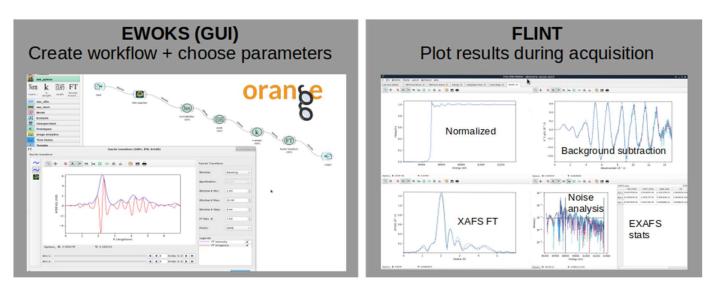


Figure 7. Ewoks workflow of EXAFS plotting during acquisition.

Workflow	exafs_plot
Beamlines	BM23
Users	Beamline scientists + users
When	Automatically after each scan
Input	EXAFS scan data (memory)
Output	fft of EXAFS spectrum (flint)
Trigger	bliss

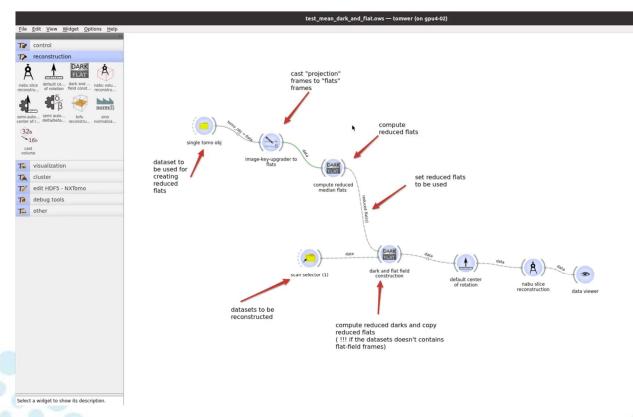
Table 4: Summary of exafs_plot workflow.

Use case #4 - Tomography

Tomwer is a library and a set of applications allowing users to treat tomographic data using a workflow. A 'canvas' application is used to create and edit the workflow and settings for each of the steps from a graphical user interface. The graphical editor is based on Orange.

New features based on the 'ewoks-orange' library have been added to ensure backward compatibility with EDF format, NXtomo keys upgrade and the projections normalisation.

Some processing ('darkref', 'nabu' processes, 'center of rotation research') have been upgraded to improve usage (Figure 8).





An important work has been done on improving handling volumes across the tomotools suite and end it with simplification at the tomwer side and addition of volume related widgets using the ewoks libraries, such as volume cast or volume selection.

Finally job submission to slurm has also been reworked to improve robustness and better distribute the workload to the compute cluster (Figure 9).

test, slurm aws (on p9-04)	_ # 😣
Elle Edit View Widget Options Help	
Te ² control	
Image: Arrow and a state of the s	
And make And mak And make And make And ma	
T- visualization	
sinogr sica radio Image: Sicos sinogr sica radio data viewer sixade sixade viewer volume simple Image: Sicos Image: Sicos volume viewer viewer Image: Sicos Image: Sicos	
Tæslurm	
sturn sturn cluster super	08
dataset	clear
Key flat	clear
ange dark-fl If flat field com-* it center of roi slice reconstruites and the flat field com-* it center of roi slice reconstruites flat field com-* it center of roi slice reconstruites flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat com-* it center of roi slice reconstruites flat flat flat com-* it center of roi slice reconstruites flat flat flat com-* it center of roi slice reconstruites flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat flat flat com-* it center of roi slice reconstruites flat flat flat flat flat flat flat flat	
i # T ` \] II 0	

Figure 9: tower: workflow with some nabu job submitted to slurm.

Tomwer and other tomotools have been recently deployed on BM18 in addition to being in production on the ID11, ID16b, ID17, ID19, BM18 and BM05 beamlines.

A series of tutorials has been made to train users how to use tomwer. They are available here: <u>http://www.edna-site.org/pub/doc/tomwer/video/canvas/</u>. In the future, they could also be shared on the pan-training.org site.

Workflow	tomwer
Beamlines	BM05, ID11, ID16B, ID17, BM18, ID19
Users	Beamline scientists + users
When	Daily
Input	raw 2d images (hdf5 or EDF)
Output	sinogram 2d image (hdf5) reconstructed 3d volume (tiff)
Trigger	bliss / user

Table 5: Summary of tomwer workflow.

Use case #5 - Dark-field microscopy

darfix (<u>https://doi.org/10.48550/arXiv.2205.05494</u>) is a Python library for the analysis of dark-field X-ray microscopy data. It provides ewoks tasks and widgets for the ewoks orange3 desktop GUI (Figure 10).

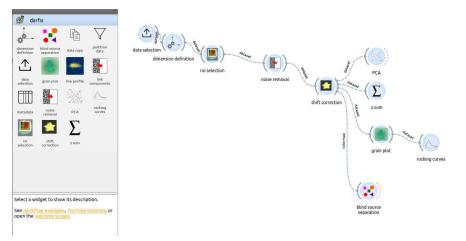


Figure 10: darfix, the workflow for DFXM.

Workflow	darfix
Beamlines	ID06
Users	Beamline scientists + users
When	Daily
Input	raw 2d images (hdf5)
Output	3d grain volumes
Trigger	user

Table 6: Summary of darfix workflow.