AN INTERACTIVE WORKFLOW TO MANAGE TOMOGRAPHY EXPERIMENTS AT ESRF

H. Payno, C. Nemoz, European Synchrotron Radiation Facility, Grenoble, F-38043, France

Abstract

At the ESRF the activity of several beamlines is based upon tomography X-ray imaging in various fields such as Paleontology, Medical Imaging and Materials Science. The instrument control and data processing systems are cloned on all the relevant beamlines, however the steps of the processing pipeline from the data acquisition to their full exploitation in high quality publications are based upon a heterogeneous stack scenario comprised of e.g. SPEC, Python, Octave, PyHST2 and MATLAB modules. The need has clearly appeared to logically sequence the operations performed by these different actors as user-friendly workflows. At the ESRF we selected a generic workflow tool, Orange, originally developed at the University of Ljubljana and designed for data mining in collaboration with the open source community. The graphical interface enables the easy inclusion/exclusion of functionalities represented by individual boxes. Each box can be managed by simple Python code generating graphical interfaces via the PyQt5 library and is defined by a set of inputs and outputs which can be linked together to produce consistent data processing workflows.

Figure 1: tomwer logo.

INTRODUCTION

Tomography experiments require a set of treatments during acquisition to control the images and to prepare the post-processing. Those treatments are time consuming if repeated frequently.

The goal of tomwer (see Fig. 1) is to offer a tool able to automate a part of the acquisition and reconstruction processes on ESRF Tomography beamlines. tomwer stands for "TOMography Workflow EsRf".

One of the objectives is also to divide these treatments into common atomic processes. Processes can be associated together to create a workflow defining a complex treatment.

Due to the complexity of the treatment and the long history to develop the routines, the software have been developed in a large variety of languages such as Octave, MATLAB, Python, etc. Due to lack of resources it is not possible to envisage rewriting all these codes. On top of this many users/developers at the ESRF do not have knowledge of all these languages, most have knowledge about Python programming. This is why it has been decide to use Python as the glue language.

To let users define their own workflows we chose the Orange [1] canvas. The user-friendly interface for defining workflow was the first reason of this choice. Moreover Orange offers some key functionality like help, checks on input and output types for each connection and the packaging.

Despite the fact that we decided to create an Orange3 add-on for tomwer, each process is accessible without the Graphical User Interface.

In order to fit some of our needs we had to fork the Orange3 original project. Only small modifications have been made, especially to allow cycling inside a workflow. We hope to move those modifications into the Orange3 core project to facilitate maintenance in the future.

The fork of the original Orange3 project can be found in [2].

The tomwer project is distributed under the MIT license. Orange3 follows the GNU General Public License v3.

The tomwer add-on and the library are also strongly based into the silx toolkit [3]. This library offers a set of tools to simplify development of data analysis applications. This work particularly benefits from the I/O functionalities and the visualization widgets.

DESIGN

Architecture

The two main packages of the tomwer project are:

- The 'core' package which groups core functionalities and has no dependency on Orange3.
- The 'widgets' package where classes inheriting from 'OWWidget' (Orange3 widget) are stored.

As we have a graphical user interface and as some treatment can be time consuming we decided to design the application to be multithreaded.

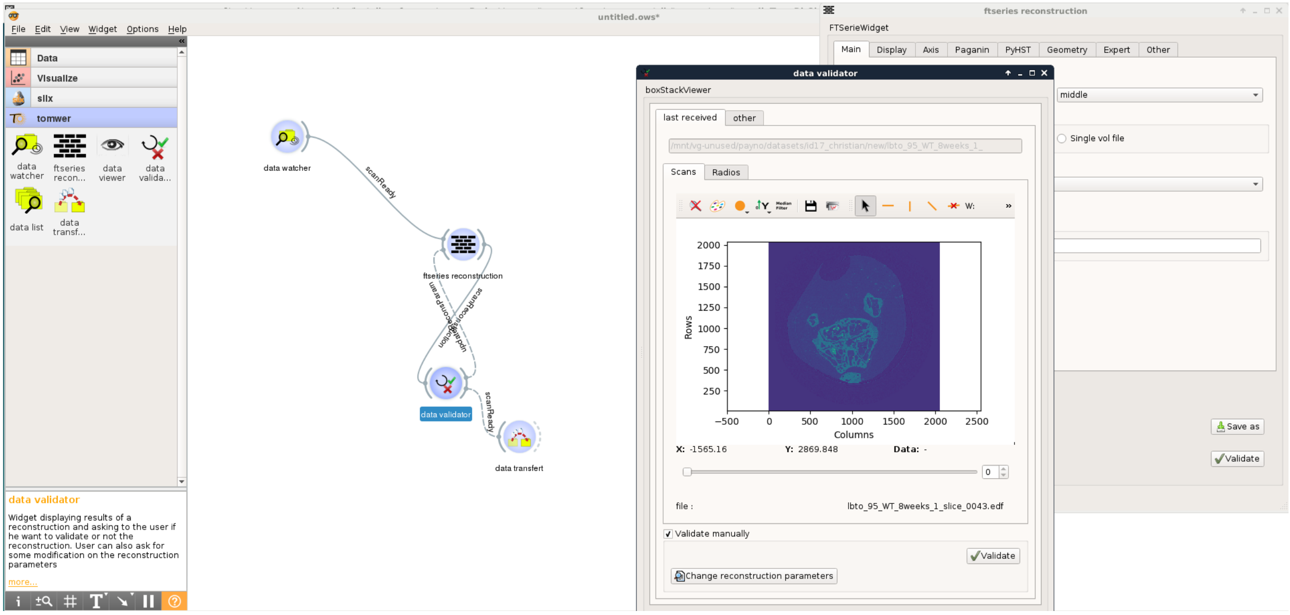
The application for tomography manages data buffering and synchronization in order to speed up treatment.

Process Definition

Each step of the workflow corresponds to a single atomic data treatment and is graphically represented by an orange3 box. This treatment is also characterized by input(s) and output(s). When the orange3 box receives an input/signal it executes the tool action and can emits one or more output/signal. This is the Orange3 design.

The list of inputs and outputs is defined in Table 1.

Signals between tools are designed by Qt signals. Connection between two tools is technically a Qt signal (output emitted) and a Qt slot (input 'received') connected together.



data validator

Widget displaying results of a reconstruction and asking to the user if he want to validate or not the reconstruction. User can also ask for some modification on the reconstruction parameters

/mnt/vg-unused/payno/datasets/id17_christian/new/lbto_95_WT_8weeks_1_

file : lbto_95_WT_8weeks_1_slice_0043.edf

Figure 2: TOMWER set up.

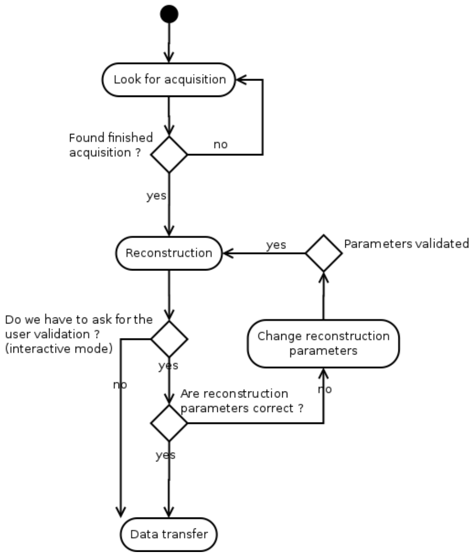

Look for acquisition
Found finished acquisition ?
no
yes
Reconstruction
yes
Parameters validated
Do we have to ask for the user validation ? (interactive mode)
Change reconstruction parameters
yes
no
Are reconstruction parameters correct ?
no
yes
Data transfer

Figure 3: Flow diagram for the workflow of Fig. 2.

- ‘scanReady’ which will carry a string with the path of the scan.
- ‘updateReconsParam’ which will carry needed information for a new reconstruction.

Workflow

The user starts a workflow by picking the tool he wants to process. Then he has to configure the parameters of each tool in an interactive way. For example in the Fig. 2 case he has to find the appropriate reconstruction parameters of the “Ftseries reconstruction” tool. To do so he will have the help of the “data validator” tool. Those tools are described later.

Table 1: Tool Input and Output

Tool	**Input(s)**	**Output(s)**
Data watcher	-	scanReady
Ftseries	scanReady updateReconsParam	scanReady
Data viewer	scanReady	-
Data validator	scanReady	scanReady updateReconsParam
Data list	-	scanReady
Data transfer	scanReady	scanReady

Once the workflow is correctly configured he moves into a “batch mode” which does not require any particular user action and waits for acquisitions and treatments to be done.

The flow diagram corresponding to the workflow designed in Fig. 2 is presented in the Fig. 3

SET OF TOOLS

The whole treatment is divided into a set of atomic processes also called tools. The list of the available tools in TOMWER v0.1 and their actions are:

Data Watcher

This tool parses all subdirectories from a root directory and check if an acquisition is recorded on one of those folders. Acquisition can be detected and marked as finished by parsing a .info file, checking received data or waiting for a particular file. Once the acquisition is marked as finished, a “scanReady” signal is emitted.

Ftseries Reconstruction

This process will be launched when receiving a “scanReady” signal. It will launch an Octave script with the user inputs (given by the interface presented in the Fig. 4 if used as an Orange add-on).

When the reconstruction is finished a "scanReady" signal is sent to let the next box handle it.

During the setup phase an "updateReconsParam" signal can be received. In this case the tool will wait until the user select and validate new reconstruction parameters. Then it will run the reconstruction again.

Data Viewer

This tool allows the user to visualize the data acquired and reconstructed. The viewer is based on the silx 'PlotWidget' class. It allows the display of the image and proposes a basic set of imaging tools such as colormap, median filter and profiling histograms.

Data Validator

This is the composition of a 'Data Viewer' and a set of control buttons.

In the batch phase it will display the last received acquisition and the reconstruction(s) made if any. In the interactive mode it allows the user to :

- validate (the acquisition). This will emit a 'scanReady' output signal.
- cancel : nothing will be done.
- change reconstruction parameters : This will emit an 'updateReconsParam' signal. For example if the box is connected to the ftseries (such as in the 2 case) it will propose to modify the reconstruction parameters and run again the reconstruction.

Data List

Data list is defining a list of tomography data paths. Those data sets will be noticed to the upstream linked processes once the user activate it ('Send' button)

Data Transfer

This will move acquisition from one folder to an-other.

REMOTE LOG

It is mandatory to be able to monitor remotely the progress of such a workflow. Especially when we are on the "batch mode" because the user starts his workflow process and then can be anywhere. This is why each tool send its own progress to a graylog server (see Fig. 5).

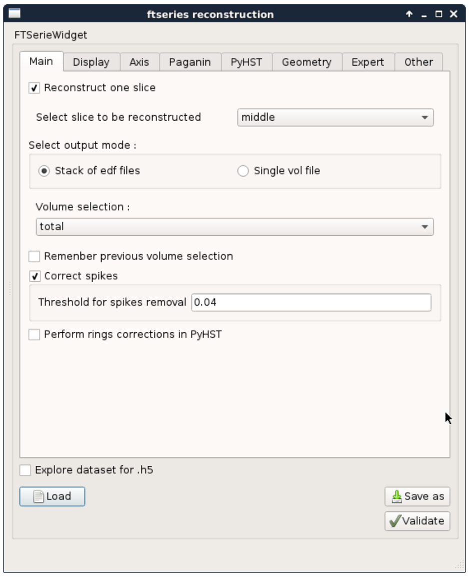
CONCLUSION

Status

A first version of TOMWER (0.1) with the tools described in this paper has been released in September 2017.

TOMWER is used in acquisitions performed at BM05 and ID19 Paleontology imaging beamlines. Unit tests and integration tests have also been included in the development of the project.

Thanks to the Orange3 canvas GUI, users learn very quickly how to use the set of tools available.



☑ Reconstruct one slice

Select slice to be reconstructed: middle

Select output mode :

● Stack of edf files ○ Single vol file

Volume selection : total

☐ Remember previous volume selection

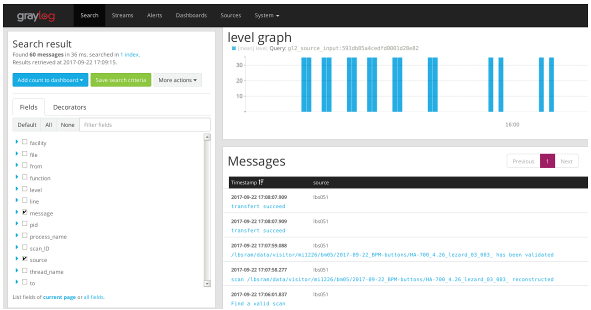
☑ Correct spikes

Threshold for spikes removal: 0.04

☐ Perform rings corrections in PyHST

☐ Explore dataset for .h5

Figure 4: Ftseries interface.



Search result

Found 60 messages in 36 ms, searched in 1 index.
Results retrieved at 2017-09-22 17:09:15.

☐ facility
☐ file
☐ from
☐ function
☐ level
☐ line
☑ message
☐ pid
☐ process_name
☐ scan_ID
☑ source
☐ thread_name
☐ to

List fields of current page or all fields.

[mean] level, Query: gl2_source_input:591db85a4cedfd0001d28e82

```
Timestamp                 source
2017-09-22 17:08:07.909   lbs051
transfert succeed
2017-09-22 17:08:07.909   lbs051
transfert succeed
2017-09-22 17:07:59.088   lbs051
/lbsram/data/visitor/mi1226/bm05/2017-09-22_BPM-buttons/HA-700_4.26_lezard_03_003_ has been validated
2017-09-22 17:07:58.277   lbs051
scan /lbsram/data/visitor/mi1226/bm05/2017-09-22_BPM-buttons/HA-700_4.26_lezard_03_003_ reconstructed
2017-09-22 17:06:01.837   lbs051
Find a valid scan
```

Figure 5: Graylog interface.

Future

Next step is to adapt the existing tools to other techniques like holotomography and fluorescence imaging particularly on the nano-imaging beamline ID16.

This workflow will be extended to the postprocessing part of the data analysis, especially to deal with series of tomography scans and massive reconstructions on OAR, the batch scheduler provided at ESRF.

ACKNOWLEDGEMENT

We would like to thanks scientists from BM05 and ID19 and especially V.Fernandez and P.Tafforeau who took time during their experiments to test the add-on and give helpful feedback for specifications. We also appreciated the time given by the Ljubljana university to give us a tour of Orange and what are the limitations of Orange3. We are also grateful to V.A. Solé for pointing us to a user friendly interface. Finally we would like to thanks M.S. del Rio for introducing us to Orange from Oasys [4] and explain us some welcome tricks.

REFERENCES

[1] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup and Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek and Lan Žagar, Jure Žbontar, Marinka Žitnik and Blaž Zupan, "Orange: Data Mining Toolbox in Python", in *Journal of Machine Learning Research*, 2013, vol 14, pp 2349–2353.

[2] Orange3, https://github.com/payno/orange3

[3] silx, https://zenodo.org/record/576042
[Python packages to support the development of data assessment, reduction and analysis applications at synchrotron radiation facilities]

[4] M. Sanchez del Rio, L. Rebuffi, J. Demšar, N.Canestrari and O. Chubar, "A proposal for an open source graphical environment for simulating X-ray optics", in *Proc. SPIE 9209, 92090X* 2014.