

Minutes of the EDNA workflow meeting at Soleil September 21st - 22nd 2009

Participants: Gwenaëlle Abeillé^a, Sandor Brockhauser^b, Alain Buteau^a, Peter Chang^c, Andy Götz^d, Karl Levik^c, Erwin de Ley^e, Sandra Pierre-Joseph^a, Majid Ounsy^a and Olof Svensson^d

Partly participating: Brigitte Gagey^a and Andrew Thompson^a

^aSoleil (France), ^bEMBL Grenoble (France), ^cDiamond Light Source (UK), ^dESRF (France), ^eIsencia (Belgium)

Morning session September 21st :

Alain Buteau started the meeting by presenting* the agenda (see Appendix 1) and the agreed goals with this meeting (see Appendix 2, gentlemen's agreement):

“The EDNA plugin architecture contains execution and control elements. The use of a workflow system, like Passerelle, for building scientific workflows is very attractive as a replacement of the “hard coded” EDNA control plugins. The EMBL Grenoble has already started to study the possibilities of using the Kepler system that is an extension project based on Ptolemy, just like Passerelle. SOLEIL expertise on using such systems, and especially Passerelle can give a big advantage when evaluating the technology, or even later if EDNA chooses that technology for further developments.”

and

“The goals for the EDNA collaboration in the September workshop is to evaluate Passerelle as a possible choice for a workflow tool for EDNA.”

Olof then presented* the EDNA project and his views of the advantages / difficulties in using a workflow tool like passerelle in EDNA. The major obstacle in Olof's point of view are the current limitations of how the data model is used by the plugins : all input and result objects for the plugins must be precisely defined by the data model, and it is not possible to provide more than one input or output object.

This presentation was followed by Majid who presented* the current use of passerelle at Soleil, together with an overview of the Soleil control system.

Finally Erwin presented* Isencia and the passerelle tool. He also presented his views on using Passerelle with EDNA with three concrete approaches for how to use Passerelle workflows to run EDNA plugins:

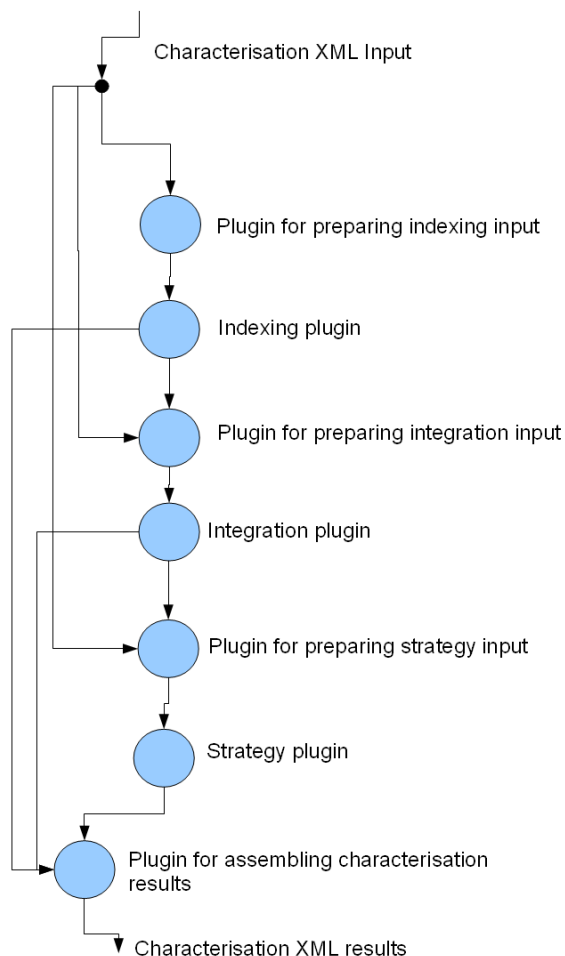
1. Invoke EDNA plugins from Passerelle
2. Directly invoke external programs from Passerelle, i.e. replace the EDNA plugins with Passerelle actors
3. Mixed approach : option 1. for existing plugins, 2. for new ones

He also presented a comparison of the EDNA plugin and Passerelle actor API. Since the two APIs are almost identical it should be straightforward to work with any of the three approaches listed above.

* See http://www.edna-site.org/wiki/index.php/EDNA_Documents for links to the presentations.

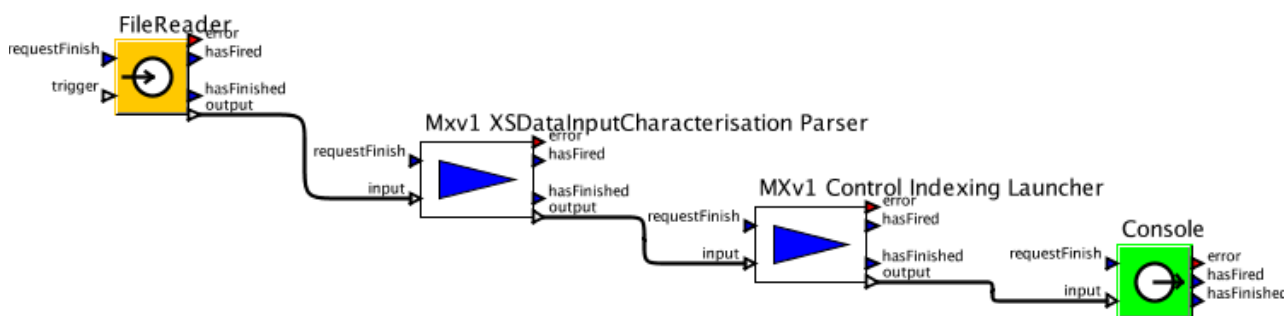
Afternoon session September 22nd :

The afternoon session started with a discussion around which workflow to try to use as example for developing a Passerelle workflow which runs EDNA plugins. Olof suggested to use the EDNA Mxv1 characterisation workflow:



We agreed to start working on this workflow. In order to facilitate the execution of the workflow we decided to start from a test XSDDataInputCharacterisation XML file. Erwin then started to develop the passerelle actors using the projector which allowed the meeting participants to follow the developments. By the end of the day Erwin had two passerelle actor skeletons working:

Passerelle Std



Morning session September 23rd :

This session was used for improving the actors and testing them with Passerelle and EDNA installed on Olof's laptop. By lunchtime the workflow was correctly executing the EDNA control indexing plugin given the input from the FileReader actor.

During this testing a bug was found in the Mxv1 EDNA application (Bug #343 in the EDNA bugzilla repository). The fact that this bug had not been found earlier is because this is the first time (a part from the executing tests) that we try to run the control indexing plugin stand-alone. We therefore believe that a workflow tool like Passerelle will be of tremendous help to use the full power of the EDNA modularity.

As the last part of the pervious session and this session mostly concerned Erwin and Olof, many other meeting participants held parallel sessions about Nexus format etc. (see the agenda in Appendix 1).

Afternoon session September 23rd :

We spent a part of the afternoon on Proxima 1 for testing the workflow in "real" conditions, i.e. by using the passerelle installation on Proxima 1 and executing the indexing process on a different computer.

The workflow did work however there was a glitch on the EDNA side which prevented the prediction images to be generated. We can though say that the test was a success as we managed to run an EDNA plugin from passerelle in a beamline environment.

We spent the last hour of the meeting concluding the experiences we had during these two days and preparing for future developments:

- We all agreed that this meeting was a success in many ways : the understanding of Passerelle and EDNA was increased amongst the participants and we managed to reach concrete results.
- The next step is to discuss the Passerelle workflow tool and EDNA coupling at the next EDNA developers' meeting at the ESRF October 12th -13th. Gwenaëlle, Sandra and Erwin will participate in this meeting, however neither Andy G nor Alain can participate due to the ICALEPS meeting in Japon.
- Alain suggested that Soleil could develop the Passerelle solutions for EDNA if agreed to use Passerelle in EDNA. Both Andy G and Olof pointed out that since Passerelle would be linked with core functionality of EDNA, all interested EDNA developers should be free to participate in these developments. The experience of using and developing Passerelle at Soleil would be very valuable in any eventual future developments.
- We agreed that the goal of the EDNA / Passerelle discussions in the next EDNA meeting are to agree on a roadmap for implmenting workflows in Passerelle which executes EDNA plugins. If we can agree on such a roadmap it will be submitted after the meeting for approval to the EDNA executive committee.

Appendix 1: Agenda

Passerelle-EDNA workshop

Sept 21-22, 2009 - SOLEIL

1 Participants

- EMBL : Sandor Brokhauser
- DIAMOND : Karl LEVIK
- DIAMOND : Peter CHANG
- ISENCIA: Erwin de Ley
- ESRF: Andy Gotz
- ESRF: Olof Svensson
- SOLEIL
 - Head of computing division : B. Gagey
 - ICA group : : Gwenaëlle Abeillé, Sandra Pierre-Joseph, Majid Ounsy , A. Buteau
 - PX1 beamline : A. Thompson

2 Monday 21th , 2009 . *Room O6.1.01*

09:30 - 09:45 INTRODUCTION (Alain)

- Welcome message
- Meeting Goals

09:45 - 10:15 EDNA Project (Olof)

- Goals and architecture
- The workflow tool problem

10:15 - 10:45 Coffee break

10:45 - 11:15 PASSERELLE (Majid)

- The tool
- Use at Soleil

11:15 - 12:00 PASSERELLE (Erwin)

- Isencia software offers
- Professional workflow use cases
- Architecture proposal for EDNA Workflows

12:00 - 14:00 lunch

14:00 - 15:00 : EDNA Test Case Elaboration (Andy Thompson)

- Discussion of a possible test case to run on PX1 beamline

15:00 - 18:00 Parallel sessions

a) Workflow Implementation (Erwin + Gwen + Sandra + Olof + Karl ? + Sandor ?)

Implementation and desktop simulation

b) Tango DeviceServer source code organisation (Alain, Andy G , xxx ?)

c) NeXus data visualisation components : SOLEIL and DIAMOND architecture and solutions (Peter + Majid+ Katy + Gregory+ Stephane +Céline)

19:00 - 22:00 DINNER

3 Tuesday 22, 2009 . Room O6.1.01

09:00 - 12:00 Parallel sessions

d) Passerelle-EDNA Software installation and tests on PX1 (Majid,Gwen,Sandra, Erwin,Olof,Sandor)

- Use case test on PX1 beamline

e) NeXus metadata : (Peter + Stephane + Alain + Andy, Celine). Room

- Bolero demo

12:00 - 14:00 lunch

14:00 - 15:00 PX1 results analysis (Andy Thompson ?)

- First outcomes

15:00 - 16:00 END SESSION

- Discussion of the further collaboration steps

16:00 - 17:00 beamlines visit ??

Appendix 2:

Gentlemen agreement on “Passerelle/EDNA software integration” between the ICA group of SOLEIL and the EDNA collaboration

1 Aim of this document

This paper defines how SOLEIL and the EDNA collaboration will collaborate on the software projects defined in this document. It gives a description of the interests of each partner, and the first identified goals and tasks, in particular the small workshop to be organised at the Soleil on September 21st - 22nd 2009.

2 Institutes motivations

2.1 SOLEIL motivation

SOLEIL uses the Passerelle[1] tool on the beamlines as an environment allowing scientists to develop their own “control workflow” for beamline alignment, data collection, etc. Unitary steps of the workflow are handled by what is called actors in the Passerelle vocabulary.

Many actors have been developed by SOLEIL mainly to access to Tango Control system, because our first target was for Controls and Data Acquisition applications.

Extending the library of available actors, to actors dedicated to data analysis, is an opportunity for SOLEIL to:

- To add the “Online data analysis” functionalities to the existing Passerelle environment
- Benefit from the expertise of the EDNA group on online data analysis

Working with others institutes on Passerelle is also an opportunity to enhance the Passerelle framework by enlarging its diffusion to a user’s community with similar needs.

2.2 EDNA collaboration motivations

EDNA will be a new generation environment for ODA (Online Data Analysis), in particular for the automation of the collection and processing of X-ray diffraction data from macromolecular crystals. The EDNA collaboration consists of seven institutes which participate in the executive committee (Diamond Light Source, EMBL Grenoble, EMBL Hamburg, ESRF, Global Phasing, MRC-LMB Cambridge and Soleil) and another six official collaborators (BESSY, Swiss Light Source, NSLS Brookhaven, MAX-Lab, University of York and the University of Sydney). The goal of the EDNA collaboration is to:

- provide a framework for on-line data analysis which is modular, test-driven (unit and execution tests) and which contains a data model framework
- provide applications for on-line data analysis, especially in the domains of synchrotron radiation macromolecular crystallography.

The EDNA plugin architecture contains execution and control elements. The use of a workflow system, like Passerelle, for building scientific workflows is very attractive as a replacement of the “hard coded” EDNA control plugins. The EMBL Grenoble has already started to study the possibilities of using the Kepler[2] system that is an extension project based on Ptolemy[3], just like Passerelle. SOLEIL expertise on using such systems, and especially Passerelle can give a big advantage when evaluating the technology, or even later if EDNA chooses that technology for further developments.

The goals for the EDNA collaboration in the September workshop is to evaluate Passerelle as a possible choice for a workflow tool for EDNA.

3 Already planned project tasks

<i>Who</i>	<i>What</i>	<i>When</i>
SOLEIL	sends a version of Passerelle to EMBL to allow EMBL to use the framework for test	W28 - done
EDNA	Gives an UML model of the EDNA data model	Done[4]
SOLEIL	Develops generic wrapper of EDNA plugins into Passerelle framework	W30-W36
SOLEIL	Develops a few actors using this wrapper	W30-W36
SOLEIL/EDNA	Small workshop to examine the possibilities and consequences of using passerelle for defining the workflows in EDNA.	Mid september
SOLEIL, EDNA	Present the result of the small workshop, preferably with a prototype of Passerelle/EDNA integration, to the official EDNA developers' meeting.	To be decided

4 To be discussed in the September meeting

4.1 The possibility of using a workflow tool for EDNA - general considerations

The workflow for the current EDNA architecture is programmed in the so called control plugins. If a workflow tool is used for EDNA, how would these control plugins need to be modified? There are two possibilities:

1. A workflow tool, like Passerelle or Kepler, is used both for defining the workflow and for executing it. This is how the Passerelle tool currently is used at Soleil. In this scheme the control plugins would be unnecessary, they would simply be replaced by the workflow tool.

The workflows could be debugged with the tool. This solution would imply that it not be possible to execute the EDNA workflow without the workflow tool.

2. A workflow tool is used to define the workflow but not to execute it. After having defined the workflow it would be passed to the EDNA framework, for example in XML format, which would then execute it. Thus the EDNA framework would not be dependent on the workflow tool in runtime however workflows cannot be debugged using the tool.

4.2 The possibility of using the Passarelle workflow tool for EDNA

Here are some of considerations need to be taken into account in incorporating a new tool, such like Passerelle, in EDNA:

- The tool must be easily available for all developers. It doesn't have to be open-source, however the cost of acquiring the tool should not be prohibitive.
- If the EDNA applications become dependent on using the tool, it's licence must allow it to be distributed together with the EDNA applications.

A part from these general considerations, the implications for using Passerelle for EDNA must be discussed. Ideally a small prototype could be developed which would run some EDNA plugins, either directly or through some generated XML (see Section 4.1 above).

[1] Passerelle article: <http://accelconf.web.cern.ch/AccelConf/ica07/PAPERS/RPPB20.PDF>

[2] Kepler: an open-source scientific workflow application, see <https://kepler-project.org>

[3] The Ptolemy project studies modeling, simulation, and design of concurrent, real-time, embedded systems, see: <http://ptolemy.eecs.berkeley.edu>

[4] URL to the EDNA data models: http://www.edna-site.org/wiki/index.php/Data_Models