

Applications to Catalysis Workflows

**Leandro Liborio¹, Patrick Austin², Alejandra González-Beltrán²,
Abraham Nieva de la Hidalga³**

¹Theoretical and Computational Physics Group, SCD, STFC

²Data and Software Engineering Group, SCD, STFC

³UK Catalysis Hub, Research Complex at Harwell

EGD meets ESG, October 5th 2023, Freiburg



PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE



ada lovelace centre



Software
Sustainability
Institute



Motivation

EuroScience
Gateway

WP5 Objectives:

Develop customizable pilot workflows
Onboard new communities

Targeting Materials Science community

Enable researchers in the physical sciences to handle data more easily

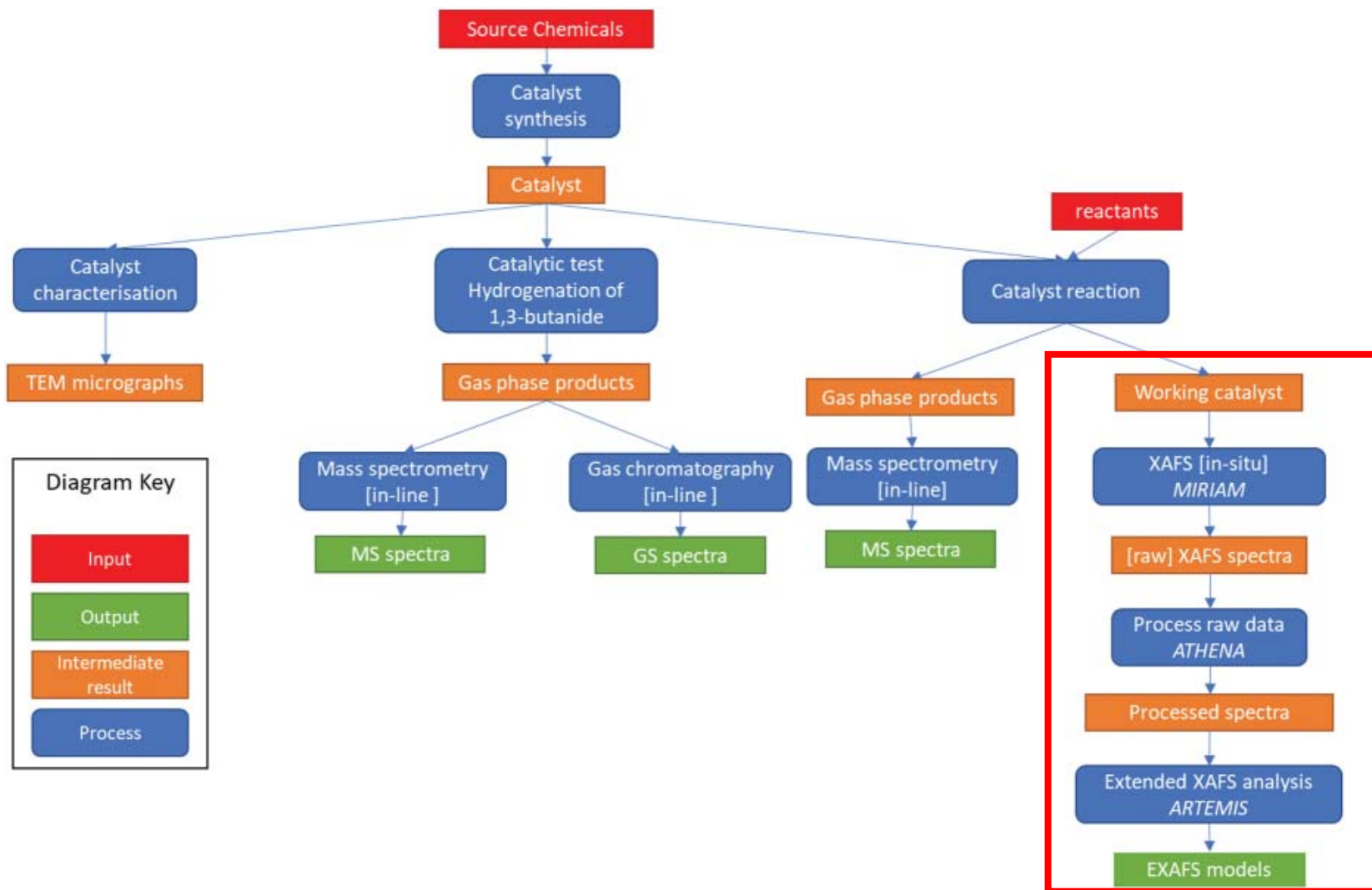


PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

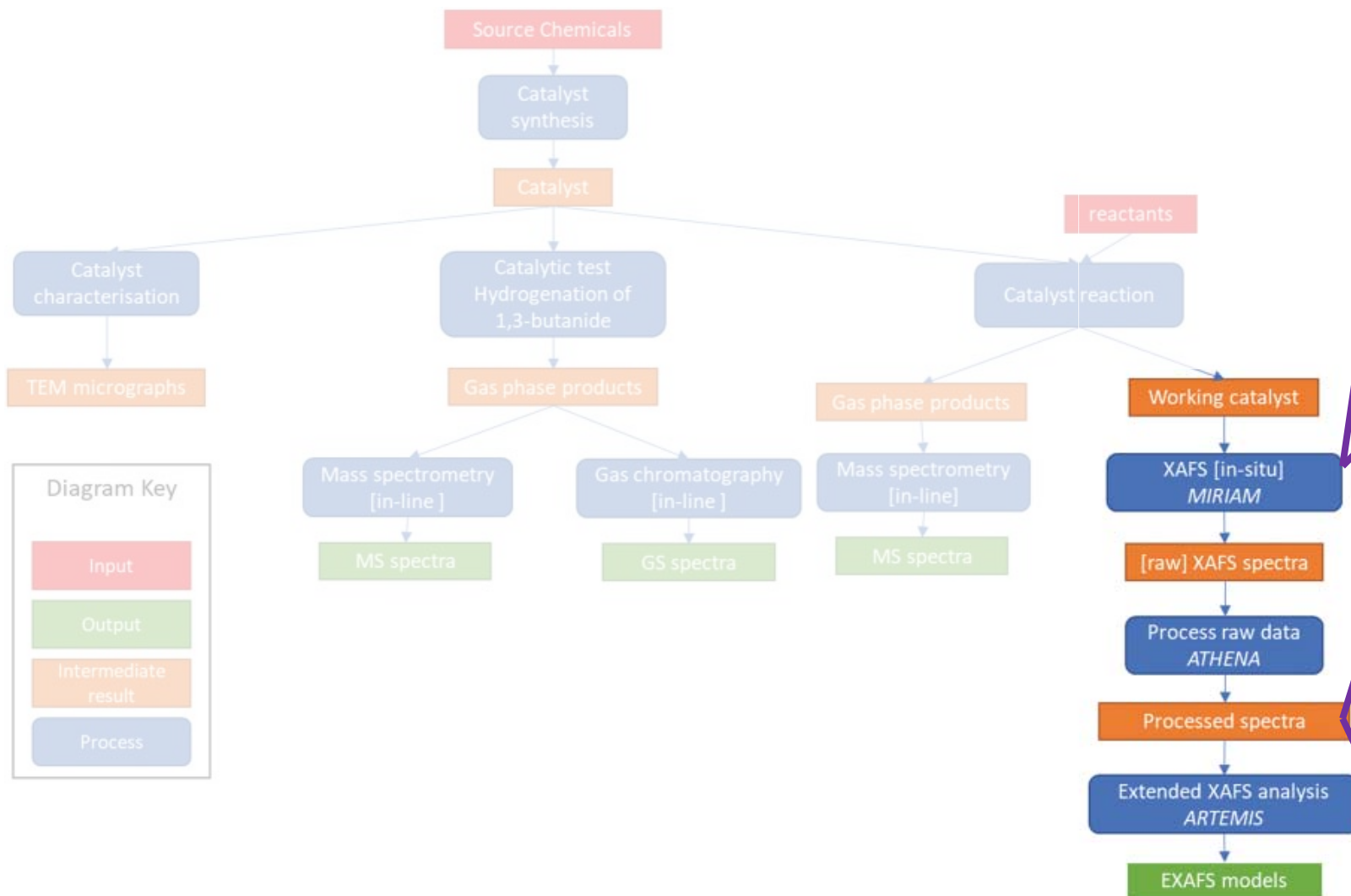
WP4, Pathfinder 1:

...creating an integrated environment to support capture, analyse and reuse data. Prototype developed for XAS data

Catalysis use case



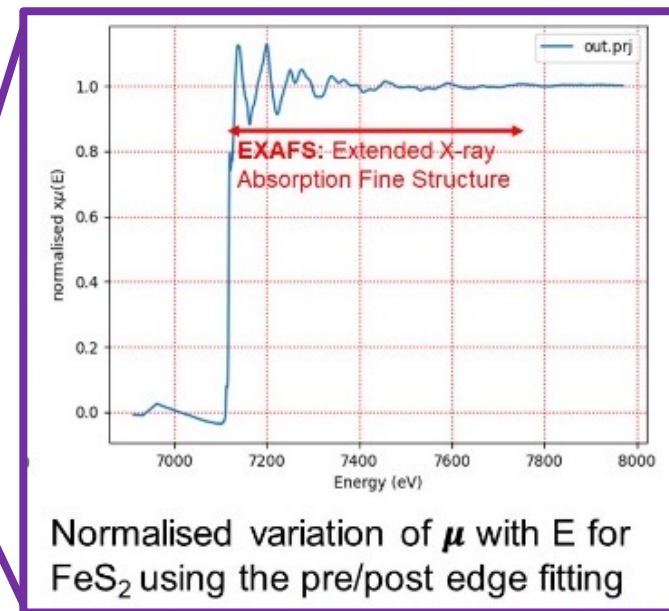
Catalysis use case



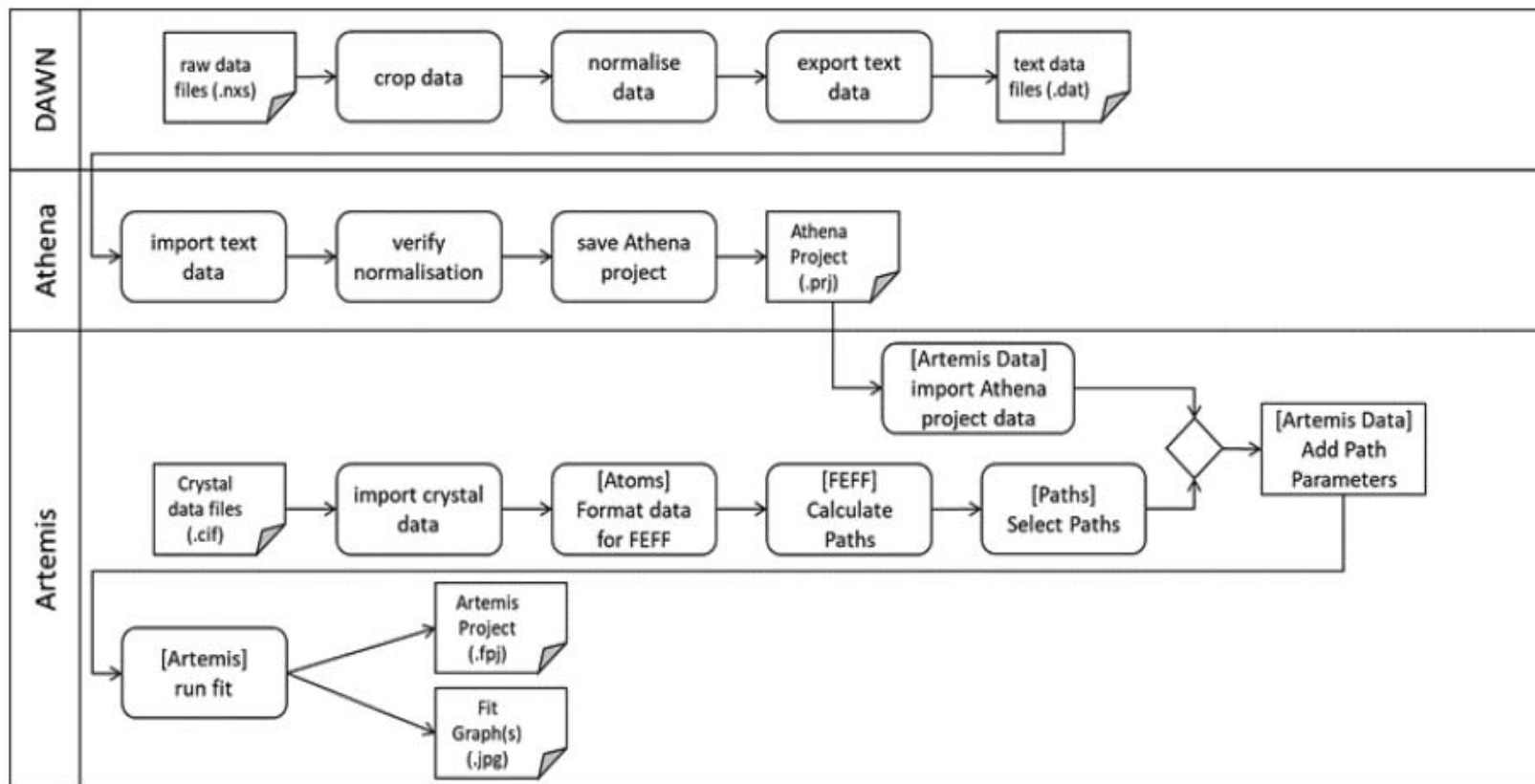
I_0 \rightarrow FeS2 \rightarrow I_t

$$\mu(E)x = -\ln(I_t/I_0)$$

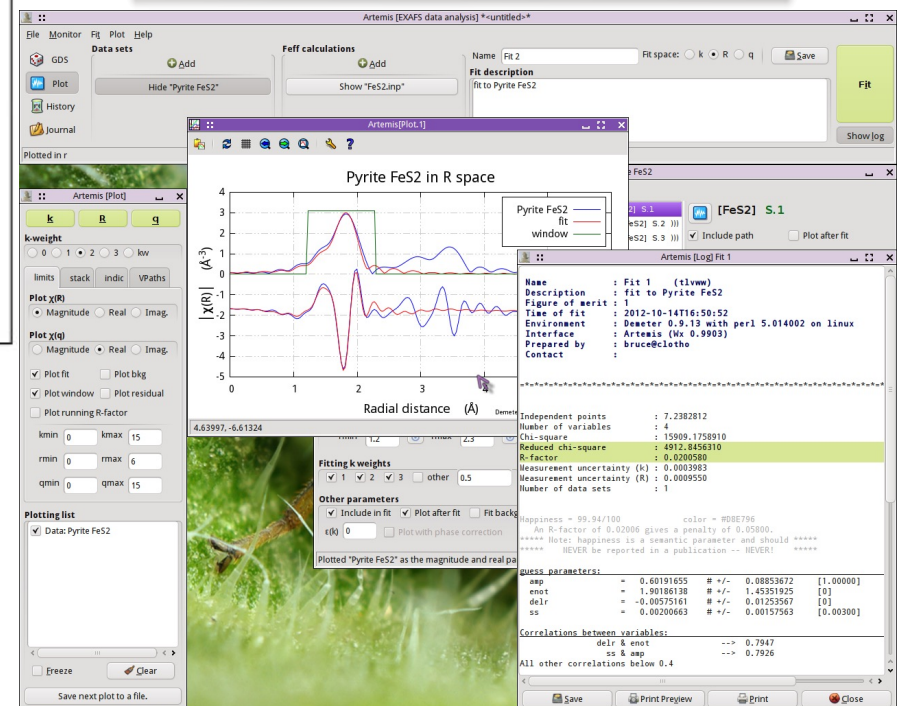
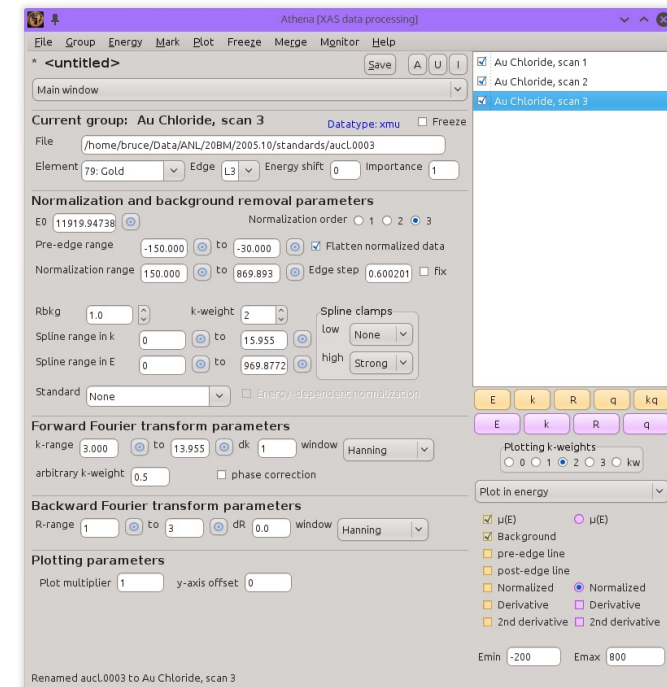
X-ray beam of energy E hits a FeS2 sample of width x . Using the initial and transmitted intensities, the absorption coefficient μ is defined.



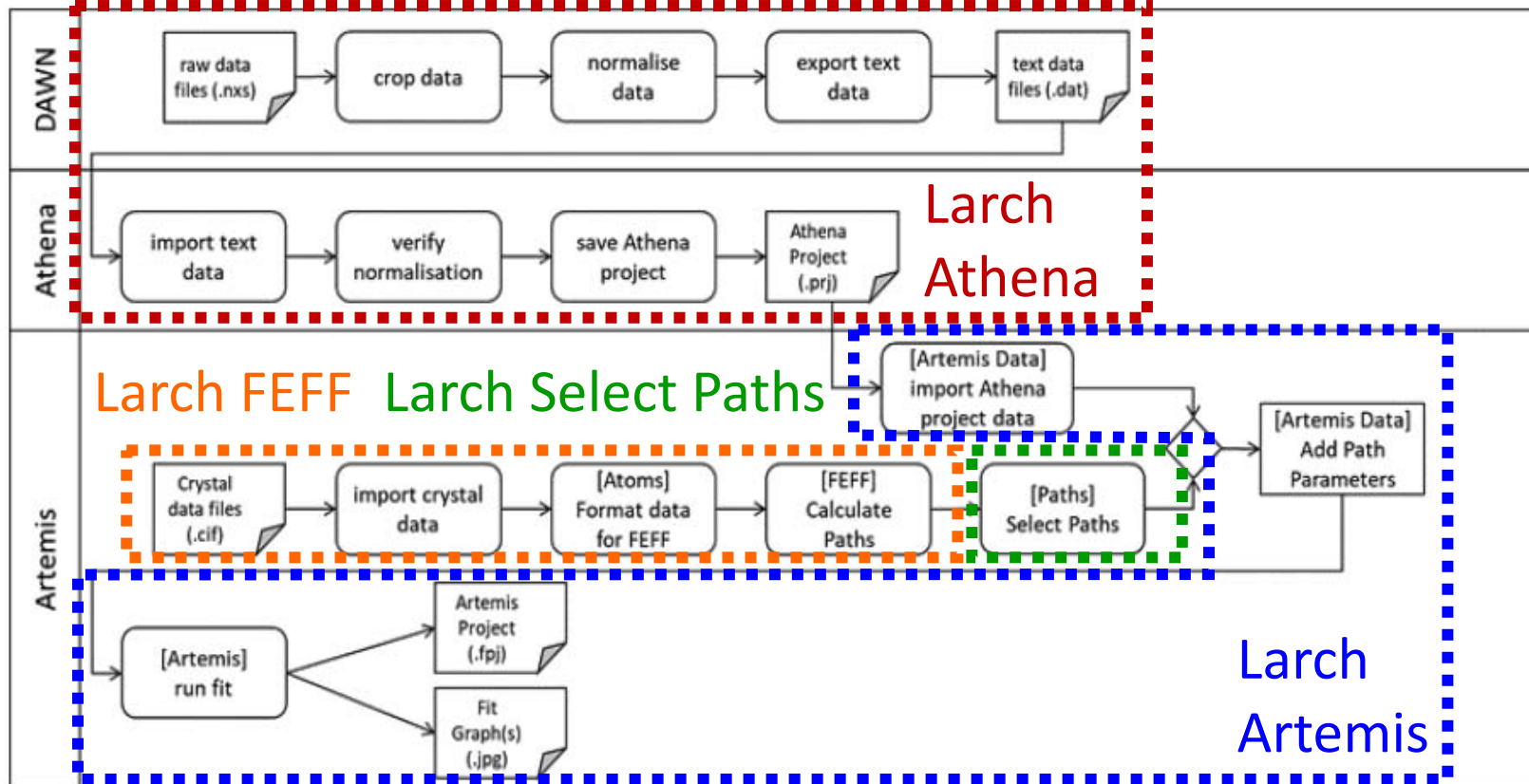
Current approach



<https://bruceravel.github.io/demeter/>



XAS in Galaxy



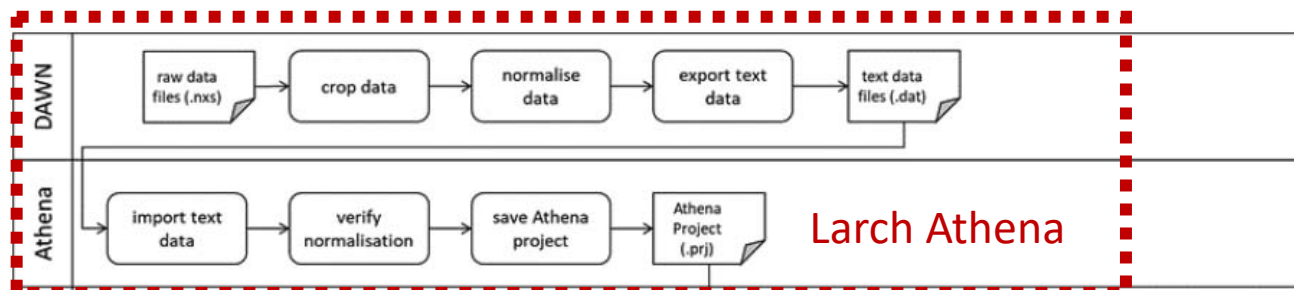
- **Larch Athena:**
 - Processing and Normalization
 - Cropping energy range
 - Outputs project file and plots
- **Larch FEFF:**
 - Load from cif and converts to FEFF input file (or loads FEFF input)
 - Outputs zipped directory of paths
- **Larch Select Paths:**
 - Selects which paths from **Larch FEFF** to use
 - Defines parameters for these paths
- **Larch Artemis:**
 - Performs fitting on FEFF paths
 - Outputs report on fitting and plots

Also implementing utility tools and tools for other XAS techniques (not shown on diagram).

<https://xraypy.github.io/xraylarch/>

Process and Normalise XAS with **Larch Athena** in Galaxy

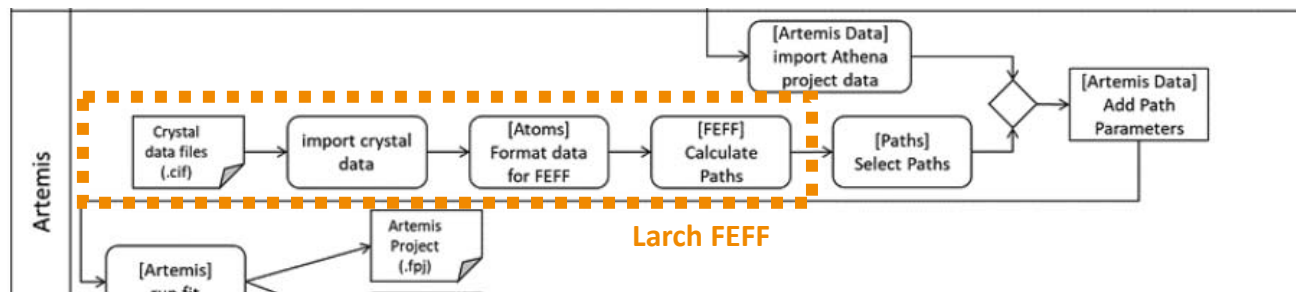
The screenshot displays the Galaxy µSR web interface. The top navigation bar includes 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User'. The main workspace shows the 'Larch Athena generate Athena projects from XAFS data' tool. The input field is set to '5: test.xmu.txt'. The tool description states: 'Normalised X-ray Absorption Fine Structure (XAFS) data, generated by DAWN.' The 'Plot graph' option is set to 'No'. An 'Execute' button is visible. Below the tool description, there are citations and requirements. The citations include: 'Newville, M. (2013). Larch: An Analysis Package for XAFS and Related Spectroscopies. *Journal of Physics: Conference Series*, 430, 012007. <https://doi.org/10.1088/1742-6596/430/1/012007>' and 'Ravel, B., & Newville, M. (2005). *Journal of Synchrotron Radiation*, 12(4), 537–541. <https://doi.org/10.1107/s0909049505012719>'. The requirements listed are: 'xraylarch (Version 0.9.66)', 'xraydb (Version 4.4.7)', 'sqlalchemy (Version 1.4.46)', and 'matplotlib (Version 3.5.2)'. The license is 'MIT License' and the creator is 'Patrick Austin'. On the left, a 'Tools' sidebar lists various tools under 'Tools Under Development' and 'Collection Operations'. On the right, a 'History' sidebar shows 'PSDI_XAS' with a size of 16.7 kB and a message: 'This history is empty. You can load your own data or get data from an external source.'



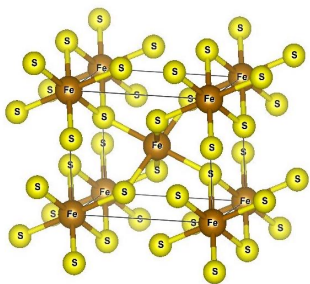
Path Generation in FeS₂ using Larch FEFF in Galaxy

The screenshot displays the Galaxy web interface. On the left, a 'Tools' sidebar lists various categories like 'Get Data', 'OTHER TOOLS', and 'WORKFLOWS'. The main area shows a table of data with three columns of numerical values. On the right, a 'History' panel lists datasets such as 'test.xmu' and '1564889.cif'. A video player at the bottom shows a progress bar from 01:21 to 08:36.

7434.1940	0.89577218	304310.00
7437.7090	0.89366937	304338.00
7441.2440	0.89233424	304342.00
7444.7980	0.89109983	304576.00
7448.3710	0.88991984	304397.00
7451.9620	0.89021992	304464.00
7455.5740	0.88787942	304333.00
7459.2040	0.88208656	304131.00
7462.8530	0.87531841	304376.00
7466.5210	0.87040248	304297.00
7470.2080	0.86516782	304046.00
7473.9150	0.85785170	304975.00
7477.6400	0.85054938	305708.00
7481.3850	0.84453901	306082.00
7485.1480	0.83863201	305961.00
7488.9310	0.83371670	306201.00
7492.7330	0.82904991	306134.00
7496.5540	0.82459446	306039.00
7500.3940	0.82079844	305977.00
7504.2530	0.81677576	305156.00
7508.1310	0.81178911	306748.00
7512.0280	0.80667240	306054.00
7515.9440	0.80196850	305949.00
7519.8790	0.79709432	305906.00
7523.8340	0.79034294	305935.00
7527.8070	0.78287009	305891.00
7531.8000	0.77566164	305816.00
7535.8120	0.77172412	305895.00
7539.8420	0.76918155	305839.00
7543.8920	0.76588805	305840.00
7547.9610	0.76149338	305863.00
7552.0490	0.75668300	305740.00
7556.1560	0.75093523	305560.00
7560.2820	0.74539760	305668.00
7564.4270	0.74088376	305785.00
7568.5910	0.73831944	305684.00
7572.7740	0.73529922	305582.00
7576.9770	0.73187551	305523.00
7581.1980	0.72922440	305558.00
7585.4390	0.72676436	305647.00
7589.6980	0.72359612	305208.00
7593.9770	0.71897714	306038.00
7598.2750	0.71252104	305315.00
7602.5920	0.70549579	305573.00
7606.9280	0.69954396	305378.00
7611.2830	0.69265029	305425.00
7615.6570	0.68689223	305488.00
7620.0500	0.68092659	305297.00
7624.4620	0.67499882	305001.00
7628.8940	0.66889046	305246.00
7633.3440	0.66341014	305391.00
7637.8130	0.65798296	305277.00
7642.3020	0.65352747	305129.00
7646.8100	0.64842668	305185.00
7651.3360	0.64433064	305107.00
7655.8820	0.64130624	304968.00
7660.4470	0.63726602	303292.00
7665.0300	0.63282645	306701.00
7669.6310	0.62799440	304001.00



Path Selection in FeS₂ using Larch Select Paths in Galaxy



FeS₂

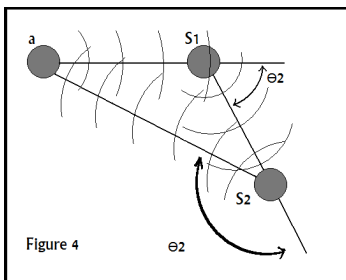


Figure 4

Multiple scattering
of emitted
photoelectron

Tools

search tools

Upload Data

UEP Method

Other Methods

OTHER TOOLS

PyMuonSuite

File Conversion

Tools Under Development

PyMuonSuite AIRSS UEP Optimise

run UEP optimisation

PyMuonSuite AIRSS Cluster

run clustering for optimised structures

MuSpinSim Combine

combine datafiles generated from MuSpinSim

MuSpinSim Generator

Generate MuSpinSim config from a structure file

Larch FEFF

generate FEFF paths from XAFS data

Larch Select Paths

select FEFF paths for XAFS data

Larch Athena

generate Athena projects from XAFS data

Larch Artemis

generate Artemis projects from XAFS data

MuSpinSim Configure

define simulation parameters

Collection Operations

WORKFLOWS

All workflows

```

* This FEFF.inp file generated by pymatgen
TITLE comment: None given
TITLE Source:
TITLE Structure Summary: Fe2 S4
TITLE Reduced formula: FeS2
TITLE space group: (Pnm), space number: (58)
TITLE abc: 3.385200 4.447400 5.428700
TITLE angles: 90.000000 90.000000 90.000000
TITLE sites: 6
* 1 Fe 0.000000 0.000000 0.000000
* 2 Fe 0.500000 0.500000 0.500000
* 3 S 0.000000 0.199900 0.378040
* 4 S 0.000000 0.800100 0.621960
* 5 S 0.500000 0.699900 0.121960
* 6 S 0.500000 0.300100 0.878040

POTENTIALS
*ipot Z tag lmax1 lmax2 xnatph(stoichiometry) spinph
*****
0 26 Fe -1 -1 0.0001 0
1 26 Fe -1 -1 2 0
2 16 S -1 -1 4 0

ATOMS
* x y z ipot Atom Distance Number
*****
0 0 0 0 Fe 0 0
-0.889035 -2.05227 -0 2 S 2.23656 1
0.889035 2.05227 0 2 S 2.23656 8
-1.33466 0.662084 -1.6926 2 S 2.2549 2
-1.33466 0.662084 1.6926 2 S 2.2549 3
1.33466 -0.662084 -1.6926 2 S 2.2549 4
1.33466 -0.662084 1.6926 2 S 2.2549 5
0 0 -3.3852 1 Fe 3.3852 6
0 0 3.3852 1 Fe 3.3852 7
END
                    
```

History

search datasets

Unnamed history

149 kB

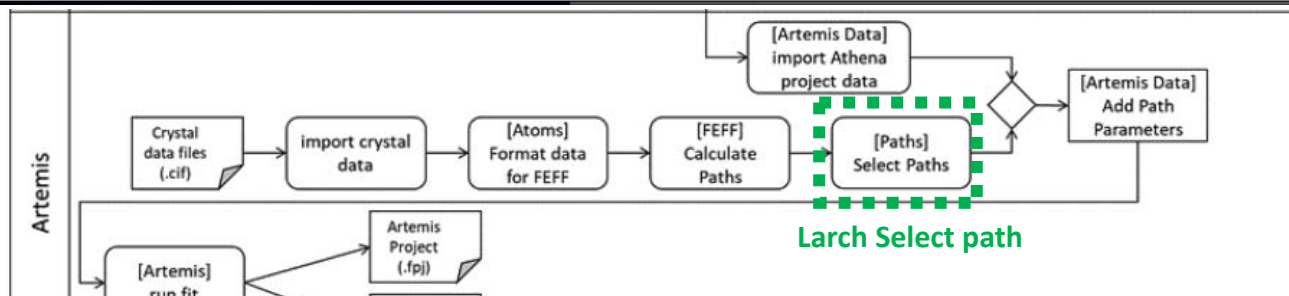
5 : FEFF input of 1564889.ci f

4 : FEFF paths of 1564889.ci f

3 : CSV summary of 156488 9.cif

2 : test.xmu

1 : 1564889.cif



Extended Analysis in FeS₂ using Larch Artemis in Galaxy

Tools

search tools

Upload Data

UEP Method

Other Methods

OTHER TOOLS

PyMuonSuite

File Conversion

Tools Under Development

PyMuonSuite AIRSS UEP Optimise run UEP optimisation

PyMuonSuite AIRSS Cluster run clustering for optimised structures

MuSpinSim Combine combine datafiles generated from MuSpinSim

MuSpinSim Generator Generate MuSpinSim config from a structure file

Larch FEFF generate FEFF paths from XAFS data

Larch Select Paths select FEFF paths for XAFS data

Larch Athena generate Athena projects from XAFS data

Larch Artemis generate Artemis projects from XAFS data

MuSpinSim Configure define simulation parameters

Collection Operations

WORKFLOWS

All workflows

Larch Artemis generate Artemis projects from XAFS data (Galaxy Version 0.9.66+galaxy0)

Execution mode

Parallel

Series

Whether to execute Athena projects in parallel or in series.

Athena project file

8: Athena project of test.xmu

Normalised X-ray Absorption Fine Structure (XAFS) data, in Athena project format. If a collection of files is provided, these will be submitted and executed in parallel.

FEFF paths file

4: FEFF paths of 1564889.cif

Zipped directory of the FEFF paths.

GDS parameters file

7: Selected paths of CSV summary of 1564889.cif

File defining the fitting parameters as a 'guess' (to be varied in the fit), 'def' (defined by an expression evaluated throughout fitting) or 'set' (evaluated at the start of fitting, then left unchanged).

SP parameters file

7: Selected paths of CSV summary of 1564889.cif

File defining the scattering paths.

Fitting Variables

Plot graph

No

Whether to plot the data.

Execute

Using Larch, perform fitting on an Athena project file, originally from the input X-ray Absorption Fine Structure (XAFS) data file.

Optionally, plot the $\chi\mu$ data along with RMR and ChiQR plots for visual inspection of the fit.

Citations

- Neville, M. (2013). Larch: An Analysis Package for XAFS and Related Spectroscopies. *Journal of Physics: Confer*

History

search datasets

Unnamed history

251 kB 10

10: Flattened plot of test.x mu

9: Edge fitting of test.xmu

8: Athena project of test.x mu

7: Selected paths of CSV summary of 1564889.cif

6: GDS values of CSV summary of 1564889.cif

5: FEFF input of 1564889.cif

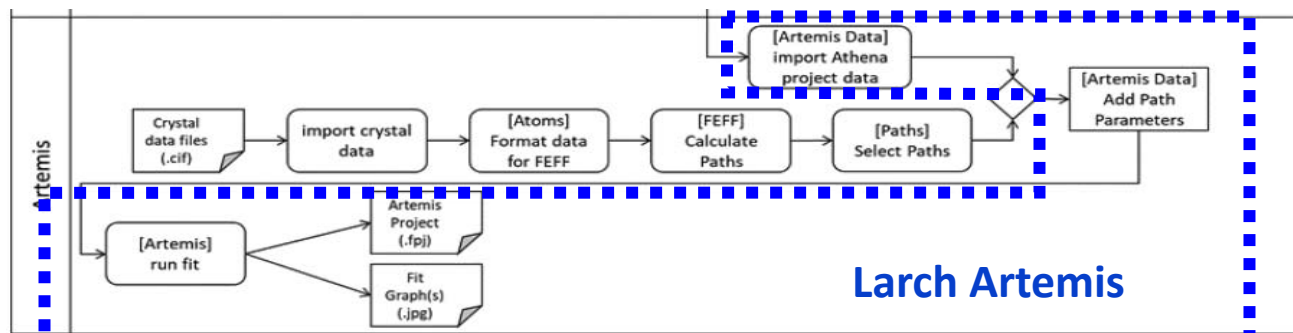
4: FEFF paths of 1564889.cif

3: CSV summary of 1564889.cif

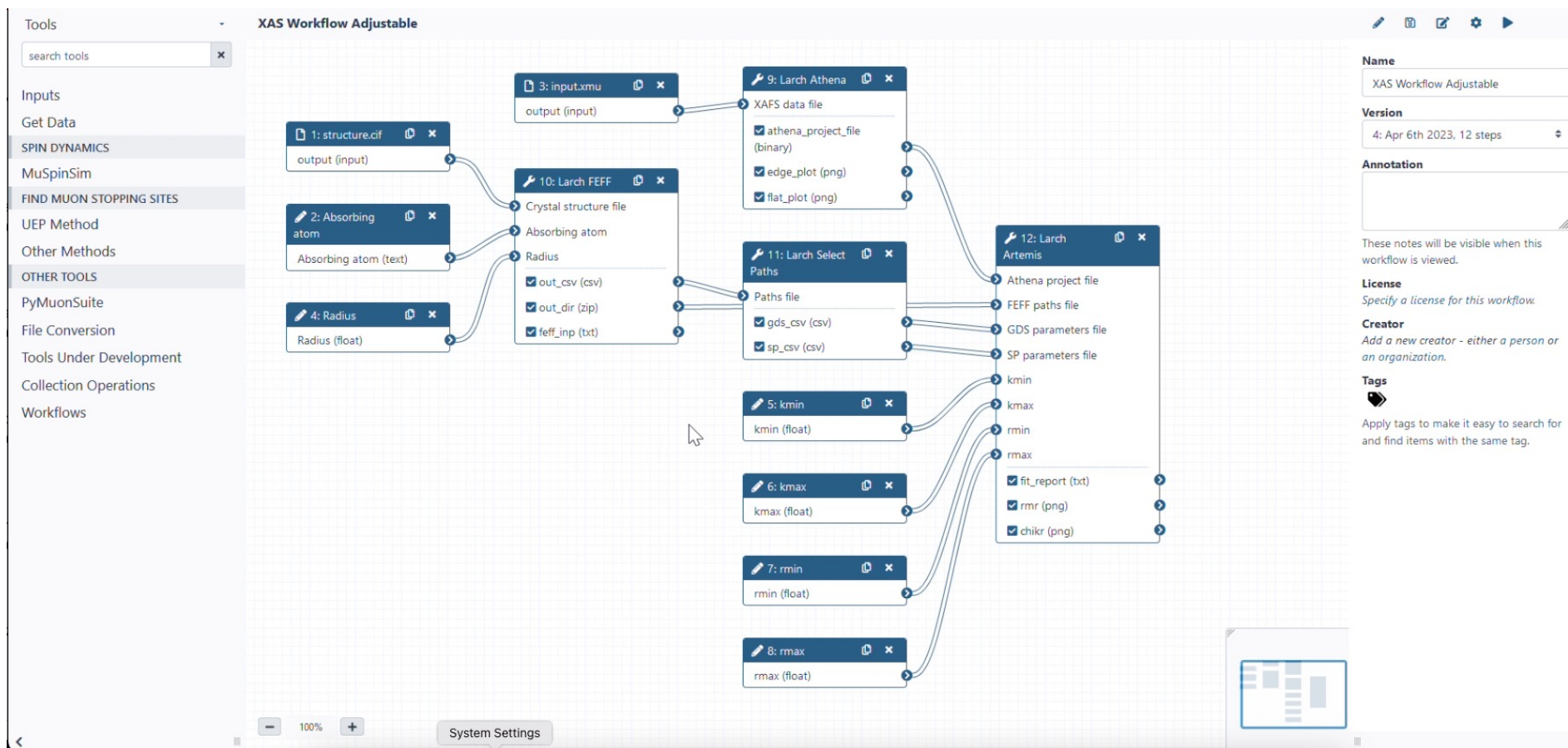
2: test.xmu

1: 1564889.cif

04:07 08:36



Saving Full EXAFS Workflow for FeS₂ in Galaxy



- The tasks that we included in Galaxy are:
 - Processing and Normalization of Raw Data
 - Extended XAFS Analysis of Data
- **Larch Athena:** Galaxy tool for Processing and Normalization of raw data.
- **Larch FEFF:** Galaxy tool for the generation of paths in a material using FEFF.
- **Larch Select Paths:** Galaxy tool for the selection of paths generated using **Larch FEFF**.
- **Larch Artemis:** Galaxy tool for extended analysis of EXAFS data.

Reproducing results

The highly surprising behaviour of diphosphine ligands in iron-catalysed Negishi cross-coupling

[Antonios M. Messinis](#), [Stephen L. J. Luckham](#), [Peter P. Wells](#), [Diego Gianolio](#), [Emma K. Gibson](#), [Harry M. O'Brien](#), [Hazel A. Sparkes](#), [Sean A. Davis](#), [June Callison](#), [David Elorriaga](#), [Oscar Hernandez-Fajardo](#) & [Robin B. Bedford](#) 

Nature Catalysis 2, 123–133 (2019) | [Cite this article](#)

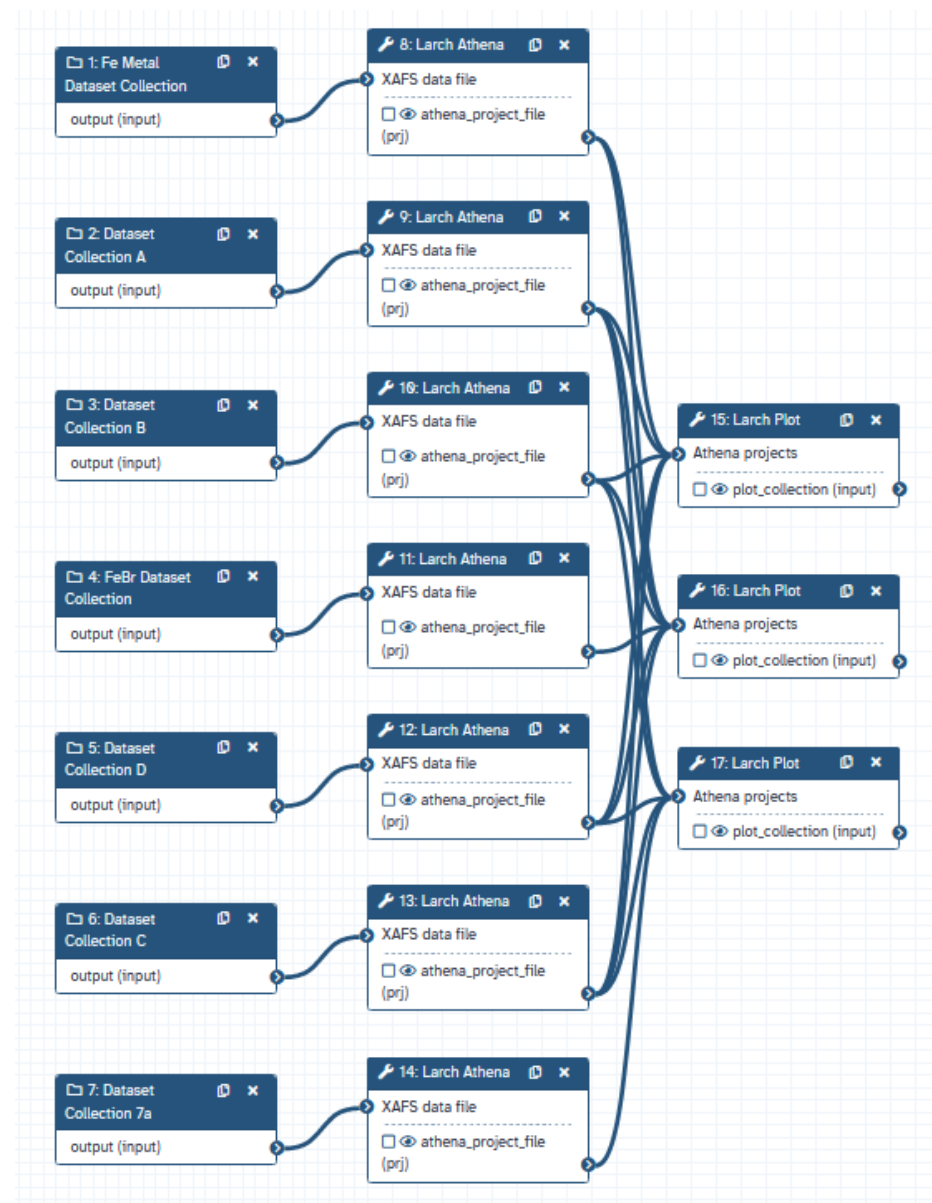
3395 Accesses | 26 Citations | 12 Altmetric | [Metrics](#)

Iron Negishi mechanistic NatCat 2018

Data supporting Nature Catalysis paper

[Complete download \(zip, 1.5 GiB\)](#)

Creator(s)	Robin Bedford, Antonios Messinis
Publication date	08 Jan 2019
Language	eng
Publisher	University of Bristol
Licence	Non-Commercial Government Licence for public sector information
DOI	10.5523/bris.1kp2f62x3klb02mfz2qymcmxmx
Citation	Robin Bedford, Antonios Messinis (2019): Iron Negishi mechanistic NatCat 2018. https://doi.org/10.5523/bris.1kp2f62x3klb02mfz2qymcmxmx
Total size	1.5 GiB



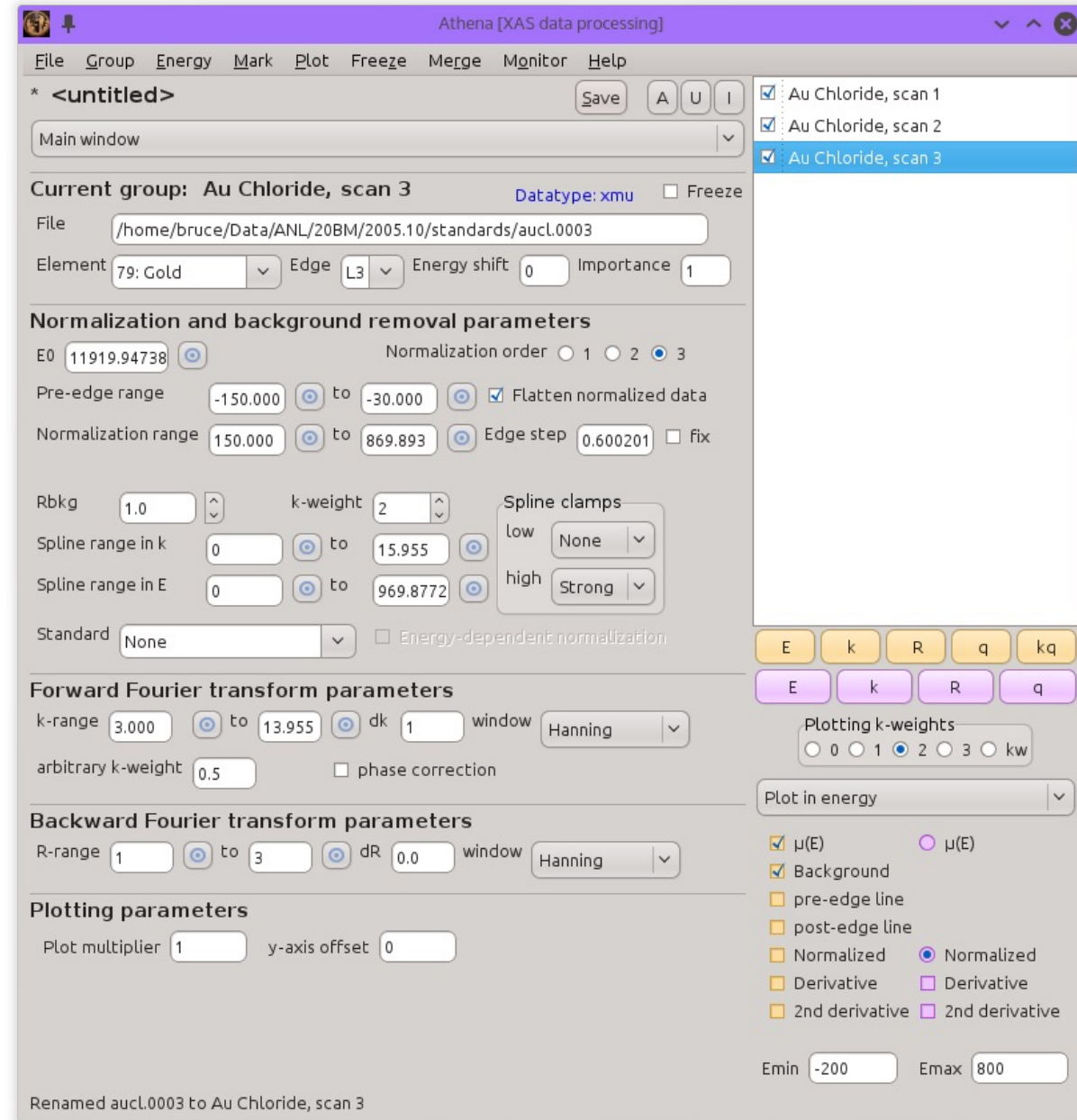
Benefits of Galaxy

- Embed parameters used in the calculation in workflow invocations
- Potential to integrate with existing or new resources:
 - Data storage at the synchrotron
 - Compute resources already managed by STFC/UKRI/PSDI
- Apply workflow to thousands of measurements of the same sample
 - ...without requiring scientists to write their own Perl scripts

Workflow Type	Average Processing Times			Estimated (3790 groups)
	Normalise	Fit	Per Group	
Manual Novice	3 minutes	21 minutes	24 minutes	~ 63 Days
Manual Expert	2 minutes	8 minutes	10 minutes	~ 26 Days
Scripted Demeter	7.68 Seconds	13.56 Seconds	21.24 Seconds	~ 23 Hours

Challenges

- Parameters sometimes missing from published papers
- Building on a Python library – not executables
 - Have to decide on tool scope and write own Python scripts to execute
- Large parameter space requiring expert knowledge
- Difficulty in embedding interactivity to allow “trial and error”
 - Should we be considering Jupyter/interactive tools?



Conclusions

Galaxy can reproduce the tasks associated to the:

- 1) Processing and Normalization of Raw Data
- 2) Extended XAFS Analysis of Data.

- We have created 5 Galaxy tools for this:
 - **Larch Athena:** Galaxy tool for Processing and Normalization of raw data.
 - **Larch FEFF:** Galaxy tool for the generation of paths in a material using FEFF.
 - **Larch Select Paths:** Galaxy tool for the selection of paths generated using **Larch FEFF**.
 - **Larch Artemis:** Galaxy tool for extended analysis of EXAFS data.
- The full workflow can be extracted from a history of individual tool executions and then be saved, reused and shared or exported as an RO-Crate object
- Galaxy provides a single interface for all tools, and can submit jobs to HPC resources without users needing to worry about the details (for example writing Slurm submission scripts)
- **We need to refine these tools by interacting with Catalysis scientists.**

The Galaxy Platform: Applications to Catalysis Workflows

Theoretical and Computational
Physics Group - SCD

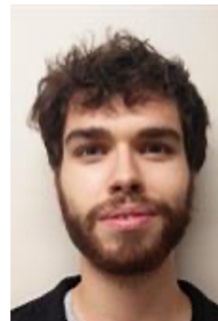


LIBORIO

Data & Software Engineering Group
SCD



GONZALEZ-
BELTRAN



AUSTIN

UK Catalysis Hub



NIEVA DE LA HIDALGA



PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Galaxy
PROJECT

al.c
ada lovelace centre



Software
Sustainability
Institute



eosc

EuroScienceGateway