

# Applications to Catalysis Workflows

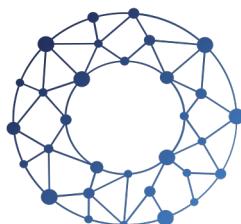
**Leandro Liborio<sup>1</sup>, Patrick Austin<sup>2</sup>, Alejandra González-Beltrán<sup>2</sup>,**  
**Abraham Nieva de la Hidalga<sup>3</sup>**

<sup>1</sup>Theoretical and Computational Physics Group, SCD, STFC

<sup>2</sup>Data and Software Engineering Group, SCD, STFC

<sup>3</sup>UK Catalysis Hub, Research Complex at Harwell

EGD meets ESG, October 5<sup>th</sup> 2023, Freiburg



**PSDI**  
PHYSICAL SCIENCES  
DATA INFRASTRUCTURE

  
ada lovelace centre





Software  
Sustainability  
Institute



 eosc

EuroScienceGateway

# Motivation

**EuroScience  
Gateway**

## WP5 Objectives:

*Develop customizable pilot workflows*

*Onboard new communities*

## Targeting Materials Science community

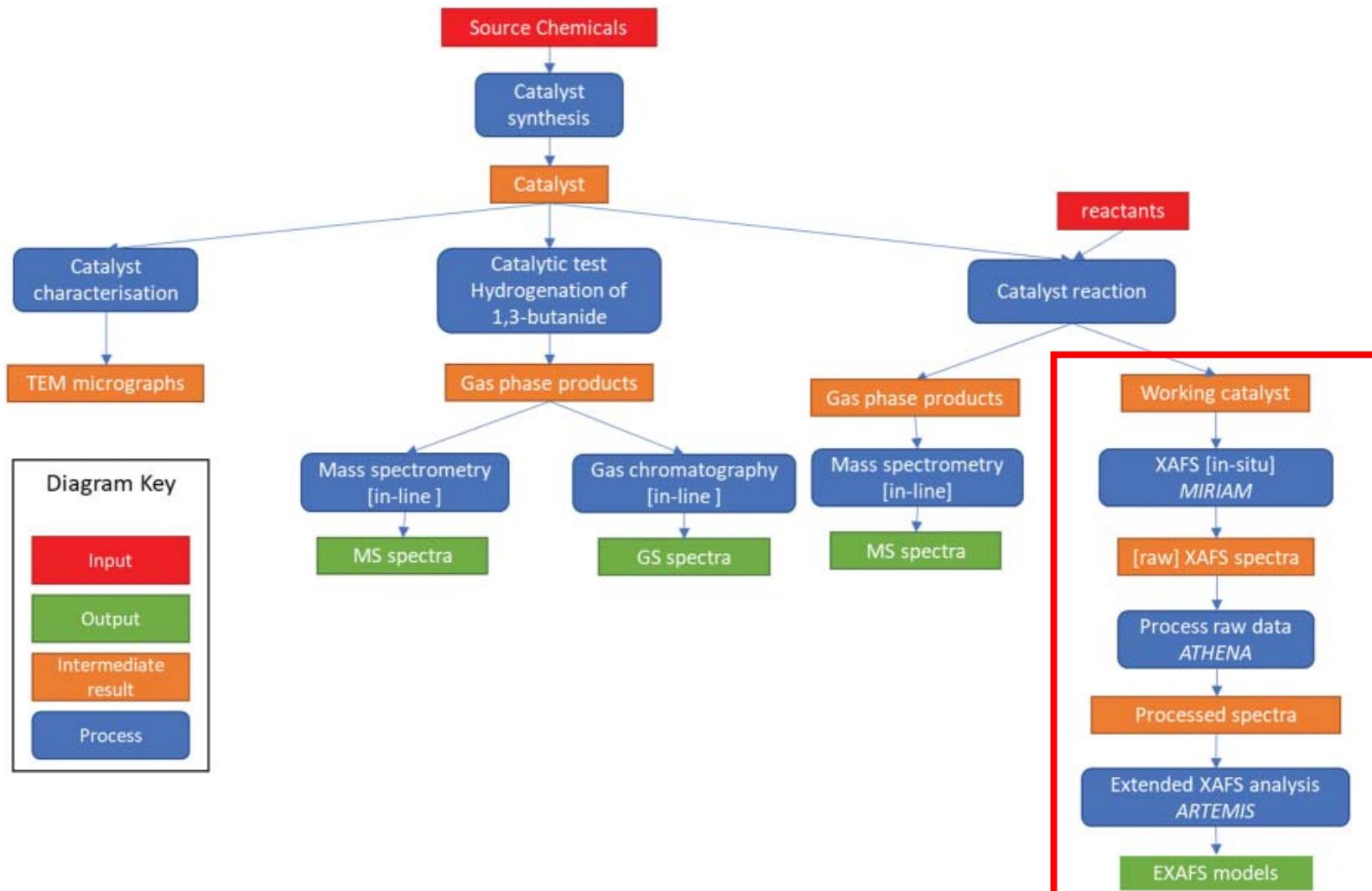
*Enable researchers in the physical sciences to handle data more easily*

## WP4, Pathfinder 1:

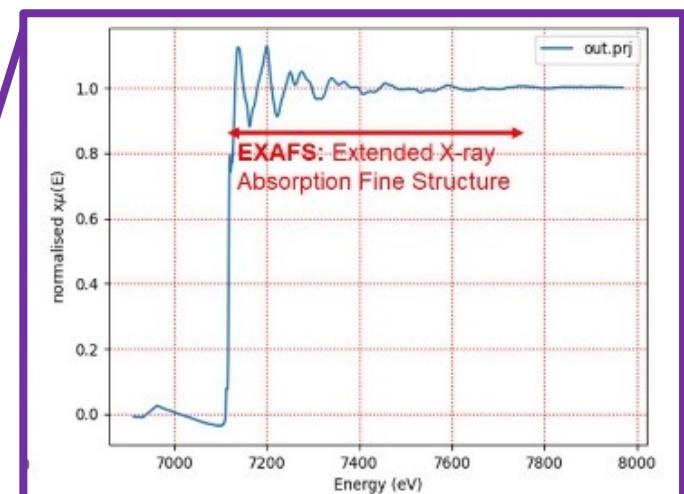
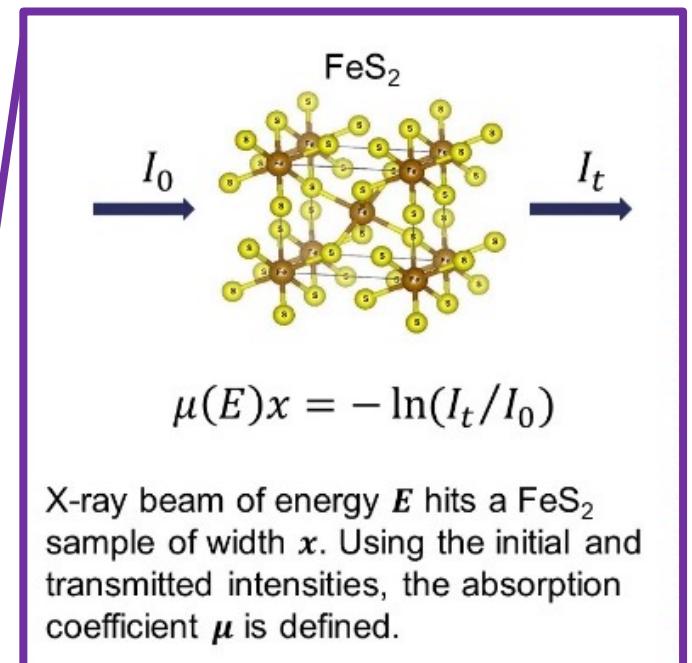
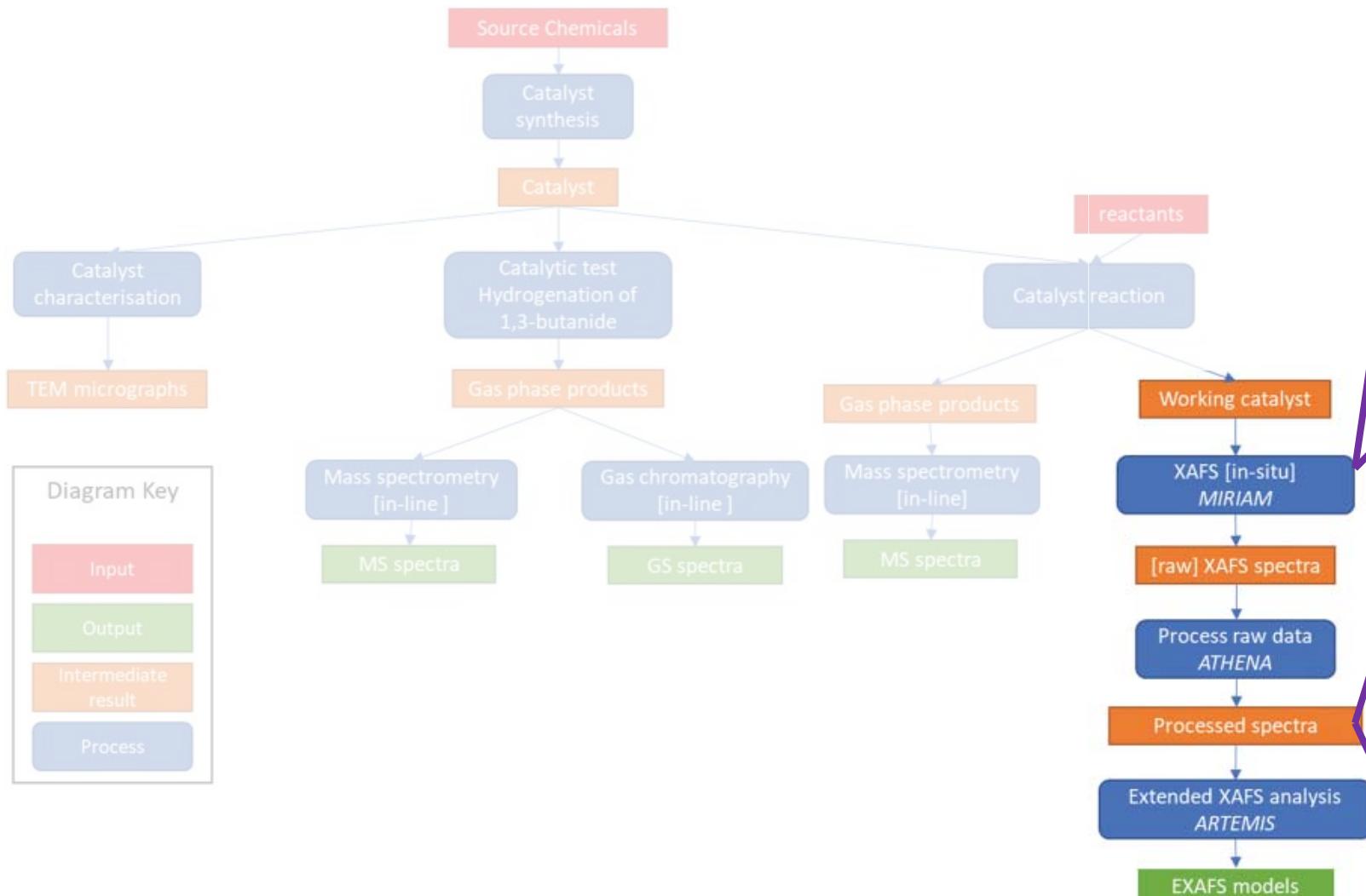
*...creating an integrated environment to support capture, analyse and reuse data. Prototype developed for XAS data*



# Catalysis use case

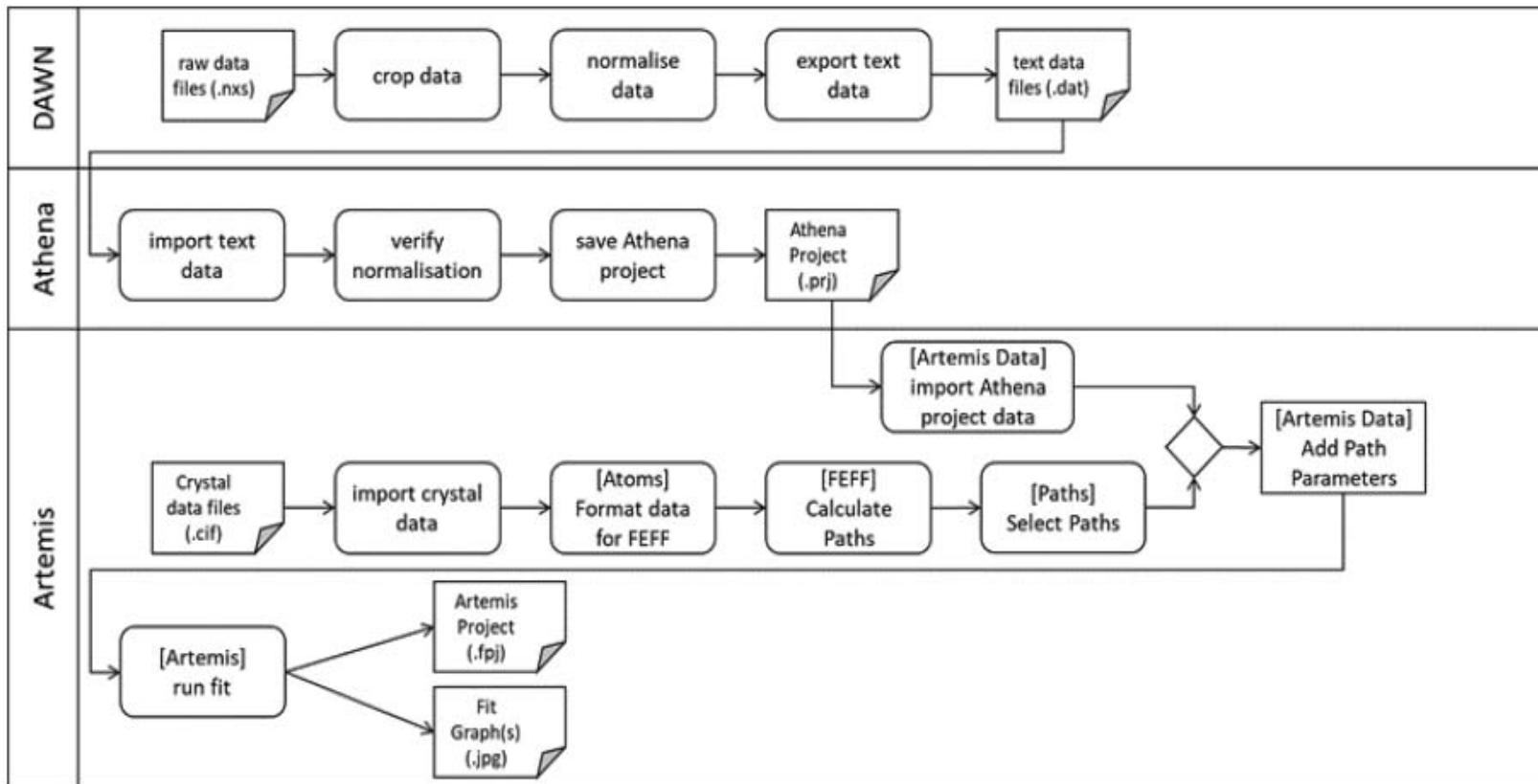


# Catalysis use case



Normalised variation of  $\mu$  with  $E$  for  $\text{FeS}_2$  using the pre/post edge fitting

# Current approach



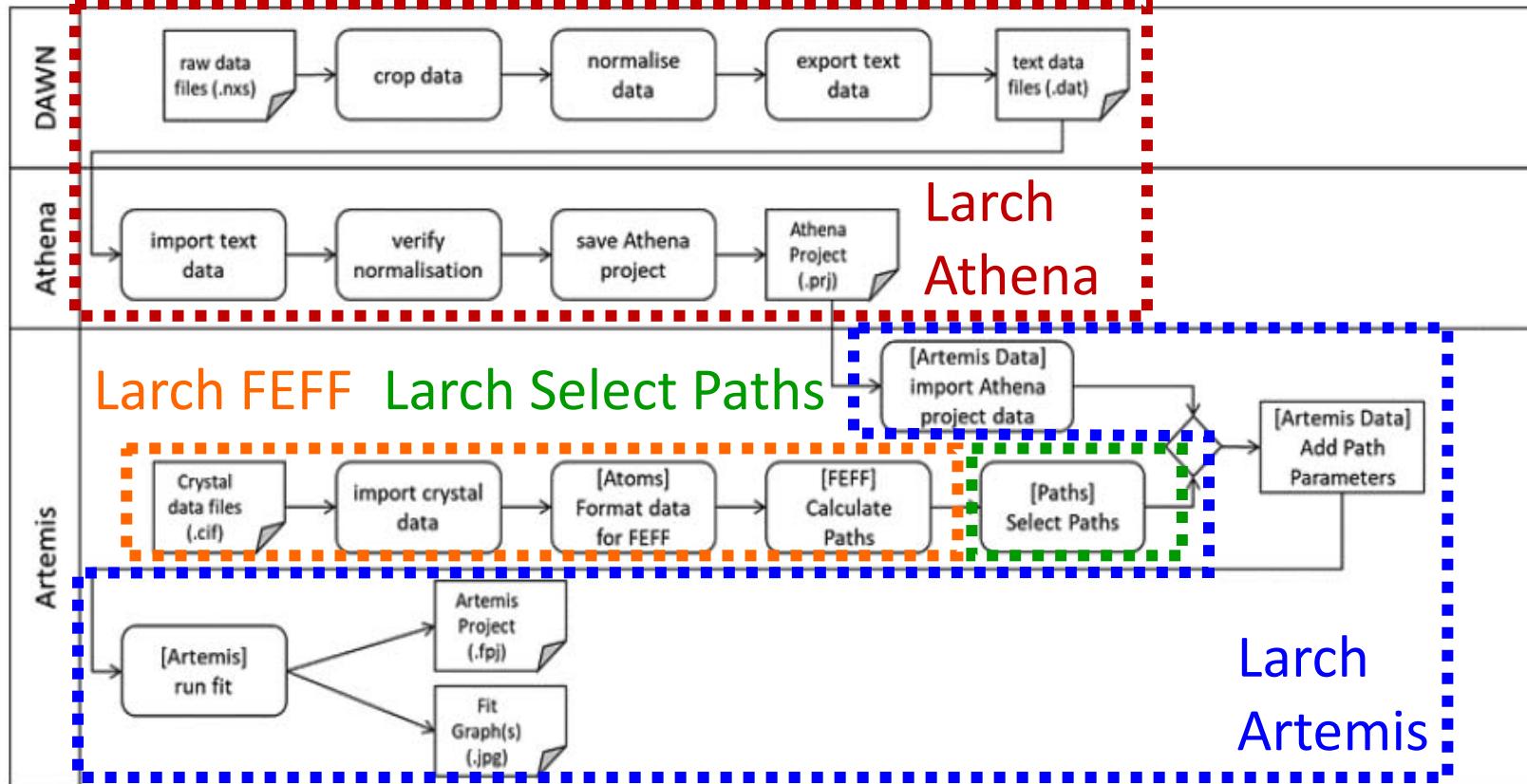
<https://bruceravel.github.io/demeter/>



Science and  
Technology  
Facilities Council



# 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/xraylight/>

# Process and Normalise XAS with Larch Athena in Galaxy

Galaxy µSR

Workflow Visualize Shared Data Help User Notifications Help Center

Using 44.1 MB

Tools

search tools

Upload Data

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

<https://muongalaxy-dev.esc.rl.ac.uk>

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

XAFS data file

5: test.xmu.txt

Normalised X-ray Absorption Fine Structure (XAFS) data, generated by DAWN.

Plot graph

No

Whether to plot the pre/post edge fitting and the normalised  $x\mu$  data.

Execute

Using Larch, create an Athena project file from the input X-ray Absorption Fine Structure (XAFS) data file. This renames the *xmu* column of the data to *mu*.

Optionally, plot the  $x\mu$  data along with pre and post edge fitting lines for visual inspection.

Citations:

- 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>
- Ravel, B., & Newville, M. (2005). A data analysis for X-ray absorption spectroscopy using FEFFIT. *Journal of Synchrotron Radiation*, 12(4), 537–541. <https://doi.org/10.1107/s0909049505012719>

Requirements:

- xraylarch (Version 0.9.66)
- xraydb (Version 4.4.7)
- sqlalchemy (Version 1.4.46)
- matplotlib (Version 3.5.2)

License:

MIT License

Creators:

Patrick Austin

History

search datasets

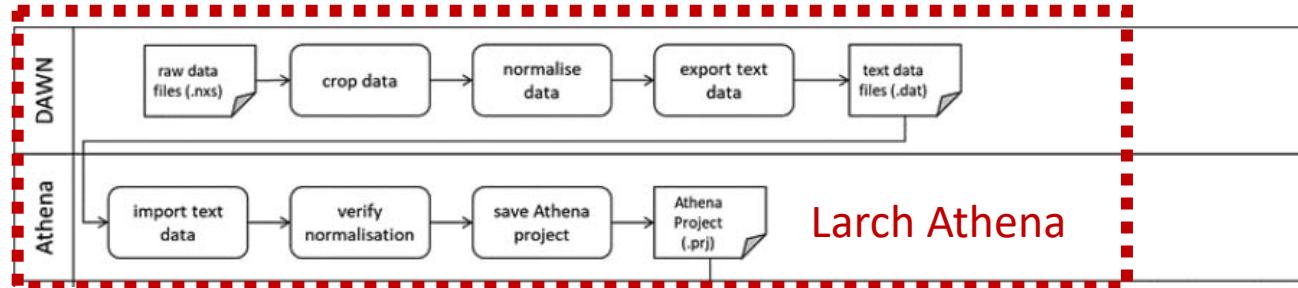
PSDI\_XAS

16.7 kB

This history is empty. You can load your own data or get data from an external source.



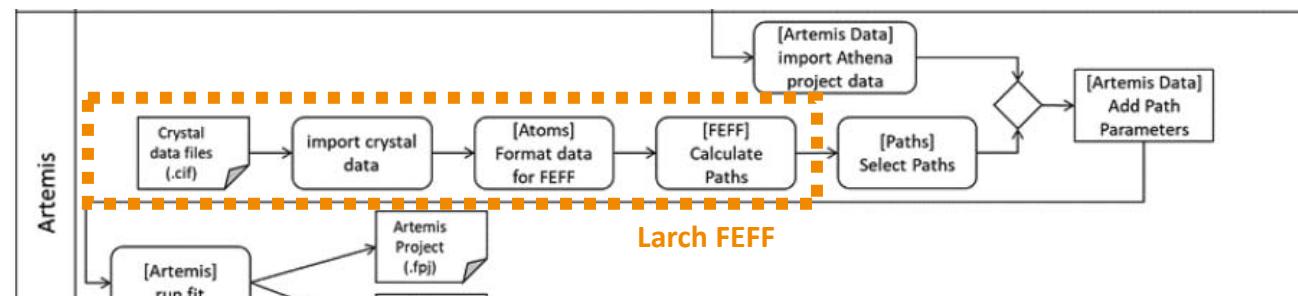
Science and Technology Facilities Council



# Path Generation in FeS<sub>2</sub> using Larch FEFF in Galaxy

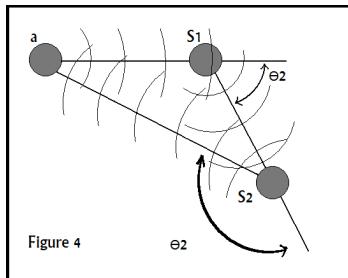
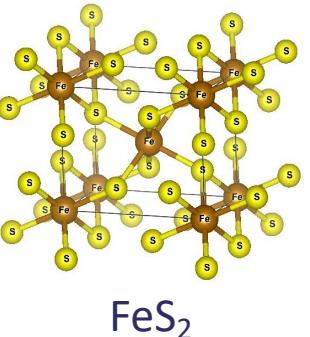
The screenshot shows the Galaxy web interface with a search bar at the top. Below it is a sidebar with a 'Tools' section containing a 'Get Data' dropdown and a 'Upload Data' button. The main area displays a table of data with columns for ID and values. To the right is a 'History' panel showing two datasets: 'test.xmu' and '1564889.cif'. A playback control bar at the bottom indicates the video is at 01:21 of 08:36.

ID	Value
7434.1940	0.89577218
7437.7090	0.89566937
7441.2440	0.89233424
7444.7980	0.89189983
7448.3710	0.88991984
7451.9620	0.89021992
7455.5740	0.88787942
7459.2040	0.88208656
7462.8530	0.87531841
7466.5210	0.87040248
7470.2080	0.86516762
7473.9150	0.85785178
7477.6400	0.85854938
7481.3850	0.84453901
7485.1480	0.83863281
7488.9310	0.83371678
7492.7330	0.82904991
7496.5540	0.82459446
7500.3940	0.82079844
7504.2530	0.81677576
7508.1310	0.81178911
7512.0280	0.80667248
7515.9440	0.80196858
7519.8790	0.79709432
7523.8340	0.79634294
7527.8070	0.78287089
7531.8000	0.77566164
7535.8120	0.77172412
7539.8420	0.76918155
7543.8920	0.76588805
7547.9610	0.76149338
7552.0490	0.75668380
7556.1560	0.75093523
7560.2820	0.74539768
7564.4270	0.74088376
7568.5910	0.73831944
7572.7740	0.73529922
7576.9770	0.73187551
7581.1980	0.72922448
7585.4390	0.72676436
7589.6980	0.72359612
7593.9770	0.71807714
7598.2750	0.71252104
7602.5920	0.70549579
7606.9280	0.69954396
7611.2830	0.69265829
7615.6570	0.68689223
7620.0500	0.68092659
7624.4620	0.67499882
7628.8940	0.66889046
7633.3440	0.66341014
7637.8130	0.65796296
7642.3020	0.65352747
7646.8100	0.64842668
7651.3360	0.64433864
7655.8820	0.64130624
7660.4470	0.63726602
7665.0300	0.63282645
7670.6110	0.62774440
7675.1910	0.62261100



Science and  
Technology  
Facilities Council

# Path Selection in FeS<sub>2</sub> using Larch Select Paths in Galaxy



Multiple scattering  
of emitted  
photoelectron



Science and  
Technology  
Facilities Council

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: Fe2  
TITLE space group: (Pnnm), space number: (56)  
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 0.378048  
\* 4 S 0.000000 0.800100 0.621960  
\* 5 S 0.500000 0.699900 0.121960  
\* 6 S 0.500000 0.300100 0.878048

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	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

Untitled history

1564889.cif

FEFF input of 1564889.ci

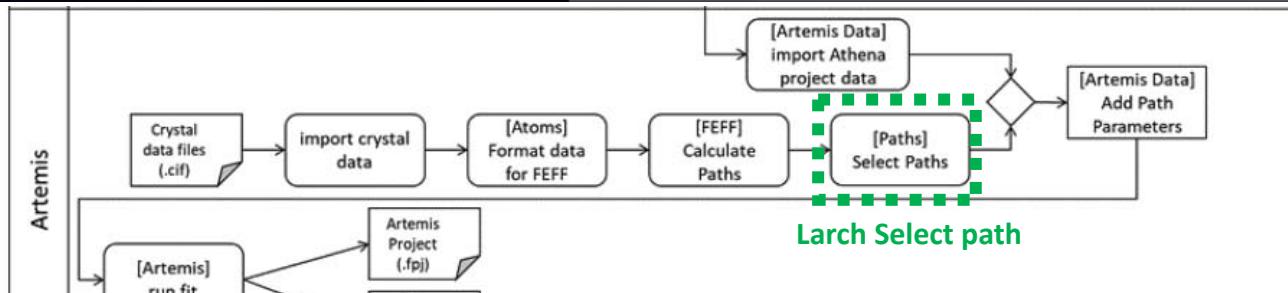
FEFF paths of 1564889.ci

CSV summary of 1564889.cif

test.xmu

1564889.cif

02:33 08:36

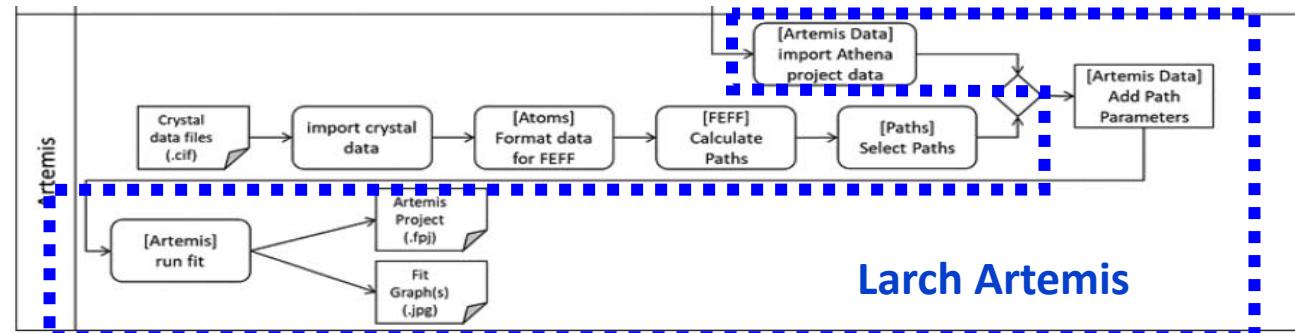


# Extended Analysis in FeS<sub>2</sub> using Larch Artemis in Galaxy

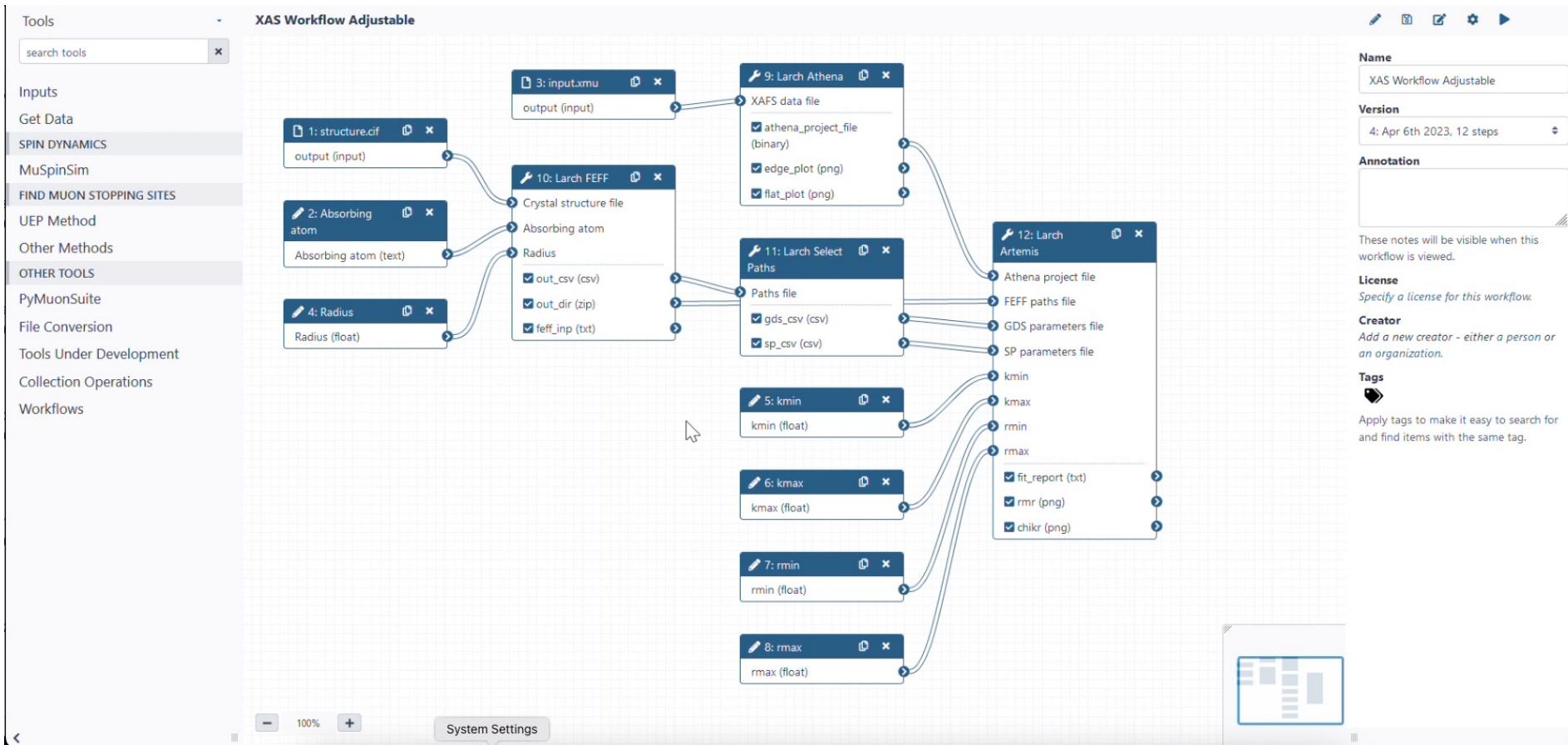
The screenshot shows the Galaxy web interface with the Larch Artemis tool selected. The left sidebar lists various tools under 'OTHER TOOLS' such as PyMuonSuite, File Conversion, and Tools Under Development. The main panel displays the configuration for the Larch Artemis tool, including execution mode (Parallel), Athena project file (8: Athena project of test.xmu), FEFF paths file (4: FEFF paths of 1564889.cif), GDS parameters file (7: Selected paths of CSV summary of 1564889.cif), SP parameters file (7: Selected paths of CSV summary of 1564889.cif), and Fitting Variables. A 'Plot graph' section is also present. At the bottom, there is a 'Citations' section with a link to a journal article and a media player bar. The right sidebar shows the 'History' panel with a list of dataset entries.



Science and  
Technology  
Facilities Council



# Saving Full EXAFS Workflow for FeS<sub>2</sub> 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

Antonis 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](#)

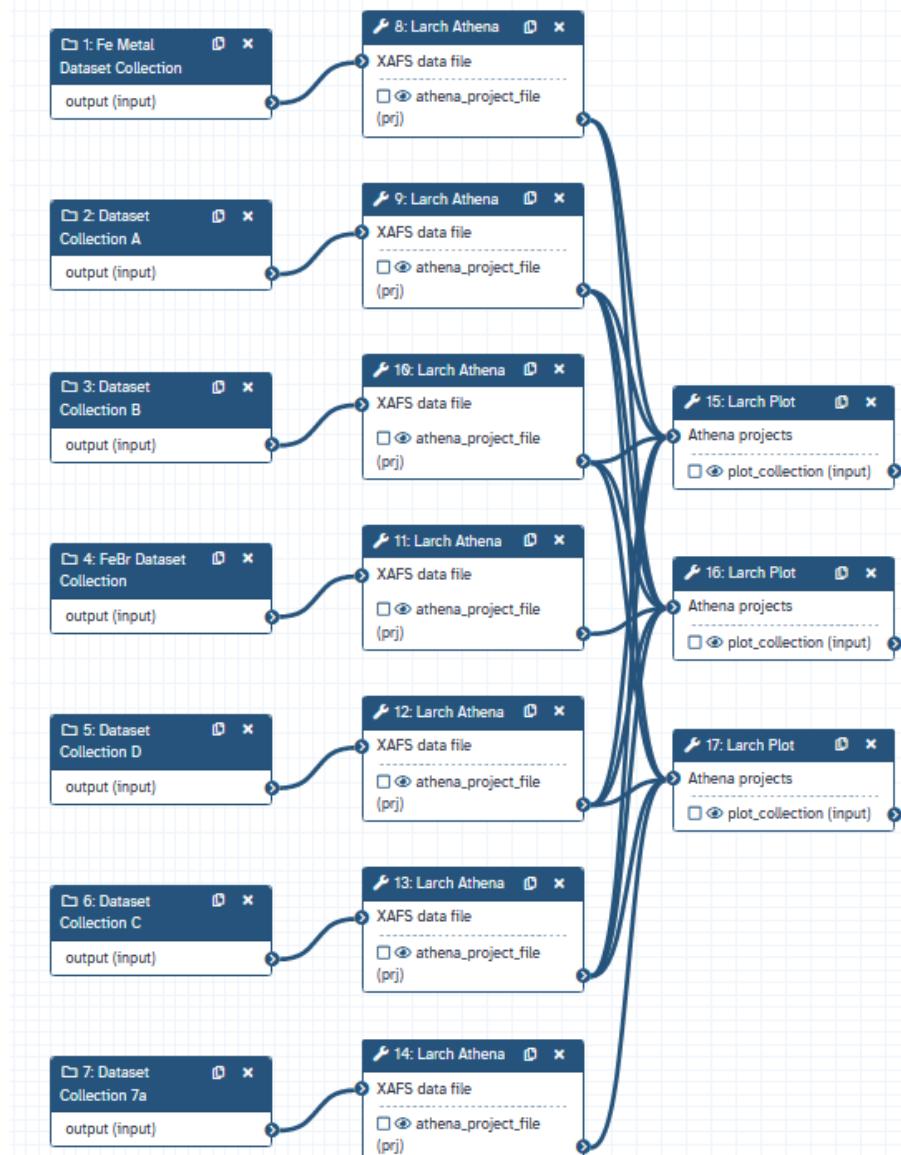
3395 Accesses | 26 Citations | 12 Altmetric | [Metrics](#)

## 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.1kp2f62x3klb02mfz2qymcmxm
Citation	Robin Bedford, Antonios Messinis (2019): Iron Negishi mechanistic NatCat 2018. <a href="https://doi.org/10.5523/bris.1kp2f62x3klb02mfz2qymcmxm">https://doi.org/10.5523/bris.1kp2f62x3klb02mfz2qymcmxm</a>
Total size	1.5 GiB



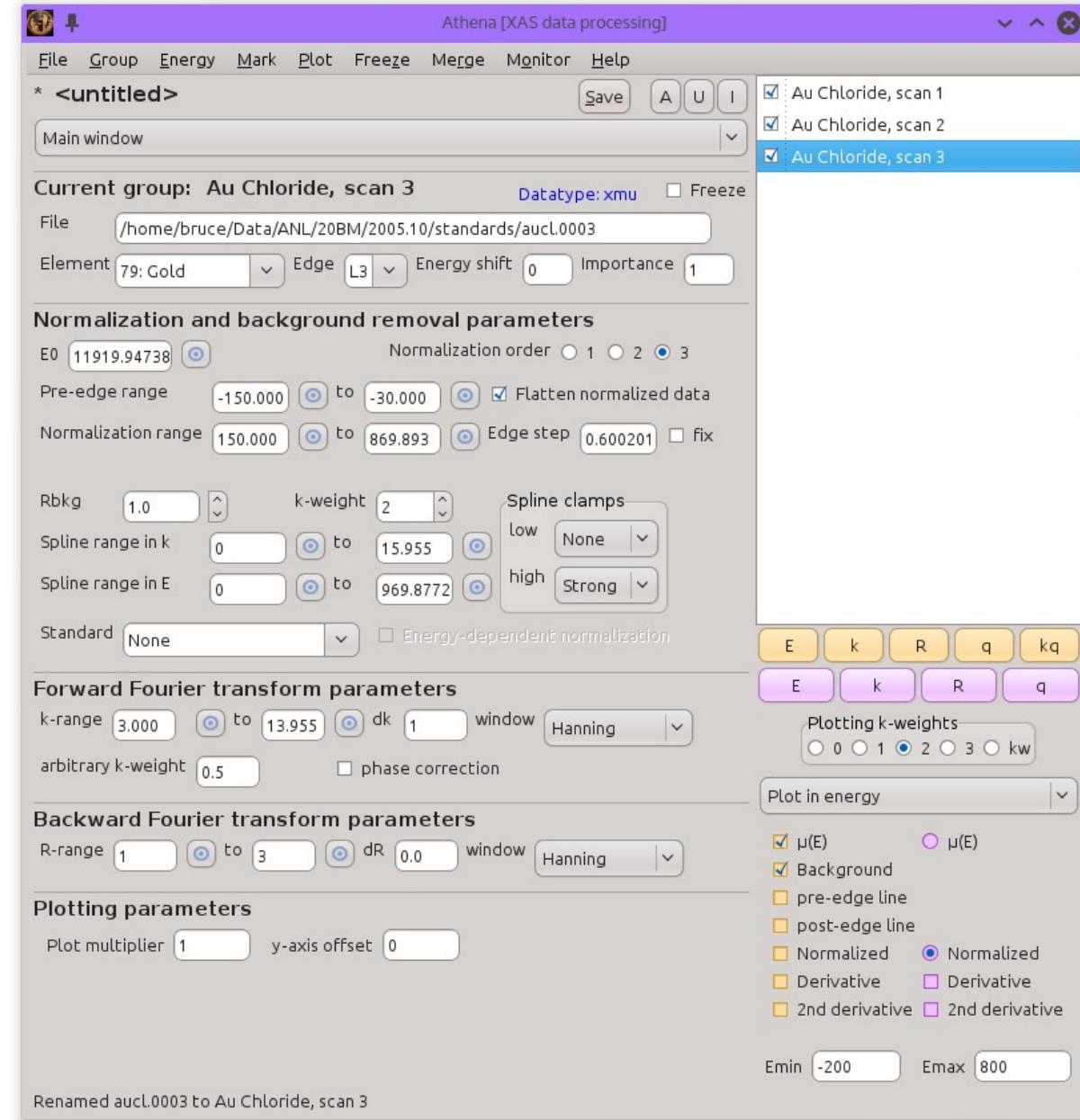
# Benefits of Galaxy

- 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



NIEVA DE LA HIDALGA



**PSDI**

PHYSICAL SCIENCES  
DATA INFRASTRUCTURE

= **Galaxy**  
PROJECT

**alc**  
ada lovelace centre



Software  
Sustainability  
Institute



Sci  
Tech  
Faci

**eosc** | EuroScienceGateway