Applications to Catalysis Workflows

Leandro Liborio[1], Patrick Austin[2], Alejandra González-Beltrán[2] ,
Abraham Nieva de la Hidalga[3]

[1]Theoretical and Computational Physics Group, SCD, STFC

[2]Data and Software Engineering Group, SCD, STFC

[3]UK Catalysis Hub, Research Complex at Harwell

EGD meets ESG, October 5th 2023, Freiburg

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

ada lovelace centre

CatalysisHub

Software
Sustainability
Institute

Galaxy
EUROPE

eosc | EuroScienceGateway

Motivation

EuroScience Gateway

WP5 Objectives:

Develop customizable pilot workflows

Onboard new communities

Targeting Materials Science community

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

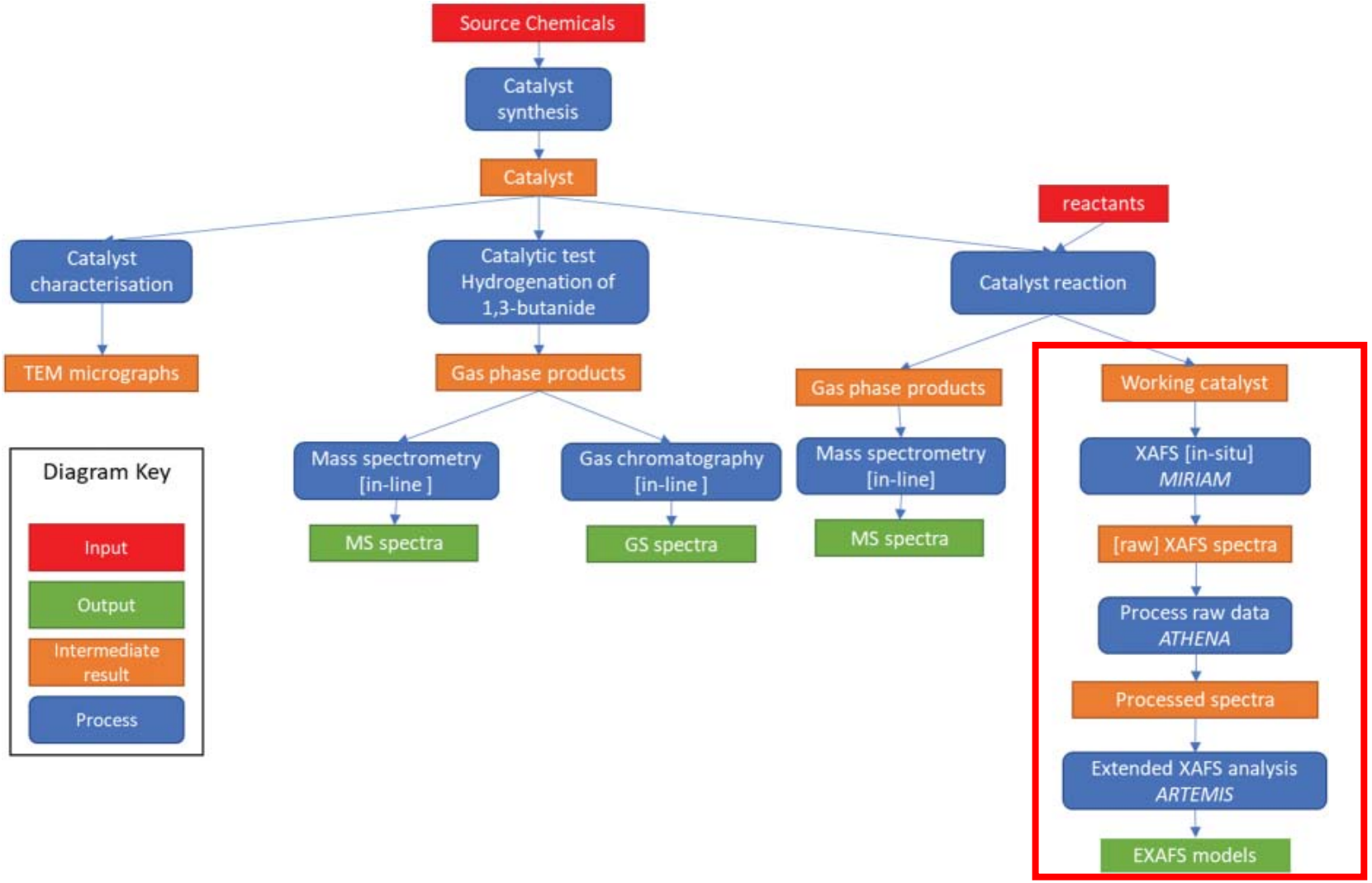
Enable researchers in the physical sciences to handle data more easily

WP4, Pathfinder 1:

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

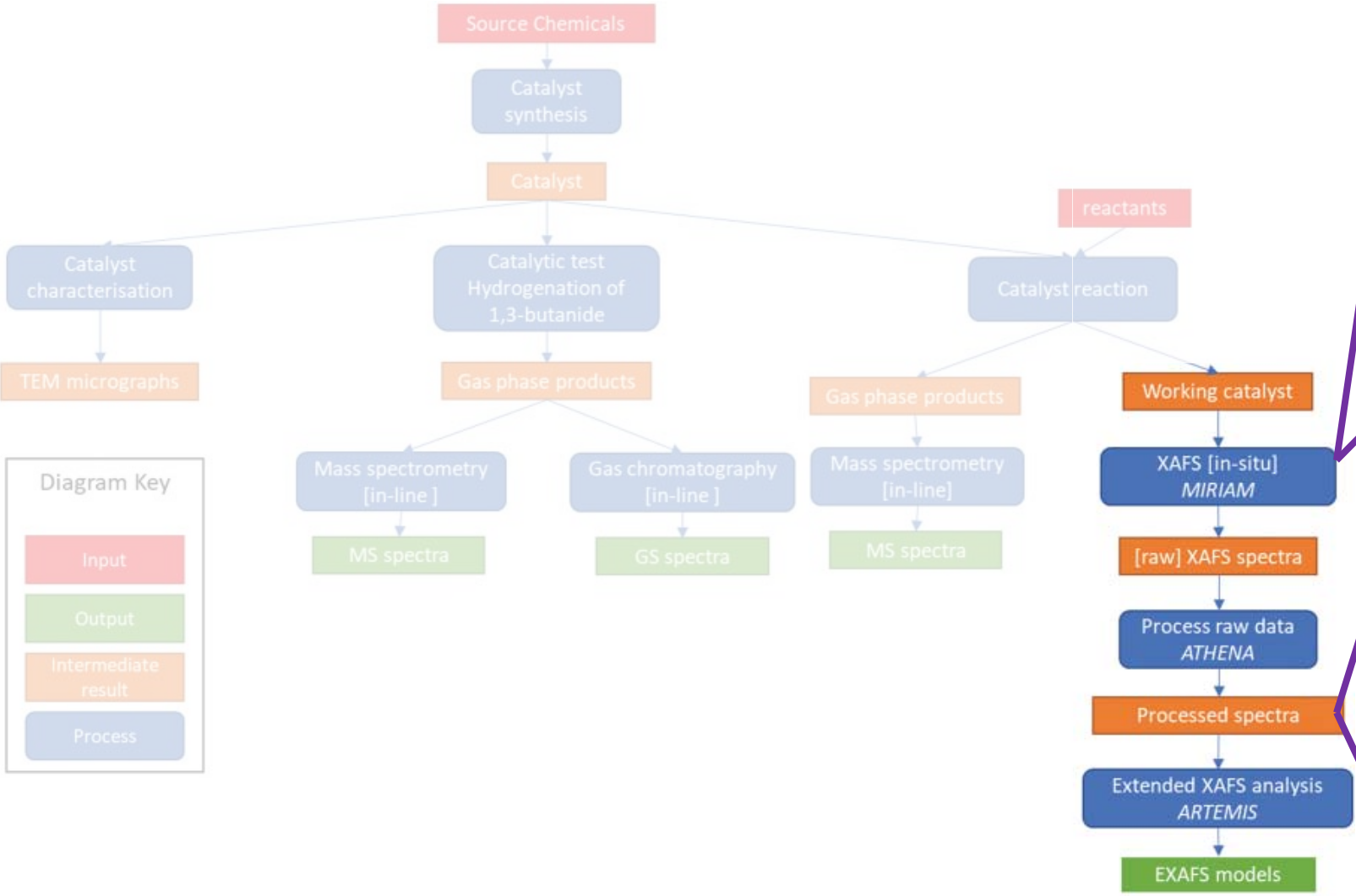
UKRI Science and Technology Facilities Council

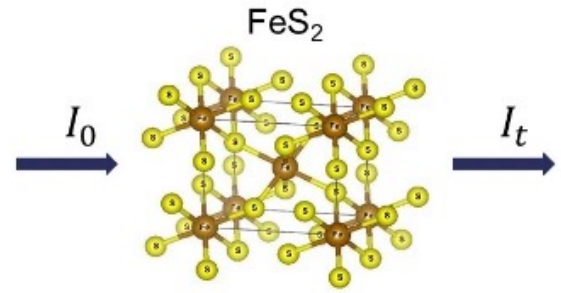
Catalysis use case



Nieva de la Hidalga, A, et al.: A workflow demonstrator for processing catalysis research data. Data Intelligence 4(2), 455-470 (2022). doi: 10.1162/dint_a_00143

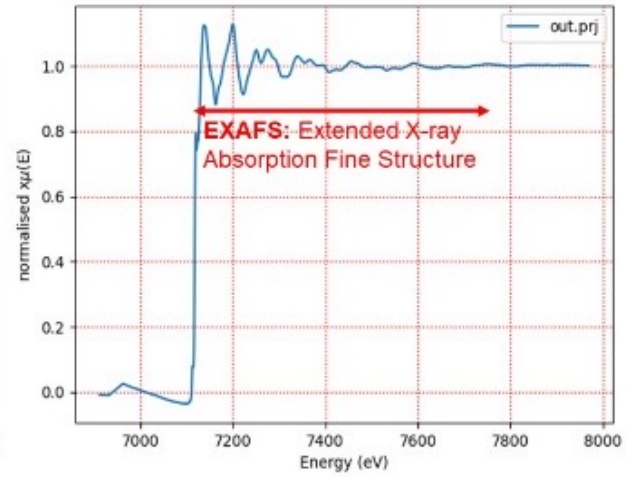
Catalysis use case





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

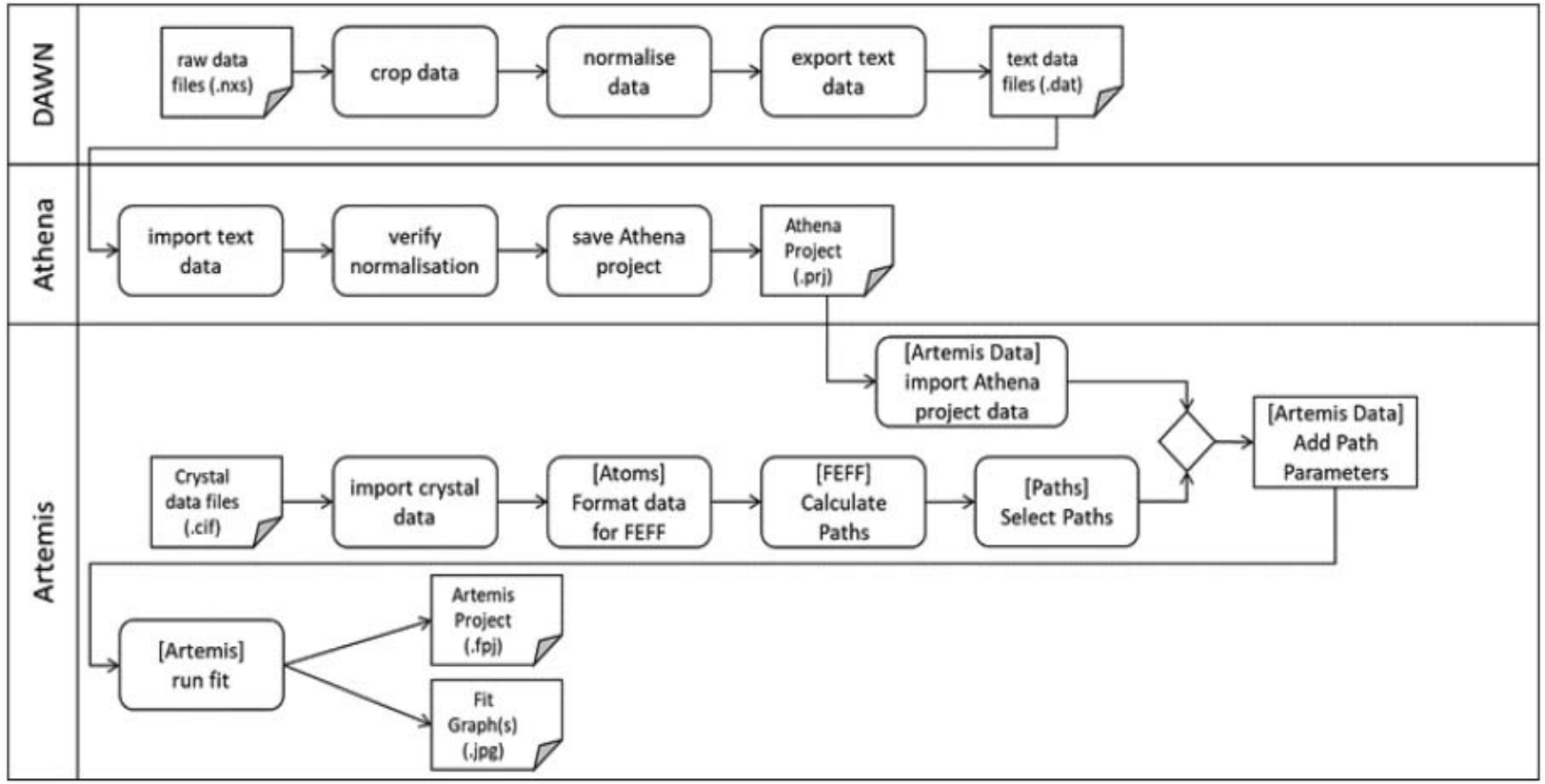
X-ray beam of energy E hits a FeS_2 sample of width x. Using the initial and transmitted intensities, the absorption coefficient μ is defined.

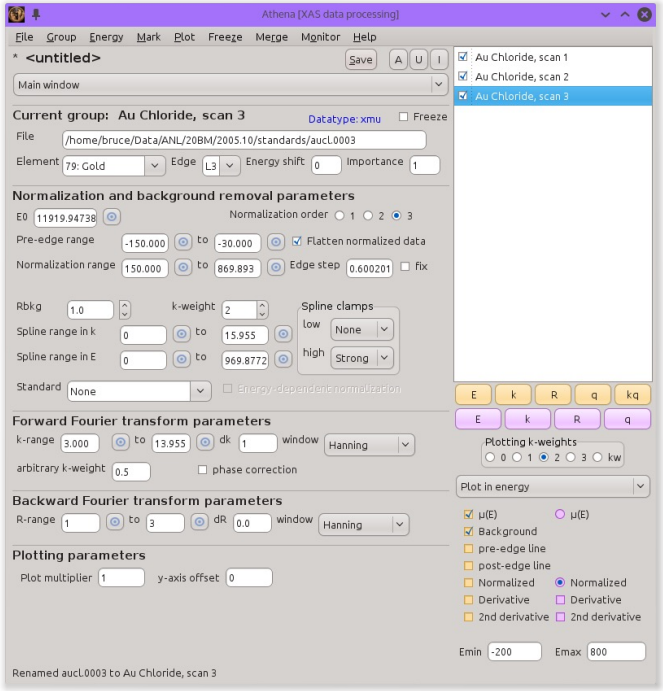


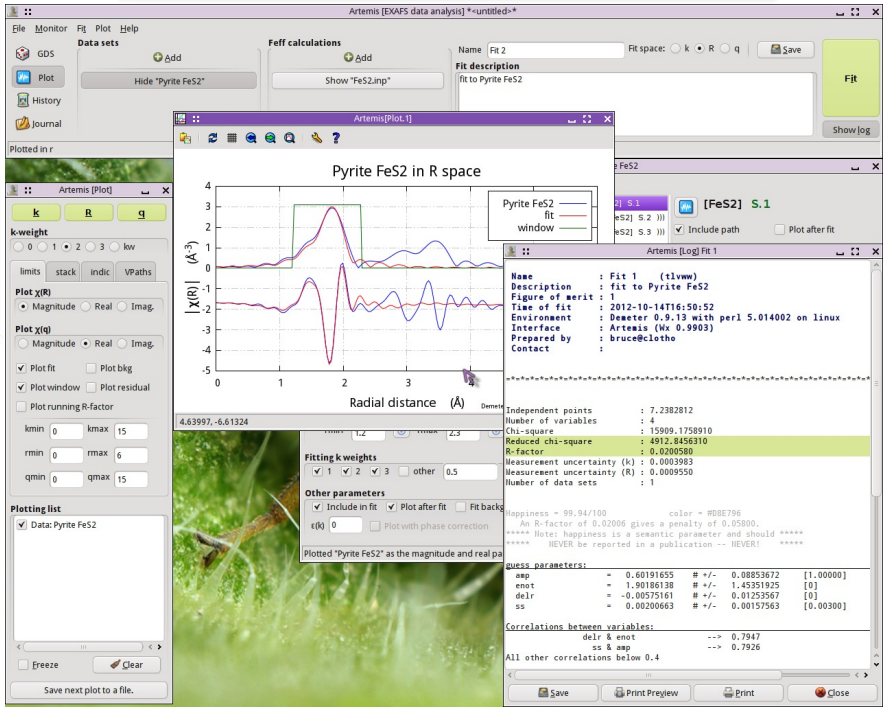
Normalised variation of μ with E for FeS_2 using the pre/post edge fitting

Nieva de la Hidalga, A, et al.: A workflow demonstrator for processing catalysis research data. Data Intelligence 4(2), 455-470 (2022). doi: 10.1162/dint_a_00143

Current approach



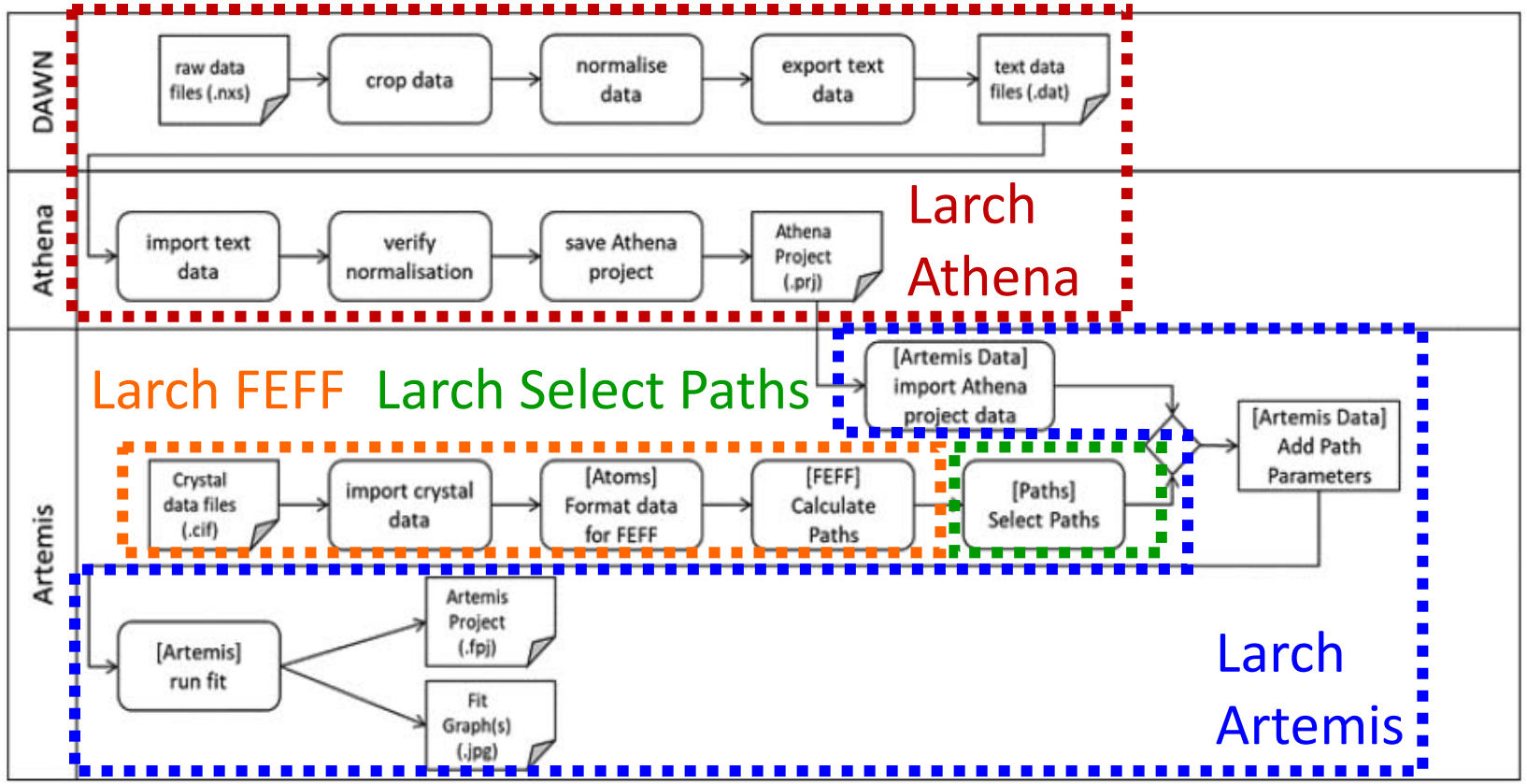




https://bruceravel.github.io/demeter/

Science and Technology Facilities Council

XAS in Galaxy



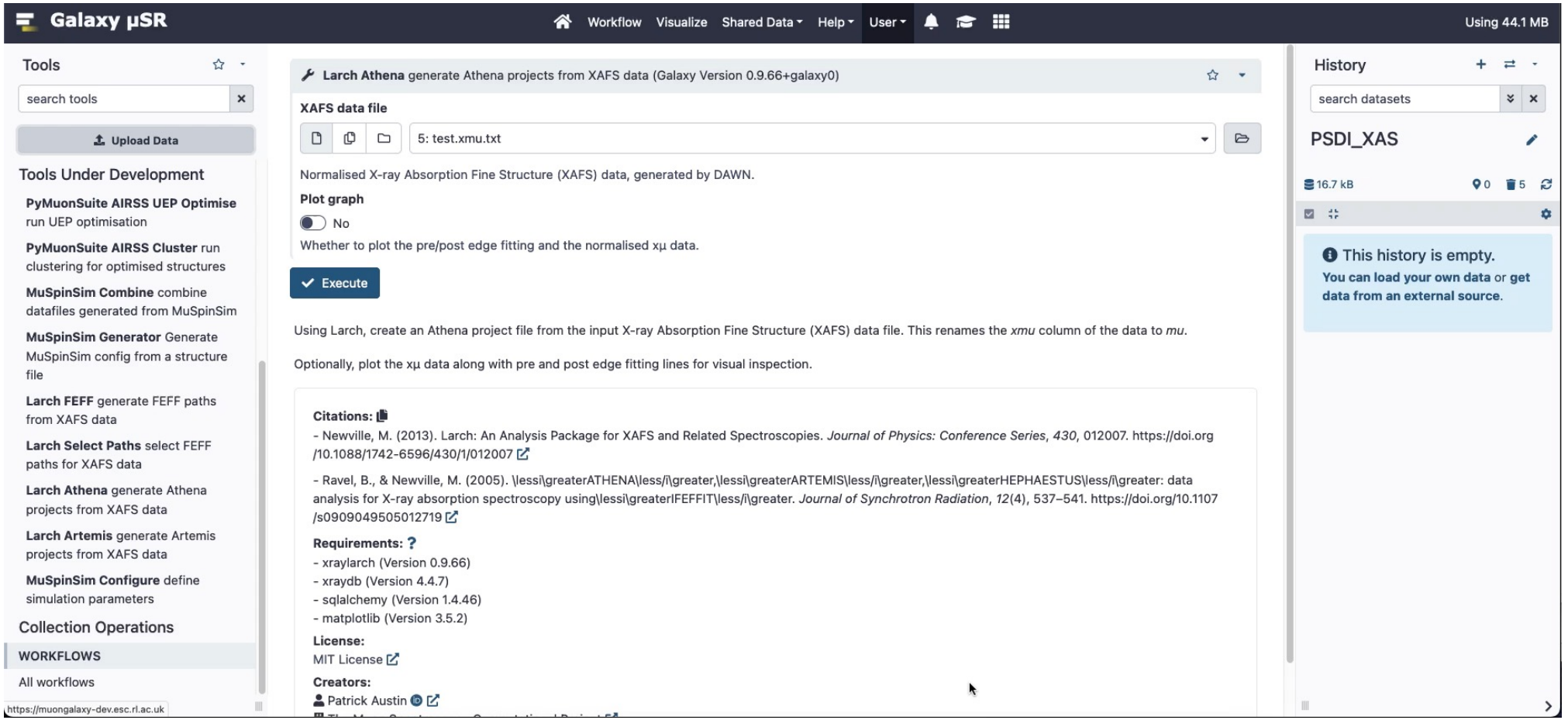
https://xraypy.github.io/xraylarch/

- 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

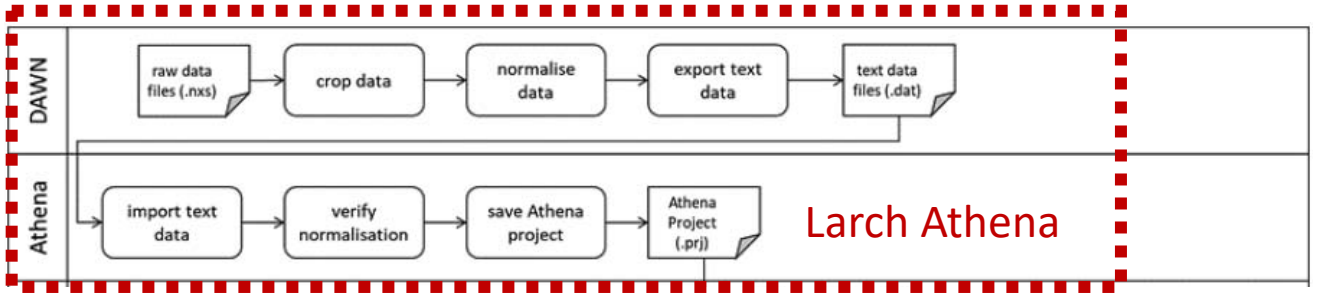
Also implementing utility tools and tools for other XAS techniques (not shown on diagram).

Science and Technology Facilities Council

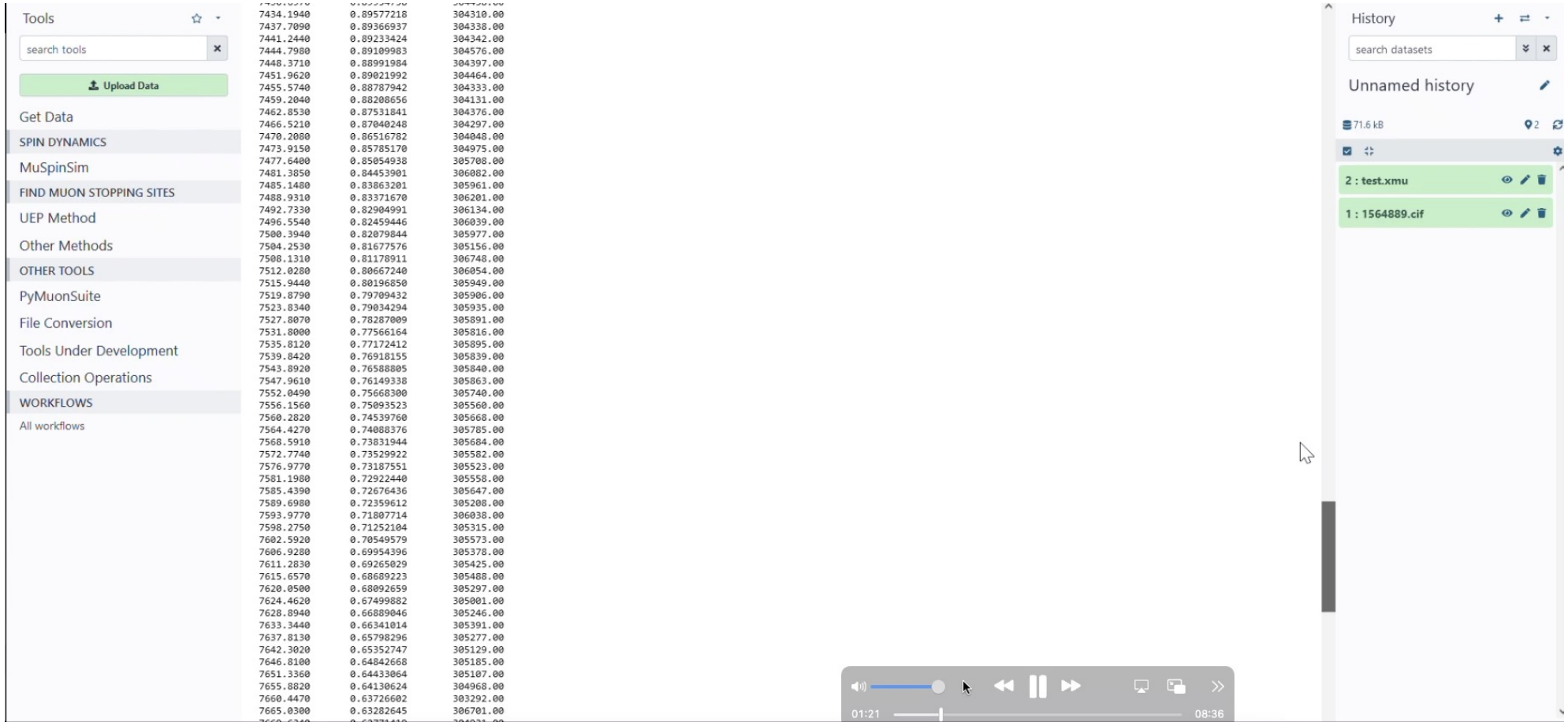
Process and Normalise XAS with Larch Athena in Galaxy

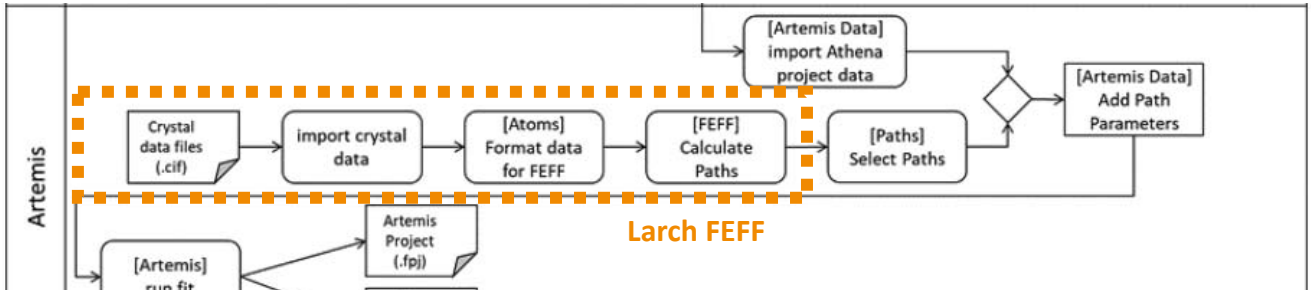


Science and Technology Facilities Council



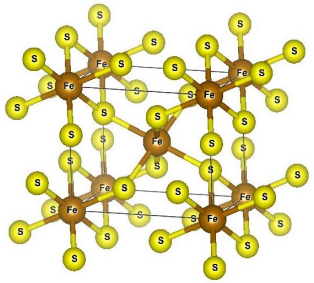
Path Generation in FeS_2 using Larch FEFF in Galaxy



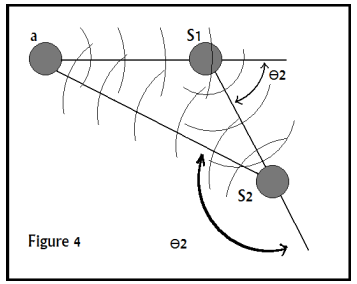


Science and Technology Facilities Council

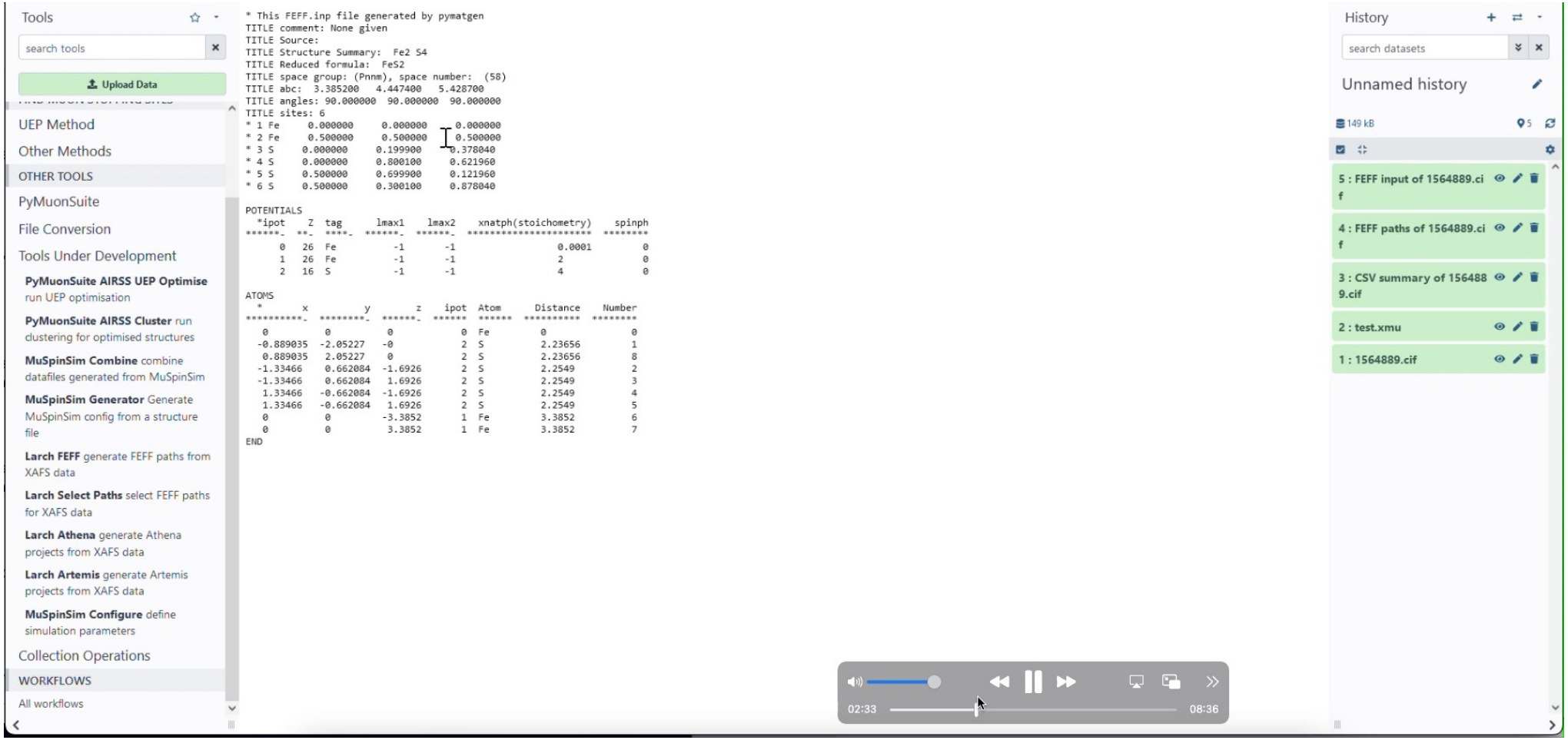
Path Selection in FeS2 using Larch Select Paths in Galaxy

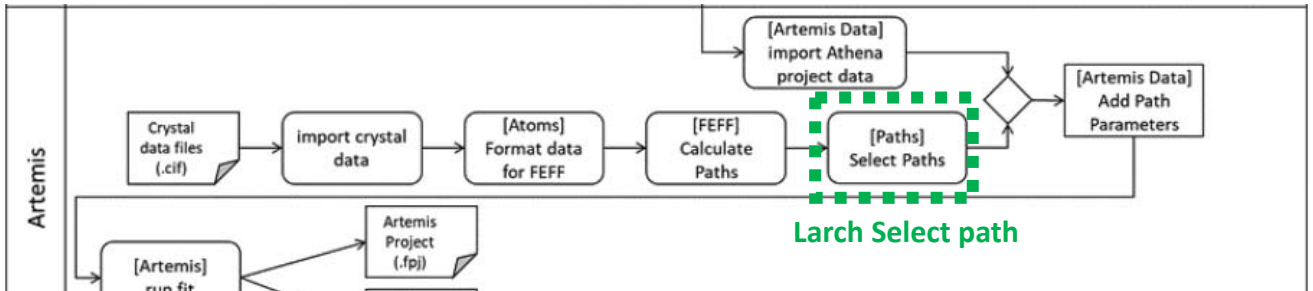


FeS_2



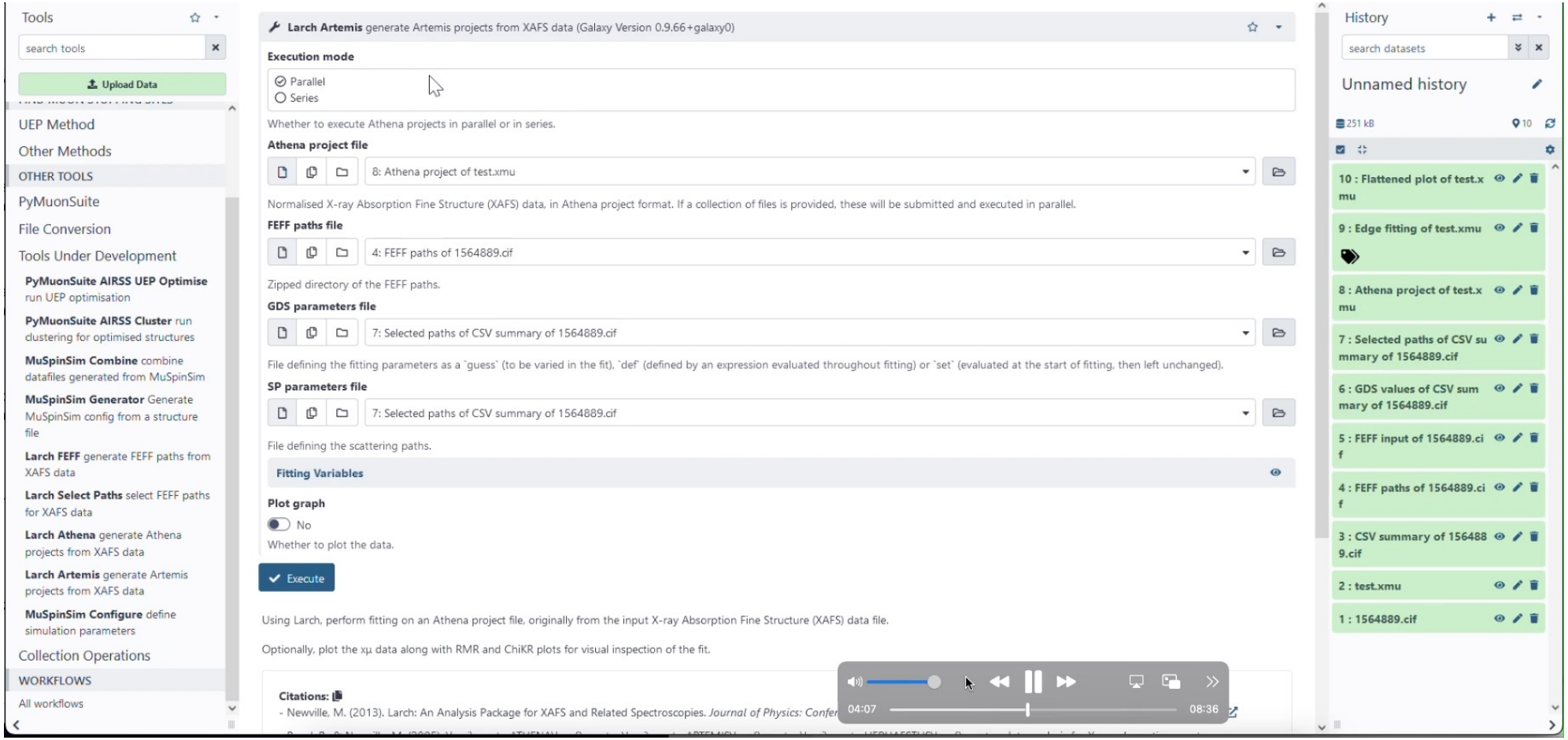
Multiple scattering of emitted photoelectron

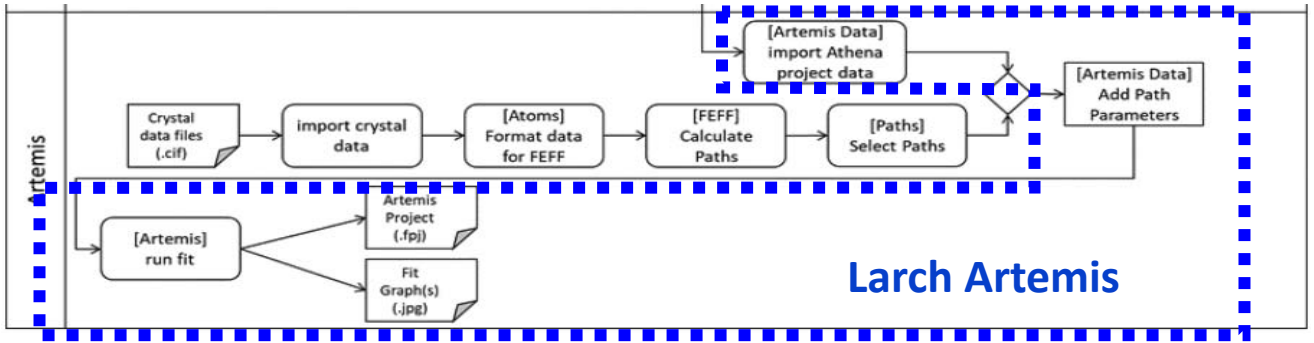




Science and Technology Facilities Council

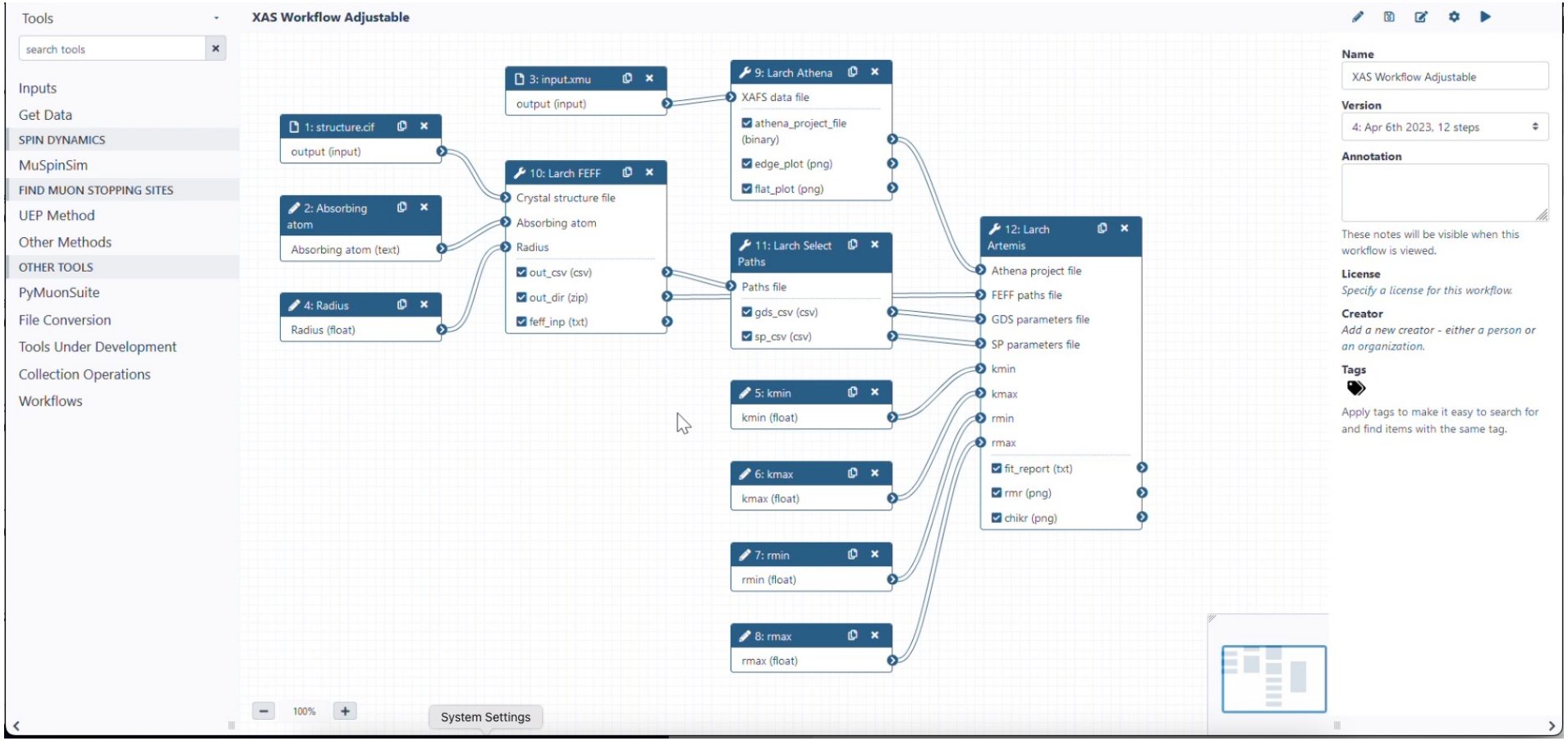
Extended Analysis in FeS_2 using Larch Artemis in Galaxy





Science and Technology Facilities Council

Saving Full EXAFS Workflow for FeS_2 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.

Science and Technology Facilities Council

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

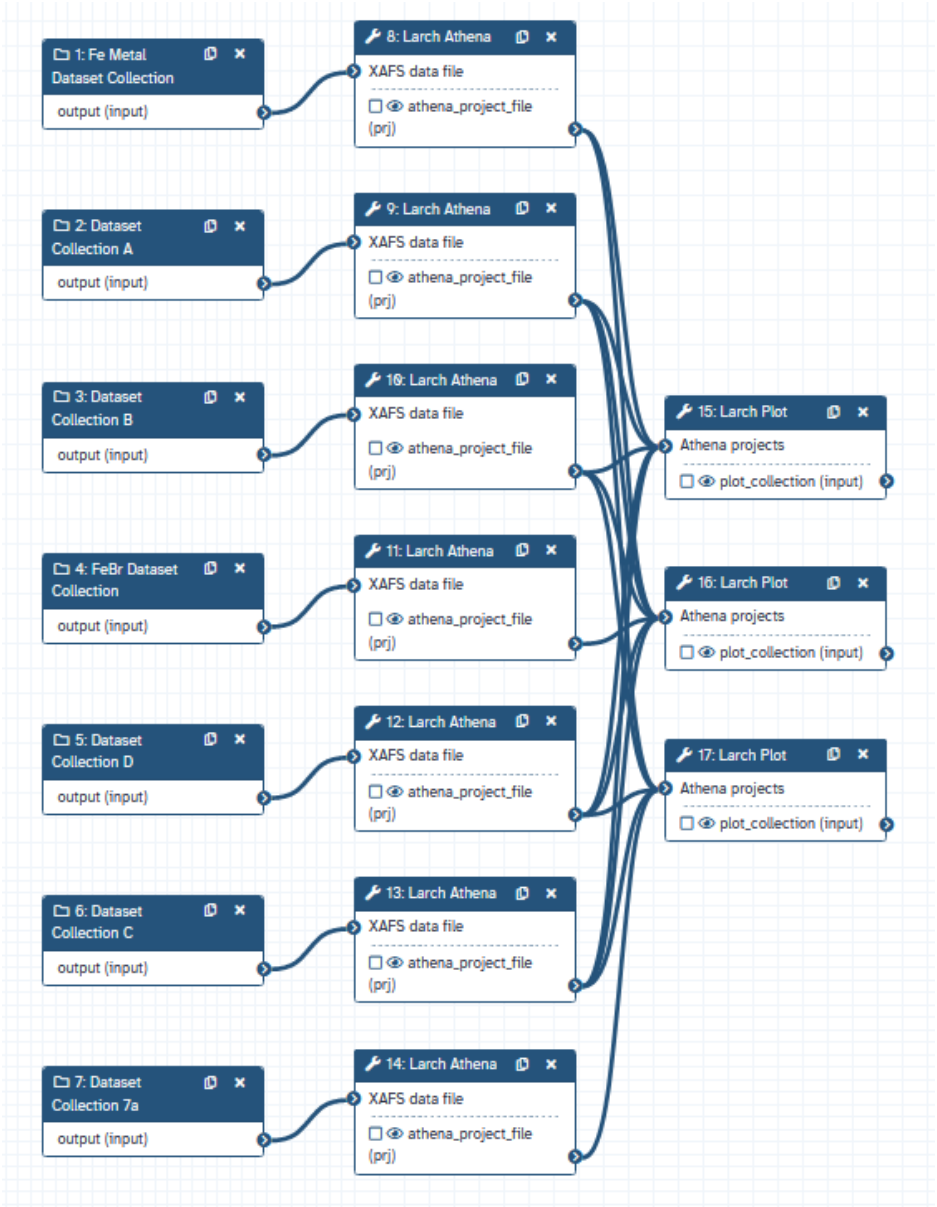
Nature Catalysis **2**, 123–133 (2019) | Cite this article

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



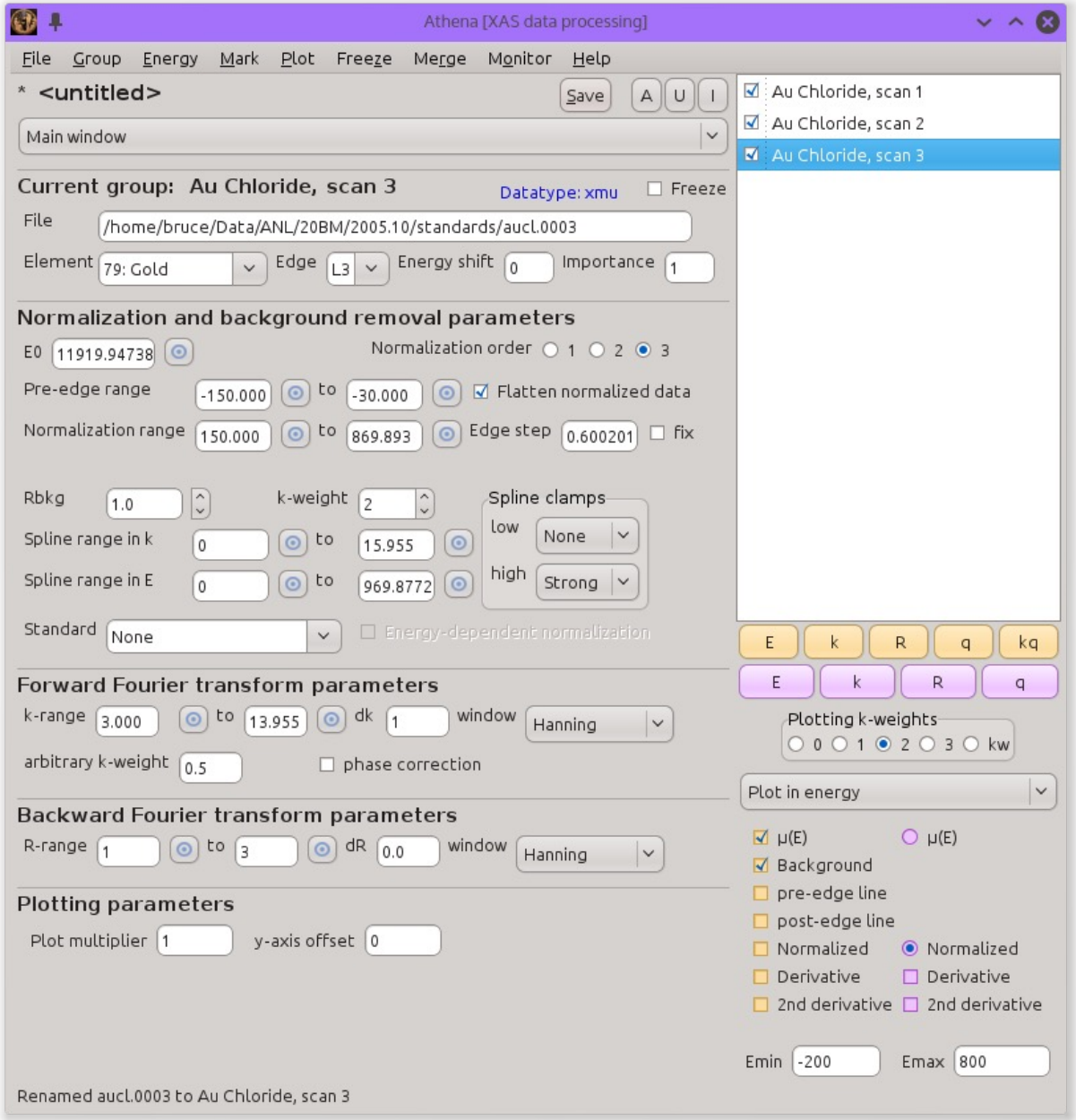
Benefits of Galaxy

- 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.**

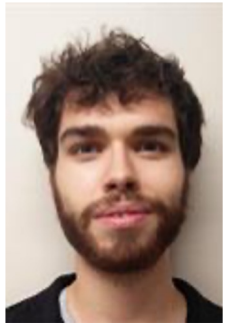
The Galaxy Platform: Applications to Catalysis Workflows

Theoretical and Computational Physics Group - SCD

LIBORIO

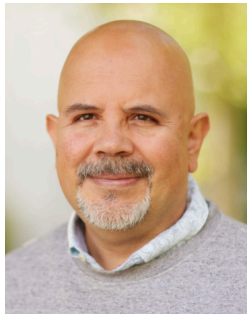
Data & Software Engineering Group SCD

GONZALEZ-BELTRAN



AUSTIN

UK Catalysis Hub



NIEVA DE LA HIDALGA

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

Galaxy PROJECT

ada lovelace centre

Software Sustainability Institute

UKRI Scie Tech Faci

eosc | EuroScienceGateway