ChemCatChem

The European Society Journal for Catalysis

Chemistry Europe

European Chemical Societies Publishing

Supported by

GECATS GERMAN CATALYSIS SOCIETY

Accepted Article

Title: Facilitating Reproducibility in Catalysis Research with Managed Workflows and RO-Crates: A Galaxy Case Study

Authors: Abraham Nieva de la Hidalga, Leandro Liborio, Patrick Austin, Subindev Devadasan, Tom Underwood, Alexander Belozerov, Martin Wilding, Nitya Ramanan, and C. Richard A. Catlow

To be cited as: *ChemCatChem* **2025**, e202401676

Link to VoR: https://doi.org/10.1002/cctc.202401676

WILEY VCH

WILEY VCH

RESEARCH ARTICLE

Facilitating Reproducibility in Catalysis Research with Managed Workflows and RO-Crates: A Galaxy Case Study

Abraham Nieva-de-la-Hidalga[a, b, f]*, Leandro Liborio[c], Patrick Austin[c], Subindev Devadasan[c], Tom Underwood[c], Alexander Belozerov[c], Martin Wilding[b], Nitya Ramanan[d], C. Richard A. Catlow[a, b, e]

[a] Dr. A. Nieva-de-la-Hidalga, Prof. C. R. A. Catlow
School of Chemistry, Cardiff University
TRH Building, Maindy Road, Cardiff, CF24 4HQ
E-mail: nievadelahidalgaa@cardiff.ac.uk

[b] Dr. A. Nieva-de-la-Hidalga, Dr. M. Wilding, Prof. C. R. A. Catlow
UK Catalysis Hub
Research Complex at Harwell, OX11 0FA Didcot, UK

[c] Dr. L. Liborio, P. Austin, S. Devadasan, Dr. T. Underwood, Dr. A. Belozerov,
Scientific Computing Department
Science and Technology Facilities Council
Rutherford Appleton Laboratory, Didcot, OX11 0QX, UK

[d] Dr. N. Ramanan
Diamond Ligh Source Ltd.
Didcot OX11 0DE, UK

[e] Prof. C. R. A. Catlow
Chemistry Department, University College London
WC1H 0AJ, London UK

[f] Dr. A. Nieva-de-la-Hidalga
School of Computer Science and Informatics, Cardiff University

Supporting information for this article is given via a link at the end of the document.

Abstract: Publishing supporting data significantly impacts researchers' productivity, especially in experiments requiring extensive tracking of data, processing steps, parameters, and outputs. A managed workflow environment, combined with RO-Crates, addresses these data management challenges. Workflows provide an alternative for handling complex data analyses by orchestrating various processing tools. The RO-Crate format, a community-driven proposal for packaging data, provenance, and workflows, facilitates publishing and reproducibility. The Galaxy workflow management system integrates workflows and RO-Crates, enabling the export of analyses, which can be shared and restored by other users. Using Galaxy, we demonstrate how to improve support for reproducibility. We tested our approach by designing an experiment using diverse supporting data from selected papers. In the experiment, we identified specific FAIRness and completeness issues hindering result reproduction, even when authors made significant efforts to document and publish their supporting data. In comparison, the proposed approach supports reproducibility by packaging datasets in RO-Crate format, streamlining the process. The Galaxy RO-Crates, published as supporting materials, enhance data sharing, transparency, and reproducibility, thus supporting the advancement of FAIR research practices in catalysis research.

List of abbreviations:

XAS: X-ray absorption spectroscopy; **XANES**: X-ray absorption near edge structure; **LCF:** linear combination fitting; **EXAFS**: extended X-ray absorption fine structure; **UKCH**: UK catalysis hub; **FTIR**: Fourier transform infra-red; **EPR**: electron paramagnetic resonance; **UI**: user interface; **FAIR**: findable, accessible, interoperable and reusable; **RO-Crate**: research object crate.

Introduction

Large scale facilities such as ESRF (France), DESY (Germany), Spring-8 (Japan) or Diamond (UK) are continuously enhancing their equipment [27, 31], enabling the design of complex experiments – such as operando experiments – that generate larger datasets [15, 16]. In this scenario, catalysis researchers face growing challenges to process, analyze and share these datasets. This points to a need for effective data management tools which can facilitate documenting the details of the processing and analysis steps to ensure reproducibility and replicability. Scientific workflow management systems provide a solution that facilitates processing of large datasets, tracking processing details (such as configurations, resources, parameters, and variables), and supporting reproducibility and replication. This paper shows how using a workflow environment and a data packaging standard can streamline processing and analysis while ensuring that published data supports reproducibility. This paper illustrates this proposal by applying a set of GALAXY XAS analysis tools to create workflows and RO-Crates to package the data of complex X-Ray absorption spectroscopy (XAS) experiments in catalysis. The examples selected cover three types of XAS analysis: for X-ray absorption near edge structure (XANES), Linear combination fitting (LCF), and extended X-ray absorption fine structure (EXAFS) analysis (see Figure 1**Fehler! Verweisquelle konnte nicht gefunden werden.**).

The challenges of reproducibility

The UK Catalysis Hub (UKCH) has been facilitating access to Diamond Light source XAS beamlines since its founding in 2013 [30]. This program has supported performing ex-situ, in-situ, and operando XAS experiments by research groups from UK

Universities and Research Institutions, producing large amounts of data that require complex processing and analysis methods.

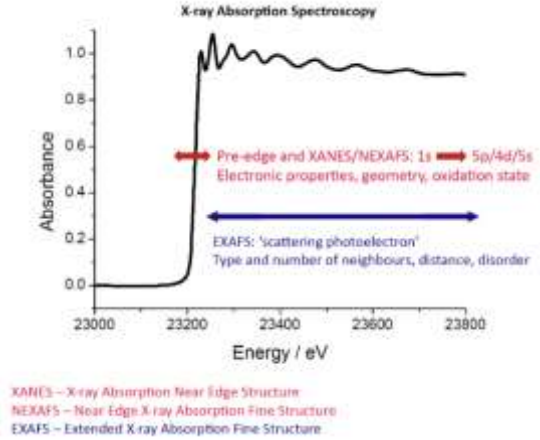

Figure 1 The XAS spectrum regions and the corresponding analysis techniques providing structural information (image from [9]).

XAS Data Analysis

X-Ray Absorption Spectroscopy comprises a set of experimental analysis techniques for characterizing materials. Figure 1**Fehler! Verweisquelle konnte nicht gefunden werden.** shows the XAS spectrum regions and the corresponding analysis techniques providing structural information. XAS is observed when X-rays are ejected from the tightly bound core electrons to the unoccupied states and the continuum. The spectral features that are typically within 50 eV above the absorption edge can be used to identify oxidation states and band occupancy (X-ray absorption near edge structure, XANES), while higher energies in the spectrum relate to local atomic structure like coordination number and distance, Debye-Waller factor, and inner potential correction (extended X-ray absorption fine structure, EXAFS) [13]. XAS techniques are often combined with other spectroscopic techniques like Raman, Fourier transform infrared (FTIR), X-ray photoelectron spectroscopy (XPS), and electron paramagnetic resonance (EPR) to provide a more detailed picture of the nature of the chemical bonds to better understand reaction mechanisms and intermediates [13].

An example of an XAS data management workflow is presented in Figure *2***Fehler! Verweisquelle konnte nicht gefunden werden.**. This workflow shows some of the common processing tools used such as DAWN [2], for data selection and cropping of input spectra; Athena for normalization and background subtraction; and Artemis for FEFF fit [23]. Additionally, Artemis and Athena rely on additional packages such as Atoms and FEFF, also shown in Figure 2**Fehler! Verweisquelle konnte nicht gefunden werden.**. X-Ray Larch (Larch) [19] is a Python based processing and analysis program that provides an alternative to the Artemis and Athena. The traditional way of analyzing XAS data with these tools requires extensive human intervention. In this scenario, visual inspection of data is often required to inform the choice of parameters used in data analysis. Moreover, data files are frequently transferred manually between different software programs.

Reproducibility Needs

Using this processing model, Catalysis researchers generate valuable and high-quality XAS research data. However, data publishing practices seem to undervalue the significance of these data. As the complexity of experiments and the size of datasets increase, the time spent processing and analyzing data also grows, making the manual approach less suitable for large datasets. In this context, the added requirement of publishing data to enable reproducibility further complicates the process and increases research time. The lack of support for reproducibility and reuse beyond the original research is challenging, and the provenance link between published results and supporting data is not always explicit.

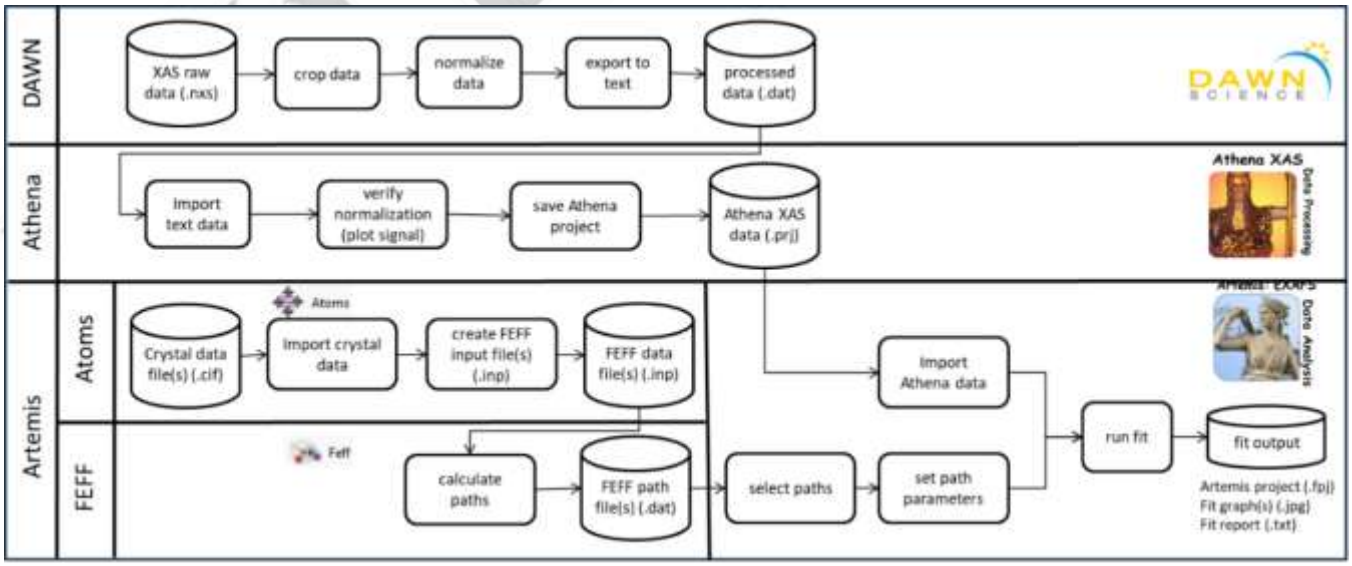

Figure 2 Schematic view of a XAS workflow. This workflow illustrates the processing and analysis of raw data using DAWN, Athena, Artemis, and two additional tools Atoms and FEFF. The workflow's complexity is evident in the number of inputs, outputs, and intermediate results, as well as the variety of file types involved.

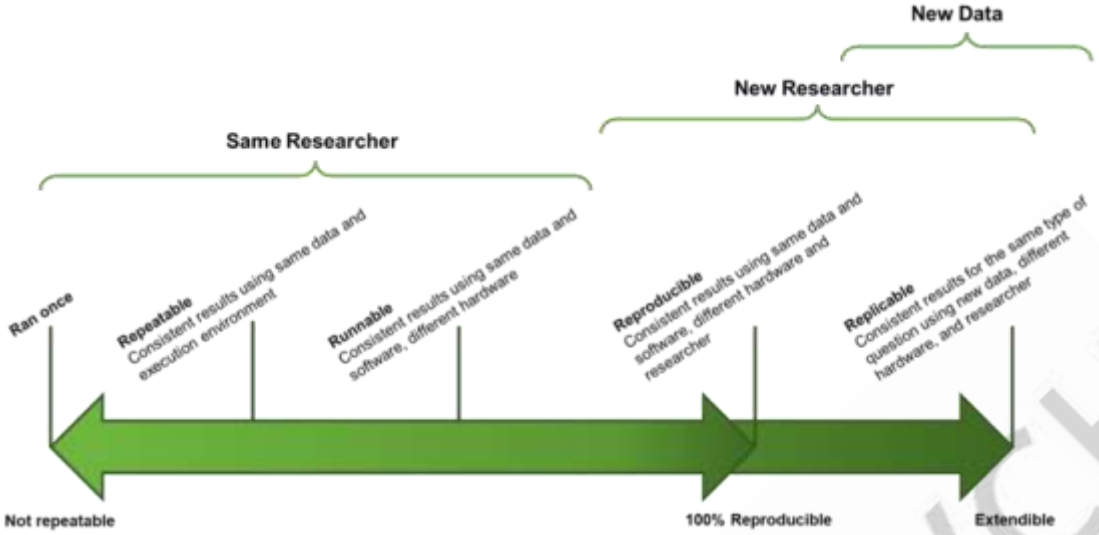

Figure 3 The levels of replicability can be seen as a spectrum. This graph shows the levels in which experiments may fall. The spectrum goes from 'not repeatable' up to "replicable" with a short explanation indicating when a experiment can be considered in each category.

Complex computational studies require increasing effort and time to progress in the scale from repeatability to runnability, reproducibility, and replicability [1]. Figure 3 (adapted from [14]) shows reproducibility as part of a wider spectrum. Most research results fall withing the repeatability and runnability levels. This is: researchers can easily obtain consistent results using their own data and computational resources. Reproducibility means that other researchers can obtain consistent results, using an equivalent set of resources (software, computing power). The results may not be exact but should not contradict/invalidate the findings of the original team. Replicability in turn means that other researchers can answer the same question with different data by reusing the experimental protocol and/or software.

Using Workflows for Catalysis Research

New experiments aimed at collecting larger quantities of data push the processing capabilities of current tools and complicate publishing of supporting data. In response to these issues, the UKCH explored the development of scripted and managed workflows tools using large datasets (See Table 1 adapted from**Fehler! Verweisquelle konnte nicht gefunden werden.** [20]). The results showed the potential of workflow tools for speeding up the processing and analysis of larger XAS datasets. However, the tools used (Demeter [23], Larch [19], Nextflow [29]) require some training in programming (Perl or Python). Moreover, the interfaces for processing in scripted and batch mode for these tools are command line based, compared to graphical tools such as Artemis and Athena. As an alternative, the Scientific Computing Department (SCD) of the Science and Technology Facilities Council (STFC) proposed using Galaxy for the creation of workflows. Galaxy provides a web UI and so does not require programming expertise whilst offering the same order of magnitude speed up from days to hours. It should also be noted that the computing resources behind Galaxy are highly configurable, and it supports job scheduling features like parallel submission/execution of jobs and execution on multiple nodes for further possible speed up [7]. The process for executing a Galaxy workflow on either a single dataset or a large collection with the same choice of parameters is the same from a user's perspective. As parameters only need to be defined once, analysis for large amounts of data can be made more efficient.

Additionally, Galaxy implements Research Object Crates (RO-Crates) as one of the formats for exporting data. An RO-Crate is a digital object designed to preserve data, configurations and parameters. This packaging facilitates publishing of data to support research works (papers, books, theses) [25].

Table 1 Speed up on processing and analysis of a large EXAFS dataset consisting of 3,790 spectra with different types of workflows.

Workflow Type	Process[a]	Fit[b]	Per spectrum[c]	Per dataset[d]
Manual Novice	3 min.	21 min	24 min	~ 63 days
Manual Expert	2 min	8 min	10 min	~ 26 days
Scripted Demeter	7.68 sec	13.56 sec	21.24 sec	~ 23 hours
Scripted Larch	0.12 sec	97.68 sec	1.6 min	~ 4 days
NextFlow –Larch	0.06 sec	6.54 sec	6.6 sec	~7 hours
Galaxy[e]	0.44 sec	7.68 sec	8.12 sec	~9 hours

[a] Average time required for preprocessing a spectrum.
[b] Average time required for a FEFF fit.
[c] Average time required for the full processing of a spectrum.
[d] Average time required for the full processing of a dataset.
[e] Run on a single 2.2GHz core

As a method for organizing and sharing research, RO-Crates ensure that all components of a study are properly documented and accessible. RO-Crates are not exclusive to the Galaxy project; they are used in various contexts as an open alternative for creating FAIR (findable, accessible, interoperable and reusable) Digital Objects in research [33]. By following these FAIR principles, RO-Crates promote organization, sharing and long-term usability of research data across different platforms and disciplines.

Galaxy: Tools and Workflows

Galaxy is a successful workflow management system with over 500,000 registered users from various research communities, including bioinformatics, chemistry, astronomy, environmental science and physics. Research communities have deployed Galaxy servers globally in the United States, Australia, and Europe. In 2023, Galaxy supported more than 11,000 users running more than 1 million jobs monthly [11]. The Galaxy platform itself is domain agnostic, allowing researchers to develop, deploy and run computation tools independently from the core software. The popularity and flexibility of Galaxy has extended its use beyond bioinformatics, its original target community. Consequently, recent efforts have added more support for subdomains [11] and promoted the development of tools in communities new to Galaxy [10] like those described here for XAS. Galaxy provides an accessible data analytics service hosted on a computing infrastructure capable of handling complex research computing tasks. Galaxy provides analytical tools that can be used individually or linked into complex workflows, with intermediate data outputs capable of triggering logic conditionals within the workflow [11].

The basic units of analysis in Galaxy are called tools. Each tool accepts data files and a set of parameters as inputs and generates data files as outputs. When a user runs a tool, a job is scheduled on the underlying computing infrastructure. The relation between the job's inputs and outputs is managed by Galaxy and presented to the user via the user interface (UI) history. Inputs are typically provided via a form-based UI, supporting Boolean options, check boxes, radio buttons, numerical inputs and free text. Additionally, Galaxy allows building interactive tools supporting user interaction through a pop-up window during job execution.

From a technical perspective, a tool in Galaxy is defined by an XML (eXtended Markup Language) file. This file specifies the tool's inputs, outputs, metadata (such as version, authors, and help text), and the underlying command to be executed when a job is submitted. Essentially, this XML file 'wraps' an existing piece of software. This separation of concerns means that the Galaxy tool developer does not need to be the developer of the underlying software and does not even need to be a programmer—only knowledge of the required Galaxy XML schema is necessary. This schema is well-documented and supported by a VSCode extension that provides schema validation and autocompletion.

Workflows in Galaxy chain multiple tools together, enabling the execution of complex processes and analyses as a single action. Workflows can be designed in advance or generated automatically from a user's history jobs and shared with collaborators via the Galaxy platform. Additionally, workflow executions are tracked by Galaxy as "Workflow invocations". Each time a workflow runs, a new invocation is generated. This cycle enables refining the processing and analysis parameters in response to the results produced. Additionally, invocations can be compared to determine if earlier runs were better or not. Once an optimal combination of parameters and inputs has been obtained, the corresponding workflow invocation can be exported as an RO-Crate, creating a reproducible data object which is ready for publishing.

XAS Galaxy Tools

The creation of Galaxy tools and workflows aimed to support a wide range of XAS processing and analysis alternatives. This effort resulted in the development of seven Galaxy tools (see Table ***2***), which can be composed into workflows to support XAS analyses. The XAS Galaxy tools wrap functionality from the Larch Python library. The starting point for this were the scripts produced for the "Scripted Larch" Workflow (reported in Table 1)**Fehler! Verweisquelle konnte nicht gefunden werden.**. The naming of the tools and their descriptions incorporate the names from the Demeter package to help users familiar with Artemis and Athena. The splitting of tasks is intended to allow generation of intermediate outputs which can be inspected to verify the results at later stages.

Table 2 Galaxy XAS tools developed by the Scientific Computer Department of the Science and Technology Facilities Council.

Tool Name	Inputs	Outputs	Description
Larch Athena	Tabular (text *.dat/*.csv/*.txt)and NeXus/HDF5 (*.nxs) data or a zip file containing data (*.zip)	Athena project file (*.prj) Image(s) of spectra plots (*.png)	Create an Athena project file from the input X-ray Absorption spectroscopy (XAS) data file.
Larch Plot	Athena project file(s) (compressed *.prj)	Image(s) of spectra plots (*.png)	Read Athena project files and plot data from them.
Larch LCF	Athena project files (compressed *.prj)	Images of fitting plot (*.png)	Performs a Linear Combination Fit (LCF) on an Athena project, using a number of other projects as components to the fit.
Larch FEFF	FEFF input (text *.inp) or crystal structure (text *.cif)	A set of scattering path files (text *.dat) produced by FEFF can be zipped or as a directory	Performs a FEFF6 calculation to find potential scattering paths in the provided structure.
Larch Select Paths	Set of scattering path files (text *.dat) produced by FEFF (can be zipped)	selected paths list (text -*_sp.csv) fitting parameters (text -*_gds.csv)	Select FEFF scattering paths to use in the fitting process.
Larch Artemis	Athena project(s) (*.prj), scattering paths (path or zipped file(s)), Select paths list (*_sp.csv) and fitting parameters (*_gds.csv)	Images of fitting plots (*.png) and fitting report (text *.feffit)	perform FEFF fitting on an Athena project file, originally from the input X-ray Absorption Fine Structure (XAFS) data file.
Larch Criteria Report	A set of fitting reports (text *.feffit) a compressed file of fitting reports (*.zip), or a CSV file of results to plot.	Image of plots comparing fit results (*.png)	Plot criteria of interest from multiple Artemis fittings.

Larch's functionality is accessible through imported library functions, rather than a command line executable. Therefore, a small Python script was created to define the functionality of each tool, the script is then executed using the command in the XML file.

The tools were initially tested with the data from the UKCH workflow demonstrator [20] and then extended to cover additional cases from the selected of publications. Reproducing their results led to the addition of analysis methods and adjustments to better support reproduction. For example, allowing setting input parameters, which had not been used previously.

Galaxy tools are versioned to ensure reproducibility based on the versions of the wrapped software and the wrapper. For example, "0.9.75+galaxy1" indicates the tool uses version 0.9.75 of the Larch library with a wrapper modified once since its "+galaxy0" release. This versioning strategy, along with Galaxy's use of containers to execute jobs, enforces a consistent and reproducible computing environment.

All previous tool versions remain available via the Galaxy Tool Shed, functioning like an "app store". Galaxy instances run in different scientific domains or regions, allowing administrators to install tools developed by others. This separation of tool development and platform administration enables more efficient tool distribution. The XAS tools are deployed on the STFC Materials Galaxy server [57] and the main European Galaxy instance [56].

Data Selection

The data used for testing and refining the XAS Galaxy tools were identified though the UKCH Catalysis Data Infrastructure (CDI) [58]. The only criterion for selection was that the datasets must have research data in a format which could be used by XAS Tools. At the time when the experiment was designed, the CDI listed 494 publications and 730 supporting datasets. Of these 140 papers (linked to 217 datasets) mention XAS data. However, only nine of these papers referenced accessible and processable XAS data [4, 5, 8, 12, 17, 18, 21, 26, 28]. Table 3 summarizes the articles and data used, providing an overview of the sources and types of data available for testing.

Results and Discussion

The reproducibility experiment was designed to test Galaxy as an alternative processing platform for XAS data. Openly available supporting XAS data was selected to test if the Galaxy tools, and their combination into workflows, could support the processing and reproduction of published results. Specifically, independently process and analyze the supporting data using our tools and attempt arrive to the conclusions of the original papers.

The primary advantage of the Galaxy XAS Tools is that, once developed, tested, and published, their learning curve is comparable to that of commonly used desktop tools, with no need for additional software installation. Replicating complex analysis and processing protocols is further supported by the capability to build custom workflows that can combine multiple processing steps. The workflows can themselves be shared with other researchers and exported as self-contained research objects. The Galaxy tools described above (Table 2), were composed into workflows to reproduce the results presented in nine papers. Figure 4 Shows a version of a small EXAFS workflow created as part of the learning materials to promote the use of Workflows and RO-Crates [49]. The workflow and RO-Crate reproduce the EXAFS fitting example for Athena and Artemis as described by Bruce Ravel [24]. Instead of using the FeS2.inp file in the original example, the workflow uses a crystal structure file (1564889.cif) from the Crystallography Open Database (COD) [54]. The discussion of the results is organized according to three workflow patterns that were identified as the models supporting the analyses presented in the selected literature.

Table 3 Details of papers and datasets used for the development of Galaxy tools. The first column indicates the paper using an ordinal number and the element named in the title. The second column indicates the data formats, and the subsequent columns indicate the types of XAS analyses.

Paper[a]	Data Formats[b]	Methods[c]	Raw[d]	Published Data ID [e]
1-Pt_3Sn [12]	text, prj	XANES, LCF, EXAFS	Yes	10.5258/SOTON/D0408 [43]
2-Diphosphine [18]	text	XANES	Yes	10.5523/bris.1kp2f62x3klb02mfz2qymcmxmx [36]
3-Au Colloids [28]	prj	XANES	Yes	10.5258/SOTON/D0921 [53]
4-Palladium [26]	prj	XANES	Yes	10.5525/gla.researchdata.654 [51]
5-$LaMnO_3$ [4]	prj, opju	XANES, EXAFS	No	10.5258/SOTON/D1128 [39]
6-Pd/Al_2O_3 [8]	prj, opj	XANES, EXAFS	No	10.5258/SOTON/D1723 [41]
7-MoO_x/Al_2O_3 [17]	text	XANES, EXAFS	No	10.5525/gla.researchdata.1092 [46]
8-Gasoline [21]	prj	XANES, EXAFS	No	10.5525/gla.researchdata.1141 [50]
9-N_2O-$LaMnO_3$ [5]	text, prj, opju	XANES, EXAFS	No	10.5258/SOTON/D1342 [40]

[a] Paper Identifier
[b] Data formats of published data: text – tabular plain text, prj – Athena project file, opj/opju – origin project, nexus – compressed HDF5 based format
[c] Analysis methods for published results, common preprocessing includes normalisation and alignment.
[d] Indicates if data is provided in raw format (i.e. closest to original data from beamline)
[e] Identifier of published data retrieved for the experiment.

10.1002/cctc.202401676

RESEARCH ARTICLE

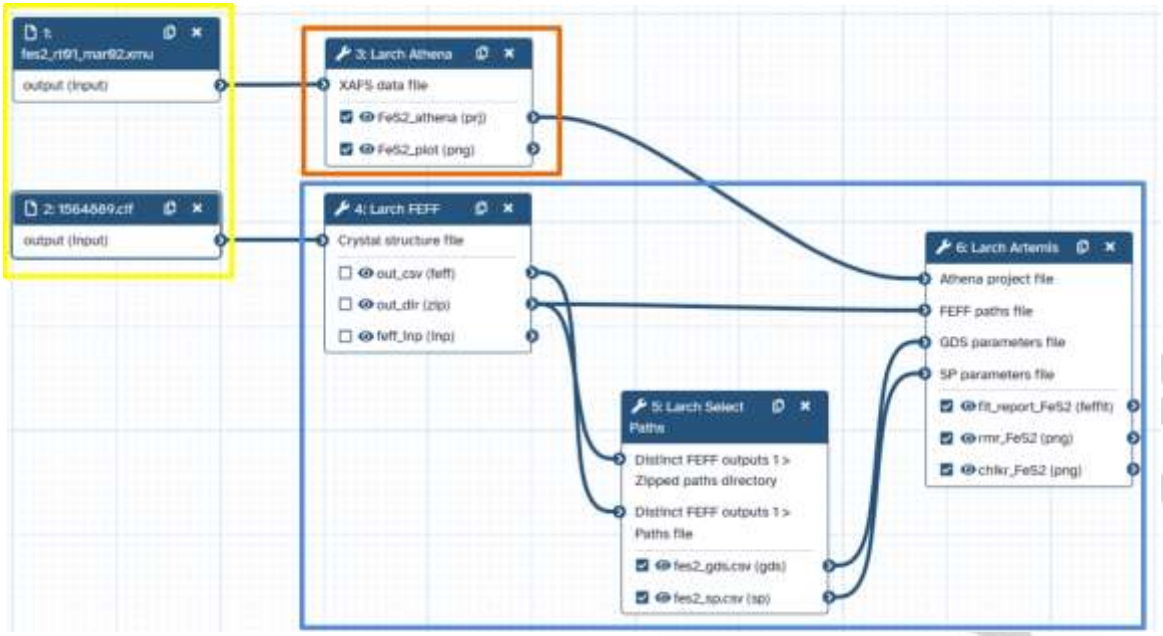

Figure 4 A graphical view of a workflow in the Galaxy environment. The example shows the textbook example of FeS2 fitting, from training materials. The icons next to component names indicate the type of element: a document icon (🗋)indicates an input and a wrench icon indicates a tool (🔧). This workflow has two inputs and four tools. The lines connecting the boxes indicate the flow of data between components. The body of the toolboxes contains two lists of data items. The top par indicates the inputs required and the bottom part the outputs produced. For instance, the box with the number 3 takes as input a XAS data file and produces two outputs: an Athena file (.prj) and an image of a plot (.png). The highlights in yellow, red, and blue have been added here to distinguish the workflow components related to XAS processing and analysis. The two boxes highlighted in yellow correspond to inputs: an XAS text data file (fes2_rt_01_mar02.xmu), and a crystal structure file (1564889.cif). The box highlighted in red is the representation of the Larch Athena tool that performs the XAS processing including, cropping, alignment, and normalization, which are stored in an Athena file. The three boxes in the blue highlight the main activities in a FEFF fit: generate FEFF paths, select paths, set fit parameters, and FEFF Fit of spectra.

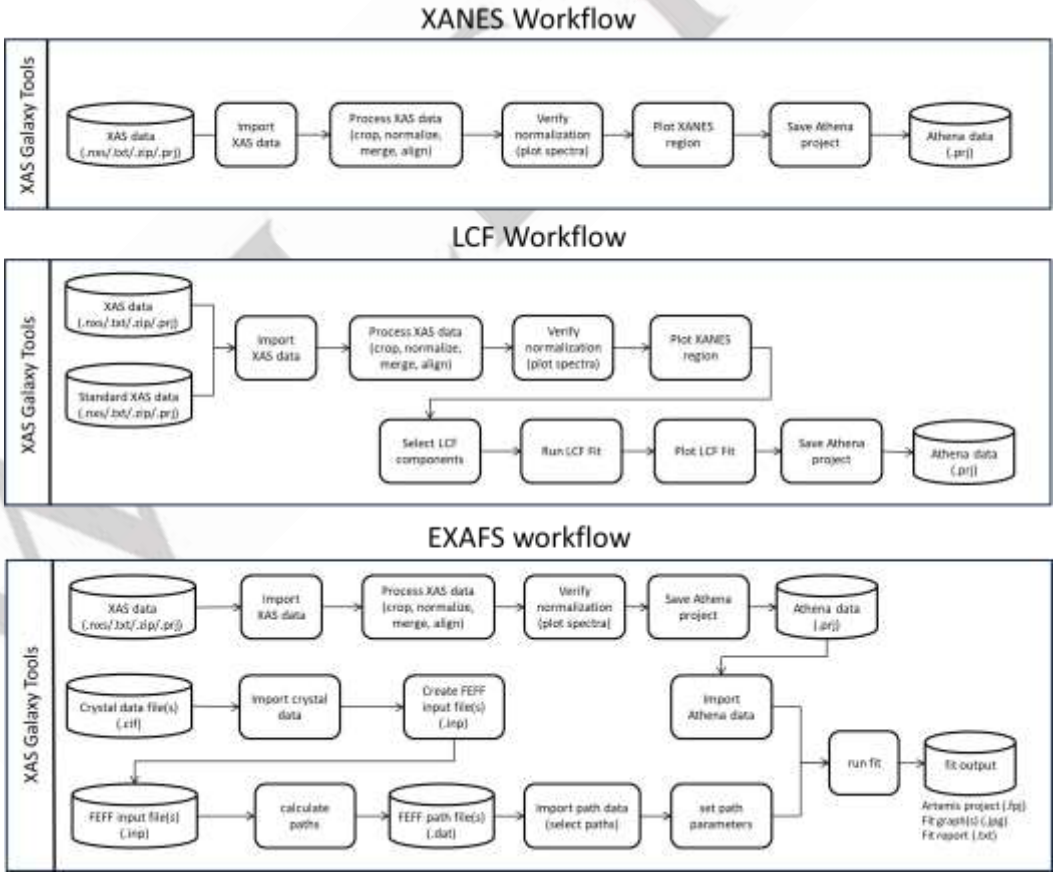

Figure 5 The XANES, LCF and EXAFS workflows supported by the XAS Galaxy tools

Common Workflow Patterns

Although the nine papers selected target different materials and analyse the data at different levels of detail, it was possible to identify three common workflows which cover the nine cases studied.

The summary of papers and datasets given in Table 3 shows that all papers report the results of XANES analyses, the second most relevant analysis technique reported is EXAFS (six papers), and the third XANES LCF (one paper). The three workflow patterns shown corresponding to these analyses is presented in Figure 5. Looking at XAS analysis this way allows identifying the tools which are commonly required for analysing the different datasets. The supporting materials elaborate on these workflows and how they were refined for reproducing the results of the selected papers and highlight some of the issues encountered. As 1-Pt_3Sn was the only paper to include all methods, its results will be used to discuss the types of difficulties encountered in the reproduction process. The details of the individual papers are discussed in the supporting information.

XANES workflow

XANES is the only type of XAS analysis included in all the articles. The outputs from XANES analysis are presented as plots in the nine papers. To compare the reproduced results to the published results we perform a visual inspection of the plots. In principle, the actual results of the normalization in Athena could be compared numerically to the results obtained using the XAS Galaxy tools. But this could not be done in all cases, because to do that we would need to have had all the data, metadata and parameters used to produce the experimental plots, and we lack much of the metadata and parameters. One of the advantages of the Galaxy tools is that they keep a record of all this information by default, which makes reproducibility quite straightforward. The reproduction of the XANES results was the most effective, as the shapes of the plots match those of the articles.

The plots in Figure 6 show a comparison of the results published in the paper and the results obtained using the Galaxy XAS tools. It is worth noting that to get the positions of these peaks to align with those from the paper, the position of the edge energy had to be manually set for the SnO_2 and Pt_3Sn spectra. The values used are captured in the corresponding digital object published as part of the supporting data for this paper. The XANES results in the rest of the papers follow a similar trend, comparing spectra to various references, highlighting the regions where peaks overlap. The XANES data can be presented as Fourier transform, normalized spectra, closeups of normalized spectra near the edge region and normalized derivative.

The main limitation of the Galaxy tools is in the presentation of results. Currently the tools produce only simple plots with all signals overlapping in the x-y axes. However, most of the authors tend to use additional software, such as OriginLab [55] to format data for publishing.

In view of these needs, the team developing the Galaxy Toos is analyzing alternatives which could support custom plotting options in Galaxy. The basic features to include are the possibility of adding offsets along the y axis, so that data can be presented in cascade, offsets along the x and y axes to allow the comparison of data in 3D cascades, as well as highlighting and labelling of overlap regions.

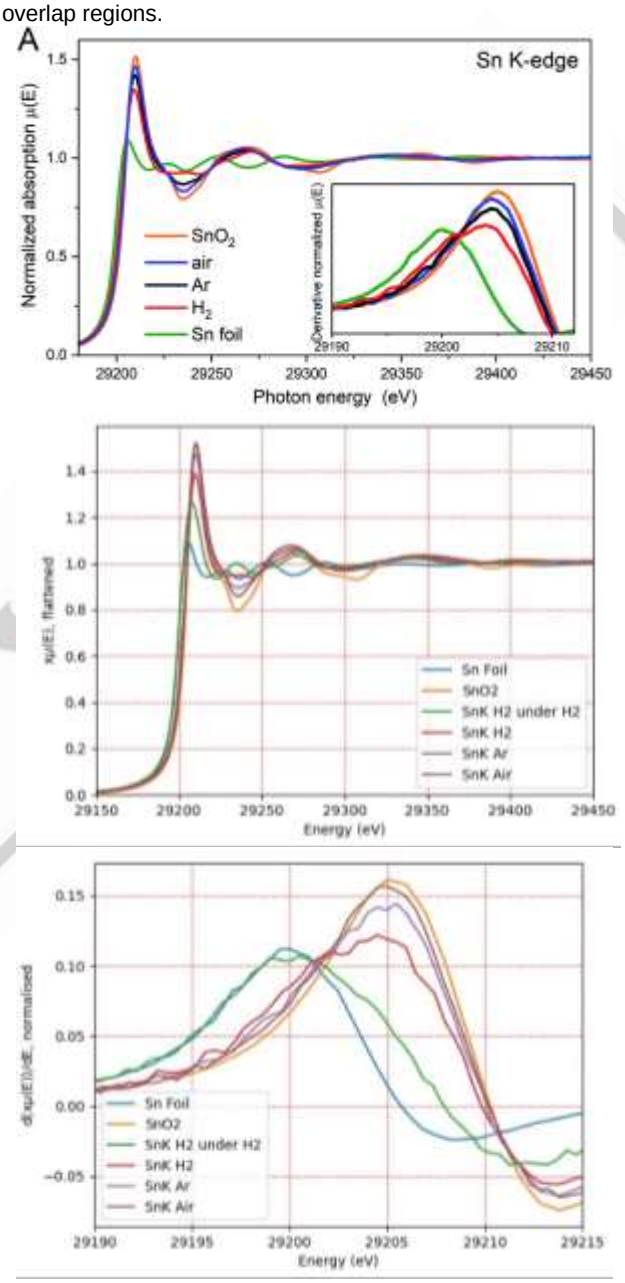

Figure 6 Comparison of XANES plots from Figure 4A in paper 1-Pt3Sn and equivalent plots generated from Galaxy. The top plot is the one included in the paper, the two plots under it are the ones produced in Galaxy. Note that Sn K edge of the H2 treated sample collected under H2 appears in the reproduction, to show how it was calibrated with respect to the other signals before using it in LCF

LCF Workflow

LCF is only performed in one of the papers (1-Pt3Sn). LCF is performed on XANES data, and its purpose is to identify the composition of pre-edge and edge features by comparing them to the spectra of at least two standards. In this case, the published

dataset included both the data and the percentages of the standards, so the reproduction was straightforward.

Figure 7 shows the comparison of the LCF results published to the results obtained with the Galaxy XAS tools. The results from galaxy are produced individually, the side-by-side placement is done manually here. Additionally, the percentages for the standards are also added manually to ensure visibility For LCF, the overall trend is consistent with the paper: the percentage of Sn decreases from H2, Ar, Air, and is higher when the sample treated with and collected under H2. However, the values reported by Galaxy for the Sn percentage were lower in all cases. This may be a combination of two factors, a propagation of error from a poor normalization and energy calibration earlier or setting a different range for performing the LCF.

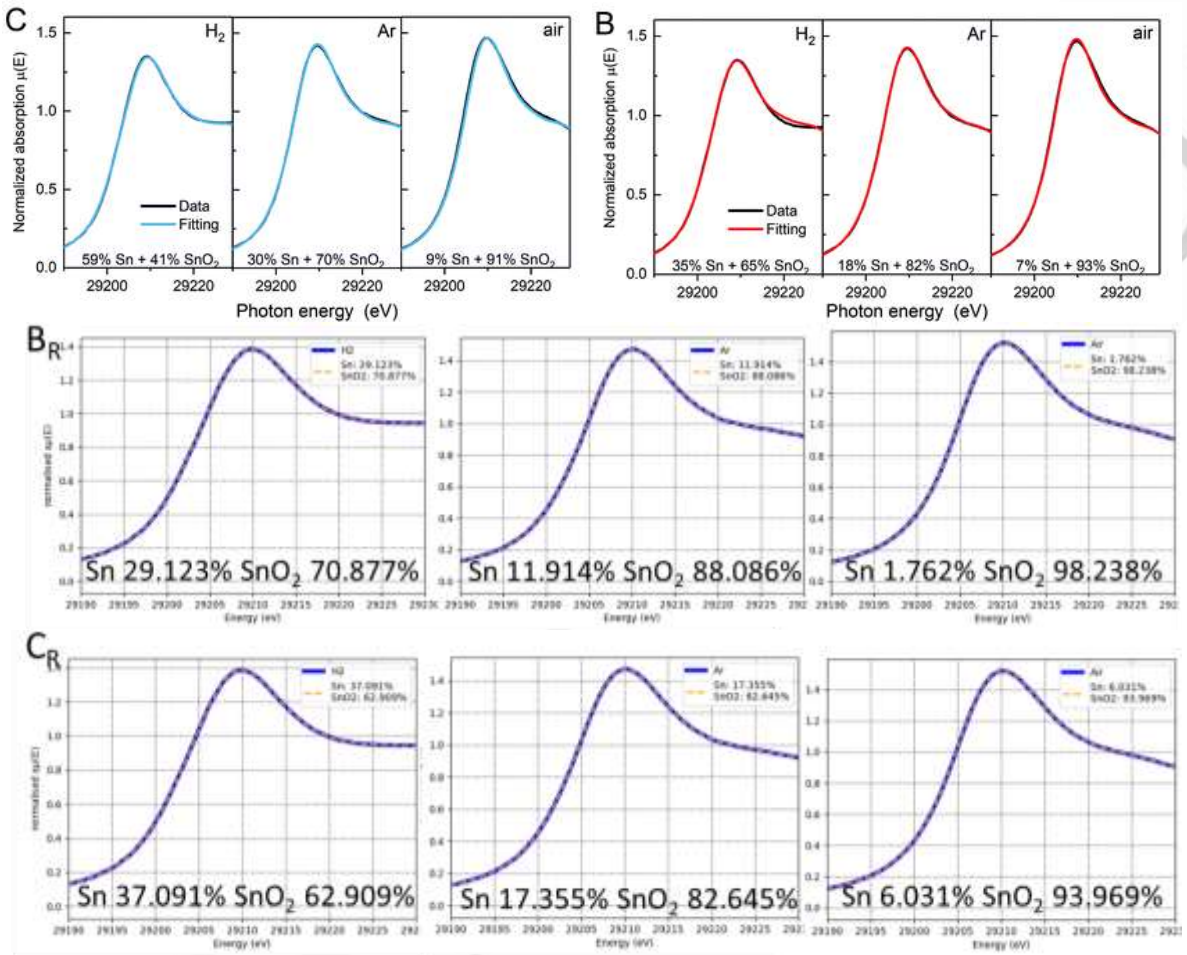

Figure 7 Comparison of LCF plots from Figure 4 in paper 1-Pt3Sn (B and C) and corresponding plots generated from Galaxy (B_R and C_R). The main differences are because the Galaxy tools generate the LCF plots individually, and there is no option for combining and aligning plots. Additionally, the legends do not scale well. For this paper, we overlaid the percentages of the standards for better visibility.

The Galaxy LCF tool does not allow generating custom plots. In this case the plots from the paper are stretched on the x axis, allowing side-by-side comparison. This placement of the plots also include the same y axis scale, showing the increase in the of peaks, this is: peak of H_2 < Ar < air. The plots from Galaxy also show this, but each is stretched to cover the same area, so the y scales do not match. Additionally, the Galaxy tool for LCF, at the time of writing, only produces a graphic output while Larch and Athena also generate a fitting report. In this case, producing an LCF report as a text file is a feature that will be included in future versions of the tool.Additionally, instead of relying on visual comparison of plots scaled arbitrarily, a better option would be to obtain the actual values for the fits so that they can be compared more effectively. The possibility of exporting the tabular values for the fits is also a feature to be included in future versions .

EXAFS workflow

EXAFS analysis results are presented in 6 papers (1-Pt3Sn, 5-LaMnO3, 6-Pd/Al2O3, 7-MoOx/Al2O3, 8-Gasoline, and 9-N2O-LaMnO3). These results were harder to reproduce for two reasons: (1) absence of crystal structure files and (2) absence of open standard readable fit files. In the first case, the crystal structure files used to create the FEFF paths for fitting were not provided or referred to, neither in the papers nor the corresponding published datasets. In the second case, the Artemis projects were not shared, instead some papers included origin files, which can only be accessed by using licensed Origin Lab software [55]. In this case, the reproductions approximate

those results provided by the publications, but the actual accuracy of the reproduction cannot be confirmed (See supporting material).

The plots in Figure 8 compare the results produced in Galaxy to those of the 1-Pt3Sn example. The most obvious discrepancies from the results of the paper show in the EXAFS fitting. This was difficult for a non-expert since mistakes made in either the normalisation of the spectra, generation of the FEFF paths or choice of fitting variables may have had an impact, but it was not obvious where the source might be. Once again, some degree of ad hoc variation was attempted to try and adjust the Fourier transform applied so that at least the experimental plot in r-space showed peaks in the same position as the paper. Even this was a challenge and was not fully successful.

In addition to visual comparison, the parameters varied when fitting the experimental data is also compared in the supporting information. In all cases the r-factor obtained in our reproduction was higher, and in some cases unphysical values were obtained (negative values for degeneracy and/or σ2).

In this case, a better approach would be to compare the signals of the fits numerically, not only the report and the plots. For this reason, future versions of the tool will also have an option to produce these as tabular data.

FAIRness and Completeness of supporting data

During the development and testing of the Galaxy tools, we worked with publicly available datasets, aiming to reproduce the experimental results from published papers. At first glance, the datasets are FAIR (Findable, Accessible, Interoperable and Reusable) [33]. The datasets are findable, i.e. clearly linked to the corresponding publications and properly identified. They are also accessible; this is, they are retrievable from established data repositories. In terms of interoperability, the data formats used for storing them are widely accepted within the XAS community. Finally, as the repositories and associated publications allow for their reuse with proper acknowledgement for the sources.

However, while the data meets at some level the FAIR principles, the reproducibility experiment uncovered various issues the prevent fully reproducing the results. The issues identified are directly related to FAIRness and completeness concerns.

The FAIRness issues are mainly due to the lack of machine-actionable metadata, which could facilitate the identification of correspondences between data and results. While this metadata is provided in all cases, it is typically in human-readable text that needs to be extracted and codified by an expert to support reproduction, limiting the applicability of automated processes.

Regarding completeness, we observed that the absence of data and metadata hindered the full reproduction of results. In addition to raw data and parameters, the dataset should include mappings to data and any complementary data, such as XAS standards and crystal structure data. Furthermore, all data and metadata should be machine-actionable, reducing the need for manual intervention.

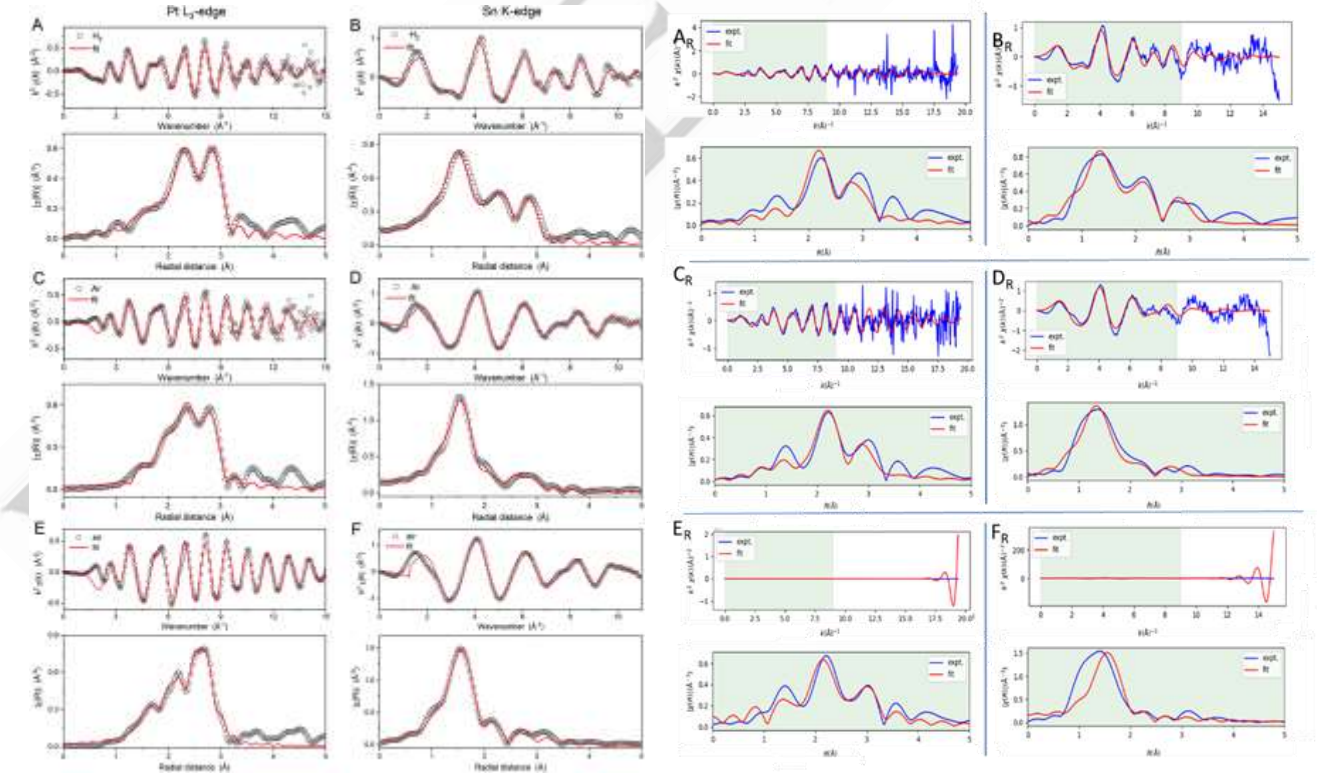

Figure 8 Comparison of EXAFS plots from figure 5 in paper 1-Pt3Sn in columns one and two from left to right, and equivalent plots generated from Galaxy in columns three and four from left to right.

Supporting reproducibility

The Galaxy platform allows exporting a workflow execution as a research object (RO) [25]. This guarantees the reproducibility of results because the tools for processing are openly available on the web and the RO contains the data, metadata, parameters and processing information required to reproduce the processing and analysis steps. In fact, a published Galaxy RO can be imported into Galaxy and executed directly, providing a shortcut for the reproduction of results. The data supporting the results presented in this article include nine research objects, one for each of the reproduction experiments performed [35, 37, 38, 42, 44, 45, 47, 48, 52].

Conclusions

The Galaxy platform and the Galaxy XAS Tools provide a consistent processing and analysis environment for XAS data. The design, development, and testing of this environment relied heavily on online published supporting data. In this paper we report the work performed testing the support for reproducibility when comparing Galaxy to current publishing practices. For this a set of papers with published data was used to provide real life examples. The selection criteria required the data to be available online in an open format, ensuring compatibility with the tools. Additionally, the data needed to be directly linked to the specific results published in the corresponding papers, ensuring both accessibility and verifiability.

The Galaxy XAS tools, the Galaxy environment and RO-Crates provide a user-friendly platform for analysis and processing XAS data. The reproduction experiment presented shows that the tools can produce results consistent with those found in existing literature. The development team was able to explore alternative processing paths, complementary techniques, and further processing alternatives. These scenarios demonstrated the value of published data for the evaluation, and development of alternative tools and methods for processing and analysis, taking advantage of the links between published data and results. Packaging data as RO-Crate is a viable publishing alternative that supports reproducibility without requiring major time investment from researchers. It is important to note that high-quality data, comparable to those used here, are generally scarce and not as well documented. We hope this example will encourage more widespread publication of supporting data and encourage the improvement of the quality of the associated metadata.

Extending the range of analysis

In addition to demonstrating how the XAS Galaxy tools support reproducing published results to an acceptable degree of accuracy, the discussion of the proposed tools with researchers and beamline scientists raised several ideas for the alternative application of these ideas. For instance, the basic XANES workflow is being further refined to facilitate the processing of the data produced during UKCH BAG sessions at Diamond light source. Additionally, the execution environment can be configured to process larger datasets using existing HPC facilities, making the Galaxy approach suitable for processing and analysis of large numbers of XAS readings, such as those resulting from in-situ and operando experiments.

Similar approaches in materials science areas

Beyond the support of processing XAS data, Neutron and Muon facilities have also implemented similar Galaxy tools. This expertise is available in the UK facilities as our SCD colleagues also collaborated in the development of Galaxy tools to process simulation data from muon experiments [6]. In the area of neutron spectroscopy, the work of colleagues from Oak Ridge National Labs [32] has demonstrated a suite of Galaxy tools and workflows. In this case, the SCD is also working on developing Galaxy tools for processing neutron diffraction experiments based on the Oak Ridge examples.

Enhanced support of catalysis research

Galaxy XAS Tools and RO-Crates enhance the reproducibility of XAS data workflows by covering various techniques. Galaxy generates detailed metadata, describing the execution environment comprehensively, which improves reproducibility and replicability. RO-Crates can be imported, inspected, and executed in the Materials Galaxy and Galaxy Europe servers [56, 57], allowing the reproduction of data processing and analysis, matching published results. Additionally, workflows and data can be reused for further analyses, such as extending XANES results to perform EXAFS analyses, or analyzing EXAFS results with different crystal structure models.

The catalysis data value chain describes the process of managing and utilizing data throughout the entire lifecycle of catalysis research, from molecular studies to the development of chemical processes [34]. While the examples in this paper are derived from catalysis articles used, the Galaxy platform and tools are applicable to a broader context. Integrating Galaxy tools and research objects into the data value chain for catalysis sciences requires the creation, adoption, and use of a custom vocabulary and ontologies for catalysis [3]. The vocabulary and ontology will be used to complement the Galaxy RO-Crate metadata with catalysis metadata, providing domain information about the experiments, such as materials, sample preparation, experiment design, and equipment. This richer semantic context will enhance the FAIRness, reproducibility and replicability of catalysis data. The final section of the supporting information for this paper[1] presents an example of how these can be achieved. The example of the extended metadata file is also included as a JSON file[2].

[1] SupportingInformationGalaxyROCrates.docx

[2] supporting_information_gasoline_filters_roc_desc.json

Efficient use of resources

In addition to the benefits highlighted in the discussion section, we also want to mention that reuse, reproduction, and replication of results contribute to reducing CO_2 emissions in national facilities. This is because reusing published data reduces the requirement to repeat costly experiments, supporting the management of large-scale facilities.

Data availability statement

The authors have cited additional references in this paper and within the Supporting Information [36-53]. This includes the nine datasets used for the experiments and the dataset containing the nine Galaxy RO which support the results on the papers selected for analysis. The data from the previous papers is available from the linked references, while the RO-Crates are al available in Zenodo.

Acknowledgements

The authors would like to acknowledge the support provided by the researchers of the UK Catalysis Hub. We particularly appreciate the comments and suggestions provided by Peter Wells, Andy Beale, Harry Drivas, Chistopher Parlet, and Mathew Potter. The funding for the research and development of the tools presented here came from the collaboration of UKCH with the STFC Scientific Computing Department in the "Physical Sciences Data Infrastructure (PSDI)" project (funded by ESPRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1). Additional funding was provided by the UK Catalysis Hub (funded by EPSRC grants: EP/R026939/1, EP/R026815/1, EP/R026645/1, EP/R027129/1, and EP/M013219/1(biocatalysis)); and by the EuroScienceGateway project, funded via the UK government's Horizon Europe funding guarantee (project number 10038963)

Keywords: Workflow Management System • Scientific Workflows • Fair Data Objects • X-Ray Absorption Spectroscopy • Data Processing and Analysis

References

1. Bakinam T. Essawy, Jonathan L. Goodall, Daniel Voce, Mohamed M. Morsy, Jeffrey M. Sadler, Young Don Choi, David G. Tarboton, Tanu Malik, Environmental Modelling & Software **2020**, Volume 134.
2. M. Basham, J., Filik, M.T. Wharmby, P.C.Y. Chang, B. El Kassaby, , M. Gerring, J. Aishima, K. Levik, B.C.A. Pulford, I. Sikharulidze, D. Sneddon, M. Webber, S. S. Dhesi, F. Maccherozzi, O. Svensson, S. Brockhauser, G. Náray, A. W. Ashton. J. Synchrotron Rad **2015**. 22
3. A.S. Behr, H. Borgelt, N. Kockmann, Journal of Cheminformatics **2024**, 16(1), p.16. doi: 10.1186/s13321-024-00807-2
4. R. H. Blackmore, M. E. Rivas, T. Eralp Erden, T. Dung Tran, H. R. Marchbank, D. Ozkaya, M. Briceno de Gutierrez, A. Wagland, P. Collier, P. P. Wells. Dalton Trans., **2020**, 49, 232-240.
5. R. H. Blackmore, M. E. Rivas, G. F. Tierney, K. M. H. Mohammed, D. Decarolis, S. Hayama, F. Venturini, G. Held, R. Arrigo, M. Amboage, P. Hellier, E. W. Lynch, M. Amri, M. Casavola, T. Eralp Erden, P. Collier, P. P. Wells. Phys. Chem. Chem. Phys., **2020**, 22, 18774-18787.
6. S. J. Blundell et. al., Psi-K Scientific Highlight of the month, Sept. 2024, 158. https://psi-k.net/download/highlights/Highlight_158.pdf
7. N Coraor, B. Grüning, H. Rasche, Galaxy Training Network 2024, https://gxy.io/GTN:T00002 Online; accessed Mon Nov 25 2024
8. D. Decarolis, A. H. Clark, T. Pellegrinelli, M. Nachtegaal, E. W. Lynch, C. R. A. Catlow, E. K. Gibson, A. Goguet, P. P. Wells. ACS Catalysis **2021**, 11, 4, 2141–2149.
9. Diamon Light Source, Techniques Available, about XAS. Available: https://www.diamond.ac.uk/industry/Techniques-Available/Spectroscopy/X-ray-spectroscopy-XAS/About-XAS.html
10. EuroScienceGateway, Available: https://doi.org/10.3030/101057388
11. The Galaxy Community. Nucleic Acids Research, **2024**. https://doi.org/10.1093/nar/gkae410
12. H. Huang, A. B. A. A. Nassr, V. Celorrio, S. F. R. Taylor, V. K. Puthiyapura, C. Hardacre, D. J. L. Brett, A. E. Russell. Faraday Discussions. **2018** V. 208. pp. 555-573.
13. A. Iglesias-Juez, G. L. Chiarello, G. S. Patience, M. O. Guerrero-Pérez. Can J Chem Eng. **2022**; 100:3–22.
14. The Informatics Technology for Cancer Research (ITCR) Training Network (ITN), Reproducibility in cancer research **2022**, Available: https://jhudatascience.org/Reproducibility_in_Cancer_Informatics/defining-reproducibility.html
15. J. D. Kammert, G. Brezicki, R. Acevedo-Esteves, E. Stavitski, R. J. Davis Rev. Sci. Instrum **2020**. 91, 013107
16. B. Kimmerle, P. Haider, J. D. Grunwaldt, A. Baiker, P. Boye, C. G. Schroer. Applied Catalysis A: General **2009.** 353, 36–45
17. S. K. Matam, C. Moffat, P. Hellier, M. Bowker, I. P. Silverwood, C. R. A. Catlow, S. D. Jackson, J. Craswell, P. P. Wells, S. F. Parker, E. K. Gibson. Catalysts **2020**, 10(12), 1370.
18. A. M. Messinis, S. L. J. Luckham, P. P. Wells, D. Gianolio, E. K. Gibson, H. M. O'Brien, H. A. Sparkes, S. A. Davis, J. Callison, D. Elorriaga, O. Hernandez-Fajardo, R. B. Bedford Nature Catalysis, **2019**, 2, 123-133 DOI: 10.1038/s41929-018-0197-z.
19. M. Newville J. Phys. **2013**, Conf. Ser. 430 012007
20. A. Nieva-de-la-Hidalga, D. Decarolis, S. Xu,S. Matam, W. Y. Hernández-Enciso, J. Goodall, B. Matthews, C. R. A. Catlow. Data Intelligence 2022 4 (2) 455–470
21. M. Panchal, J. Callison, V. Skukauskas, D. Gianolio, G. Cibin, A. P. E. York, M. E. Schuster, T. I. Hyde, P. Collier, C. R. A. Catlow, E. K. Gibson. J. Phys.: Condens. Matter **2021**, 33, 284001. DOI: 10.1088/1361-648x/abfe16.
22. Physical Sciences Data Infrastructure, Homepage, can be found under https://www.psdi.ac.uk/
23. B. Ravel, M. Newville Journal of Synchrotron Radiation **2005** 12 537–541
24. B. Ravel, GitHub code and data, from: https://github.com/bruceravel/demeter/tree/master/examples/recipes/FeS2
25. S. Soiland-Reyes, P. Sefton, M. Crosas, L. J. Castro, F. Coppens, J. M. Fernández, D. Garijo, B. Grüning, M. la Rosa, S. Leo, E. Ó Carragáin, M. Portier, A. Trisovic, RO-Crate Community, Paul Groth, Carole Goble. Data Science 2022, vol. 5, no. 2, pp. 97-138. doi: 10.3233/DS-210053
26. C. Stewart, E. K. Gibson, K. Morgan, G. Cibin, A. J. Dent, C. Hardacre, E. V. Kondratenko, V. A. Kondratenko, C. McManus, S. M. Rogers, C. E.

Stere, S. Chansai, Y. C. Wang, S. J. Haigh, P. P. Wells, A. Goguet. ACS Catalysis **2018**, 8, 9, 8255–8262 DOI: 10.1021/acscatal.8b01509.

27. A. Taylor, Synchrotron radiation news **2023**, Vol. 37, No. 1, 2-3
28. G. F. Tierney, D. Decarolis, N. Abdullah, S. M. Rogers, S. Hayama, M. Briceno de Gutierrez, A. Villa, C. R. A. Catlow, P. Collier, N. Dimitratos and P. P. Wells. Nanoscale Advances **2019**. V. 1. pp. 2546-2552. DOI: 10.1039/C9NA00159J.
29. Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. Nature Biotechnology **2017**, 35(4), 316–319. doi:10.1038/nbt.3820
30. UK Catalysis Hub, " BAG at Diamond Light Source", can be found under https://ukcatalysishub.co.uk/bag-at-diamond-light-source/, 2024
31. H. Wagner, Synchrotron radiation news **2023**, Vol. 36, No. 1, 2-2
32. G. R. Watson, T. A. Maier, S. Yakubov, P. W. Doak. Frontiers in High Perf. Computing **2024**. 2:1390709. doi: 10.3389/fhpcp.2024.1390709
33. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. Sci Data **2016**, 3, 160018, doi: 10.1038/sdata.2016.18
34. C. Wulf, M. Beller, T. Boenisch, O. Deutschmann, S. Hanf, N. Kockmann, R. Kraehnert, M. Oezaslan, S. Palkovits, S. Schimmler, and S.A. Schunk,. ChemCatChem **2021**, 13(14), pp.3223-3236. doi: 10.1002/cctc.202001974
35. [dataset] P. Austin, A. Nieva-de-la-Hidalga, L. Liborio, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. Pt3Sn nanoparticle electrocatalysts. Zenodo, doi: 10.5281/zenodo. 13940793.
36. [dataset] R. Bedford, A. Messinis. *Iron Negishi mechanistic*. University of Bristol doi: 10.5523/bris.1kp2f62x3klb02mfz2qymcmxmx
37. [dataset] A. Belozerov, A. Nieva-de-la-Hidalga, L. Liborio, P. Austin, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. LaMnO3 catalytic behaviour. Zenodo, doi: 10.5281/zenodo. 14013574
38. [dataset] A. Belozerov, A. Nieva-de-la-Hidalga, L. Liborio, P. Austin, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. MoOx on Al2O3 Cyclic Operation. Zenodo, doi: 10.5281/zenodo.13919888
39. [dataset] R. H. Blackmore, M. E. Rivas, T. Eralp Erden, T. Dung Tran, H. R. Marchbank, D. Ozkaya, M. Briceno de Gutierrez, A. Wagland, P. Collier, P. P. Wells. *Dataset for 'Understanding the mechanochemical synthesis of the perovskite LaMnO3 and its catalytic behaviour'*. University of Southampton Institutional Repository. doi: 10.5258/SOTON/D1128
40. [dataset] R. H. Blackmore, M. E. Rivas, G. F. Tierney, K. M. H. Mohammed, D. Decarolis, S. Hayama, F. Venturini, G. Held, R. Arrigo, M. Amboage, P. Hellier, E. W. Lynch, M. Amri, M. Casavola, *Dataset for 'The electronic structure, surface properties, and in situ N2O decomposition of mechanochemically synthesised LaMnO3'* University of Southampton Institutional Repository, doi:10.5258/SOTON/D1342
41. [dataset] D. Decarolis, A. H. Clark, T. Pellegrinelli, M. Nachtegaal, E. W. Lynch, C. R. A.Catlow, E. K. Gibson, A. Goguet, P. P. Wells. *Dataset for 'Spatial profiling of a Pd/Al2O3 catalyst during selective ammonia oxidation'* University of Southampton Institutional Repository, doi: 10.5258/SOTON/D1723
42. [dataset] S. Devasan, A. Nieva-de-la-Hidalga, L. Liborio, P. Austin, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. Ultra-dilute Au colloids. Zenodo, doi: 10.5281/zenodo.13840848
43. [dataset] H. Huang, A. B. A. A. Nassr, V. Celorrio. *Dataset for Effects of heat treatment atmosphere on the structure and activity of Pt3Sn nanoparticle electrocatalysts: a characterisation case study*. University of Southampton Institutional Repository, doi: 10.5258/SOTON/D0408
44. [dataset] L. Liborio, A. Nieva-de-la-Hidalga, P. Austin, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. Pd on Al2O3 Catalyst for Ammonia. Zenodo, doi: 10.5281/zenodo.13941579.
45. [dataset] L. Liborio, A. Nieva-de-la-Hidalga, P. Austin, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. Catalysts in gasoline particulate filters. Zenodo, doi: 10.5281/zenodo. 14009131.
46. [dataset] S. K. Matam, C. Moffat, P. Hellier, M. Bowker, I. P. Silverwood, C. R. A. Catlow, S. D. Jackson, J. Craswell, P. P. Wells, S. F. Parker, E. K. Gibson. *Investigation of MoOx/Al2O3 under Cyclic Operation for Oxidative and Non-Oxidative Dehydrogenation of Propane*. University of Glasgow Enlighten Research Data, doi: 10.5525/gla.researchdata.1092
47. [dataset] A. Nieva-de-la-Hidalga, L. Liborio, P. Austin, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. Diphosphine ligands in iron. Zenodo, doi: 10.5281/zenodo.13842780.
48. [dataset] A. Nieva-de-la-Hidalga, L. Liborio, P. Austin, S. Devasan, T. L. Underwood, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. Palladium-Catalyzed CO Oxidation. Zenodo, doi: 10.5281/zenodo.13842852.
49. [dataset] A.Nieva de la Hidalga, Reproducing the FeS2 Example from the Demeter Tutorials. Zenodo, doi: 10.5281/zenodo.13987511
50. [dataset] M. Panchal, J. Callison, V. Skukauskas, D. Gianolio, G. Cibin, A. P. E. York, M. E. Schuster, T. I. Hyde, P. Collier, C. R. A. Catlow, E. K. Gibson. *Operando XAFS investigation on the effect of ash deposition on three-way catalyst used in Gasoline Particulate Filters and the effect of the manufacturing process on the catalytic activity*. University of Glasgow Enlighten Research Data, doi: 10.5525/gla.researchdata.1141
51. [dataset] C. Stewart, E. K. Gibson, K. Morgan, G. Cibin, A. J. Dent, C. Hardacre, E. V. Kondratenko, V. A. Kondratenko, C. McManus, S. M. Rogers, C. E. Stere, S. Chansai, Y. C. Wang, S. J. Haigh, P. P. Wells, A. Goguet. *Unraveling the H2 Promotional Effect on Palladium-Catalyzed CO Oxidation Using a Combination of Temporally and Spatially Resolved Investigations*. University of Glasgow Enlighten Research Data, doi: 10.5525/gla.researchdata.654
52. [dataset] T. L. Underwood, A. Nieva-de-la-Hidalga, L. Liborio, P. Austin, S. Devasan, A. Belozerov, M. Wilding, N. Ramanan, C. R. A. Catlow. N2O decomposition on LaMnO3. Zenodo, doi: 10.5281/zenodo.13941613.
53. [dataset] G. F. Tierney, D. Decarolis, N. Abdullah, S. M. Rogers, S. Hayama, M. Briceno de Gutierrez, A. Villa, C. R. A. Catlow, P. Collier, N. Dimitratos and P. P. Wells. Dataset for 'Extracting Structural Information of Au Colloids at Ultra-Dilute Concentrations: Identification of growth during Nanoparticle Immobilization.' University of Southampton Institutional Repository, doi: 10.5258/SOTON/D0921
54. [cryistal data] Synthetic control over polymorph formation in the d-band semiconductor system FeS2, the Crystallography Open Database (COD), url: https://www.crystallography.net/cod/cif/1/56/48/1564889.cif
55. [Software] Origin. OriginLab Corporation, Northampton, MA, USA.
56. [Software] Galaxy Europe (version 24.1.4.dev0) Available: https://usegalaxy.eu
57. [Software] STFC Materials Galaxy (version 24.1.2) Available: https://materialsgalaxy.stfc.ac.uk
58. [Research infrastructure] UK Catalysis Hub, Catalysis Data Infrastructure. Available: http://cdi.ukcatalysishub.org/