

Integrating Multimodal Imaging Data Flow in Paleontology

Nicholas Edwards, SLAC National Accelerator Laboratory, Menlo Park, CA, USA; Samuel M. Webb, SLAC National Accelerator Laboratory, Menlo Park, CA, USA; Michael B. Toth, R.B. Toth Associates LLC, Oakton, VA, USA

Abstract

Over approximately the last 20 years, advances in scientific instrumentation have made many analytical techniques more readily accessible to the broader scientific community. This has empowered paleontologists, archeologists, conservators, scholars and preservation specialists to acquire new data types from multiple techniques, revealing important information about cultural and natural heritage objects, such as ancient documents and fossils. As such, the volume of data and metadata that can be acquired has dramatically increased, putting increased strain on researchers to process, analyze, interpret, archive and manage this data. This paper discusses the advances of synchrotron x-ray fluorescence (XRF) imaging and multispectral imaging in paleontological research to demonstrate the challenges of handling the data that can now be generated for such an object. It also addresses how facilities such as synchrotrons are generating big data, and how data from multiple techniques can be analyzed in a multimodal framework.

Fossil Record

For organisms to become fossilized, their remains must undergo numerous favorable processes. They must be deposited in an environment that inhibits degradation and predation, then survive subjection to millions of years of geological processes. Therefore, to confidently interpret once-living organisms from their fossilized remains, paleontologists require knowledge and tools spanning biology, environmental science, geology, and both inorganic and organic chemistry.

Fossils are most recognizable as bones, skeletons, and teeth. These tissues were thought to be more readily preserved due to their inherent physical robustness, allowing them to survive mechanical and chemical processes over millions of years. However, many examples of soft tissue preservation exist, including skin, fur, feathers, and even fully soft-bodied organisms. These fossils usually occur under narrow, unique geological conditions and time ranges, called sites of exceptional preservation or Konservat-Lagerstätten.

It was once thought that the original composition of fossil tissues—hard or soft—was highly altered or replaced, having been subjected to a common series of taphonomic (fossilization) processes. Therefore, they were considered useful only for their visible morphology. The environments in which these tissues are preserved are quite diverse and vary widely in their taphonomic histories. The style of preservation, and what is preserved, can vary dramatically, even within the same sedimentological niche. Clearly, a generalized taphonomic model is inadequate.

Two decades of technological advances and increased accessibility to a wide range of analytical techniques—including scanning electron microscopy, Fourier transform infrared spectroscopy, CT scanning, multispectral imaging, and synchrotron X-ray fluorescence imaging—have allowed fossils to be studied in far greater detail than previously possible. There is now a realization that of the existence of a chemical fossil record, that can be captured with advanced digitization and scientific equipment, providing

highly informative insights into the physiology, behavior, and evolution of ancient organisms.

The presentation of fossils in museums can convey an appearance of abundance, but many of these objects are extremely rare and fragile, guarded vigilantly by their curators. As such, chemical analyses of fossils have historically been limited due to the spatial constraints and sample preparation requirements of many instruments (e.g., the vacuum chambers of scanning electron microscopes and electron microprobes) and/or the destruction of samples during analysis (e.g., dissolution and extreme heating of material for inductively coupled plasma mass spectrometry and gas chromatography-mass spectrometry). Therefore, techniques that provide information without requiring sampling or resulting in permanent alteration are highly desirable.

Non-destructive imaging at different energy levels offers unique opportunities to study these objects. With advanced imaging systems, image processing and data management, scientists can capitalize on these technologies to support research into important and rare fossils, revealing crucial information about ancient organisms never seen before. In particular, the ability to directly compare and correlate data from multiple analytical methods is key to extracting the most information.

Synchrotron XRF Imaging at SSRL

X-ray fluorescence (XRF) imaging is a technique that provides 2D images of chemical elemental distribution by bombarding a sample with a concentrated high-energy particle beam (e.g. electrons or photons) to ionize atoms, inducing X-ray emission – the atom itself emitting (fluorescing) an X-ray photon. The energy and wavelength of the fluoresced X-rays are characteristic of the atom, and therefore the chemical element, from which they originate.

These fluoresced X-rays can be detected with an appropriate X-ray detection system (such as solid-state detectors). Data can be collected as an XRF spectrum (X-ray counts vs. energy) at a single location, showing the elemental inventory as a series of peaks at the known X-ray energies of the elements. The larger the peak, the greater the quantity of that element present. It is also possible to scan a particle beam across a sample, or scan the sample across the beam, to generate an XRF image of elemental distribution. This is exceptionally useful to spatially resolve and correlate elemental distribution with specific structures or features in a sample.

Widely available and well-known electron beam systems, such as scanning electron microscopes, primarily used for imaging microscopic structures, commonly have detectors installed to collect Energy Dispersive Spectroscopic data [1]. However, electron beam methods have limitations, including high background signal and the requirement for a vacuum chamber, which restricts sample size to approximately 100×100 mm. They also require elemental concentrations of thousands of parts per million (ppm), resulting in many hours to scan areas of just a few square millimeters. Some scanning electron microscopes can be built with integrated X-ray photon sources which improves the sensitivity to 10s-100s ppme.

Synchrotron radiation generated by a synchrotron particle accelerator provides extremely bright X-rays that scientists use for a wide range of research, probing matter on the scales of atoms and molecules.[2] The high brightness, flux, and monochromaticity of synchrotron radiation provides numerous advantages that make it especially amenable to studying cultural and natural heritage artifacts.[3] Specifically, because of the penetrating nature of photons versus electrons, experiments can operate under ambient conditions, removing the major spatial restrictions of electron beam methods and allowing large samples to be mounted without the need to remove material or modify an object. The high flux combined with the latest X-ray detection systems results in high sensitivity (as low as 1 ppm) and greatly reduced acquisition times compared to other methods, allowing areas on the order of hundreds of square centimeters to be imaged in hours rather than days.

As synchrotron XRF instrumentation has advanced, the amount of data that can be acquired has grown significantly. At the Stanford Synchrotron Radiation Lightsource (SSRL) in 2005, an early XRF imaging system designed to study large objects in short time frames was able to achieve extremely fast scanning speeds. But this speed sacrificed the ability to collect more substantial in depth data sets. This was initially adequate for the pioneering imaging of cultural heritage objects such as the Archimedes Palimpsest at SSRL, where rapid imaging was all that was required to reveal scrubbed off texts. Imaging this manuscript required a multimodal imaging program including the higher energy levels of an X-ray synchrotron to make text visible that had been obscured by forgeries.[4] That pioneering work required development of an object mount to safely hold and scan the object in front of the X-ray beam, operational workflows, and data management protocols that have been used ever since for synchrotron XRF scanning of a broad range of objects, from the earliest printed works to fossils.

The system generated simple raster (.ras) image files where each pixel contained an X-ray count value for the binned energy range, which could be used to generate TIFF images of elemental distributions. That system utilized a single-element XRF detector (a single detector window with ~50 mm² active detection area), collecting 16 chemical element data channels (photon energy bins) simultaneously. The detector system was limited to a count rate of ~300,000 counts per second before saturation occurred. This generated issues in later years when the system was applied to fossils, which often contained very high concentrations of some elements. Additionally, this system could not satisfactorily contribute to testing increasingly complex hypotheses, which demanded more robust statistical and quantitative data.

Today, the capability and usability of XRF imaging systems have dramatically improved. Detector systems can now be coupled to encoded sample stages, where each pixel has a sample stage coordinate associated with it, allowing for direct image correlation and improved targeting of features for further analysis. Multi-element detectors are now available, increasing the active area and commensurately increasing the number of photons that can be detected in the same amount of time compared to single-element detectors. For XRF imaging at SSRL, 1-, 4-, and 7-element detectors are employed depending on the required application. Each detector element collects up to 48 user-defined data channels (chemical element ROIs). Additionally, advances in detector system electronics have enabled count rates as high as 3 million counts per second, as well as the ability to record the full XRF spectrum Multi Channel Analyzer (MCA) data concurrently for every pixel in an XRF image (4,096 data points per spectrum multiplied by the number of pixels in the image). At SSRL, Xspress3 systems from

Quantum Detectors (UK) are employed which generate files in Hierarchical Data Format version 5 (HDF5) format. This is an open-source, binary file format designed to store and organize massive, complex and heterogeneous scientific data [5], and a standard format for US national labs. All these factors have dramatically expanded the amount of information that can be gleaned from an XRF image, allowing for in-depth statistical analysis and the ability to accurately fit the data for elemental quantification for a broad range of studies, including some of the earliest printed works.[6]

More recently, the addition of optional high-resolution optical and fluorescence microscopy at the SSRL XRF imaging beamlines has increased file sizes and complexity, in terms of processing power requirements, to analyze these various data types as correlatable datasets. No commercially available software is available that can combine and analyze diverse data types, so most user facilities use custom software developed for their own particular data formats. At SSRL, the MicroAnalysis Toolkit is used as a multi-purpose data-processing program for X-ray data, and is capable of reading multiple data formats from other facilities and commercial instruments. [7]

A typical XRF experiment requires mounting a sample vertically on motorized stages with the sample surface oriented at 45° to the fixed horizontally oriented incident X-ray beam and the detector mounted at 90° to the beam. Fig. 1 shows a simplified plan-view schematic and photo of beamline 7-2 at SSRL. This beamline is the "large-scale" instrument where the majority of fossil and cultural heritage objects are analyzed. Today, many paleontological research groups utilize synchrotron radiation and a host of other analytical techniques to study ancient life with these types of fossil mounts.[8]

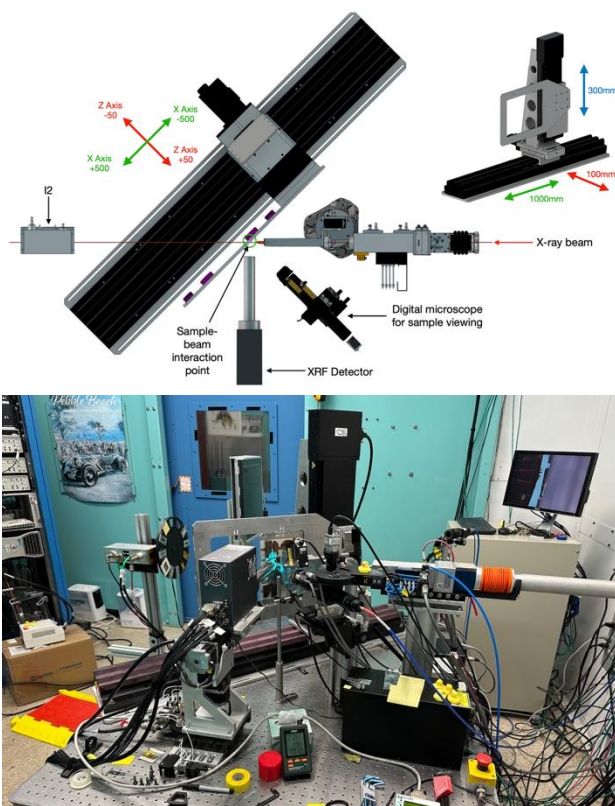


Figure 1. Plan view diagram and photo of SLAC SSRL Beamline 7-2 with components for XRF, UV and other optical data collection. (Nicholas Edwards)

Synchrotron XRF Imaging of fossils at SSRL

The fossils studied include some of the earliest known vertebrates, dinosaurs, early birds, invertebrates, and plants. In multiple cases, the distribution of chemical elements within fossils has been shown to correlate with specific tissues, remain distinct from the surrounding geological matrix, and in some instances, indicate the existence of specific biosynthetic pathways.

For example, in 2010, synchrotron XRF imaging enabled the first comprehensive chemical characterization of feather and bone chemistry across an entire *Archaeopteryx* fossil,[9] a pivotal specimen discovered just after Darwin published *On the Origin of Species*, solidifying his theory of evolution and providing the first indicator that birds are descendants of dinosaurs. Prior to the application of synchrotron imaging, *Archaeopteryx* fossils (to the best of the authors' knowledge) were not permitted to be subjected to chemical analysis owing to their exceptional rarity (only 14 specimens exist) and cultural significance.

Another study in 2012 enabled XRF imaging of an entire 120-million-year-old early bird (*Confuciusornis sanctus*) with preserved feathers.[10] The presence of copper in the preserved feathers, combined with similar observations in other fossils and modern organisms, provided the first indication that elemental composition of soft tissues could be derived from the chemistry of once-living organisms and could indicate the presence of specific biosynthetic pathways—in this case, pigmentation.

A follow-up study in 2016 used XRF imaging to investigate the relationship of elements such as calcium, copper, zinc, and sulfur to melanin pigment patterns in modern bird of prey feathers.[11] This study showed that pigment patterns observable in visible light correlated with the distribution and variation of these important nutrient trace elements and that the relative amounts of these elements could be indicative of specific coloration.

More recently, a 2026 study used synchrotron XRF imaging combined with fourier transform infrared spectroscopy to investigate two Silurian (~444–423-million-year-old), soft-bodied early jawless vertebrates, whose anatomy has long been disputed due to the ambiguous origins of the biomineralized skeleton.[12] XRF imaging provided the first unequivocal evidence for biomineralized scales in these organisms, indicated by elevated calcium and phosphorus. A close correlation between zinc and copper distribution in the structure believed to be eyes, provided evidence for retinal pigments and complex eyes, along with potential traces of other sensory anatomy not visible under visible light. These findings revealed important new information about early vertebrate anatomy, indicating that both bone and complex eyes originated in stem-group vertebrates.

A review of all XRF paleontological research is beyond the scope of this paper, which will focus on synchrotron XRF imaging data from a 90-million-year-old octopus from Lebanon (BHI-2315). Synchrotron XRF imaging is ideal for such a fossil due to the previously discussed benefits. Octopuses are almost entirely soft-bodied and inhabit ocean environments that are not readily conducive to soft-tissue preservation. This makes octopus fossils exceptionally rare and even less likely to be made available for chemical analysis. Additionally, the entire specimen is too large to fit in any other type of instrument at ~30 cm in length and 25mm thick in places.

Fig. 2 shows an illustrative photograph of the fossil with a grey-scale image of zinc distribution below. Detailed discussion of what these distributions reveal is beyond the scope of this paper and is retained by the authors in preparation for a future publication.

The imaging data set consists of a series of XRF images acquired to probe different features of this fossil at various resolutions. Due to the current implementation of limits on both the array size and memory for the imaging system, there is a maximum of ~2000 pixels per line of image data. This currently constrains the relationship of map resolution and width. i.e., a 2000 pixel row of with 100 micron resolution per pixels has a maximum physical width of 200 mm. For a large sample such as this, to image the whole surface required separating the imaging into 2 parts. At 80 micron resolution, the two parts were 1990x1189 and 1378x1189 pixels respectively, generating a total image of just over 4 megapixels for each of the 16 elements. The total time for collecting these 2 large areas was ~18 hours. For higher resolution imaging in two smaller areas of interest, 35 micron resolution was used generating ~0.2 megapixel images. Combined with some calibration data, the resulting size for the binned 16-element data set was ~4GB, and the full XRF spectrum MCA data was ~77GB. With planned upgrades to increase the pixel line size, file sizes and pixel density can be expected to increase in the future.

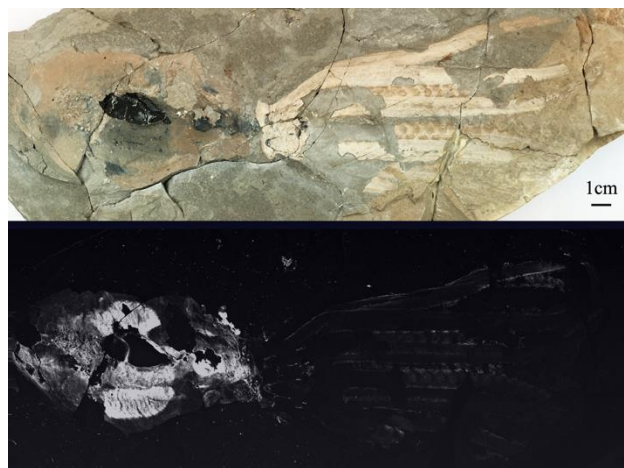


Figure 2. Image of cephalopod fossil produced with natural light and greyscale image of Zinc via XRF scanning at SLAC SSRL.

Multispectral Imaging of Fossils

Digital multispectral imaging has been used to study cultural heritage objects for over 2 decades. This has included system development and integration of not only the latest imaging and illumination technology, but also the post-processing, work processes and operational skills for efficient digitization and digital studies. Current systems use a medium-format, high-pixel-count camera to take a series of 16-bit TIFF images, each illuminated by a specific wavelength of light from narrowband light-emitting diodes (LEDs) in ultraviolet (UV), visible and infrared (IR) spectra. Post-processing of the captured image cube of monochrome images expands the utility of multispectral imaging. To fully capitalize on multispectral imaging, integrated open-source statistical software is used to digitally process and combine the resulting image stack to reveal features in the object that are not visible in natural light.

The image files contain both captured and flattened registered images, as well as “flats” calibration images and processed images of groups or “stacks” of images collected in UV, visible and IR wavelengths. Multiple narrowband illuminations and at times filters were used during imaging. Each image stack of the cephalopod fossil comprised fifteen images in each wavelength of light

including five shorter wavelengths with filters to detect fluorescence emitted by the fossil chemical components (Fig 3).

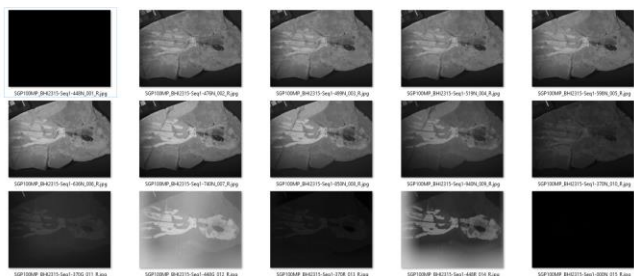


Figure 3. Sequence or 'stack' of captured narrowband multispectral images of cephalopod fossil.

Narrowband multispectral imaging of fossils during manuscript imaging in the Stanford University Library in March 2018 demonstrated this technology's potential to support research into paleontology. An integrated multispectral imaging system with semi-automated system operations and post-processing tools that produced pseudocolor images (Fig. 4) provided paleontologists and other researchers with quality optical images of fossil specimens.

This multispectral imaging allows rapid research input developed in-house, minimizing the need for external technical support and delays in getting results. It can be used for assessment of fossils on site or in a lab, including prior to more advanced XRF imaging. Scientists are evaluating the results from this imaging to determine if it offers an initial imaging assessment to identify fossil samples and regions of interest with potential for XRF imaging. This could assist in prioritizing and justifying transportation of a fragile object to an expensive research instrument. Initial multispectral imaging offers new opportunities for collaborative interdisciplinary research into the fossils, with data access and collaboration in support of global research teams.

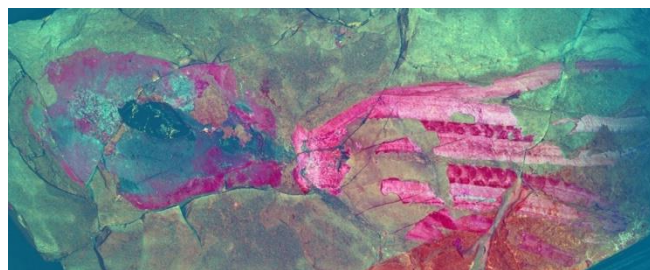


Figure 4. Image of cephalopod fossil produced with PCA post-processing of multispectral images captured in multiple wavelengths.

Integrating Multimodal Image Collection

Integrating multimodal data sets is highly informative but can also be extremely labor intensive. There are a few ways in which this workload can be mitigated and data quality improved. First, one can attempt to collect as many different data types from a single experiment. This would minimize sample movement thus removing the need for complex data alignment processes. However, most instruments are not multimodal, as they require specific conditions in order to perform optimally. Second, one can try to standardize the sample mounting method between instruments where possible. That would at least maintain sample orientation, again reducing the data alignment processing requirements. SSRL has engineered sample mounting systems that allow samples to be transferred between multiple beam lines with different capabilities, a ThermoFischer FTIR imaging microscope, and a Sigray Attomap XRF microscope,

all without unmounting a sample from the primary holder. Fiducial markers with a known pattern placed on samples could also provide an origin point to allow the same areas of interest to be found quickly and with greater precision when moved between instruments. It would improve alignment workload as a known pattern can be used to correct image alignment and distortions.

The imaging beamlines at SSRL have integrated several imaging and scientific analysis modalities into the data capture chain at the beam lines. Having multiple data collection modalities at the same point of data collection as XRF helps to minimize time spent in the post-processing steps to do processes such as image registration and alignment. The primary source of visible data imagery is via a motorized zoom and focus microscope with an 8MP CCD camera, oriented perpendicular to the sample and aimed at the beam-sample interaction point. This camera provides the live video feed for setting up the XRF scan, but mosaics can also be collected using the same map grid area as an XRF scan. A 50% overlap of images is used to maintain even illumination when stitched. Several lighting options are available, from white LED backlighting to reflectance imaging using an 8-channel Lumencor LED solid state light engine, capable of illumination from the UV to IR. An optical filter can be rotated into the optical path of the microscope via motorized filter wheel. While any 12.5 mm filter can be used, the beam lines have a typical selection of a UV cutoff (> 400 nm), and band pass filters typically used for biological imaging (DAPI, GFP, mCherry). These options allow for the collection of autofluorescence images as well as imaging of biological fluorescence markers. The different lighting and filtering options are programmable and be run in a queue to collect a series of different hyperspectral optical images.

In addition to the CCD camera, the optical path is sent through a beam splitter before the optical filter to a fiber optic port. This port is connected to an OceanOptics fiber spectrophotometer, which can be used to collect the entire UV-visible reflectance spectrum at each data collection pixel.[13] The spectrophotometer can be configured with an adjustable input slit to trade light intensity with spectral resolution. Moderate resolution (~ 2 nm band pass) requires some reduction in the overall intensity and this requires a slightly longer data collection dwell time (~ 100 ms as opposed to 5-10 ms).

Data Management

The true power of the data collected from these various imaging systems can only be unlocked with global access to the data and in a format that is not only physically accessible, but also scientifically useful to those without prior knowledge or experience with the techniques. Converting all the images to TIFF, including SSRL XRF elemental scans offers a common image format that is accessible to a broad range of systems and users, and can be processed with widely available image processing tools. To fully capitalize on these images all the capture information must be stored in the metadata and in the images.[14]

The advanced imaging data set comprises narrowband multispectral images and XRF scans, as well as other scientific data. Multispectral data comprise image sets captured in the labs of the Stanford University Libraries Conservation Services. The XRF data set provides sequences or stacks of captured and registered images converted to TIFF and JPEG thumbnail images with metadata. These images are retained as archival images and can be read with most image viewers. The multispectral image data set includes:

- Captured Image data consisting of captured TIFF image files converted from proprietary or open source standard images. each of the illuminations and hdf5 images from the detectors.

- Digitally processed images that have been produced through the application of computer algorithms to combine and enhance captured images to enhance visibility of features of interest. Multispectral images were statistically processed with Principal Component Analysis (PCA) and/or hdf5 images processed using the MicroAnalysis Toolkit.
- JSON Files containing metadata about the image capture, the object, and the image data.

The XRF data collection system saves several data files during a data collection run. The main data are stored in a hdf5 format data file which contains binned ROI data for each element of interest, and other data channels at the beam line, as well as basic information on the grid size, data collection step, and incident X-ray energy. A human readable DAT text file is also provided. A series of HDF5 files, one for each line of the XRF map, are also created from the Xspress3 system that contains the full X-ray fluorescence spectrum (MCA spectrum) collected at each pixel. This data can be used in post-processing to perform re-binning if a new element of interest is required, or detailed spectral shape fitting to resolve overlapping emission peaks. This latter data set can be many gigabytes in size depending on the size of the XRF image. A running logbook file is created by the data collection system that saves the data collection parameters for each XRF image, as well as a more detailed log that collects all of the data collection sequence data. These logs can be used to maintain a full record of the data collection processes during a particular beam session. A file is also created that records the starting positions of the beam line motors and the photon fluxes throughout the beam line for each of the major configurations in which the beam line was used.

The XRF data is initially stored on the local drive of the beam line computer and automatically mirrored to a centralized networked data storage system as the data is collected. This ensures full data fidelity in the case of any network interruptions. The larger MCA data is saved locally and tar/zip archived into a single file for ease of transfer and storage. Users can access their data via secure FTP for remote download.

Data Integration Potential

With access to a broad range of data, post-processing and correlation of multiple data types from different technologies can provide additional critical information for analysis. This can include combining and integrating images created from different XRF elemental or spectral optical responses, as well as integrating images from the different imaging modalities.

Post-processing of the raw XRF data can take several forms and produce new data channels that contain MCA fitting analysis of X-ray emission peaks, PCA analysis, concentration analysis, among many others. The MicroAnalysis Toolkit, used for data integration and analysis, is able to read data formats from a large number of sources, including most synchrotron XRF beam lines (from US and international sources), commercial FTIR and Raman instruments, imzML format for mass spectroscopy imaging, JPG, GIF, BMP, TIFF and qTIFF. These data can all be processed and registered with the toolkit and common statistical analyses and correlations can be performed. Data can be output in many formats, including the SSRL hdf5 format, raw Python Numpy matrices, JPG, TIFF, and OMETIFF. [15] The latter is an open-source format based on the standard pyramidal TIFF format that supports multi-layer TIFF data and is generally accessible by most image readers.

With funding agencies now requiring data in FAIR (findable, accessible, interoperable, reusable) formats, and the current trends of making data “AI-ready”, the largest challenge is the integration

of data with the appropriate metadata to make relevant data machine findable and associated with other related data. The SSRL imaging group has been addressing these needs within the umbrella of the US Department of Energy, Biological and Environmental Research (DOE-BER) to integrate X-ray imaging data with other modes of data and information. Here, the LAMBDA schema[16] format is used, which defines much of the experiment and facility-based metadata using a linkML data structure. Sample contextual metadata can be included as well, utilizing schema slots from NMDC definitions,[17] which are layered on the GSC MixS standard [18]. While these schema and standards are very biologically focused, many of the slots are multi-purpose and work is on-going to expand and use standards from other fields as well. At the time of this publication, involvement in the program is voluntary, but as the program develops, this could become a standard data archiving process.

At SSRL a set of Python scripts are used to assemble all experiment and machine related metadata from the beam line experiment logs, as well as sample and preparation metadata from the user to make a combined metadata manifest. This ideally contains all of the relevant metadata to define the samples, their preparations, and the analysis and can be used to link the XRF data to other modes of data collected. Data is processed following standard pipelines to extract images of individual elements in units of concentration (fg/pixel) and stored in OMETIFF format. The use of fg ensures that the individual pixel data retains sufficient significant precision when stored in the integer-based TIFF format. OMETIFF is chosen at this point for simple usage by machine-learning based tools, but future development and agentic tools may choose to access the raw MCA data files rather than relying on human processing or other data processing pipelines. The processed data files and the metadata manifest is stored in an accessible location on the SLAC Shared Science Data Facility (S3DF) and connected to the appropriate data lakehouse management system as needed (i.e. the DOE-BER BRIDGE program) for data retrieval by human or machine-base interfaces. Future work plans to leverage DOE Genesis infrastructure projects such as the American Science Cloud for data manipulation and agentic workflows.

The rise of powerful data analysis tools, including Artificial Intelligence (AI), may ease the integration of image data in different formats, such as hdf5 data from SSRL and 16-bit TIFFs from multispectral imaging. The methods used for integrating data from different imaging and scientific modalities have applications for data from different facilities, techniques and disciplines.

The imaging systems, data management and work processes cited in this paper offer potential for the natural and cultural heritage community to tap the scientific and technical communities more robustly funded developments in technology and methodology. Even at the local level of an under-resourced archive, museum or library, these developments may provide insights into methods for preparing data to capitalize on ongoing developments in imaging, data processing and analysis.

The increase in accessibility to techniques and the commensurate increase in data types and size, can be overwhelming to a researcher. Cross-disciplinary collaboration remains critical to ensure all these data can be coalesced into useful, reliable and publishable research. The development and increased use of more powerful AI and machine learning tools to reveal information contained within ever expanding data sets will offer new opportunities to integrate different types of data and metadata from different modalities. Agreement on and use of data exchange guidelines ranging from the LAMBDA schema to the International

Image Interoperability Framework (IIIF) can help ensure continued preservation and open access to a broad range of data from ever larger and more complex data systems by diverse groups of end users from multiple research disciplines.

Acknowledgements

This paper and the application of advanced technical imaging builds on the collaboration of people from multiple disciplines. The utility of advanced imaging is dependent on the patient digging and paleological research of Pete Larson of the Black Hills Institute of Geological Research (South Dakota, USA), Prof. Roy Wogelius (University of Manchester, UK), Prof. Phillip Manning (Natural History Museum Abu Dhabi, United Arab Emirates, and University of Manchester, UK). The imaging methods cited in this paper were initially developed by Dr. Uwe Bergmann (University of Wisconsin, Madison, USA) for the XRF scanning work on the Archimedes Palimpsest. Dr. William Christens-Barry (Equipoise Imaging, Ellicott City, USA) played a critical role in the development of narrowband multispectral imaging and post-processing systems, with the illumination panels he developed becoming the model for those used in several commercially available systems. Doug Emery (University of Pennsylvania, USA) developed many of the basic multispectral imaging metadata schemas and concepts in his pioneering role as data manager on the Archimedes Palimpsest project. Bill Sellers (University of Manchester, UK) developed some of the first machine learning algorithms to detect undertext in XRF palimpsest scans. Cultural heritage objects could only be imaged at SSRL and the Stanford University Libraries with the professional support of Elizabeth Ryan and Kristen St. John and others in Conservation Services.

Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

References

- [1] J.I. Goldstein et al. Scanning Electron Microscopy and X-ray Microanalysis (2017). <https://doi.org/10.1007/978-1-4939-6676-9>
- [2] H. Wiedemann, Synchrotron Radiation, Springer Berlin, Heidelberg (2003).
- [3] U. Bergmann, et al. Chemical Mapping of Ancient Artifacts and Fossils with X-Ray Spectroscopy in Book - Synchrotron Light Sources and Free-Electron Lasers: Accelerator Physics, Instrumentation and Science Applications. Eds: E.J. Jaeschke, et al. Springer Nature. DOI: 10.1007/978-3-319-04507-8_77-1. (2019)
- [4] U. Bergmann, Imaging with X-ray fluorescence, in The Archimedes Palimpsest (Cambridge University Press, Cambridge, UK, 2011)
- [5] N.P. Edwards et al., A new synchrotron rapid-scanning X-ray fluorescence (SRS-XRF) imaging station at SSRL beamline 6-2. J. Synchrotron Radiat. 25, 1565–1573 (2018)
- [6] S. Minhal Gardezi et. al., "X-Ray Fluorescence Imaging of Early Moveable Type Print" in From Jikji to Gutenberg: The origins of book printing from moveable metal type, eds. C. Baker and R. Silverman, The Legacy Press, pp 605-612 (2026)
- [7] S. M. Webb, The MicroAnalysis Toolkit: X-ray Fluorescence Image Processing Software. AIP Conf. Proc. 1365, 196 (2011).
- [8] Prado et al, Synchrotron radiation in palaeontological investigations: Examples from Brazilian fossils and its potential to South American palaeontology, Journal of South American Earth Sciences, Volume 108, 2021, 102973,
- [9] U. Bergmann et al., Archaeopteryx feathers and bone chemistry fully revealed via synchrotron imaging. Proc. Natl. Acad. Sci. U. S. A. 107(20), 9060–9065 (2010)
- [10] R.A. Wogelius et al., Trace metals as biomarkers for eumelanin pigment in the fossil record. Science 333(6049), 1622–1626 (2011)
- [11] N.P. Edwards et al., Elemental characterisation of melanin in feathers via synchrotron X-ray imaging and absorption spectroscopy. Sci. Rep. 6, 34002 (2016)
- [12] Reeves et al, Early vertebrate biomineralization and eye structure determined by synchrotron X-ray analyses of Silurian jawless fish. Proc Biol Sci 1 January 2026; 293 (2063): 20252248. <https://doi.org/10.1098/rspb.2025.2248>
- [13] X. Wu, S. Sanchez-Cortes, I. Kakoulli, Carminic Acid Based Red Dye from Scale Insects Detected in Red Ruby-Crowned Kinglet Feathers by Surface-Enhanced Raman Scattering, ChemPlusChem 86, 1074. (2021)
- [14] D. Emery, M. B. Toth and W. Noel, "The convergence of information technology and data management for digital imaging in museums", Museum Management and Curatorship, 24: 4, 337 – 356 (2009).
- [15] OME, "OME Data Model and File Formats 5.6.3 documentation", The OME-TIFF format (2018), retrieved on 21 June 2026, <https://docs.openmicroscopy.org/ome-model/5.6.3/ome-tiff/>
- [16] Dept. of Energy, "Lakehouse-enabled AI-ready multi-modal bioimaging data architecture (LAMBDA)" retrieved on 21 June 2026, <https://lambda-doe.org/>
- [17] NMDC, "National Microbiome Data Collaborative Schema Documentation", retrieved on 21 June 2026, <https://microbiomedata.github.io/nmcd-schema/>
- [18] Genomic Standards Consortium, "MixS GSC Project", retrieved on 21 June 2026, <https://www.gensc.org/pages/projects/mixs-gsc-project.html>

Author Biography

Dr. Nick Edwards is a Staff Engineer at the Stanford Synchrotron Radiation Lightsource (SSRL) at SLAC National Accelerator Laboratory in Menlo Park, California. He supports the XRF imaging beam line experimenters and beam line development with a particular interest in designing sample holders for a wide range of objects.

Dr. Samuel M. Webb is a Lead Scientist at SSRL at the SLAC National Accelerator Laboratory in Menlo Park, California, where he operates the X-ray Imaging Program across four beam lines at SSRL. He has his Ph.D. in Civil and Environmental Engineering from Northwestern University, a Master of Science in Environmental Engineering and Bachelor of Science in Geochemistry from Caltech.

Michael B. Toth, is President of R.B. Toth Associates LLC, and has been a SLAC SSRL User since his initial work there on the Archimedes Palimpsest in 2006. In the intervening 20 years he has imaged a range of cultural heritage objects at SSRL, from palimpsests to early Eastern and Western printed works.