Eur. Phys. J. Plus (2026) 141:445
https://doi.org/10.1140/epjp/s13360-025-07133-y

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

Multimodal manuscript representation: toward an open repository for manuscript research data

Hans Clausen[1,a], **Simon Brenner**[2,b], **Manfred Schreiner**[3,c], **Willfried Vetter**[3,d]

[1] Department of Digital Humanities, University of Graz, Elisabethstraße 59, 8010 Graz, Austria
[2] TU Wien, Favoritenstraße 9-11/193-1, 1040 Vienna, Austria
[3] Academy of Fine Arts Vienna, Augasse 2–6, D1.17.2–17, 1090 Vienna, Austria

Received: 21 November 2024 / Accepted: 29 November 2025

Abstract Research on historical manuscripts is increasingly supported by technical disciplines, which are used, for example, to digitally restore degraded text or to analyze the composition of inks, pigments, binders, and the materials constituting the manuscript's support. The Centre of Image and Material Analysis in Cultural Heritage (CIMA) in Vienna is an inter-university research institution that fosters collaboration between philology and the natural sciences. CIMA has generated a large amount of diverse research data and findings concerning numerous manuscripts. These data exist in various formats and are distributed across different storage media at the participating institutions. This paper describes the conceptual framework and technical implementation of a digital research data repository designed for the long-term preservation and dissemination of CIMA's large volumes of heterogeneous data. The repository provides the research data related to the examined manuscripts for download and reuse by interested users, in accordance with open-access principles. Furthermore, the diverse data are visualized through a user interface, presenting different research insights derived from a single physical manuscript. This multimodal representation of a manuscript is intended to serve as an accessible starting point for interdisciplinary investigation, data exploration, and teaching.

1 Introduction, objective and scientific approach

Manuscript research is increasingly benefiting from advances in the digital age and the resulting interdisciplinary collaborations between the humanities, computer science, and the natural sciences [1, 2]. Digital photography, advanced imaging techniques, and various spectroscopic methods have become valuable tools for the investigation of cultural heritage. The *Centre of Image and Material Analysis in Cultural Heritage* (CIMA) in Vienna [3] exemplifies such cooperations and was established as an inter-university research institution.[1] CIMA was founded to strengthen and expand a long-standing collaboration[2] by establishing a mobile laboratory that offers its services to universities, libraries, museums, and other institutions in Austria and abroad. Within this collaboration, computer vision specialists use multispectral imaging (MSI) to enhance the legibility of damaged written text, thereby facilitating the work of philologists. Moreover, nondestructive and noninvasive material analysis aims to determine the composition of inks, pigments, and binders used for texts and illuminations—an area of particular interest to art historians. Additionally, spectroscopic analysis enables the characterization of the writing support, providing further information that can assist in determining the manuscript's place and time of origin. The technical approaches complement traditional philological, art-historical, and conservation methods, enabling CIMA to study manuscripts from the perspectives of the various disciplines and with regard to diverse physical properties.

The methodological diversity within CIMA has resulted in a large volume of diverse datasets and research findings related to numerous investigated manuscripts. The data and results generated by the participating institutions have been stored on various media in multiple formats. These include extensive volumes of digital photographic images, image-processing data, and raw spectroscopic data, as well as material analysis and conservation reports, and manuscript descriptions and transcriptions. This diverse accumulation of data and research findings has been utilized in a range of scientific publications, including descriptions of innovative research methods [5, 6] and material-analytical studies on selected manuscripts [7]. Philological studies—such as the critical edition

[a] e-mails: hans.clausen@uni-graz.at; hans.clausen@posteo.de (corresponding author)
[b] e-mail: sbrenner@cvl.tuwien.ac.at
[c] e-mail: m.schreiner@akbild.ac.at
[d] e-mail: w.vetter@akbild.ac.at

[1] University of Vienna (Philology), TU Wien (Computer Science), Academy of Fine Arts Vienna (Chemistry and Material Science), Danube University Krems (Restoration and Conservation).

[2] Within Vienna Archaeographic Forum [4] and several Austrian Science Fund projects.

Published online: 22 April 2026

of *Psalterium Demetrii Sinaitici* [8]—which combine classical philological approaches with the aforementioned technical methods, have demonstrated the added value of CIMA's interdisciplinary research. These publications have greatly benefited from the data produced through technical analysis. However, while such traditional forms of publication effectively communicate research results, they are constrained by media limitations and cannot provide the original data in a form that allows full reuse—for example, for the reproduction of published results, independent research, machine learning, or teaching purposes. To overcome these limitations, researchers at CIMA have collaborated with colleagues from the Digital Humanities at the University of Graz to enhance dissemination through a digital research data repository. This repository was developed as part of the project *Digital Transformation of the Austrian Humanities* (DiTAH) [9]. Its objective is to provide interested users with access to the extensive datasets and thereby enable them to gain comprehensive insights into the investigated manuscripts.

This paper describes the conceptual framework and technical implementation of the repository. The repository is designed around the following key features:

First, the repository ensures the long-term preservation of CIMA's extensive and heterogeneous body of data and research results [10].

Second, it provides open access to the data in a curated form. All data and the associated metadata shall be available in open, standardized formats according to the FAIR principles [10].

Third, it visualizes the diverse data through a user interface, presenting varied research insights derived from a single physical manuscript. This multimodal representation of a manuscript serves as an accessible starting point for interdisciplinary investigation, data exploration, and teaching [11].

To enable the appropriate archiving and integrated presentation of CIMA's research data as multimodal manuscript representations, first, a thorough understanding of the diverse methods applied by CIMA for manuscript investigation is required. Accordingly, Chapter 2 outlines the research objects and methods relevant to the repository—particularly the technical ones. A deeper understanding of these methods, their disciplinary contexts, and the resulting data allows the identification of potential strategies for interlinking related data. This enables the generation of coherent multimodal representations of the investigated manuscripts and the development of an appropriate repository design. The conceptual framework and implementation are described in Chapter 3.

2 CIMA's research objects and methods

CIMA is specialized in the investigation of textual heritage on parchment (handwritten codices, book scrolls, or charters) in a fragmentary state. Comparable initiatives can be found in Hamburg [12] and, to some extent, in California [13]. For this purpose, CIMA has developed a comprehensive analysis package of hardware and software [14, 15], whose application is not limited to written material alone, but also extends to paintings, graphics, archeological artworks, and other objects [16].

Most investigated manuscripts originate from the Middle Ages and are written in various languages and scripts. The primary focus so far has been on Slavic (Glagolitic and Cyrillic), Greek, and Latin manuscripts from the Austrian National Library and several Austrian monasteries dating from the 8th to the fourteenth century. Within this selection, two categories of objects are prioritized: poorly preserved sources or manuscripts containing overwritten text (palimpsests), which pose particular challenges to philological investigation; and manuscripts featuring remarkable colored decorations (initials, miniatures, etc.), which are of interest not only from art-historical and philological perspectives but also from art-technological and chemical point of view.

Usually, philologists select the manuscripts of interest for investigation based on their research interests. CIMA supports these analyses with technical methods, which must be nondestructive and noninvasive, so that the objects are not altered during or after analysis.

2.1 Multispectral imaging

As described above, the computer vision specialists aim to digitally restore damaged writings by applying multispectral imaging (MSI). To further increase the contrast of degraded characters and enhance the readability of overwritten or damaged texts, several post-processing techniques are applied to the MSI images. As a noninvasive and nondestructive imaging technique, MSI can be applied to any object, including very fragile ones.

Usually, MSI is applied to manuscripts with multiple text layers (palimpsests), as demonstrated by the well-known *Archimedes Palimpsest Project* [17]. In addition to palimpsests, poorly preserved writing can also have other causes. Loss of legibility may result from deliberate erasure or accidental damage, such as soiling, mold, exposure to heat or water. Additionally, archival degradation due to natural aging is often accelerated by intensive handling and inadequate storage conditions.

MSI is an imaging technique that captures images of the same scene across multiple narrowband spectral ranges. The human eye, with its red, green, and blue receptors, is sensitive only to radiation within the 380–700 nm range, known as the visible spectrum (VIS). MSI provides additional information about an object by extending this range into the ultraviolet (UV) and infrared (IR) spectral ranges. This imaging technique is particularly effective when the object under study consists of materials with differing optical properties. Good examples are historical manuscripts made of parchment written with iron-gall inks. Under UV illumination,

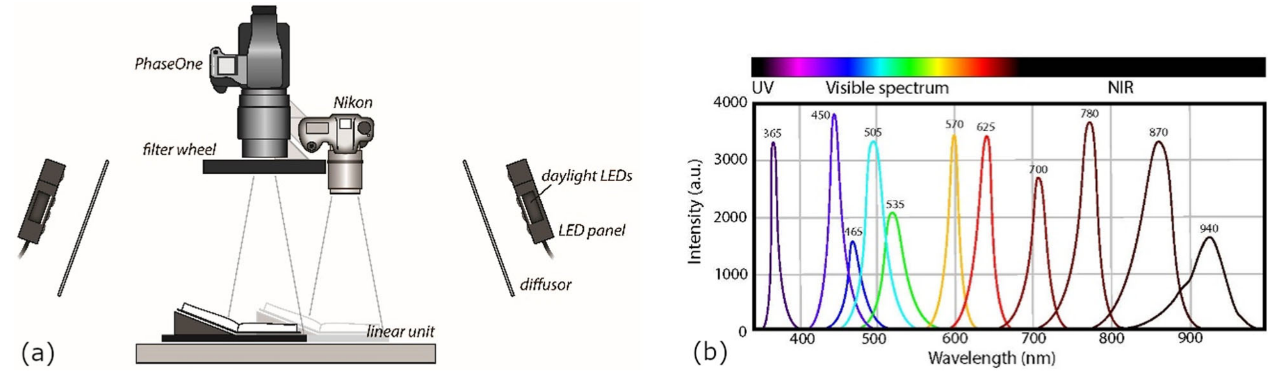

Fig. 1 MSI acquisition setup for the digital reconstruction of damaged writings: **a** The manuscript under investigation is placed on a board mounted on a linear stage, which allows for automatic shifting between the two camera positions. Illumination is provided by an Eureka!Light™ LED system featuring 11 discrete wavelengths. For visible light imaging, additional white LED panels are mounted to the left and right of the Eureka panels. Furthermore, two diffusers are positioned between the light sources and the object to ensure uniform illumination across the surface. **b** The Eureka!Light™ LED system provides illumination across a spectral range from 365 to 940 nm. When combined with the dedicated cameras, it enables image acquisition in the ultraviolet (UV), visible (VIS), and infrared (IR) spectral ranges. C. Rapp, G. Rossetto, J. Grusková, G. Kessel, New Light on Old Manuscript. https://doi.org/10.1553/978OEAW91575s373

the parchment typically fluoresces, whereas iron-gall ink does not [18]. This difference significantly enhances the contrast between the text characters and the background in UV fluorescence images compared to those taken under visible light. Faded writings often appear most clearly in UV fluorescence images, while darkened scripts or texts written with carbon-based inks are usually better revealed under infrared illumination.

To apply MSI to manuscripts in situ, a new portable system was designed and built within CIMA (Fig. 1). It comprises two different cameras, two multispectral LED panels, and various accessories. The cameras are: (1) a traditional RGB camera (Red Green Blue—Nikon D4, 4928 × 3280 pixels), used for capturing UV fluorescence and visible light images; (2) a Phase One IQ260 achromatic multispectral camera (8984 × 6732 pixels) with spectral sensitivity extending from the UV to the near-IR range (350–950 nm). For special cases, such as manuscripts written with carbon ink or charcoal, an IR camera (Osiris, Opus Instruments) [19] is available, with a sensitivity range of 900–1700 nm and a resolution of 4096 × 4096 pixels. In two cases, additional optical filters are employed: a 400 nm long-pass filter for UV fluorescence imaging and a 365 nm short-pass filter for UV reflectography. The complete setup of the portable system is shown in Fig. 1a, and the spectra of the LED panels are illustrated in Fig. 1b.

For a single manuscript page (folio), a series of at least 13 images is captured using the Phase One camera, supplemented by two additional images taken with the Nikon D4, resulting in a total of 15 images. To enable more advanced use of the MSI data, several post-processing steps must be applied. These include the alignment of all captured images (referred to as registration) [20], as well as the application of various enhancement techniques to further enhance the visibility of the writing [21]. The most commonly used enhancement techniques are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA). Each of these represents a distinct algorithm that can be applied to multispectral image sets to reveal information not visible in the individual raw images. To improve the legibility of handwritten texts, these techniques are applied both on palimpsests and documents containing a single layer of writing. Figure 2 illustrates an example of the separation of the different layers of palimpsest writing.

2.2 Material analysis

Due to the development of new instruments for material analysis and the miniaturization of equipment over the past decades, user-friendly application of noninvasive and nondestructive techniques in libraries and museums has become possible [22–26]. The most frequently used methods are X-ray Fluorescence analysis (XRF) for elemental analysis, and Fourier Transform Infrared spectroscopy in External Reflection mode (ER-FTIR) as well as Raman spectroscopy for the compound-specific material analyses [27, 28]. These chemical investigations are expected to expand our knowledge about the materials used in manuscripts and—when collected and compared in a database—to contribute to the dating and localization of these objects. Furthermore, these techniques are, as required, nondestructive and even noninvasive, and can be performed at any location due to the portable design of the instruments.

All manuscripts investigated so far were first analyzed using XRF. Either a custom-built XRF instrument [29] or the portable XRF analyzer Elio from Bruker Optics®, Germany [30], has been applied. Both instruments were specifically designed for applications in the fields of art and archeology. They are equipped with a Rhodium X-ray tube with a maximum voltage of 50 kV together with a silicon drift chamber detector and two laser pointers for positioning. For the measurements of the manuscripts, the instruments are mounted on a tripod (Fig. 3). Usually, a tube voltage of 40 kV and a tube current of 0.8 mA with a 100 s acquisition time of each spectrum were chosen for the analyses of the inks and pigments. The beam diameter for both instruments is approximately 1 mm.[3]

[3] In the repository, the XRF spectra of the self-built XRF are named as PXRF, whereas the spectra obtained with the Elio are described just as XRF.

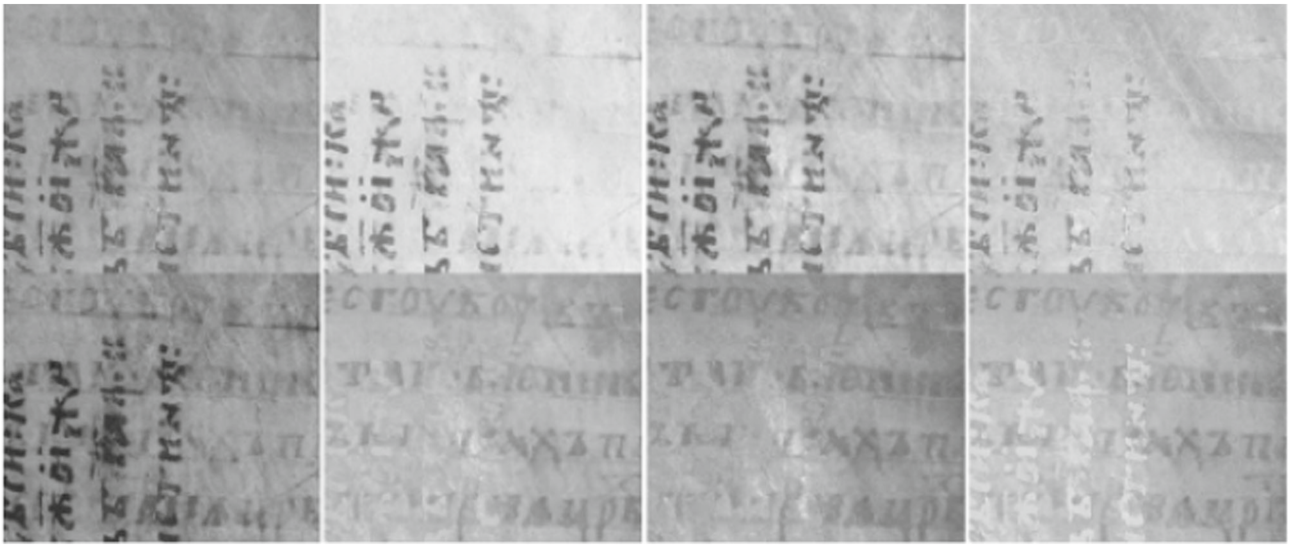

Fig. 2 Increased legibility of the palimpsest is achieved through the application of specific algorithms to the multispectral images: the first column shows an unprocessed white light image (top) and an unprocessed UV fluorescence image (bottom), both captured with the Nikon D4 camera. The subsequent columns present the separation results obtained by different analytical techniques: Linear Discriminant Analysis (LDA) in the second column, Principal Component Analysis (PCA) in the third, and Independent Component Analysis (ICA) in the fourth. The first row displays the separated overwriting, while the second row shows the underwriting. The ancient text (underwriting, upper right region of the analyzed image section) is more clearly visible in the LDA and ICA images compared to the PCA image

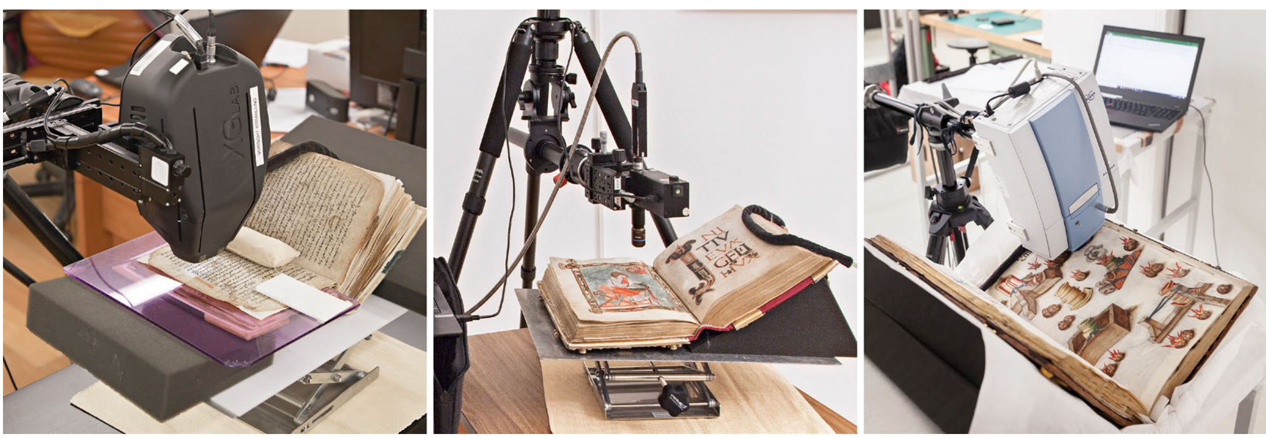

Fig. 3 Investigation of the materiality of various manuscripts in the Austrian National Library (ÖNB), Vienna, using complementary spectroscopic methods: X-ray fluorescence analysis with the ELIO instrument (left); Raman spectroscopy performed with the Pro-Raman-L-Dual-G system (center); and ER-FTIR measurements with the ALPHA Spectrometer (right)

In many cases, the inks and pigments of the text and the miniature painting can be deduced from the elements determined while taking into account the history of materials. The elemental analysis of single areas (point analysis) by XRF combined with visual inspection can often yield a certain characterization of the compounds of a pigment or ink [31, 32]. For example, evidence of mercury (Hg) in red areas implies the use of the pigment vermillion (HgS). However, sometimes elemental characterization does not allow the unambiguous identification of the chemical compound; for this reason, molecular or compound-specific techniques such as Fourier Transform Infrared spectroscopy in External Reflection mode (ER-FTIR) (Fig. 3) and, in some cases, Raman spectroscopy (Fig. 3) are used to identify organic and inorganic chemical compounds, in addition to XRF.

For this purpose, a portable ALPHA FTIR spectrometer from Bruker Optics®, Germany, with a spot size of approximately 2 mm in diameter is available. External reflection spectra (specular and diffuse reflection) can be collected in situ in the range of 7500–375 cm^{-1} with a resolution of 4 cm^{-1} over 64–128 scans using a DTGS detector. Reflectance spectra are calculated by dividing the detected intensities reflected from the analyzed area by those reflected from a gold mirror. The spectra obtained are transformed into absorption index spectra applying the Kramers–Kronig algorithm [33], which often allows a comparison to database spectra measured in transmission mode. The interpretation of the spectra is carried out by comparison to the IRUG [34] and IR Hummel Industrial Polymers [35] databases as well as a database containing spectra obtained from the material collection

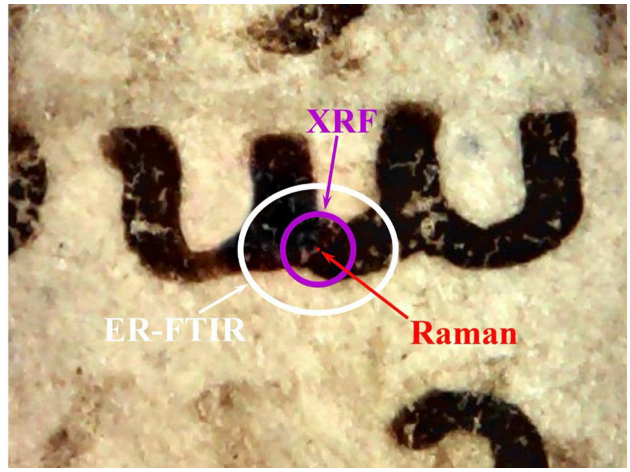

Fig. 4 Investigation of the materiality of a manuscript using complementary spectroscopic methods: Beam diameters of XRF (1 mm), ER-FTIR (approx. 2 mm), and Raman spectroscopy (few µm)

of ISTA.[4] In cases, in which Kramers–Kronig transformation does not yield usable results, mock-ups with reference materials on parchment are prepared and measured in external reflection mode (e.g., lead tin yellow powder, gelatin, solution 1%). The software package OPUS 7.5 is used for controlling the ALPHA instrument as well as the data acquisition and evaluation. The strength of FTIR lies in its ability to detect many organic and inorganic pigments, as well as other materials such as binders, glues, waxes or degradation products (oxalates). On the other hand, some materials frequently used in manuscripts, such as the semiconductor pigments vermilion and minium, cannot be detected.

Furthermore, Raman spectroscopy [36] is applied as a complementary technique to XRF and ER-FTIR. The in situ measurements are carried out with the fully integrated and transportable instrument Pro-Raman-L-Dual-G from Enwave Optronics, USA. The excitation sources applied for the investigations are diode lasers at 785 nm (~350 mW) and 532 nm (~50 mW) with narrow line widths of 2.0 cm^{-1} and 1.5 cm^{-1}, respectively. The instrument is based on a two-dimensional CCD array detector, which is temperature regulated (− 60 °C). The integrated microscope is equipped with a 1.3 Mpixel CMOS camera with an in-line LED illumination. Analyses are performed using a 50 × LWD (long working distance) objective lens, and the spectra are evaluated by comparison to an ISTA reference database. Raman spectroscopy has proven to be particularly useful for the characterization of inks (e.g., carbon black and iron-gall inks) and the detection of the semiconductor pigments vermilion and minium, for which FTIR provides no results.

Unfortunately, the analyzed area is only several microns in size, which hampers a clear comparison of ER-FTIR spectra with those obtained by Raman spectroscopy or XRF. In general, for the evaluation and interpretation of the results, it must be taken into account that the analyzed area and the excitation depth differ considerably among the three methods (Fig. 4).

2.3 Manuscript description, codicology and conservation

The scientific description of a manuscript facilitates an initial assessment for scholars and can support its philological and historical investigation. Typically, general manuscript descriptions provide information about the manuscript as a whole, including the language, script, and content of the manuscript, as well as its number of folios and their size. Such descriptions have a long-standing tradition in scholarly research, with established conventions regarding their form and content [37].

For a more in-depth description of a manuscript, selected objects are examined within CIMA by experts in codicology and conservation. Such investigations require a broad range of expertise and the application of diverse methods. A thorough understanding of the support materials—such as parchment, leather, and textiles—is essential, together with knowledge of their historical production processes, including the preparation of inks, binders, and codices. In addition, the natural aging processes of these materials and their effects on the manuscript are important considerations during examination. In cases where a manuscript is damaged, its exposed inner structure is analyzed prior to conservation. Investigations may employ noninvasive and cost-effective techniques, such as visual inspection under transmission or raking light. These methods require expert interpretation [38]. The determination of materials by codicologists and conservators can be greatly informed by the results of the material-analytical methods described above (Chapter 2.2).

2.4 Critical edition

Within the context of CIMA, the selection of manuscripts for examination is typically carried out by philologists. The philological examination of the text begins with the transcription of the manuscript and often culminates in the production of a critical edition.

[4] Institute of Science and Technology in Art, Academy of Fine Arts Vienna, database with spectra obtained in transmission and reflection mode.

The approach to textual criticism, as well as the presentation of the text in an edition, is determined by the specific research focus of the respective scholar.

Critical editions produced within the framework of CIMA [8] have greatly benefited from the various methods described above. MSI assists philologists in recovering the complete text of damaged manuscripts. Codicological and conservation assessments, complemented by material-analytical investigations, provide a deep understanding of the structure and material characteristics of a manuscript. The additional examination of manuscripts provides valuable contextual information—including the determination of their place and time of origin—and thereby supports philological and historical research.

3 Repository

The investigation of manuscripts using the research approach and technical methods described above results in a variety of media products in different modes [[39], pp. 46–54], including images (MSI and enhanced imaging techniques), spectral data (material analyses), and texts (research reports, manuscript descriptions, and transcriptions). Each of these media products can be understood as an individual digital artifact of a specific research practice, each representing a particular aspect of the investigated physical manuscript. Ranging from individual measurement data to more extensive scientific findings, this diverse information should be made accessible to users by archiving the corresponding digital artifacts in the repository for long-term preservation and providing them for download in accordance with open-access principles. The true added value of CIMA's interdisciplinary research approach, however, lies in the interlinking of all digital artifacts available for a manuscript. Therefore, the individual digital artifacts are interlinked [[39], pp. 41–46] and visualized as multimodal manuscript representations via the repository's user interface. The following sections outline the conceptual framework and the implementation of the repository.

3.1 Conceptual framework

The conception of the multimodal manuscript representation unfolds within a field of tension. On the one hand, established (digital) forms of publication in manuscript studies should be considered (Chapter 3.1.1). On the other hand, the multimodal manuscript representation is designed to embody CIMA's specific research approach by interlinking all digital artifacts into a single virtual object, which is presented via the user interface as the digital surrogate of an individual physical manuscript (Chapter 3.1.2).

3.1.1 Established forms of publication

In line with CIMA's disciplinary focus—which encompasses philological, conservation, art-historical, and art-technological priorities—three well-established forms of publication can essentially be identified, which increasingly overlap and benefit from the possibilities offered by web technologies:

Critical edition: The foundation of editorial work is the transcription of the manuscript. Depending on the disciplinary focus of the edition, various aspects of the transmitted manuscript are highlighted in the edited text, such as textual, diplomatic, or content-related features. In digital editions, it has become standard to publish the edited text alongside a digital facsimile of the manuscript in a text-image synoptic view. Moreover, medievalists have increasingly employed digital media to produce new types of scholarly editions [40], including those that integrate the results of MSI to recover damaged texts [41].

Cataloging: Manuscript collections are often indexed in catalogs to provide an overview and to facilitate the identification of items of interest. Cataloging can range from basic metadata and brief summaries to detailed manuscript descriptions, including codicological and conservation information. The process increasingly benefits from the availability of online repositories [42].

Spectral database: Spectroscopic examination of manuscripts aims to expand our understanding of the materials used. When collected in a database and compared, these data provide a valuable resource for manuscript research [34].

The user interface of the multimodal representation of manuscripts should generally align with the described established forms of dissemination.

3.1.2 Integration of individual research artifacts

The type, file format, and scholarly status of the digital artifacts of a manuscript depend on the methods applied and the preferred working practices of the researchers. To be included in the repository, digital artifacts must first meet certain fundamental requirements: they may need to be converted into an open, standardized file format suitable for long-term preservation, enriched with metadata describing the (measurement) techniques used, and, if necessary, manually consolidated, automatically processed, or subjected to scholarly revision.

In Digital Humanities, work typically begins with modeling—in this case, with the question of how to model historical manuscripts. However, the various digital artifacts already constitute models, each reflecting a specific methodological perspective on the manuscripts. At this stage, the challenge lies less in creating representations of the manuscripts than in integrating existing representations (digital artifacts) into a coherent multimodal representation. The MSI images, spectroscopic measurements, manuscript descriptions, and text transcriptions differ significantly in form and content, which makes their interlinking and integrated representation both technically and conceptually challenging.

To interlink digital artifacts to a multimodal manuscript representation, first a conceptual frame of reference must be identified: Initially, all individual types of digital artifacts can be linked to a specific manuscript. Furthermore, the various digital artifacts of a manuscript can be associated with precise locations within the physical structure of the object. While descriptions and transcribed texts typically refer to the manuscript as a whole, MSI images capture specific manuscript pages (folios), and material analysis focus on selected points on a page. Consequently, measured spectra can be assigned to exact points in the image of a manuscript page. Thus, all digital artifacts maintain a spatial relationship to the physical manuscript, alongside the underlying methodologies. The manuscript, as a three-dimensional object, serves as a shared conceptual frame of reference, within which each artifact can be positioned and interconnected for multimodal manuscript representation.

Following the described approach of integration of digital artifacts, the following different types of digital artifacts related to a manuscript are considered for integration into the multimodal manuscript representation:

Manuscript description and manuscript text (application/tei + xml) [43]: As described in Chapter 2.3 the manuscript description contains information that pertains to the entire manuscript. Therefore, the manuscript description is well suited for representing the manuscript as a whole. A manuscript is composed of individual pages, known as folios. The text of a manuscript (Chapter 2.4) is distributed across the individual manuscript folios. The transcribed text, with its content spanning multiple pages, pertains to the entire manuscript. TEI (Text Encoding Initiative) is a widely used, open format for the digital representation of both manuscript text and manuscript descriptions. For the multimodal manuscript representation, the description of a manuscript and its transcribed text were integrated into a single TEI document.

Images (image/tiff) [44]: As described in Chapter 2.1, for each page of the manuscript up to 15 images of MSI (plus possible image derivatives of enhanced techniques) are available. Each image typically corresponds to the same entire manuscript page. Therefore, each of the multiple MSI images can be understood as a virtual layer representing specific (optical) properties of the manuscript page. The full-page image in natural colors is the most intuitive digital representation of a manuscript page. It can be used as a reference image to which the other (registered) images can be aligned. Further digital artifacts that correspond to specific locations within the represented manuscript page can be linked to the reference image by specifying the corresponding image coordinates. From a technical perspective, the images should be in a lossless format to preserve the complete photometric information and spatial resolution recorded during acquisition, enabling potential further image editing and processing. For this purpose, 16-bit TIFF images are the preferred choice.

Spectroscopic measurements (chemical/x-jcamp-dx) [45]: As described in Chapter 2.2, the material-analytical investigations are conducted at selected points on the surface of a manuscript page where the material composition is of particular interest. Up to three complementary spectroscopic methods may be employed, yielding a maximum of three spectra per selected point. The measured spectra can be linked to the reference image of the manuscript page by specifying the corresponding image coordinates, allowing the location of the spectroscopic examination to be visually marked within the manuscript page. The original measurements may be in proprietary, device-specific formats. For use in the repository, they are converted into the open JCAMP-DX format, independent of the device used.

Terminology used for materials (application/skos + xml) [46]: Based on the measured spectra, the material analysts can identify the material and its composition for the examined area. As part of the repository's development, materials and chemical elements commonly found in manuscripts were represented using a hierarchical taxonomy. This taxonomy enriches the represented materials with example images and scientific information about their production and use. The location of the spectroscopic examination can be annotated with the results of the examination using the terms from this taxonomy, i.e., with the identified materials and elements. In this way, the identified materials and elements can be viewed together with the spectra within the manuscript page. The taxonomy was realized with the help of a SKOS (Simple Knowledge Organization System) document.

Terminology used for conservation findings (application/skos + xml) [46]: As part of the repository's development, conservation-related terminology was represented using a hierarchical taxonomy. The manuscript description, manuscript pages or sections of a manuscript page can be annotated with terms from this taxonomy, i.e., the corresponding findings. The taxonomy was also implemented using a SKOS document.

3.1.3 Technical interlinking of the digital artifacts

One way of combining the various digital artifacts of a manuscript is by organizing the CIMA research data according to file naming conventions and using appropriate folder structures within the repository [47]. Depending on the chosen structure, all digital artifacts of a manuscript or individual manuscript pages can be grouped and made available for download. However, such a basic grouping is not sufficient for the type of repository outlined above.

To enable a multimodal representation of the manuscript, the existing digital artifacts must be interlinked according to the relationships described in Chapter 3.1.2. This includes linking the description of the entire manuscript to its individual folios (and their text), which in turn must be associated with natural-color and other MSI images. Additionally, spectroscopic measurements and material terms must be mapped to specific locations within these images, among other connections.

This complex interlinking among the digital artifacts must be implemented through the use of supplementary digital artifacts. To ensure maximum transparency, users of the repository should be able to access and download these supplementary artifacts, enabling them to understand the decisions made in the process of interlinking. These artifacts must be provided in open, standardized formats and made accessible in the same manner as the primary digital artifacts.

To meet these requirements, a combination of two supplementary artifacts—one representing the entire manuscript (TEI document) and another representing individual manuscript pages (METS document)—is employed to realize the multimodal manuscript representation.

TEI document: In addition to its role in containing the manuscript description and text transcription (see Chapter 3.1.2), the TEI file serves for the multimodal manuscript representation in the following way. The TEI file offers several possibilities for linking the manuscript description and text transcription represented within the file to external digital resources [48–50]. On the one hand, certain segments of the manuscript description that address conservation aspects are annotated with corresponding terms from the conservation SKOS taxonomy (application/skos + xml). This allows scholarly explanations stored alongside the terms in the taxonomy to become an integral part of the manuscript description. On the other hand, individual pages of the text transcription are each linked to a high-resolution color image of the page (image/tiff), enabling the presentation of the manuscript's individual pages as a synopsis of text and facsimile.

However, the referencing and integration of external digital resources in TEI are primarily limited to linking the represented text to images (digital facsimiles) and annotating the manuscript description with textual data (material and conservation taxonomy). The independent interlinking of multiple digital artifacts—as required for the comprehensive representation of individual manuscript pages—is not supported. Consequently, within the context of the repository, TEI documents function as supplementary artifacts for representing the manuscript as a whole, but not for representing individual pages and their various (technical) digital artifacts.

METS document [51]: For the representation of individual pages of the manuscript, the multispectral images (image/tiff) should be organized as virtual layers, each representing specific (optical) properties of the page. Additionally, measurement points must be located on the reference image and linked to the corresponding spectroscopic measurements (chemical/x-jcamp-dx) as well as the relevant terms from the material taxonomy (application/skos + xml). This results in the complex interlinking of numerous digital artifacts. METS documents provide all the necessary functionalities for managing and interlinking these digital artifacts, along with their associated metadata [[51], p.15].

For this reason, the multimodal representations of the individual manuscript pages are each realized with a METS document. Using the *File section* [[51], p.46], all digital artifacts associated with a page are initially indexed. On the one hand, each digital artifact is linked to its technical metadata referenced in the *Administrative Metadata* section [[51], p.39]. On the other hand, the indexed digital artifacts are referenced and integrated into a virtual page using the *Structural Map* section [[51], p.57]. The *Structural Map* section provides a vocabulary to address the artifacts indexed in the *File section* as a whole or in parts — such as an entire spectroscopic measurement, a specific area of an image, or a term from a taxonomy—and link them in various ways. In this manner, the digital artifacts are integrated into complex structures and become components of the multimodal representation. As an established, open format, METS documents have proven to be suitable as a supplementary artifact for representing the complex structure of the individual manuscript pages.

Bidirectional linking of the TEI and METS document: With the help of a TEI document, manuscripts are represented as a whole. This representation includes a manuscript description annotated with terms from a conservational taxonomy (application/skos + xml) and a segment for the manuscript text segmented by manuscript pages, each of which is linked to external color images (image/tiff), as shown in the upper part of Fig. 5. Using a METS document, each individual manuscript page is represented as a complex structure consisting of color images, multispectral images (image/tiff), linked spectroscopic measurements (chemical/x-jcamp-dx), and terms from a material taxonomy (application/skos + xml), as shown in the middle and lower part of Fig. 5.

Together, the TEI and METS documents provide a multimodal representation of the entire manuscript, reflecting CIMA's interdisciplinary research approach.[5] To associate the individual pages (METS) with the overall manuscript (TEI) and represent the manuscript as a coherent sequence of folios along with all corresponding digital artifacts, it is necessary to establish explicit links between the two file types. TEI enables the connection of the encoded manuscript to other existing digital surrogates, thereby allowing the METS-based page representations to be explicitly mapped onto the overarching manuscript structure. Reciprocally, METS allows for the inclusion of related items, making it possible to link the TEI-based manuscript representation to the individual manuscript pages.

3.2 Implementation

The technical infrastructure required for implementing the repository must meet several critical requirements. First and foremost, it must ensure the long-term preservation of the digital artifacts and enable their download via the web. Additionally, it must be capable of processing both the supplementary digital artifacts (TEI, METS) and the associated referenced and linked digital artifacts, facilitating their presentation as parts of the intended multimodal representation of the manuscript within the repository's graphical user interface.

[5] The implementation of the multimodal manuscript representations using TEI and METS documents can be examined in detail through the examples provided in the repository: https://gams.uni-graz.at/mmmr

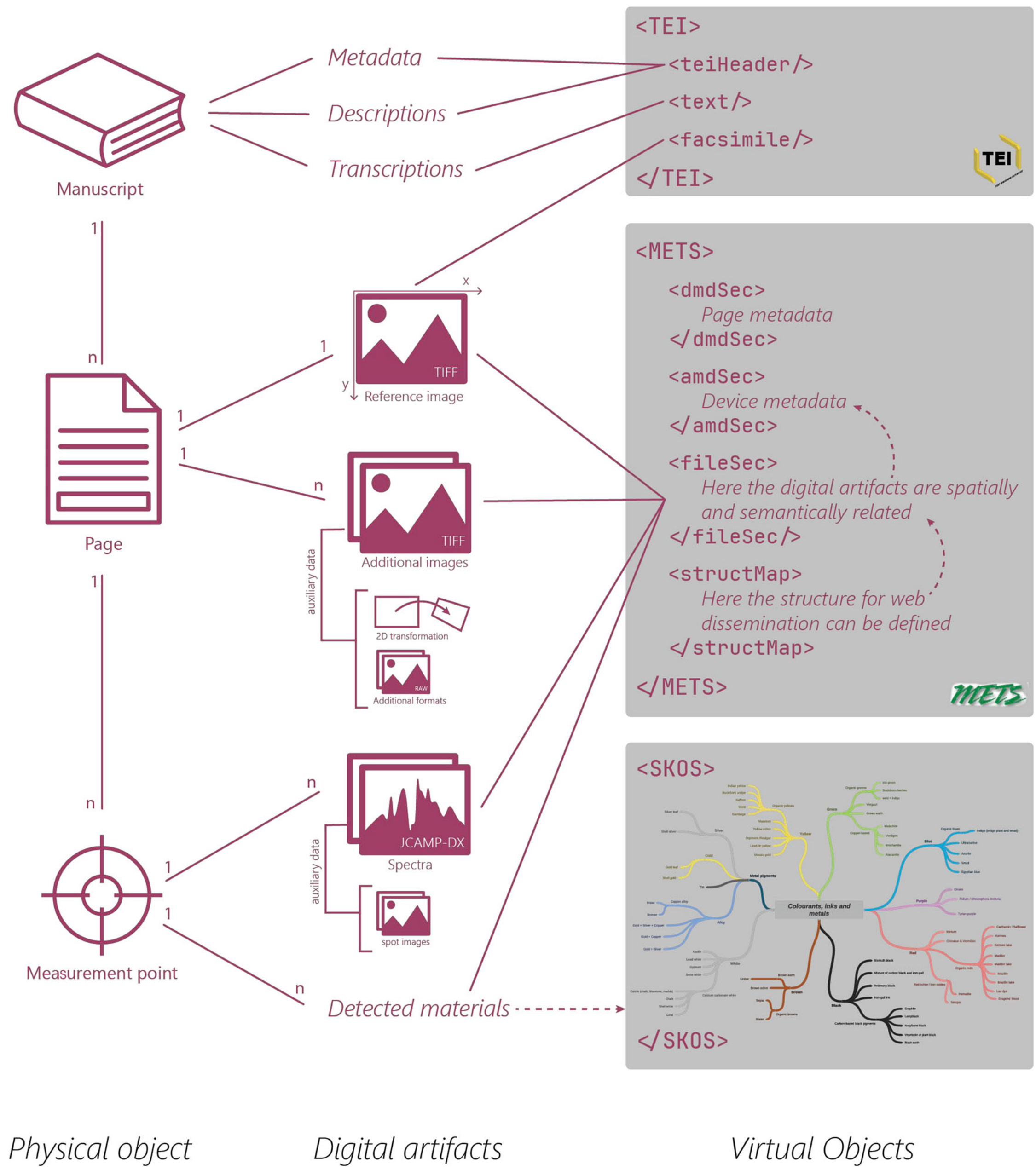

Fig. 5 Integration of digital artifacts into multimodal manuscript representations using the additional artifacts TEI, METS, and SKOS

3.2.1 Technical infrastructure

The implementation of the repository is based on GAMS (Humanities' Asset Management System) [52], a repository infrastructure certified with the Core Trust Seal [53] for Digital Humanities projects. It was developed by the Institute for Digital Humanities at the University of Graz, Austria. GAMS is built on the FEDORA Commons architecture [54], which is designed for the persistent preservation and management of digital resources deemed worthy of conservation. Within GAMS, digital artifacts of cultural heritage objects are organized and stored in virtual objects using specific content models. These models are tailored to ensure appropriate data maintenance and the provision of metadata suitable for the representation of a particular type of cultural heritage objects and its research domain. For instance, data derived from textual objects are managed with a content model specifically designed for textual content. The virtual object not only stores the digital artifacts of the investigated real object but also provides services for web distribution in various output formats, such as HTML. Regarding long-term preservation, the data stored within a GAMS object are archived in open, sustainable formats, often as XML documents. The virtual objects are assigned persistent identifiers for citation purposes through a handle server and are enriched with metadata following the Dublin Core (DC) [55] standard to enhance accessibility. Since the repository's objects are accessible on the web via the Protocol for Metadata Harvesting (OAI-PMH), they can also be harvested by metadata harvesters and are included in virtual libraries, such as Europeana [56].

3.2.2 Content models

Each GAMS object, regardless of its content model, groups and archives all digital artifacts associated with the intended representation of the corresponding cultural heritage object as datastreams, each of which can be accessed via the web. Each content model includes a specific type of primary datastream (e.g., TEI, METS, SKOS), which primarily represents the real object. Additionally, each object can contain additional datastreams (digital artifacts), which are referenced by the primary datastream and are intended to be part of the representation (multimodal manuscript representation). Furthermore, various system datastreams may be present, such as metadata about the virtual object itself or necessary stylesheets for processing the primary datastream. Each GAMS object provides various REST API-based dissemination methods [57], which can be used to request individual datastreams or to process the primary datastream into an HTML representation (visual user interface of the multimodal manuscript representation) using an XSLT stylesheet.

As described in Chapter 3.1.3, the manuscript should be represented in its entirety by a TEI document, and its individual pages should each be represented by a METS document, with each functioning as a supplementary artifact representing the structure of the multimodal manuscript representation as described. For this reason, the manuscripts in the repository are represented, on the one hand, by a virtual object of type cm:TEI, with the TEI document as the primary datastream, and on the other hand, by virtual objects of type cm:Spectral, each with a METS document as the primary datastream. In addition, two objects of type cm:SKOS are created: one containing the material taxonomy and the other containing the conservation taxonomy, with a SKOS document as the primary datastream in each case. These content models used for the repository are presented below, along with a listing of important datastreams.

- cm:TEI (entire manuscript)
  - Primary Datastream: Document representing the entire manuscript with the help of a manuscript description and transcription (including the referencing of external digital artifacts, see: datastreams) in format TEI (application/tei + xml)
  - Datastreams: Natural color images (RGB) of the individual manuscript pages in format TIFF (image/tiff)
- cm:Spectral (single manuscript page)
  - Primary Datastream: Document representing a manuscript page as a complex structure (including the referencing of external digital artifacts, see: datastreams) in format METS (application/mets + xml)
  - Datastreams: Natural color image (RGB) and multispectral images (grayscale and RGB) of the manuscript page in format TIFF (image/tiff); Various spectroscopic measurements generated at locations on the manuscript page surface in format: JCAMP-DX (chemical/x- jcamp-dx)
- cm:SKOS (taxonomy)
  - Primary Datastream: Document representing the scientific taxonomy in the format SKOS (application/skos + xml)

3.2.3 Data ingest, processing and web access

The GAMS objects and their primary datastreams, especially the numerous METS documents for the manuscript pages, can become complex and would be tedious to create manually. For this reason, the GAMS repository provides client software that allows semi-automatic ingest of objects and their data based on the available digital artifacts. For GAMS objects of type cm:Spectral (manuscript pages), an ingest process has been defined that is optimized for the multimodal representation described. Provided that the existing digital artifacts are organized in a given local directory structure, they are automatically stored as datastreams within the generated object during the ingest process. Additionally, the METS primary datastream is automatically populated based on the existing directory structure and the included digital artifacts.

To dynamically generate a graphical web interface (HTML) via URL calls in a web browser, stylesheets were developed for the three object types: cm:TEI, cm:Spectral, and cm:SKOS. The mutual linkages between cm:TEI and cm:Spectral objects are recorded in their respective primary datastreams (Chapter 3.1.3). During the generation of the web presentations, these linkages are rendered as hyperlinks, enabling navigation between the web presentation of the entire manuscript (cm:TEI) and those of its individual pages (cm:Spectral), and vice versa.

Both the annotations of the entire manuscript in the TEI document and the annotations of individual pages in the corresponding METS documents, using terms from the SKOS taxonomies, are automatically transformed into RDF statements [58] and integrated into a graph database. Using a query object, this graph database can be queried in reverse to find manuscripts or individual pages with corresponding annotations. The taxonomy objects (cm:SKOS) are presented in a graphical web interface (HTML) that visualizes the hierarchical structure of the taxonomy and enriches individual terms with explanatory details. Users can select specific terms and use them to query the graph database as described.

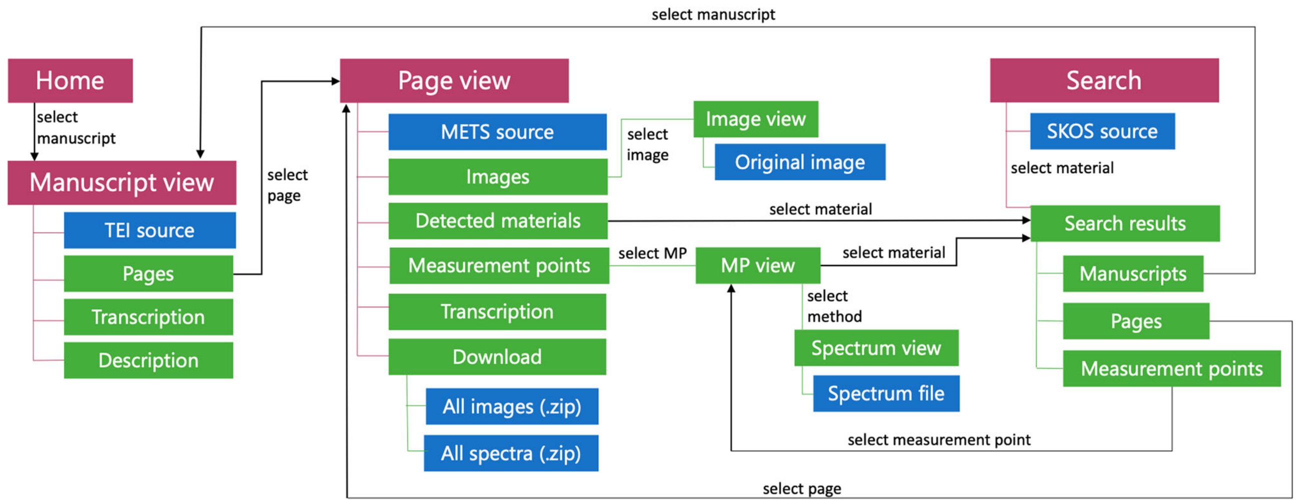

Fig. 6 Navigation overview of the graphical web interface: Magenta blocks indicate views, green blocks indicate tabs/dynamic displays within views and blue blocks indicate download options for source digital artifacts

3.3 Graphical user interface

The graphical user interface for the multimodal manuscript representation essentially consists of three views (Fig. 6, magenta boxes): A view of the entire manuscript, a view of the individual manuscript pages, and a search interface that allows users to search for manuscripts based on various parameters. These individual views are closely interconnected, enabling users to explore the manuscripts through all available digital artifacts (Fig. 6, green boxes) in a seamless and integrated manner. All digital artifact-s—including the supplementary artifacts used for linking the others—can be downloaded (Fig. 6, blue boxes).

The interface of the Manuscript View (Fig. 7) organizes the content of the manuscript into three distinct sections: (1) Pages—which can be filtered by detected materials or conservation findings; (2) Manuscript Description—providing a detailed description of the manuscript (Fig. 7, right side); and (3) Transcription—presenting an image-text synopsis (Fig. 7, left side).

The interface of the Page View (Fig. 8, left side) organizes the digital artifacts associated with the selected manuscript page into several sections: (1) Images—displays all available images of the page using the interactive OpenSeadragon viewer; (2) Detected Materials/Measurement Points—lists material-analytical point measurements, showing the identified materials and spectral graphs for each selected point; and (3) Transcription—presents an image-text synopsis of the page.

The interface of the Search View (Fig. 8, right side) enables users to find terms from the material or conservation taxonomy. For each selected term, a definition is provided alongside search results that list occurrences of the material within the archived manuscripts, organized by manuscript, page, or measurement point. Additionally, manuscripts in the repository can be searched using combined parameters such as date of origin, place of origin, language, material composition, and conservation findings.

4 Results and discussion

The resulting repository *M3R—Multimodal Manuscript Representation* is accessible at the following URL: https://gams.uni-graz.at/context:mmmr[6]

With regard to the outlined requirements for the repository, the following observations can be made:

The repository ensures the long-term preservation of CIMA's extensive and heterogeneous body of data and research results.
It provides open access to the digital artifacts belonging to a manuscript in a curated form. All data and the associated metadata are available in open, standardized formats according to the FAIR principles.
The individual digital artifacts associated with each manuscript are integrated and disseminated as a multimodal representation via an interactive graphical web user interface (Fig. 6, 7, and 8).

The user interface aligns, to the following extent, with the forms of publication outlined as objectives in chapter 3.1.1:

Critical Edition: The manuscript folios can be displayed in a text–image synoptic view, juxtaposing the transcribed text with the digital facsimile of the text carrier. The concept of multimodal manuscript representation makes it possible to produce a critical edition of the manuscript text using the TEI document and to present it within the user interface.[7]

[6] The collection of manuscripts in this repository is currently under work. At present, initial example data are available in the repository, allowing the functionality to be assessed.

[7] This has not yet been implemented in the current state of the repository; however, the prerequisites for it are provided by the repository's design.

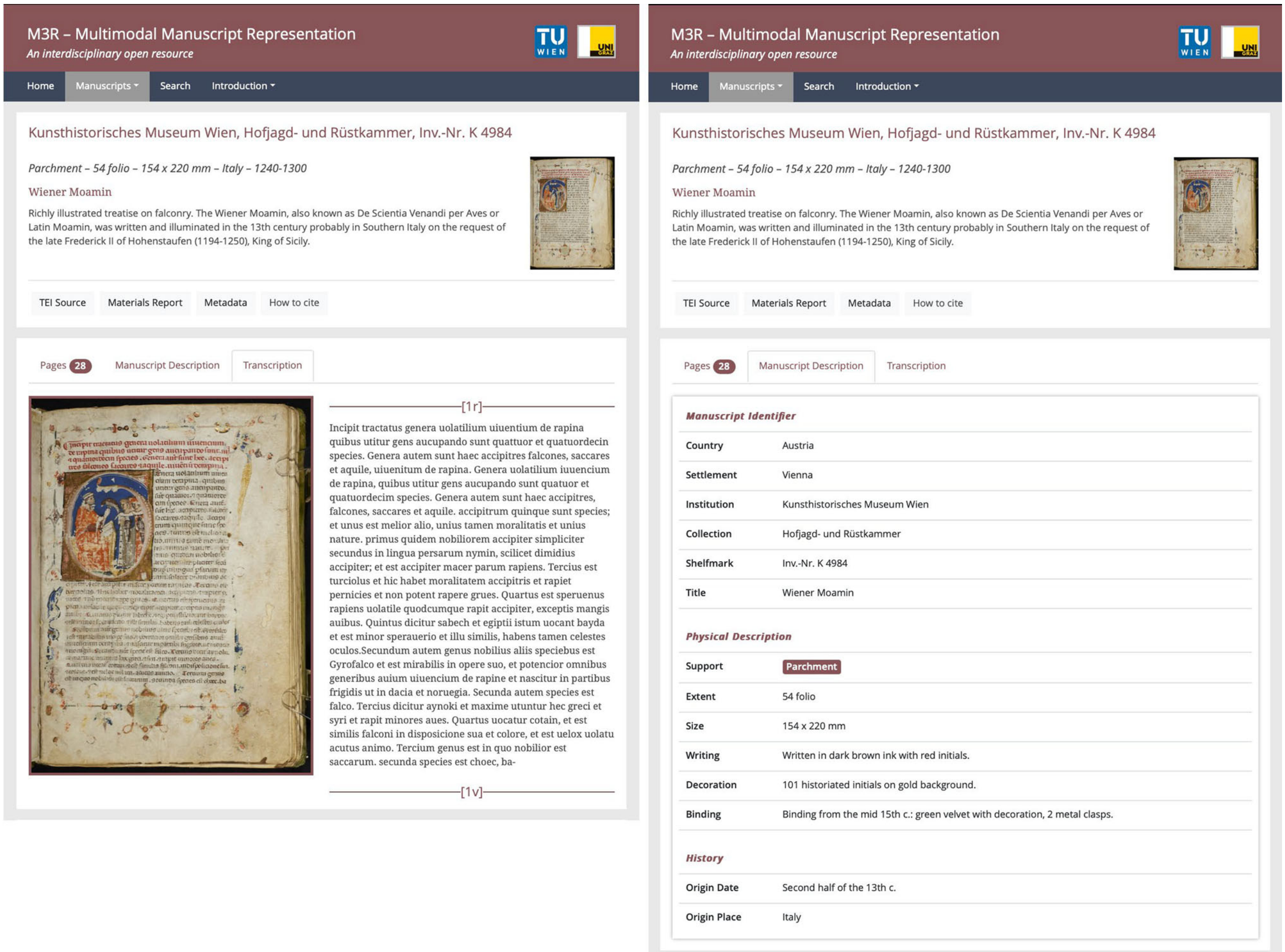

Fig. 7 Manuscript View: (left) The transcription of the manuscript as an image-text synopsis. (right) the manuscript description in structured form

Cataloging: The manuscripts in the repository are each accompanied by a detailed manuscript description. In addition to the fundamental metadata, these descriptions include codicological and conservation-related aspects. Both the basic metadata and the conservation-related terms derived from the established taxonomy are stored in a shared database, thereby enabling the filtering of the manuscript corpus based on this information.

Spectral Database: The materials identified in the manuscripts are also stored in the database as terms of the established taxonomy. This enables corpus-wide searches for the materials used. The search results provide not only the corresponding manuscripts but also the measurement points on the manuscript folios, thereby granting access to all recorded spectra of a given material.

The user interface of the multimodal manuscript representation thus implements and interconnects three established forms of digital dissemination of manuscript research. The interlinked presentation of the digital artifacts via the graphical user interface facilitates the contextualization of the underlying methodologies, making it particularly suitable for use in educational settings. For instance, spectroscopic measurements are linked to their corresponding materials, complemented by their visual localization on the manuscript page and enriched with disciplinary knowledge about the identified materials. Similarly, conservation-related findings from the manuscript descriptions are connected with scientific background information. However, the design of the individual forms of publication and the visual interlinking of the digital artifacts is not yet fully completed and still offers potential for improvement:

At present, the rendering of the manuscript text is limited to a simple text–image synoptic view, with texts not yet transcribed, let alone critically edited. However, the realized framework for the digital edition of the manuscript text—with its integrated TEI documents and digital facsimiles, dynamically rendered through XSLT stylesheets—already provides all the necessary prerequisites for producing and publishing critical editions as part of the multimodal manuscript representation. In this context, detailed spatial linkages between the digital facsimile and the edited text are conceivable, for example, connections to individual lines or to zones of the manuscript page that exhibit particular text-genetic or diplomatic phenomena [49]. The repository has thus implemented a placeholder for critical editions, which can be filled with comparatively little effort according to specific editorial interests and individually for each manuscript—even retrospectively. This enables the linking of a critical text edition with the research results from MSI and material analysis.

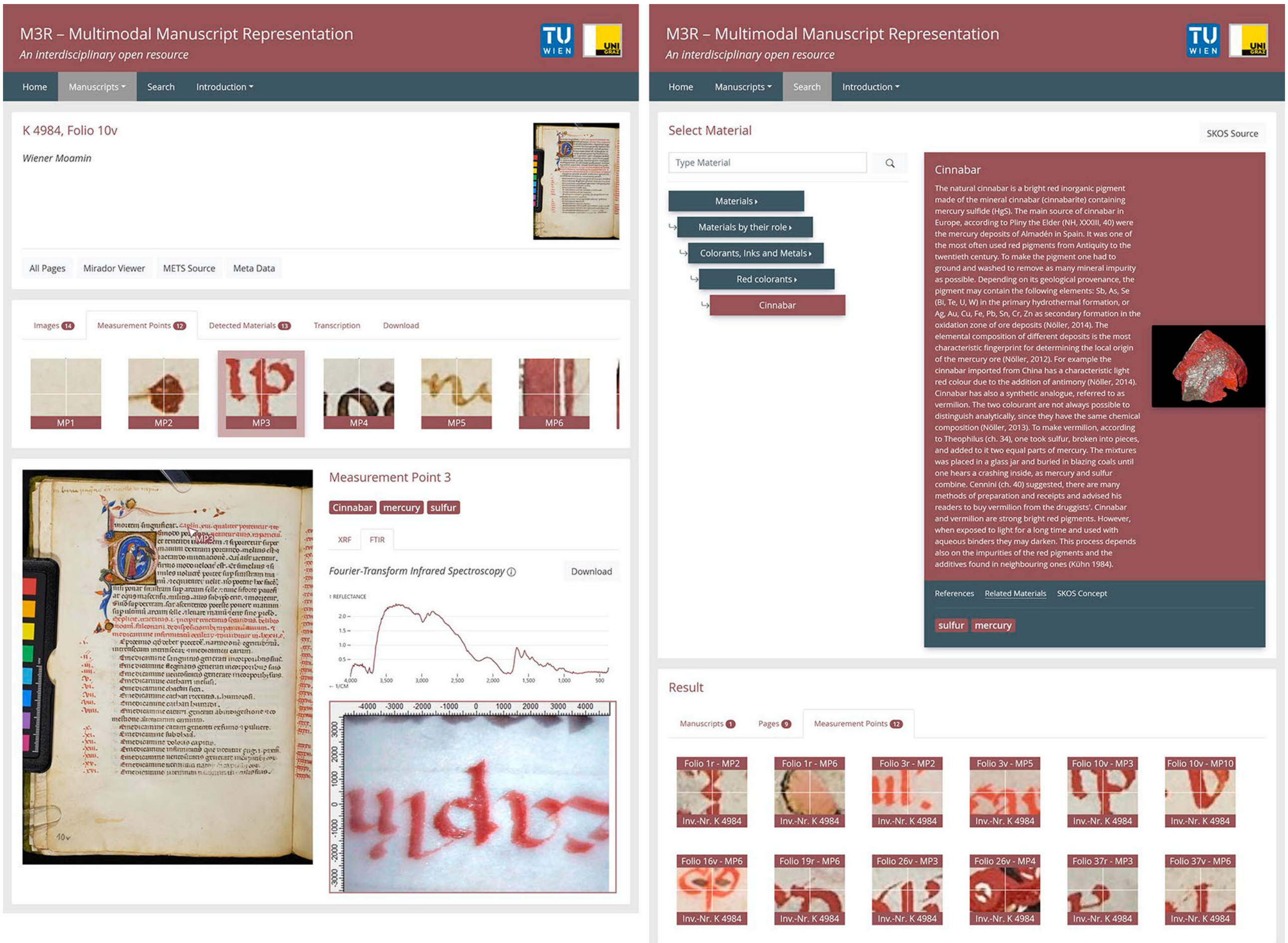

Fig. 8 Page View and Search View of the multimodal manuscript representation: (left) The measurement section of the Page View displays the identified materials along with the corresponding spectral graph for each measurement point. (right) The Search View displays the results for the material term 'cinnabar,' listing the corresponding manuscripts, pages, and measurement points

The manuscripts are currently indexed using the fields for date of origin, place of origin, and language from the manuscript descriptions, as well as codicological and material-related terms. Expanding the indexed fields would enable more precise searches within the repository. Possible additional information could include the holding institution, the number and format of the folios, or content-related information.

The search results for manuscripts containing a specific material allow users to access all recorded spectra for that material. At present, however, these must still be retrieved manually for each individual measurement point. The results view should therefore not only list the measurement points but also provide a consolidated download of all spectra, along with the corresponding term from the material taxonomy, thereby facilitating the reuse of spectral data. Potential applications for reuse include machine learning techniques—for example, training a model on spectral data and material classifications to enable the automated classification of materials from their spectra [59]. Likewise, deep learning methods could be applied to the automated determination of a manuscript's date of origin or to writer identification. In this context, it would also be advantageous to enable the download of all MSI images for a given manuscript type, as the available images can currently only be retrieved on a per-folio basis. A consolidated download of both spectroscopic and MSI data would provide an expanded foundation for machine learning approaches [60, 61].

Another objective in developing the multimodal manuscript representation was to link and visualize all available digital artifacts so that, collectively, they function as a virtual surrogate of a physical manuscript. While all digital artifacts are interconnected and accessible via hyperlinks within the user interface, the implementation of multiple publication formats currently results in several distinct views of a single manuscript. A more seamless integration of these views could be envisaged while preserving the described functionalities. Firstly, all MSI images of a folio could be overlaid as virtual layers and aligned with the reference natural-color image, rather than displayed side by side, allowing users to show or hide the layers as needed [41]. In a second step, the existing view of measurement points could also be incorporated into this composite page representation by aligning them with the same reference image. Finally, these layers of information could replace the reference image in the existing text–image synoptic view, thereby becoming an integral part of the intended critical edition. In this way, all digital artifacts would not only remain linked via hyperlinks but could also be experienced collectively within a single, unified view. This approach would visually relate the different aspects of the multimodal manuscript representation, providing a more comprehensive and integrated understanding of the manuscript. The described linkages of the digital artifacts have already been implemented in the TEI and METS documents (Chapter 3.1.3), so that only their visualization in the user interface would need to be adapted.

The proposed design for a research data repository is part of a broader set of projects dedicated to manuscript research, particularly the systematic online indexing and digital editing of historical manuscripts. To enhance the visibility of this form of research data dissemination and to foster academic discussion, it is advisable to establish links to other web resources. One suitable approach would be to register and link the manuscripts available in the repository within *Handschriftenportal* [42] using the existing indexing data from the manuscript descriptions.

Another important consideration for ensuring effective research data dissemination is the use of open, widely adopted standards. The selected data formats (TEI, METS, SKOS, TIFF, JCAMP-DX) reflect this approach. For the digital presentation of objects, the IIIF Presentation API [62] has become a widely adopted standard within the scholarly community. This standard allows, following the principles of Linked Open Data, the description of complex compound objects. Since the IIIF Presentation API is primarily designed for online viewing—that is, for the presentation of digital objects—rather than for the machine-readable representation of complex research data, it cannot replace TEI or METS documents. However, the TEI and METS documents can be used to automate the generation of IIIF Presentation instances for multimodal representations. During the development of the repository, a web visualization based on the IIIF Presentation API and corresponding viewers [63] was considered and tested. However, it proved unsuitable for the required functionalities, which led to the implementation of a custom web-based visualization within the repository. Nevertheless, a IIIF Presentation instance of the multimodal manuscript representations remains valuable for interoperability, data exchange, and integration into other web platforms, such as *Handschriftenportal* [42]. A corresponding workflow has already been established and only needs to be finalized.

The manuscript representation was developed within CIMA in accordance with the research approaches and digital artifacts produced. A deliberate effort was made to adopt a highly generic method of data integration, based on the shared reference framework of the physical object. This approach is intended to facilitate the integration of additional digital artifacts from manuscript research. Potential examples include the incorporation of 3D data, such as RTI or photogrammetry. Although such integration may require adjustments to the repository, the approach chosen here is well prepared for this purpose and may serve as a model for similar efforts in future.

The repository currently contains only a few exemplary datasets. For many manuscripts, raw data and preliminary results are already available. However, processing these data to a level that meets the scientific and technical standards defined for this repository presents a significant challenge, especially since the data originate, as described, from different researchers and institutions. This underscores the critical importance of careful planning in interdisciplinary projects that handle diverse and complex research data, regardless of the chosen form of final dissemination.

Acknowledgements This work was funded as part of the interdisciplinary research project DiTAH (Digital Transformation of Austrian Humanities) by the Austrian Federal Ministry of Education, Science and Research. The authors acknowledge the financial support by the University of Graz. The authors gratefully acknowledge the museums, archives and collections for providing access to their unique codices, maps or charters and support during the investigation of the manuscripts, charters and graphic objects. This paper is dedicated to Prof. Dr. Dr. h.c. Heinz Miklas, professor for Slavic languages at the University of Vienna/Austria, who initiated the interdisciplinary and inter-university research work of handwritten codices, book scrolls or charters. Unfortunately, he passed away in December 2023.

Funding Open access funding provided by University of Graz. Bundesministerium für Wissenschaft und Forschung

Data availability The example datasets generated within the CIMA and DiTAH projects and analyzed during the study are available in the Multimodal Manuscript Representation repository at https://gams.uni-graz.at/context:mmmr. In future, additional data will be incorporated into the repository.

Declarations The authors hereby disclose any financial or nonfinancial interests directly or indirectly related to the work submitted for publication. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. V. Brinkmann, O. Primavesi, M. Hollein, CIRCUMLITIO: The Polychromy of Antique and Medieval Sculpture, in Schriftenreihe der Liebieghaus Skulpturensammlung. (Hilmer Verlag, Munich)
2. F. Cappa, G. Piñar, S. Brenner, B. Frühmann, W. Wetter, M. Schreiner, P. Engel, H. Miklas, K. Sterflinger, Int. Biodeterior. Biodegrad. (2022). https://doi.org/10.1016/j.ibiod.2021.105342

3. The Centre of Image and Material Analysis in Cultural Heritage, https://cima.or.at/. Accessed 18 Nov 2024
4. Vienna Archaeographic Forum, http://waf.caa.tuwien.ac.at/. Accessed 18 Nov 2024
5. S. Brenner, R. Sablatnig, F. Cappa, W. Vetter, B. Frühmann, M. Schreiner, H. Miklas, Virtual Conservation and Restoration via Multispectral Imaging and Spectroscopy, Исследование и реставрация рукописей (St. Petersburg, Russia, September 2019)
6. W. Vetter, Charakterisierung von Pigment/Bindemittel-Systemen im Bereich der Kunst mittels FTIR- und UV/Vis/NIR-Spektroskopie unter besonderer Berücksichtigung zerstörungsfreier Methoden. (2014) https://doi.org/10.34726/hss.2014.7821
7. F. Cappa, B. Frühmann, W. Vetter, M. Schreiner, Material Analysis of the Vienna Moamin. Metalla Sonderheft 9, 89–92 (2019)
8. H. Miklas, R. Sablatnig, M. Schreiner, I. Tarnanidis (eds.), *Psalterium demetrii sinaitici*, vol. 2 (Holzhausen Verlag, Vienna, 2021)
9. Digital Transformation of the Austrian Humanities, https://www.ditah.at/. Accessed 18 Nov 2024
10. I. Dillo, L. de Leeuw, CoreTrustSeal. Mitt. VÖB (2018). https://doi.org/10.31263/voebm.v71i1.1981
11. D. Klimešová, Data, Information and Knowledge Transformation, Proceedings of the 10th WSEAS International Conference on AUTOMATION & INFORMATION. (2009)
12. Centre for the Study of Manuscript Cultures, https://www.csmc.uni-hamburg.de/. Accessed 18 Nov 2024
13. Early Manuscripts Electronic Library, California (EMEL). https://www.emelibrary.org/. Accessed 15 Dec 2025
14. M. Lettner, R. Sablatnig, 10th International Conference on Document Analysis and Recognition. (2009) https://doi.org/10.1109/ICDAR.2009.51
15. M. Lettner, R. Sablatnig, Multispectral imaging for analysing ancient manuscripts, Proceedings of 17th European Signal Processing Conference. (2009), pp. 1200–1204
16. C. Fischer, I. Kakoulli, Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. Stud. Conserv. (2006). https://doi.org/10.1179/sic.2006.51.Supplement-1.3
17. R.L. Easton, K.T. Knox, W.A. Christens-Barry, Multispectral imaging of the Archimedes palimpsest. (2003) https://doi.org/10.1109/AIPR.2003.1284258
18. S. Pentzien, I. Rabin, O. Hahn, J. Krüger, F. Kleber, F. Hollaus, M. Diem, R. Sablatnig, Can Modern Technologies Defeat Nazi Censorship?, 2nd Asian Conference on Computer Vision: ACCV workshop on e-heritage. (2012) p. 13–24
19. OPUS INSTRUMENTS, www.opusinstruments.com. Accessed 18 Nov 2024
20. M. Diem, R. Sablatnig, Registration Of Ancient Manuscript Images Using Local Descriptors. Proceedings of the 14th International Conference on Virtual Systems and Multimedia. (2018) pp. 188–192
21. F. Hollhaus, M. Gau, R. Sablatnig, Enhancement of Multispectral Images of Degraded Documents by Employing Spatial Information, 12th International Conference on Document Analysis and Recognition. (2013) https://doi.org/10.1109/ICDAR.2013.36
22. B. Doherty, A. Daveri, C. Clementi, A. Romani, S. Bioletti, B. Brunetti, A. Sgamellotti, C. Miliani, The Book of Kells: A non-invasive MOLAB investigation by complementary spectroscopic techniques. Spectrochim. Acta A Mol. Biomol. Spectrosc. (2013). https://doi.org/10.1016/j.saa.2013.06.020
23. C. Miliani, D. Domenici, C. Clementi, F. Presciutti, F. Rosi, D. Buti, A. Romani, L. Laurencich Minelli, A. Sgamellotti, Colouring materials of pre-Columbian codices: non-invasive in situ spectroscopic analysis of the Codex Cospi. J. Archaeol. Sci. (2012). https://doi.org/10.1016/j.jas.2011.10.031
24. M. Aceto, A. Agostino, G. Fenoglio, M. Gulmini, V. Bianco, E. Pellizzi, Non invasive analysis of miniature paintings: Proposal for an analytical protocol. Spectroc. Acta A Mol. Biomol. Spectrosc. (2012). https://doi.org/10.1016/j.saa.2012.02.021
25. I. Nastova, O. Grupče, B. Minčeva-Šukarova, M. Ozcatal, L. Mojsoska, Spectroscopic analysis of pigments and inks in manuscripts: I. Byzantine and post-Byzantine manuscripts (10–18th century). Vibrat. Spectrosc. (2013). https://doi.org/10.1016/j.vibspec.2013.05.006
26. M. Bicchieri, M. Monti, G. Piantanida, F. Pinzari, A. Sodo, Non-destructive spectroscopic characterization of parchment documents. Vibr. Spectrosc. (2011). https://doi.org/10.1016/j.vibspec.2010.12.006
27. S. Bruni, S. Caglio, V. Guglielmi, G. Poldi, The joined use of n.i. spectroscopic analyses: FTIR, Raman, visible reflectance spectrometry and EDXRF: to study drawings and illuminated manuscripts. Appl. Phys. A (2008). https://doi.org/10.1007/s00339-008-4454-x
28. O. Hahn, Analyses of Iron Gall and Carbon Inks by Means of X-ray Fluorescence Analysis: A Non-Destructive Approach in the Field of Archaeometry and Conservation Science, Restaurator. Int. J. Preserv. Libr. Arch. Mate. (2010). https://doi.org/10.1515/rest.2010.003
29. V. Desnica, M. Schreiner, A LabVIEW-controlled portable x-ray fluorescence spectrometer for the analysis of art objects. X-Ray Spectrom. (2006). https://doi.org/10.1002/xrs.906
30. MICRO-XRF SPECTROMETER ELIO, https://www.bruker.com/en/products-and-solutions/elemental-analyzers/micro-xrf-spectrometers/elio.html. Accessed 18 Nov 2024
31. H. Schramm, B. Hering, *Historische Malmaterialien und ihre Identifizierung* (VEB Deutscher Verlag der Wissenschaft, Berlin, 1989)
32. N. Eastaugh, V. Walsh, T. Chaplin, R. Siddall, *Pigment Compendium: A Dictionary and Optical Microscopy of Historic Pigments* (Elsevier eBook, Oxford, 2008)
33. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry* (John Wiley & Sons Inc, New Jersey, 2007)
34. Infrared & raman users group, www.irug.org. Accessed 18 Nov 2024
35. D.O. Hummel, *Industrial polymers* (Wiley-VCH, New Jersey, 2005)
36. A.S. Lee, V. Otieno-Alego, D.C. Creagh, Identification of iron-gall inks with near-infrared Raman microspectroscopy. J. Raman Spectrosc. (2008). https://doi.org/10.1002/jrs.1989
37. R. Clemens, T. Graham, *Introduction to manuscript studies* (Cornell University Press, London, 2007)
38. P. Engel, MANUSCRIPT CONSERVATION: A CHANCE OR AN ENDANGERMENT?, in Cistercian Horizons. Collected Essays, ed. by C. F. Barreira, C. Casanova, M. F. Andrade (Trivent, Romenia, 2024), p. 291
39. L. Elleström, The Modalities of Media II: An Expanded Model for Understanding Intermedial Relations, in Beyond Media Borders, Volume 1, ed by L. Elleström (Springer Nature, EBOOK, 2020), p. 3
40. A. Böhm, J. Eibinger, H. W. Klug, C. Steiner, CoReMA: Cooking Recipes of the Middle Ages, in Digitale Edition in Österreich (BoD, Norderstedt, 2023)
41. The Jubilees Palimpsest Project: Old Manuscripts, New Technology, https://jubilees.stmarytx.edu, Accessed 18 Nov 2024
42. Handschriftenportal, https://handschriftenportal.de/info/about. Accessed 18 Nov 2024
43. TEI, https://tei-c.org/. Accessed 18 Nov 2024
44. TIFF, https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf. Accessed 18 Nov 2024
45. JCAMP-DX, http://www.jcamp-dx.org/. Accessed 18 Nov 2024
46. SKOS, https://www.w3.org/2004/02/skos/. Accessed 18 Nov 2024
47. Data Management, https://libraries.mit.edu/data-management/store/organize/. Accessed 18 Nov 2024
48. TEI Guidelines, https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html. Accessed 18 Nov 2024
49. TEI: combining transcription with facsimile, https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PH-transcr. Accessed 18 Nov 2024

50. TEI: global attributes for simple analyses, https://tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AIATTS. Accessed 18 Nov 2024
51. METS Primer, https://www.loc.gov/standards/mets/METSPrimer.pdf. Accessed 18 Nov 2024
52. GAMS, https://gams.uni-graz.at/context:gams?mode=&locale=en. Accessed 18 Nov 2024
53. Coretrustseal, https://www.coretrustseal.org/. Accessed 18 Nov 2024
54. FEDORA Commons, https://fedorarepository.org/. Accessed 18 Nov 2024
55. Dublin Core, https://www.dublincore.org/. Accessed 18 Nov 2024
56. Europeana, https://www.europeana.eu/en. Accessed 18 Nov 2024
57. REST, https://ics.uci.edu/~fielding/pubs/dissertation/top.htm. Accessed 18 Nov 2024
58. RDF, https://www.w3.org/RDF/. Accessed 18 Nov 2024
59. A. Towarek, L. Halicz, S. Matwin, B. Wagner, Machine learning in analytical chemistry for cultural heritage: a comprehensive review. J. Cult. Herit. (2024). https://doi.org/10.1016/j.culher.2024.08.014
60. F. Wahlberg, T. Wilkinson, A. Brun, Historical manuscript production date estimation using deep convolutional neural networks. (2016) pp. 205–210, https://doi.org/10.1109/ICFHR.2016.0048
61. B. Bilici Genc, E. Bostanci, B. Eskici, H. Erten, B. Caglar Eryurt, K. Acici, D. Ketenoglu, T. Asuroglu, Development of a new non-destructive analysis method in cultural heritage with artificial intelligence. Electronics (2024). https://doi.org/10.3390/electronics13204039
62. IIIF Presentation API, https://iiif.io/api/presentation/3.0/. Accessed 29 Aug 2025
63. Mirador Viewer, https://projectmirador.org/. Accessed 29 Aug 2025