Annotation of digital music notation documents: surveying needs for a generalised implementation

Kevin R Page
University of Oxford
Oxford, United Kingdom
kevin.page@oerc.ox.ac.uk

David Lewis
University of Oxford
Oxford, United Kingdom
david.lewis@oerc.ox.ac.uk

Laurent Pugin
RISM Digital Center
Bern, Switzerland
laurent.pugin@rism.digital

Abstract

The ability to annotate music notation documents offers a powerful affordance to musicologists using digital libraries, and in the organisation and discovery of annotated sources within a music digital library. In this paper we first assess the current state of the art for annotating digital scores, then report on a survey conducted into existing uses and future needs elicited from the music library community. Analysing the survey results, we distinguish between extensions which might provide generalised annotation services for music notation software, versus application-specific interfaces and visualisations using such annotation services. Drawing upon the Web Annotation Data Model, we frame this distinction in terms of annotation targets and bodies, whereby specialist or customised bodies might utilise common shared mechanisms to address targets. We demonstrate the value of the latter by, for the first time, implementing support for annotation targets in the popular and widely used *Verovio* open source music engraving software, adding visual indications for enumerations and ranges encoded using the MEI `<annot>` element, and which can be manipulated in the resultant SVG image. We conclude that common mechanisms for specifying and implementing annotation targets are not only possible, but a practical and useful foundation for music digital library tools and infrastructure.

CCS Concepts

• **Applied computing** → **Sound and music computing**; **Annotation**; • **General and reference** → Computing standards, RFCs and guidelines; • **Information systems** → *Digital libraries and archives*; **Music retrieval**.

Keywords

Music Annotation, Music Notation, Music Encoding, MEI, Web Annotation Data Model, Linked Data, Digital Musicology.

ACM Reference Format:
Kevin R Page, David Lewis, and Laurent Pugin. 2025. Annotation of digital music notation documents: surveying needs for a generalised implementation. In *12th International Conference on Digital Libraries for Musicology (DLfM 2025), September 26, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3748336.3748346

DLfM 2025, Seoul, Republic of Korea

ACM ISBN /25/09
https://doi.org/10.1145/3748336.3748346

1 Introduction and motivation

Digital annotations refer to highlights, references, links or other selections made on digital documents or media. Annotations are a key mechanism for enabling the use and reuse of digital materials across a wide range of applications, while also enhancing the findability and accessibility of that media via its annotations.

Efforts to integrate annotation technology into online cultural collections have resulted in the development of technical standards. For instance, the International Image Interoperability Framework (IIIF)[1] has notably incorporated annotation services for images, and has been widely adopted by digital libraries, galleries, and museums.

While the importance of annotations to music notation is similarly acknowledged, there is less of a consensus on how best to integrate them into interoperable software applications and infrastructure. Annotations for music can encompass the association of textual observations with regions of a work; cross-reference between musical passages or from a musical passage to some other, non-musical, material; or categorical or structured annotations, such as metrical or harmonic labels, which may be fully or partially integrated into the music notation itself. Within this diversity, the work presented in this paper seeks to identify and implement initial areas suitable for generalisation, grounded in needs and use cases rather than assuming a universal solution.

2 Annotation of musical notation documents

For the purpose of this paper, we define an annotation of music notation as associating some information with elements of the musical score. Thus the main participants in this annotation are 'some information', defined very loosely, and notation 'elements' of the score. The latter may be explicit graphical symbols, but they may also take more abstract forms, such as a temporal range (e.g. 'from bar 2, beat 3 to bar 4, beat 1') which need not be embodied in any specific symbols.

The scope of this work is limited to annotations associated with elements *within* a musical score. We exclude document-level annotations, which although important and interesting, are not the subject of this study. Similarly, we exclude annotation of image facsimiles of music documents. Both of these cases can be implemented using existing, less music-focussed, technologies (such as IIIF), and we positively anticipate cross-referencing and combining multiple annotation sources to be of future benefit.

The context for annotating music notation can be broadly placed into four categories: **Performance and pedagogy**, often drawn as overlaid music notation augmenting or altering existing markings; **Analysis and criticism**, explanatory labelling of or commenting

[1] https://iiif.io/

on regions of music; **Editions**, describing, qualifying and giving evidence for editorial decisions; and **Performance studies**, where a score is annotated with analysis of performances, optionally with direct links to media.

Musicians will be familiar with annotating music by drawing on a score–often with a pencil or pen on paper, and increasingly with a stylus on a tablet or touchscreen. This approach is intuitive and immediate, and clearly very important. It may involve circumscribing notational elements, usually with a circle, with extra information written or drawn beside the circle; or direct notational interventions, with music notation being overwritten into the printed text. The former of these patterns–circling notation–is common where annotation is being created, but less common where it is being typeset for readers. Conventionally, few analytical or pedagogical texts use circling in music examples or teaching scores, although brackets and boxes may be used. This suggests that intuitive notations for creating annotations may be different from those that are judged most effective for communicating them. In the following sections, we seek to qualify these judgements within the scope of our work by first reviewing existing digital implementations (Section 3), then surveying the perceptions of professional practitioners (Section 4).

3 Reviewing existing implementations

3.1 Annotation encoding implementations

Annotation may involve well-established practices, either integrated into existing music notation (such as fingerings or chord labels) or as established additions (such as harmonic labels). Examples of this kind of annotation in existing notation software and encodings includes the realisation of harmonic labels using `**harm` in the text-based format Humdrum[11], and `<harm>` ([12], §10) in the XML format of the Music Encoding Initiative, MEI [4].

Since annotations may arise from novel research, other implementations cater for a general case in which the information being added could not have been prescribed, including approaches for digital and computational musicology. Examples of generalised structures being used for specialised annotations include Dezrann [8] (bespoke JSON), CRIM [19] (bespoke objects with EMA URLS, some use of Web Annotations), work by IReMUS/CEDRIC/CNAM [3] (novel Linked Data) and that built around the Music Encoding and Linked Data (MELD) framework [15] (Linked Data web annotations with MEI). Although most of these can and do operate on music notation and audio/visual media, they are primarily symbolic in focus. In the primarily audio domain, examples include JAMS [10] (JSON), the Schubert Winterreise Dataset [25] (bespoke CSV), Sonic Annotator [1] (Linked Data), and the Polifonia music annotation pattern [6] (Linked Data based on JAMS).

3.2 Graphical annotation implementations

Drawing on scores is the most common form of analogue notation annotation, and this is mirrored in the digital world. Software includes generic PDF annotators, such as GoodNotes, or specialised tools for particular music use cases, as with Nkoda's layered, sharable annotations [21], and the equivalent functionality in Enote [18] and ForScore.

Existing implementations largely rely on proprietary software and encodings, but if a standard format was available for drawn annotations (comparable, for example, with those in IIIF) the transfer of annotations between packages would be possible. This is only meaningful if the documents also share a common means for addressing notation any annotation refers to.

Figure 1: Verovio visualisation of MEI harmonic labels

3.3 Symbolic annotation implementations

Verovio [17] is an open-source software library designed for engraving and displaying symbolic music notation in a digital format. It allows users to render music scores encoded in various formats into a visually appealing form for websites, print, or integration into interactive digital applications. Verovio begins by parsing the encoded music data, with MEI XML as its primarily input and internal format, then interprets the structural elements of the encoded music, before performing layout calculations to arrange the elements on a musical staff or system according to predefined and customizable formatting rules. Verovio then generates a graphical representation of the music notation, with its most powerful graphical output being SVG [5], an XML-based W3C image standard through which Verovio can preserve the addressability of notation elements between symbolic and graphical encodings. Where specialised elements exist in MEI for some annotation-like structures, Verovio already provides basic visualisation support (e.g. Figure 1).

The Dezrann software system and libraries [8] use Verovio for rendering, and so is compatible with the symbolic formats provided there. Given automatic or manual synchronisation, itself a set of annotations, Dezrann can also visualise audio and facsimile images. Annotations of different types are visualised differently, showing boxes, highlighting, vertical lines to label points in time and labelled horizontal ribbons above or below the stave to represent time regions. Data is stored as JSON files in a consistent, but bespoke, format, and annotations are delimited by staves, bars and (crotchet) beats. Once a particular annotation type is established, data entry is usually graphical. Modes of visualisation generally focus on boxing and bracketing. Some similar functionality is shown by the IREMUS tools (Tonalities) [2], using the Polifonia data model.

The Citations in Renaissance Imitation Masses project (CRIM) [19] is a notable user of the EMA (Enhancing Music Notation Addressability) URL schema and API [22]. The targets of analytical assertions are specified as EMA URLs, such that visualisations can either use the EMA web service API directly, showing extracts from the music with relevant examples, or provide the necessary identifiers to inform the highlighting of a full score rendered in Verovio, with the notes that are the target of observations being highlighted in the score.

Music Encoding and Linked Data (MELD) [23] is the JavaScript framework behind a selection of web applications (e.g. [13]), all of which use some form of annotation to describe relationships within multimedia resources. Verovio again renders scores, with the resulting SVG used to inform either interaction or visualisation, including bounding boxes [24], note colouring and in-line music notation [15]. Some MELD applications include the ability to author annotations, and the Beethoven in the House annotator uses MELD libraries for loading annotations alongside the Music Annotation Ontology [14] to coordinate annotation of versions.

In the following Section 4, we survey the use of score annotation by music librarians and musicologists, complementing the above review of existing software. In Section 5 we will return to implementation considerations in light of the survey results, identifying and relating common annotation structures according to the conceptual Web Annotation Data Model, which informs our own generalised implementation using MEI and Verovio in Section 6.

4 Surveying annotation use and needs

To better understand current and future annotation needs, beyond those published in the literature and realised through existing tools, we undertook a survey of music library professionals. Through an online questionnaire, we requested participation from people who annotate music, whether digitally or on paper. An introduction and motivation was presented at the conference of the International Association of Music Libraries, Archives and Documentation Centres (IAML), during which the questionnaire was launched. The survey was also publicised at the Digital Libraries for Musicology Conference (DLfM) and online through the IAML, digitalmusicology, and musicology-all mailing lists and on Twitter/X from the `@TMusicology` account. The survey remained open between 26th June and 4th October 2024. 20 responses were received.

Multiple factors might account for the low number of responses received. Given the limited reach of current specialised applications[2], few music librarians and musicologists count digital annotation among their activities. While we also ask for those who annotate on paper, the digital nature of our interests may have discouraged others. The survey was also quite long, with open questions, requiring a commensurate time commitment.

We collected no personal or professional background information, because this would not affect our interpretation of survey questions, and to align with local policy on minimal data collection. This limits our ability to estimate the coverage of our responses, whether in sectoral or geographical terms. Nonetheless, since even such a small survey shows wide variation in needs and practices, a substantially larger quantity of responses would be needed for reliable quantitive analysis. Our intention instead was to gather observational data to support a qualitative understanding of the diversity of practice already underway, and this perspective is taken in the following sections.

4.1 Survey results

This section presents a summary and analysis of survey responses, marked for later reference in the form (SA*n*). A detailed coding of survey responses is provided in the supplementary material accompanying this paper.

[2]We note reported use of digital tools in the literature is at the same institutions developing their respective software, such as Dezrann, MELD, Tonalities, etc.

4.1.1 Deficiencies in the capture of digital annotations . A set of frustrations from respondents concern the difference between the flexible affordances offered by pencil and paper (or, to a slightly lesser extent, of software that allow freehand drawing on scores or images) and constrained annotation capture in existing music engraving and editing software. Limiting factors include: an insufficient range of standard symbols or notations available for annotation (SA1); poor typesetting of annotations once captured (SA2); insufficient flexibility selecting elements within a notation document for annotation, often limited to single elements (SA3); annotation entry being slower than with pencil or stylus (SA4).

Where survey respondents used engraving software for annotation, such as Sibelius, this was a greater source of dissatisfaction than specialist interactive applications. Frustrations included cluttered output, poor variety of symbols, poor anchoring, undesirable spacing and a lack of functionality for layering (and so hiding and showing) annotations. Whilst specialist software ensures musically appropriate notation, respondents found themselves restricted (compared with drawing) in what they could express, and with few compensatory benefits. In some cases, this meant producing PDF editions with engraving software, then annotating the PDF in alternative software.

4.1.2 Use-dependent preferences for visually indicating annotations. Respondents expressed variable preferences for what they considered effective for visually indicating annotations (SA5), much of which is attributed to different communication needs and use cases. For example, whether it is desirable for the user to interact with on-screen elements in order to see annotations will depend on how the score will be viewed by the final user–including whether it will be printed. Respondents deployed an indicatively wide variety of annotation formatting across their different activities–embracing colours, circles, brackets, boxes, marginalia, tables and many more.

The preferred way to *view* annotations is also dependent on use (SA6), and here the digital brings an advantage. The appearance of an analogue annotation is often optimised for 'data entry'–for example simplifying the pencil movements required to draw the indication–and then fixed, whereas an electronic score can adapt the visualisation of previously captured annotations to serve alternative objectives at the time of viewing.

Across respondents, there is a general dislike for annotations that changes the spacing and layout of a score in a way that distorts musical logic (SA7). i.e. a preference for musical typesetting to be prioritised when generating a layout incorporating annotations.

Finally, although there is confirmation of generic annotations (text and boxes) and standard notation (fingerings, phrasings, chord symbols), respondents also report specialised or bespoke symbols as being needed and used (SA8). It seems unlikely that an attempt to codify all annotations would be successful, even if it were desirable. Freehand drawing is currently the clearest (and least technical) way for such symbols to be used in a general-purpose system.

4.1.3 Layering annotations. The concept of layers–groups of annotations that are associated with one another, and which can be hidden, customised, displayed or shared as one–is popular with respondents when it occurs, and strongly desired when it does not (SA9). Layers provide a generic user-interface model for implementing a variety of conceptual aggregations: they can represent the annotations from a single music lesson, from a single annotator, or stand for a particular analytical or text-critical unit. While this removal of unwanted analytical clutter is another instance of visualisations being adapted to the preferences of the user (SA6, above), we distinguish this specific case from the general one due to its broad and consistent occurrence across survey responses.

4.1.4 High-quality typesetting for publication including annotations. Several respondents, especially those who distribute their annotations to performers or students, particularly value a version of their edition that is as clear and well-laid out as possible, distinguishing the final adjustment of annotations for static publication (SA10) from their dynamic display during capture (SA7, above).

Drawing on a score is only guaranteed to be trivially reproducible where the layout and typesetting are static. The same is true, to a lesser extent, for marginalia. The semantics of any annotation whose targeting of music notation is solely by spatial juxtaposition tends to be dependent on layout. Although this is also true for musical symbols such as hairpins, and has proven largely solvable there, it does require appropriate treatment.

4.1.5 Annotation provenance. Two situations described by survey contributors indicate a need for tracking responsibility for annotations (SA11). Firstly, encoders may be transcribing prior annotations, and recording the circumstances of the original annotations may be relevant for their work. Secondly, annotation may be collaborative, distributed or simply multi-layered, and tracking and attributing activity may be required for analysis, layered display (SA9, above), or to give appropriate credit.

4.1.6 Connecting and navigating annotations. Multiple respondents described the desirability of connecting annotations across multiple notational regions (SA12), for example: connecting formal or thematically related regions; connecting associated materials or media; connecting related annotations. This would be particularly useful for high-level visualisations (for thematic or formal structure), or for rapid navigation and juxtaposition of relevant material.

4.1.7 Specific annotation indications and visualisations. Of the visual mechanisms for denoting various aspects of annotations, four were offered with sufficient frequency to warrant reporting here: Boxes and brackets (SA13), to delimit an area; Notation colouring and circling (SA14), to denote specific elements; Insertion of music notation (SA15); and Insertion of text (SA16) placed somewhere close to the notation, or separately, via coordinating symbols.

There was difference in preferred strategies for interacting with applications bearing annotations. A notable minority suggested display-on-interaction views (SA17) as a way to avoid visual clutter – for example, clicking or hovering over a passage might trigger a pop-up window with the annotation text. Other respondents have a requirement for printed materials (SA10, above), for which such interaction would be ineffective.

There were also varied suggestions for exploration of a score (SA18), where annotations can provide musically salient navigation points. In other cases, a user may want to list and browse all annotations. In either case, a notable minority of users would resist the assumption that the score is the primary roadmap, decorated with annotations–instead wanting annotations to provide an alternative structure for exploring the score.

4.2 Survey implications for generalised annotation implementations

Consolidating the survey results, and with consideration to existing solutions (§3), we propose the following general requirements for future implementations of digital annotation of music notation documents. Emerging clearly and strongly from the survey, five requirements address the visualisation of annotations and their relationship with the score layout:

R1. Respect for music typesetting, the logic of which should not be unduly impaired by annotations (SA2, SA7, SA10). Excess space (as in Figure 1) should be considered disruptive.

R2. Visual separation of annotations, which should be obviously distinct from the score (SA7, SA10).

R3. Switchable visibility of annotations, ideally selectively i.e. of groups of annotations (SA9).

R4. Printable/performable editions should be possible including visible annotations (SA10).

R5. Annotation attribution should make it possible to indicate the provenance of an annotation, including that an annotation may have its origins in a non-digital medium (SA11). In combination with **R3** is should be possible to switch visibility of annotations by provenance (SA9, SA11).

It is notable that most of the remaining survey results reflect what could be considered 'application specific' needs (SA1, SA5, SA6, SA8, SA12, SA15, SA16, SA17, SA18)–wherein the implementation or even definition of a requirement cannot be separated from a thorough and detailed consideration of the task for which annotations are being deployed. It is not clear that a *generalised* software solution could plausible meet such a diversity of ensuing requirements, several of which are contradictory; rather, a process focussed on user experience (UX) should be expected to produce streamlined designs optimised for every application separately[3]. The resulting, customised, user interfaces might even address issues of speed and efficiency from the perspective of a user (SA4). However, the consolidation of software requirements for *application-specific* annotations is beyond the scope of the work presented here.

Although we cannot conceive of a general implementation fulfilling all possible annotation needs, if we accept the requirement for annotations to respect music typesetting (**R1**), if follows there must be some form of software interface through which application-specific annotation solutions can communicate with engraving software (to ensure the sympathetic placement of annotations). Furthermore, other annotation patterns must interact with notation elements as generated by a music engraving engine: the visual modification of notation referred to by an annotation (SA14); the selection of arbitrary notation elements for annotation (SA3); the specification of notation elements to calculate the geometries of boxes and brackets (SA13); and, indirectly, connecting annotations across multiple notational regions (SA12). It becomes apparent that most of our annotation requirements above are either wholly (**R1**), or partially (**R2**, **R3**, **R4**) reliant upon a functional requirement for music engraving software to encapsulate a minimally sufficient model of annotations as associated with notation elements, and to

[3]An example of a design-led process, albeit without annotation authoring, is [7].

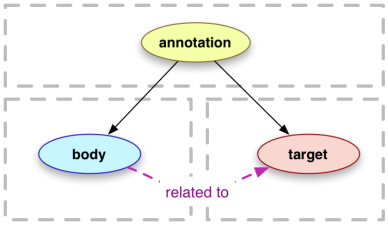

Figure 2: The Web Annotation Data Model (from [20]).

expose data structures to annotation applications by which they can manipulate the notation elements associated with their annotations.

Our proposed solution is, therefore, at this boundary between music engraving software and application-specific annotation processing. It is here that we focus our work towards generalisable approaches and implementations, described in the rest of this paper.

5 Reflecting on existing implementations: distinguishing *Targets* from *Bodies*

Moving outside of the music digital library community, there have been significant efforts to agree standardised encodings for annotations, and create generalised software frameworks which publish and consume annotations according to those standards.

One widely-implemented approach is the World Wide Web Consortium (W3C) Web Annotation Data Model[4] [20]. The Model provides abstract structures for annotating generic, web-addressable objects with generic, web-addressable annotations; associated Vocabulary and Protocol standards specify serialisations for RDF and JSON-LD, with RESTful interactions to retrieve and store them.

It is pertinent that multiple data encodings relevant to music annotation are compatible with web-addressability when published online–including both MEI and MusicXML (as XML encodings), as well as the SVG graphical encoding output by Verovio[5].

At the core of the Web Annotation Data Model is the distinction between the annotation Target and the annotation Body (Figure 2).

The **Target** of a Web Annotation specifies the 'area of interest', which can be anything with a URI. In a digital music notation document, we would expect this Target to be the musical notation to which the annotation applies.

The Web Annotation **Body** specifies the information associated with the Target by the annotation, in whatever form it takes, including arbitrary text. For an annotation associated with a digital music notation document, we might expect the Body to encode the text, symbol, or visualisation to be combined with the Target notation when indicating the annotation.

Finally, the model distinguishes the **Annotation** itself, which we can augment with information about its provenance and perhaps its type, or how it should be interpreted. Associating information about the creation of an annotation–for example, attribution of authorship, creation date and motivation–is important, especially for scholarly applications, and its visualisation in combination with annotation bodies–directly or indirectly–was noted as important by survey respondents (§4.1 (SA11))[6].

In addition to aligning our work with a tried-and-tested abstract model and terminology, we use the Web Annotation Data Model's distinction between Targets and Bodies to frame a closer examination of approaches to annotation implementation, as follows.

5.1 Music Annotation Bodies

The Body encodes an annotation's raison d'être which, as previous sections have explored, encompasses a wide and varied range of potential information. Rather than attempting to enumerate all possible body contents observed in implementations, we instead reframe existing approaches taken to presenting such content, confirming they can be encapsulated within the notion of a Body (and thereby distinguished from a Target):

Textual content This approach allows for easy integration into visualisation applications (see, for example, the Beethoven in the House annotator [13]). It is directly supported in JAMS and within the MEI `<annot>` element (see §6.1).

Notational content Many notation elements can be regarded or used as annotations. One common example is fingering indications (§4.1 (SA8)). In MEI, these could be encoded directly in the edition, optionally using tags such as `<supplied>` to show their editorial nature.

Generic data structures If annotations are primarily quantitative, the annotations can be characterised by a range of primitive data types (controlled vocabulary labels, numbers, vectors of numbers). This approach is used by Sonic Visualiser.

Bespoke data structures Dezrann has been extended to accommodate specialised data structures as they are required by particular annotators. Thus, the semantics of the format evolve over time. Early MELD applications also operated on tailored data structures.

Arbitrary data structures The Web Annotation Data Model permits the association of anything with a URI as an annotation body. However software would almost certainly require additional constraints to process and visualise annotation bodies.

5.2 Music Annotation Targets

Whatever information an annotation indicates, or visualisation it takes, or use it is put to, it must have a target–some circumscribed area of interest within one or more documents. Considering existing implementations, we distinguish three mechanisms for indicating a target region in a document of digital music notation: graphical, enumeration and range, each with their own uses and limitations.

5.2.1 Graphical targets. Based on some coordinate system over the rendered notation, recorded shapes delimit the target regions. This approach is favoured by tablet based systems for performers, and does not require formats of notation that express notational semantics, allowing annotation of PDFs or images. For use with symbolic formats, the shape of the demarcation must be recorded accurately, and with a fidelity and geometry that permits comparison to the graphical extent of notation elements. Unless all relevant elements

[4] Superseding the earlier Open Annotation Data Model.

[5] Although the work reported here focuses solely on digital music notation documents, given IIIF implements annotations according to the Web Annotation Data Model, it is technically possible to implement annotations spanning images and symbolic scores, with digital score facsimiles hosted on IIIF-compatible image services.

[6] Alongside Web Annotations, such information provenance is supported to varying degrees by JAMS and Sonic Annotator, amongst others.

are entirely within and all non-relevant ones entirely outside the shape, ambiguities of its extent are likely. This goes beyond purely graphical challenges for drawing the shape where a score can be reformatted, scaled or reflowed. Graphical targeting is insufficient as a generalised mechanism for targeting, since it introduces elements inherently beyond the control of the engraver's layout engine.

5.2.2 Enumeration targets. Notation elements affected by the annotation are explicitly listed. This can be achieved directly–inline–by marking rows (in tabular formats, such as HUMDRUM) or creating parent elements (in XML-based formats, such as MEI), or indirectly–standoff–by reference. In XML, any element with an `id` can be indicated using the fragment component of a URL. This approach is unambiguous and easy to compute, but it cannot indicate regions that are not delimited by elements (such as beats 2-4 of a bar with only a semibreve in it). The approach of using `xml:ids` is also vulnerable to changes to the source file, especially if these attributes are automatically generated, unless a mechanism for their guaranteed perpetuation is present.

5.2.3 Range targets. Taking the 'axes' for 'coordinates' of a score to be metrical (horizontally) and part/staves/layers (vertical), this approach specifies a musical region in a way that is less reliant on the precise details of the format or the encoding. The model used to specify range is crucial, and existing implementations can struggle with annotations where the target is non-metrical (such as articulation or grace notes) or with music notations for which one or other of the axes is less appropriate (for example, notations lacking measures, polyrhythms or scores with non-sequential parts). A time-based range target (usually based on seconds) is near-universal for audio-oriented models (such as JAMS and Polifonia) and common in notational models too (using beats, as in EMA, Dezrann).

For maximum generality, it is necessary for a range mechanism to encompass a set of temporal boundaries for different musical layers in a polyphonic texture. It is also important that boundaries can be specified in units that are appropriate to the notation – for example, bar and beat units may not be relevant for mensural, chant, non-western or unmeasured scores.

There is potential to translate between these mechanisms. For example, ranges could be translated into graphical selections by engraving software marking their enclosed locations in graphical output. In general, however, the interpretation of each of these mechanisms is complicated by differing degrees of specification and ambiguity, and translation should be assumed to be interpretative.

5.3 Music Target selection and maintenance

Annotations can also be indirect and subtle, with implications for their target. Not all annotations are 'about' the encoding or the edition itself–they can be directed at an abstract concept embodied by the notation, such as a theme, a phrase or a musical gesture. Annotations may also be connective, joining musical instances together, or connecting musical and non-musical elements. Whilst the Web Annotation Data Model supports multiple targets to an annotation, there is no specific discussion of annotating relationships between these. Where necessary, extension is possible.

The Music Annotation Ontology (MAO) illustrates the possibilities offered by accommodating the annotation of such musical abstraction. It provides for a step of indirection, allowing annotations to target musical ideas that are then embodied in the notation files, images or recordings. This allows, for example, describing the behaviour of multiple arrangements of the 'same' musical idea [14]. The MAO has been used for annotation in the Beethoven in the House Annotator [13] and in mei-friend [9, 16]. This model does not currently address directed relationships, such as between a derived work and its source.

Reflecting on the practical implications of maintaining targets, we can identify several potential limitations. A solely *inline* targeting implementation, whether graphical, enumerative or ranged, will require the annotation authoring process to have write access to the document, or to copy and republish it. The latter creates implications for keeping track of versions of the 'same' document or edition; and a heavily annotated document would either be duplicated many times (potentially with repeated alterations) or evolve into a single, central, very complicated file with many annotation structures held within. While a standoff approach decouples the annotation implementation from ownership and access of the notation document, it may be desirable for any derivatives or copies of that notation document to 'inherit' its annotations–again entailing an additional service to track versions of the 'same' document.

The approaches to targeting scores introduced above are broadly symbol- or performance-time- based, with little reference to the logical. In cases where a particular passage recurs, and where this is indicated with repeat or da segno signs, it is not clear how to distinguish between targeting all recurrences and specific ones; since this cannot be trivially indicated in the underlying music notation document, it may require an intermediate target representation. This neatly illustrates the dependence any targeting strategy has on the information affordances of the underlying notation encoding.

6 A generalised approach to annotation support

To demonstrate the viability of a generalised approach, we have implemented processing and visualisation of Enumeration and Range targeting in the *Verovio* music engraving library. Verovio was selected due to its wide adoption across the community, support from RISM, and incorporation within a variety of music notation applications; and for use of standards-based, web-addressable, encodings for input and output, simplifying application-specific body processing for future tools. Consequentially, Verovio's pre-existing rudimentary (if invisible) support for the MEI `<annot>` element made that our prime candidate for our encoding.

We closed Section 4 with a distinction between digital annotation requirements which are application-specific, and those which can be considered foundational across multiple music implementations–and thereby generalisable. We identified the passing of information between a music engraving engine and an annotation processor as the functional boundary between notation-generic and application-specific domains. Section 5 introduced the industry-standard Web Annotation Data Model as a lens through which to compare implementations of music notation annotation, refining our distinction between application-specific *Body* encodings, and potentially generalisable strategies for *Target* selection in music notation documents.

The functionality provided by existing annotation tools indicates that, in practice, it is unlikely any single application will require a

Listing 1: A simple enumerated annotation. The `xml:id` of three syllables in the lyrics are targets for an observation.

```
<annot xml:id="SeaBreakers" type="score"
  plist="#v100jcki #vdmu55 #v3swz5r">Words 'Sea-breakers'
  introduces quotations from Sea Slumber Song (Op. 37/1).
</annot>
```

Listing 2: A range annotation delimited using elements. The `xml:id` of a note at the start and the end of the target is specified, along with the included staves.

```
<annot xml:id="SeaSlumbers" type="score"
  startid="#n18po5kn" endid="#n172z7ie" staff="3 4">
  Quotes Sea Slumber Song from Sea Pictures.
</annot>
```

complete implementation of all generic targeting mechanisms. Specific annotation applications are more likely to be satisfied by partial implementation; for example, the suitability of Graphical or Range selection being determined by the method of input. Nonetheless a more comprehensive coverage of targeting mechanisms expands the potential building common music digital library infrastructures supporting multiple annotation applications and interoperability.

6.1 MEI and the <annot> element

It is possible for several notational annotations to be applied directly in an MEI music encoding. Many examples raised in our survey have their own specialised elements already, such as harmonic labels, fingerings, accidentals or editorial interventions. If there is need for a generic approach to annotations (for example, for making a set of annotations visible or hidden), then this range of forms they take in MEI will cause complications. Similarly, these notational annotations are largely inline by default.

The alternative and explicit MEI <annot> element is highly expressive, compatible or comparable with other annotation models (such as the Web Annotation Data Model), while remaining integrated with the MEI standard for notation. <annot> has many pointing mechanisms associated with it, making it very flexible.

The MEI Guidelines document several roles for <annot> ([12], §9.2.13, §11.2.4.1, §13.1.1, etc.).These can be divided into two categories: notes about aspects of the encoding or catalogue information–which we would not expect to visualise in the notation; and those about the notation, its source or the act of editing it–which we might. These categories may be implicit in the context where the element appears, however this is not clear. In comparison, the Text Encoding Initiative format (TEI) has an additional <note> element which is unavailable in MEI. We introduce a special value for the `type` attribute to indicate that `<annot type="score">` is a score annotation and may be drawn.

For the purpose of selecting annotation targets, <annot> can be treated as an MEI ControlEvent ([12], §1.3.2). That is, it can enumerate elements that constitute targets by `xml:id` (using the `plist` attribute, example Listing 1), and it can delimit ranges metrically (`tstamp` and `tstamp2`, example Listing 3) or by symbol (`startid` and `endid`, example Listing 2).

Listing 3: A range annotation delimited using metrical position, starting in the containing bar and lasting for three bars and two crotchet beats.

```
<annot xml:id="Enigma" type="score"
  tstamp="1" tstamp2="3m+2" staff="3 4">
  Quotes Enigma variations theme.
</annot>
```

The vertical extent of an annotation within a score may be specified using the `staff` attribute to enumerate targeted staves. Where a single voice in a single, polyphonic stave is the target, this may be specified using the `layer` attribute, but there is currently no mechanism to specify this in multiple staves (for example, to specify the lower voices of *divisi* flutes and oboes in a conductor's score). If no staves are specified, the vertical extent is ambiguous.

The body of the annotation–that is the information intended to attach to the music notation–is, seemingly, only indicated by child elements to <annot>. A variety of child elements are permitted, presumably to enable expressivity. Since the generic pointer element <ptr> is allowed, arbitrary URIs can be associated with the <annot>, albeit any meaning or implications remain undocumented.

6.2 Generalised annotation support in *Verovio*

Earlier versions of Verovio recognise <annot> and insert an invisible grouping element (<g>) into any SVG output. If passed documented options, the library will also copy the `plist` attribute value into a data attribute in the SVG. Since Verovio draws SVG elements with the same identifiers as the source MEI, this preserves the references from the original annotation, in turn supporting interactive applications built using these SVG images (meeting **R3**).

Our extension adds a new object class to Verovio's parser, initially based on the presence of a 'score' type in the <annot> element (see above). Since there is a semantic difference between enumerated and range-based annotations, they are treated differently by the engraver to meet **R1**. In both cases a basic visualisation of the target is included (Figure 3), following **R2**, whilst also providing the foundation for client applications to build more complex views and modes of interactions (**R3**).

For enumerations, we highlight individual symbols. More specifically, each element that is listed in the `plist` element is given a `data-plist-referring` attribute with a value that is the identifier of the referring annotation. If an element is the target of multiple annotations, this value will be a set of identifiers. This makes it easier for client applications, using CSS, to apply a colour or other style to annotated elements, either in a generic way, or based on attributes of particular annotations. It also means that navigating from elements to their annotations is as straightforward as it is from the annotation group object to the targets.

For range annotations, a ribbon is drawn over the stave as a child of the annotation's group element. If the annotation specifies multiple staves, multiple ribbons are drawn; whilst if no staves are specified then the ambiguity is preserved, and the ribbon is drawn at the top of the system. Again, the ribbon can be given colour and other styling using CSS, but can also be used either as an anchor for user interaction, or to provide a geometry of the annotation for client applications to overlay their own visualisations.

Figure 3: An example of Range-based annotations as ribbons in bars 115-117 (encoded in Listing 2) and 117-120 (Listing 3). An enumerated annotation targets the lyric ‘sea-breakers’ in bars 114-5 (Listing 1).

As with other drawn elements, ribbons gracefully span systems and lines, and avoid clashing–with each other or with other symbols–by extending in vertical space above their stave (**R1**).

7 Conclusions and future work

In this paper we have assessed existing provision for annotation of music notation documents, reflecting and comparing these with current and future needs returned in a survey of expert users. We proposed an alignment with the Web Annotation Data Model as a route to interoperability between music annotation applications, and within the wider digital annotation ecosystem, including IIIF standards already in use for digital libraries; and explored implications of the Model’s distinction between Targets and Bodies in the context of music notation annotations. We appraised requirements for music annotation tools in terms of whether they were application- (use-) specific, or might be generalised as common functionality which could support multiple applications, with the latter a potential foundation for future music digital library infrastructure. We re-framed this division in potential for generalisation, according to the abstract distinction between and annotation Targets (generalisable for notation) and Annotation bodies (application-specific), noting this was reflected in the information boundary between music engraving engines and annotation software built upon them. To validate this approach we successfully implemented Enumeration and Range targeting for the MEI `<annot>` element in the Verovio music engraving library, enabling generalised mechanisms for annotation manipulation in Verovio outputs.

We see this as a beginning for advancing music notation annotation, rather than an end. While our prototyping demonstrated functionality and usability, in order to build our implementation without modifying the MEI schema we temporarily used `type` in the `<annot>` element, an attribute available for project-specific use. We are working with the MEI community to contribute a new permanent alternative, with options including a `func` attribute, or transposing `motivation` from the Web Annotation Data Model.

Moving beyond data representation and visualisation, we have not realised a user interface for authoring compatible annotations. We note the annotation functionality already offered by the mei-friend editor, which uses Verovio for MEI visualisation and processing. Extending mei-friend to incorporate our new implementation is planned, and imminent, future work.

We have not yet considered the (re-)processing of annotations for printed output (**R4**). This may prove similar to the capability in Verovio to position individual symbols, something that is used for final output but is much less likely to be used for interactive, in-browser applications. Nor have we addressed the less clearly defined information boundary for Graphical targets, and any implications for increased coordination complexity between the engraving engine and annotation application in this case.

In the meantime, we hope that our provision of native target manipulation can simplify the implementation burden for existing applications based on Verovio, motivate new music applications to add notation annotations, and begin a conversation about the possibilities of annotation exchange between applications. In building these beginnings of generalised functionality for notation annotation targets, might we also dare to imagine the possibility of new innovations for processing and visualising annotation bodies, taking advantage of digital capabilities for adaptative visualisations and translations, and finally moving beyond shallow imitation of the physical world?

Acknowledgments

Implementation of annotations in *Verovio* was made possible by the UK Arts and Humanities Research Council through their funding of 'Annote: Digital Notation Annotation for RISM' (project reference AH/Z506278/1), with additional refinement supported through the Leverhulme Trust Research Project Grant 'Elgar's Themes: New Pathways for Analysis, Interpretation and Engagement' (project number RPG-2024-080). We are grateful to our colleagues from *Elgar's Themes* for their feedback, suggestions, and encouragement regarding annotation examples; in particular, the extract from Edward Elgar's *The Music Makers*, op. 69, which was proposed and reduced to vocal score by Dr. Frankie Perry. Our thanks to Dr. Andrew Hankinson for his *Verovio* code reviews, and to our *Annote* project invited experts: Prof. Ichiro Fujinaga, Prof. Dr. Christine Siegert, and Dr. David Weigl.

References

[1] Chris Cannam, Mark Sandler, Michael O. Jewell, Christophe Rhodes, and Mark d'Inverno and. 2010. Linked Data and You: Bringing Music Research Software into the Semantic Web. *Journal of New Music Research* 39, 4 (2010), 313–325. doi:10.1080/09298215.2010.522715

[2] Anne-Emmanuelle Ceulemans, Christophe Guillotel-Nothmann, Thomas Bottini, Philippe Cathé, Achille Davy-Rigaux, Marco Gurrieri, and Félix Poullet-Pagès. 2024. Tonalities: An application for the collaborative online analytical exploration of digital scores: a demonstration based on secular polyphonic works by Guillaume Dufay. In *Second international conference on computational and cognitive musicology*.

[3] Samira Si-said Cherfi, Christophe Guillotel, Fayçal Hamdi, Philippe Rigaux, and Nicolas Travers. 2017. Ontology-based annotation of music scores. In *Proceedings of the 9th Knowledge Capture Conference*. 1–4.

[4] Tim Crawford and Richard Lewis. 2016. Review: Music Encoding Initiative. *Journal of the American Musicological Society* 69, 1 (April 2016), 273–285. doi:10.1525/jams.2016.69.1.273

[5] Erik Dahlström, Patrick Dengler, Anthony Grasso, Chris Lilley, Cameron McCormack, Doug Schepers, and Jonathan Watt. 2011. *Scalable Vector Graphics (SVG) 1.1 (Second Edition)*. W3C Recommendation. W3C. https://www.w3.org/TR/SVG11/.

[6] Jacopo de Berardinis, Albert Merono Penuela, Andrea Poltronieri, and Valentina Presutti. 2022. The Music Annotation Pattern. In *The Semantic Web–ISWC 2022 21st International Semantic Web Conference: 13th Workshop on Ontology Design and Patterns (WOP2022)*.

[7] Laurence Dreyfus, David Lewis, and Kevin Page. 2025. A digital companion for musicological scholarship: the Lohengrin TimeMachine. *Journal of New Music Research* (2025). doi:10.1080/09298215.2025.2487100

[8] Mathieu Giraud, Richard Groult, and Emmanuel Leguy. 2018. Dezrann, a web framework to share music analysis. In *International Conference on Technologies for Music Notation and Representation (TENOR 2018)*. 104–110.

[9] Werner Goebl and David M. Weigl. 2024. mei-friend: An Interactive Web-based Editor for Digital Music Encodings. *Journal of Open Source Software* 9, 98 (2024), 6002. doi:10.21105/joss.06002

[10] Eric J. Humphrey, Justin Salamon, Oriol Nieto, Jon Forsyth, Rachel M. Bittner, and Juan Pablo Bello. 2014. JAMS: A JSON Annotated Music Specification for Reproducible MIR Research. In *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014, Taipei, Taiwan, October 27-31, 2014*, Hsin-Min Wang, Yi-Hsuan Yang, and Jin Ha Lee (Eds.). 591–596.

[11] David Brian Huron. 1994. *The humdrum toolkit: Reference manual.* Center for Computer Assisted Research in the Humanities.

[12] Johannes Kepper and Perry D. Roland. 2025. *The Music Encoding Initiative Guidelines (Version 5.1)*. Technical Report. Music Encoding Initiative. https://music-encoding.org/guidelines/v5/MEI_Guidelines_v5.1.pdf

[13] David Lewis, Elisabete Shibata, Andrew Hankinson, Johannes Kepper, Kevin R. Page, Lisa Rosendahl, Mark Saccomano, and Christine Siegert. 2023. Supporting Musicological Investigations With Information Retrieval Tools: An Iterative Approach to Data Collection. In *Proceedings of the 24th International Society for Music Information Retrieval Conference*. ISMIR, 795–801. doi:10.5281/zenodo.10265407

[14] David Lewis, Elisabete Shibata, Mark Saccomano, Lisa Rosendahl, Johannes Kepper, Andrew Hankinson, Christine Siegert, and Kevin Page. 2022. A model for annotating musical versions and arrangements across multiple documents and media. In *Proceedings of the 9th International Conference on Digital Libraries for Musicology* (Prague, Czech Republic) *(DLfM '22)*. Association for Computing Machinery, New York, NY, USA, 10–18. doi:10.1145/3543882.3543891

[15] David Lewis, David Weigl, and Kevin Page. 2019. Musicological Observations During Rehearsal and Performance: a Linked Data Digital Library for Annotations. In *Proceedings of the 6th International Conference on Digital Libraries for Musicology* (The Hague, Netherlands) *(DLfM '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3358664.3358669

[16] Anna Plaksin. 2023. Understanding the needs of music editors in a digital world. Adding support for editorial markup to the mei-friend editor. In *Proceedings of the 10th International Conference on Digital Libraries for Musicology* (Milan, Italy) *(DLfM '23)*. Association for Computing Machinery, New York, NY, USA, 40–48. doi:10.1145/3625135.3625149

[17] Laurent Pugin, Rodolfo Zitellini, and Perry Roland. 2014. Verovio: A library for Engraving MEI Music Notation into SVG.. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*. ISMIR, 107–112. doi:10.5281/zenodo.1417589

[18] Klaus Rettinghaus and Alexander Pacha. 2022. Building a Comprehensive Sheet Music Library Application. In *Music Encoding Conference Proceedings 2021*, Stefan Münnich and David Rizo (Eds.). Humanities Commons, 41–48. doi:10.17613/s315-2y29

[19] Emiliano Ricciardi. 2024. Citations: The Renaissance Imitation Mass Project (CRIM). *Journal of the American Musicological Society* 77, 3 (12 2024), 863–875. doi:10.1525/jams.2024.77.3.863

[20] Robert Sanderson, Paolo Ciccarese, and Benjamin Young. 2017. *Web Annotation Data Model*. W3C Recommendation. W3C. https://www.w3.org/TR/2017/REC-annotation-model-20170223/.

[21] Laura KT Stokes. 2024. Nkoda. *Music Reference Services Quarterly* 27, 2 (2024), 91–95.

[22] Raffaele Viglianti. 2015. A Specification for Addressing Encoded Music on the Web. In *Music Encoding Conference Proceedings 2015, 2016 and 2017*, Giuliano Di Bacco, Johannes Kepper, and Perry Roland (Eds.). Bavarian State Library (BSB), 47–49. doi:10.15463/music-1

[23] D Weigl and K Page. 2017. A framework for distributed semantic annotation of musical score:" Take it to the bridge!". In *18th International Society for Music Information Retrieval Conference (ISMIR 2017)*. International Society for Music Information Retrieval.

[24] David M. Weigl, Werner Goebl, David J. Baker, Tim Crawford, Federico Zubani, Aggelos Gkiokas, Nicolas F. Gutierrez, Alastair Porter, and Patricia Santos. 2021. Notes on the Music: A social data infrastructure for music annotation. In *Proceedings of the 8th International Conference on Digital Libraries for Musicology (DLfM '21)*. Association for Computing Machinery, New York, NY, USA, 23–31. doi:10.1145/3469013.3469017

[25] Christof Weiß, Frank Zalkow, Vlora Arifi-Müller, Meinard Müller, Hendrik Vincent Koops, Anja Volk, and Harald G. Grohganz. 2021. Schubert Winterreise Dataset: A Multimodal Scenario for Music Analysis. *J. Comput. Cult. Herit.* 14, 2, Article 25 (May 2021), 18 pages. doi:10.1145/3429743