



Sopracasa, Alessio, Jan Bigalke, Numa Buchs, and Sima Meziridou. 2024. "Creating a Sigillographic Search Engine for Byzantium: Preliminary Results," in "Digital Approaches to Medieval Sigillography," ed. Martina Filosa, Claes Neufeind, and Claudia Sode, special collection. *Digital Medievalist* 17(1): 1–32. <https://doi.org/10.16995/dm.15235>.



## Creating a Sigillographic Search Engine for Byzantium: Preliminary Results

**Alessio Sopracasa**, Sorbonne University/Monde Byzantin, CNRS UMR8167, FR, [alessio.sopracasa@college-de-france.fr](mailto:alessio.sopracasa@college-de-france.fr)

**Jan Bigalke**, Cologne Center for eHumanities, University of Cologne, DE, [jbigalk1@uni-koeln.de](mailto:jbigalk1@uni-koeln.de)

**Numa Buchs**, Monde Byzantin, CNRS UMR8167, FR, [numa.buchs@college-de-france.fr](mailto:numa.buchs@college-de-france.fr)

**Sima Meziridou**, Monde Byzantin, CNRS UMR8167, FR, [sima.meziridou@college-de-france.fr](mailto:sima.meziridou@college-de-france.fr)

Many of the documents that were created in the Byzantine Empire (the former Eastern Roman Empire, 4th–15th centuries) are no longer extant, but the seals that accompanied the documents have survived in large numbers, providing information for various research fields within Byzantine studies. The ANR/DFG project DigiByzSeal aims to use digital presentation to enable new understandings of Byzantium and its written culture by transforming Byzantine sigillography. The project focuses on SigiDoc, which provides an XML-based and EpiDoc-compliant data model for the digital scholarly edition of Byzantine seals. The DigiByzSeal team is working on the first scholarly digital edition of approximately 4,000 seals, which will be freely accessible and based on open-source software. The goal is to create a centralized hub for Byzantine sigillography, with a unified federated search interface for all seals encoded with SigiDoc, based on a highly customized and enhanced instance of EFES (EpiDoc Front-End Services). This paper presents the preliminary results of the development of SigiDoc, and of the unified search interface based on it, and discusses the methodology and the challenges faced thus far, while showing how the project promotes open, shared, and accessible information and overcomes the issue of accessibility and lack of interoperability in Byzantine sigillography.

La plupart des documents créés dans l'Empire byzantin (ancien Empire romain d'Orient, IVe-XVe siècles) a disparu, mais les sceaux qui accompagnaient ces documents ont survécu en grand nombre, fournissant des informations pour divers domaines de recherche au sein des études byzantines. Le projet ANR/DFG DigiByzSeal vise à utiliser la présentation numérique pour renouveler la compréhension de Byzance et de sa culture écrite en transformant la sigillographie byzantine. Le projet se concentre sur SigiDoc, qui fournit un modèle de données basé sur XML et conforme à EpiDoc pour l'édition scientifique numérique des sceaux byzantins. L'équipe DigiByzSeal travaille sur la première édition scientifique numérique d'environ 4 000 sceaux, qui sera librement accessible et basée sur des logiciels open-source. L'objectif est de créer un portail centralisé pour la sigillographie

*Digital Medievalist* is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



byzantine, avec une interface de recherche fédérée unifiée pour tous les sceaux encodés avec SigiDoc, basée sur une instance hautement personnalisée et améliorée d'EFES (EpiDoc Front-End Services). Cet article présente les résultats préliminaires du développement de SigiDoc et de l'interface de recherche unifiée basée sur celui-ci et discute de la méthodologie et des défis rencontrés jusqu'à présent, tout en montrant comment le projet favorise une information ouverte, partagée et accessible et surmonte le problème d'accessibilité et de manque d'interopérabilité qui caractérise la sigillographie byzantine.

---

## Introduction

§ 1 The chosen object of this paper is Byzantine seals, coin-like objects made mostly of lead, whose two sides display iconographic depictions, inscriptions, or monograms (see **Figure 1**): these small discs of lead were used to authenticate and identify the sender of a document or a letter, especially in the context of the civil, military, and ecclesiastical administration of the Byzantine Empire (4th–15th centuries), both in its capital, Constantinople, and in its provinces (a large bibliography on Byzantine sigillography and related resources is available [Dumbarton Oaks Seals Bibliography 2024]; see also the seals editions listed by the *Prosopography of Byzantine World* [Jeffreys et al. 2017c]). In the framework of Byzantine studies, seals are a research object *per se*: very few of them have survived still attached to the document they authenticated, but the detached seals themselves have survived in large numbers (with a current estimate of around 100,000 specimens), and their increasing number gives us access to an increasing amount of information (prosopographical, administrative, geographical, and iconographical, among others). Byzantine sigillography shares with other fields, such as Byzantine epigraphy, the absence of certain foundations that would allow it to rise fully to the role of a discipline. Two elements in particular are lacking. On the one hand, there is an absence of shared rules about the preparation of scholarly editions (with the exception of a few widespread habits), and thus, more generally, no handbook; and on the other hand, there is no corpus of Byzantine seals, no *Corpus Sigillorum Ævi Byzantini*; there are many (very expensive) publications about collections that each contain several hundred seals, but in fact most editions are scattered in journals, edited books, *Festschriften*, and so forth. Nonetheless, the first steps that can help lead us to answers to the many open questions about Byzantine seals are currently underway in the DigiByzSeal project, an ANR/DFG project that will allow Byzantine sigillography to undergo the “digital turn.”



**Figure 1:** Seal of Roussel de Bailleul, 1070/72–1078 (A. Sopracasa collection, no. 7; published in Sopracasa and Prigent 2017, 715–723).

## 1. The quest for common and shared rules: Towards SigiDoc 2.0

§ 2 To address the first of the shortcomings mentioned above—the absence of shared rules about scholarly editions—and in order to address the issue of problems of accessibility (to both published and unpublished material), sustainability, and interoperability of Byzantine sigillographic material, SigiDoc was conceived, with its 1.0 version being released in September 2021. SigiDoc provides an XML-based and TEI-compliant data model for the digital scholarly edition of Byzantine seals: its aim is to establish a set of rules and guidelines that can become a *de facto* encoding standard. By adapting EpiDoc (Elliott et al. 2024)—and more broadly TEI—to Byzantine sigillography, we established a set of recommendations taking into account the most common features related to the contemporary scholarly edition of Byzantine seals.

### 1.1 A critical mass of data

§ 3 Within the framework of DigiByzSeal, the standard established in SigiDoc is currently being updated and further developed, and, after an intermediate version 1.1 with some minor changes, its 2.0 version is currently in the process of being released. SigiDoc 1.0 was based on the personal experience of its developers, and only a very small number of seals (about 40) had been encoded. For that reason, it was of the utmost importance in updating it to take into account a large, but manageable, amount of data. The DigiByzSeal team plans to work on about 4,000 seals: of these, about 3,000 are unpublished, housed in two French public collections (the Zacos collection at the *Bibliothèque nationale de France* and the collection of the *Institut français d'études byzantines*, both in Paris) and one German private collection (that of Robert Feind, in Cologne); the remainder of the seals come from two published collections, the Seyrig collection (Cheynet, Morrisson, and Seibt 1991) and the Tatiş collection (Cheynet 2019). It should be noted that, even though these collections are among the most important in the world, and scholars have been awaiting their publication, in some cases, for decades, the edition of these seals is not, in itself, the main focus of this project: these seals provide the data, case studies, and a long list of issues that are necessary for us to be able to enhance, amend, and expand SigiDoc. Half of the planned seals from the *Institut français d'études byzantines* are ready to be encoded, about 200 from the Feind collection have already been encoded, and the entirety of the paper published seals of the Seyrig collection has been digitally enhanced, while thus far only the editions of the legends have been encoded for the seals in the Tatiş collection.

## 1.2 The challenges of digitally enhancing a paper publication: Some remarks

§ 4 Working on unpublished material is extremely different from working on published seals. With the former, the encoder can apply the SigiDoc recommendations more comfortably because they are working with an unexplored object, whereas when it comes to the latter, the encoder has to deal with the previous editor's choices, which raises questions of the nature and the extent of any digital enhancement. The conversion of a publication into a digital form could be an appropriate moment for the creation of a revised edition; this would perhaps be especially useful for those publications that are not recent, as they could thus benefit from the contribution of material that had been subsequently published, in particular data from parallel seals, which would allow for the improvement of the edition and, in certain cases, also of our understanding and therefore of the historical commentary. But even taking a very conservative approach, and not producing a new edition, it is not possible to preserve the editor's choices and at the same time to consistently use the SigiDoc recommendations because the former are not always consistent. Thus some *ad hoc* solutions must often be proposed, as we will see.

<p>..  Π.Τ..   ΡΙΤΕΠ.. ΠΠΔΡΣΠ. ΦΛΑΓΤ... -Τ· -  [... κ]ριτ(ῆ) ἐπ[ὶ τ(οῦ)] ἰππ(ο)δρ(όμου) (καὶ) Π[α]φλαγ(ονίας) τ[ῶ ...]τι.</p>
---

**Figure 2:** Example of the transcription of inscription from a seal from the Tatiş collection (Cheynet 2019, no. 3.44).

§ 5 **Figure 2** gives an example of how an editor has transcribed the legend struck on the reverse of an 11th-century seal in the Tatiş collection. This edition raises several problems from a SigiDoc point of view. First, at the beginning of the legend, the diacritical signs used reveal the difficulty of providing an appropriate interpretation because the beginning of the inscription was left outside the support when the seal was struck; but on the other hand, solutions such as the use of white space do not tell us much, especially if this is being read by a machine. Second, again at the beginning, the diplomatic transcription (first line of **Figure 2**) and the editorial interpretation (second line of **Figure 2**) do not seem to be aligned with each other. And third, the fact that the editorial interpretation is edited on one single line hinders the alignment with the diplomatic transcription and, especially in the case of lost and restored texts, does not show the reader where, in the editor's opinion, the lines begin and end.

§ 6 For these reasons, the solution currently adopted by the project is the following:

## START CODE

```

<div type="textpart" subtype="face" n="rev" rend="linear" xml:lang="grc">
  <ab>
    <lb n="1" /><gap reason="lost" quantity="1" unit="line" />
    <lb n="2" /><gap reason="lost" extent="unknown" unit="character" /> <orig>π</orig>
      <gap reason="lost" extent="unknown" unit="character" /> <orig>τ</orig> <gap
        reason="lost" extent="unknown" unit="character" />
    <lb n="3" /><gap reason="lost" extent="unknown" unit="character" /> <expan><supplied
      reason="lost"><abbr>κ</abbr></supplied><abbr>ριτ</abbr><ex>ῆ</ex></expan> ἐπ
    <supplied reason="lost">ἰ</supplied> <supplied reason="lost"><expan>
<abbr>τ</abbr><ex>οὔ</ex></expan></supplied>
    <lb n="4" /><expan><abbr>Ἰππ</abbr><ex>ο</ex><abbr>δρ</abbr><ex>όμου</ex></expan>
    <expan><abbr><am><g ref="#kai" /></am></abbr><ex>καὶ</ex></expan> <expan>
<abbr>Π</abbr><supplied reason="lost"><abbr>α</abbr></supplied>
    <lb n="5" break="no" /><abbr>φλαγ</abbr><ex>ονίας</ex></expan> τ<supplied
      reason="lost">ῶ</supplied> <gap reason="lost" quantity="3" unit="character" />
    <lb n="6" />τι
  </ab>
</div>

```

## END CODE

After its transformation using the EpiDoc stylesheets partially modified by SigiDoc, this inscription can be visualized as seen in **Figure 3**.

1	[ - - - - - ]
2	[...?...]Π[...?...]Τ[...?...]
3	[...?...κ]ριτ(ῆ) ἐπ[ὶ τ(οὔ)]
4	Ἰππ(ο)δρ(όμου) (καὶ) Π[α-]
5	φλαγ(ονίας) τ[ῶ ...]
6	τι

**Figure 3:** Example of the transcription of inscription from a seal from the Tatş collection (Cheynet 2019, no. 3.44) after transformation.

§ 7 This example, and many other similar ones, demonstrate why one of the major endeavours of DigiByzSeal is the enhancement of the SigiDoc guidelines, starting with the section devoted to the encoding of the legend, known as “leidenisation” (Sopracasa et al. 2023): as a text-bearing object, a seal requires the application of conventions for the representation of non-verbal information such as lacunae, abbreviations, and so on through the use of symbols, brackets, and dots. The set of conventions adopted by SigiDoc is directly inspired by EpiDoc, with variations intended to make some of them more suitable for Byzantine sigillographic routines. Note that this does mean that some solutions, such as those adopted at lines 1–3 in the example in **Figure 3**, are more familiar to epigraphers and are not widely used (to say the least) by Byzantine

sigillographers; but their adoption was necessary in order to get a clearer and unambiguous edition, as the previous example clearly demonstrates.

### 1.3 Updating the guidelines and the template

§ 8 The current updating of the Guidelines relates to better explanations of when the various encoding recommendations should be used, giving many more examples (which come from the increasing number of digital editions produced by the project). For example, the use of `<gap/>` for lost lines has turned out to be not so self-evident, because `<gap/>` is also used for lost characters, which raises the question of which way of indicating a missing line or missing characters is the most appropriate in specific circumstances. What we recommend is to give as much information as possible: if a line is entirely lost or illegible, but the editor is able to state with certainty the number of characters that formed it, or can give a reasoned estimation, then the solutions would be to code either:

```
<gap reason="lost" quantity="5" unit="character" />
```

or

```
<gap reason="lost" quantity="5" unit="character" precision="low" />
```

Otherwise, the part of the inscription relating to the lacuna should be encoded as a lost or illegible line:

```
<gap reason="lost" quantity="1" unit="line" />
```

In this case it is important to note that, if the loss involves a known quantity of lines—more than one—the tag should be repeated separately for each line. As well as the issues with `<gap/>`, the use of combined tags, such as `<supplied/>` for lacunae and `<expan/>` for expansion of abbreviations, needed to be clarified.

§ 9 But the encoding of a seal’s legend is just one part of a SigiDoc edition: the SigiDoc template, which provides the structure for an edition, has been prepared to encode all the necessary data, such as the seal’s physical description, its history and dating, the edition of its legend, the critical apparatus, any commentary and bibliography, and so on. The stylesheet responsible for the transformation of the template into a web page readable by the final user has undergone several changes in the update: some field names have been changed in order to make them clearer (for example, in describing the appearance of the text, “lettering” has become “epigraphy”), while others have been moved in order to be better grouped with similar fields. In addition, to enforce consistency, best practice for the use of the template involves controlled vocabularies for elements and attributes: some lists of terms or attributes have been expanded (for example, for the “object type” field, the original “seal” has been joined by “marker,” “seal [uniface],” and “token”), replaced (“lettering” was renamed “epigraphy,” as



noted), or simply deleted, and the new attributes' values have been added to the SigiDoc schema. More guidance on how and when to use each vocabulary item is currently being added to the Guidelines.

§ 10 Even more importantly, some redundancies have been deleted from the template itself in the upgrade to version 2.0: some data was stored in such a way that it needed to be encoded twice, once for the visualization after transformation through the stylesheet and once for the machine, so that it would be available for uses such as indexing and searching. This can be seen in the following example, from the unpublished seal Feind S-112:

#### START CODE

```
<origDate type="analysis" notBefore="1081" notAfter="1100" evidence="prosopography"
resp="#MC">
  <seg xml:lang="en">Late 11th c.</seg>
  <seg xml:lang="de">Spätes 11. Jh.</seg>
  <interp type="datingCriteria" xml:lang="en">Prosopography</interp>
  <interp type="datingCriteria" xml:lang="de">Prosopographie</interp>
</origDate>
```

#### END CODE

As is clearly visible here, the element `<interp/>`, designed for displaying the information on a web page after transformation, encodes the same data as the `@evidence` attribute of the `<origDate/>` element. In the newly “polished” template, the use of `<interp/>` is now deprecated and the information, for both indexing and displaying, will be taken from `@evidence`, which has a predefined list of values added to the schema. For displaying on the web page, these values will be presented with an appropriate orthography (e.g., a value such as “historical-context” will be printed as “Historical context”) and translated into the modern language used by the editors of the corpus.

§ 11 In some other cases, redundancies existed because of the need to take into account the two faces of a seal: that is, some information relates to the whole object, while other information is about only the obverse or the reverse. Consider the following example. In the current SigiDoc template, the elements that are responsible for the description of the appearance of the whole seal, visualized after transformation in a field “General layout,” are:

#### START CODE

```
<layout n="whole">
  <rs type="layout">
    <seg xml:lang="en" />
  </rs>
</layout>
```

#### END CODE



Inside the `<seg/>` element the layout of the whole seal should be described according to a necessarily controlled list of terms describing the three recurrent features of the Byzantine seals: iconography, legend, and monogram (with the addition of an “undetermined” option). However, this same information is also encoded for each of the faces of the seal, using the same terms:

#### START CODE

```
<layout n="r">
  <rs type="layout">
    <seg xml:lang="en" />
  </rs>
</layout>
```

#### END CODE

The only change here is the value of the `@n` attribute, which is “r” for obverse or “v” for reverse. This redundancy has been solved in the update by deprecating the use of the element `<rs @type="layout"/>` inside `<layout @n="whole"/>` and keeping it only in the `<layout>` element for obverse and reverse. On the final web page, the field “General layout” will be filled in with the information encoded in the latter two elements, again according to the same controlled vocabulary, such as “Iconography, Legend,” “Monogram, Iconography,” and so on.

§ 12 The few other redundancies have been solved in similar ways, thus allowing the complexity of the schema to be reduced. As a result, the editors can work more easily within the XML edition files, and the tagless editor (see Section 1.5) is also simplified by these measures. Improvements such as these seem simple from a technical point of view, but they require a lot of discussion within the team and, because they affect the data and the way it is collected, they must be done carefully, as they have a profound impact on the work of the editors.

### 1.4 A chained schema

§ 13 Another key component of SigiDoc is currently in the process of being modified: the schema. In its 1.0 version, the SigiDoc schema was aligned with EpiDoc schema 9.2: the latter was considered in its entirety; some of its parts were identified as the most appropriate for SigiDoc, and, in some cases, these were then adapted to the specific requirements of Byzantine sigillography. SigiDoc XML files have been successfully validated against EpiDoc schema 9.2 and the `tei_all` schema. However, since the intention is to maintain alignment with TEI through EpiDoc, the need arose to find an effective and reliable way of managing the SigiDoc schema.

§ 14 The chosen method is the preparation of a chained ODD (“One Document Does it all”) (TEI Consortium 2024a). The preliminary work was already done in

preparing SigiDoc schema 1.0: this file was basically the file of EpiDoc schema 9.2, but modified for SigiDoc, and all the parts that were simply selected or were modified for SigiDoc had already been highlighted. A SigiDoc ODD chained to the EpiDoc schema is currently being edited accordingly, and it will be the first chained ODD of the EpiDoc schema: the work in progress is also aimed at aligning the SigiDoc schema with version 9.5 of EpiDoc (the latest version available at the time of writing this article), both subsetting and supersetting it.

§ 15 A very small example of this concerns the element `<supplied/>`, used in the edition of a seal’s legend to encode some restored text. This is a case where the SigiDoc schema is more restrictive than EpiDoc’s schema when it comes to the attributes, so it is worth presenting the whole “genealogy” of it, starting with the `tei_all.rng` schema below (available at TEI Consortium 2024b):

#### START CODE

```
<define name="supplied">
  <element name="supplied">
    <a:documentation
xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">(supplied) signifies text
supplied by the transcriber or editor for any reason; for example because the original
cannot be read due to physical damage, or because of an obvious omission by the author or
scribe. [11.3.3.1. Damage, Illegibility, and Supplied Text]</a:documentation>
    <ref name="macro.paraContent" />
    <ref name="att.global.attributes" />
    <ref name="att.editLike.attributes" />
    <ref name="att.dimensions.attributes" />
    <optional>
      <attribute name="reason">
        <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">one
or more words indicating why the text has had to be supplied, e.g. overbinding, faded-ink,
lost-folio, omitted-in-original.</a:documentation>
        <list>
          <oneOrMore>
            <data type="token">
              <param name="pattern">[^\p{C}\p{Z}]+</param>
            </data>
          </oneOrMore>
        </list>
      </attribute>
    </optional>
    <empty />
  </element>
</define>
```

#### END CODE

Compared to the broadness of this element in the TEI schema, the EpiDoc customization (following the EpiDoc ODD 9.5; see Bodard et al. 2023) introduced a major restriction, making the attribute `@reason` required and not optional (as it is in the TEI schema), and with a closed list of constraining values. In addition, EpiDoc

explicitly introduced an optional usage of the attribute `@evidence`, with a list of suggested values, while in the TEI schema, `@evidence` was generically part of the `editLike` attributes. The EpiDoc customization is:

#### START CODE

```
<elementSpec ident="supplied" module="transcr" mode="change">
  <attList>
    <!-- suggested attribute List for supplied/@evidence -->
    <attDef ident="evidence" mode="replace" usage="opt">
      <datatype maxOccurs="1">
        <dataRef key="teidata.word" />
      </datatype>
      <valList type="semi">
        <valItem ident="parallel">
          <desc xml:lang="en" versionDate="2020-02-17">Characters supplied from a parallel
            version of this text, absent from this one</desc>
        </valItem>
        <valItem ident="previouseditor">
          <desc xml:lang="en" versionDate="2020-02-17">Characters visible to a previous
            editor, now lost/unclear due to damage</desc>
        </valItem>
      </valList>
    </attDef>
    <!-- constrained attribute List for supplied/@reason -->
    <attDef ident="reason" mode="replace" usage="req">
      <valList type="closed">
        <valItem ident="lost">
          <desc xml:lang="en" versionDate="2020-02-17">Restoration of text completely lost
            from the surface</desc>
        </valItem>
        <valItem ident="omitted">
          <desc xml:lang="en" versionDate="2020-02-17">Supplement of text erroneously
            omitted from the surface by the scribe/cutter</desc>
        </valItem>
        <valItem ident="subaudible">
          <desc xml:lang="en" versionDate="2020-02-17">Editorial supplement of words
            subaudible to the native reader</desc>
        </valItem>
        <valItem ident="explanation">
          <desc xml:lang="en" versionDate="2020-02-17">Editorial explanation of oblique
            text in original</desc>
        </valItem>
        <valItem ident="undefined">
          <desc xml:lang="en" versionDate="2020-02-17">Text restored (e.g. from earlier
            witness or alternate copy) without specification of current state of this
            text</desc>
        </valItem>
      </valList>
    </attDef>
  </attList>
</elementSpec>
```

#### END CODE

The SigiDoc ODD inherited the changes made by EpiDoc and has introduced one more value for `@evidence`, without changing the behaviour of the list:

## START CODE

```

<elementSpec ident="supplied" module="transcr" mode="change">
  <attList>
    <!-- suggested attribute added to supplied/@evidence -->
    <attDef ident="evidence" mode="change" usage="opt">
      <valList type="semi" mode="change">
        <valItem ident="similar" mode="add">
          <desc xml:lang="en" versionDate="2023-06-26">Characters supplied from a similar
            but not strictly speaking parallel version of this text.</desc>
        </valItem>
      </valList>
    </attDef>
  </attList>
</elementSpec>

```

## END CODE

§ 16 The Schematron language integrated into the ODD schema has also allowed additional constraints, in particular the use of conditional expressions. It sometimes happens, in particular with widely used attributes such as @type and @subtype, that a specific value of the one attribute implies the presence a selected list of values in the other, and a <constraintSpec/> element allows the encoding to focus on the appropriate values, as in the following example, relating to the element <provenance/>:

## START CODE

```

<constraintSpec ident="provenance" scheme="schematron" mode="add">
  <!-- constraining values of provenance/@subtype according to the @type's values -->
  <constraint>
    <sch:pattern>
      <sch:rule context="tei:provenance[@type='found'][@subtype]">
        <sch:assert test="@subtype = ('discovered','rediscovered','first-seen','first-
          recorded')">With provenance/@type='found', @subtype's values are
            'discovered','rediscovered','first-seen','first-recorded'.</sch:assert>
      </sch:rule>
      <sch:rule context="tei:provenance[@type='observed'][@subtype]">
        <sch:assert test="@subtype =
          ('seen','recorded','identified','photographed','autopsied','squeeze-
          taken','rubbing-taken','ink-transfer-taken','reported','built-
          into','reappeared')">With provenance/@type='observed', @subtype's values are
            'seen','recorded','identified','photographed','autopsied','squeeze-taken',
            'rubbing-taken','ink-transfer-taken','reported','built-into',
            'reappeared'.</sch:assert>
      </sch:rule>
      <sch:rule context="tei:provenance[@type='not-observed'][@subtype]">
        <sch:assert test="@subtype = ('lost','stolen','destroyed','drawn','reported-
          lost','reported-stolen','reported-destroyed')">With provenance/@type='not-
          observed', @subtype's values are 'lost','stolen','destroyed','drawn',
            'reported-lost','reported-stolen','reported-destroyed'.</sch:assert>
      </sch:rule>
      <sch:rule context="tei:provenance[@type='transferred'][@subtype]">
        <sch:assert test="@subtype =
          ('moved','sold','given','loaned','auctioned','donated')">With
            provenance/@type='transferred', @subtype's values are 'moved','sold','given',
            'loaned','auctioned','donated'.</sch:assert>
      </sch:rule>
    </sch:pattern>
  </constraint>
</constraintSpec>

```

## END CODE

All of the relevant values are described in an `<attList/>` element right after the Schematron constraint cited here.

### 1.5 The authoring environment: A tagless editor

§ 17 A major step towards the creation of an authoring environment that is able to support the editorial work and consistency in the encoding endeavour has been the development of a tagless editor (see **Figure 4**) based on a BaseX database providing APIs in combination with a Vue.js Javascript front end, which is currently under development (SigiDoc Editor 2024). This tool, associated to the SigiDoc schema, template, and stylesheet, allows metadata (it is not yet possible to encode the legend) to be entered into the SigiDoc template without seeing the template itself, thus making it accessible to those who do not have specific training in XML and the SigiDoc standard. In addition, the tool provides a pre-visualization of the compiled fields, and more broadly, of the final web page. Note that the provision of this tool is by no means an invitation to avoid learning XML and gaining training in SigiDoc, because encoding is not a mere action of filling in empty fields. Rather it should be seen as an effective way to verify the completeness and the consistency of encoding that has already been done.

Statements	
<b>titlestmt</b>	
Title	
Seal of Michael, vestarches and oikonomos of the Nea	en REMOVE
Sceau de Michel, vestarque et économe de la Néa	fr REMOVE
Add	
<b>Siegel Editor</b>	
CS	Add
<b>EditionStmt</b>	
Edition:	
Digitale Version	de REMOVE
Digital conversion	en REMOVE

<b>Seal of Michael, vestarches and oikonomos of the Nea</b>
<b>Artifact</b>
Type Seal
General layout Iconography
Matrix Sigidoc ID -
<b>Physical description</b>
Type of impression Original impression
Material Blei
Dimension(mm) Diam. 5
Weight (g) 500
Axis (clock) 4
Overstrike orientation (clock) 1
Channel orientation (clock) 2
Execution Struck

**Figure 4:** The interface of the tagless editor. On the left are the fields to be filled in with the edition data; on the right, the pre-visualization of them.

§ 18 The editor will support the editing of data by automatically integrating the various authority lists that are created and implemented during the editorial process (see Section 2.2). This will simplify the encoding thanks to the ability to use auto-completion and the suggestions that can be automatically made for at least a part of the annotation work. In order to be effective, it is important for this tool not to be too restrictive and intrusive to the point of overly restricting the variety of data that a user will need to enter. On the other hand, the fact that the controlled vocabularies

for elements and attributes developed for SigiDoc are integrated into this tool, coupled with the automatic validation offered by the editor, will mean a reduction in the number of errors during the creation of the XML files.

## 2. Towards a virtual corpus of Byzantine seals

§ 19 SigiDoc, in its various aspects (Guidelines, schema, template, stylesheet, controlled vocabularies), is an instrument that allows a single publishing endeavour to be connected to others.

### 2.1 EFES: Indexes and search form

§ 20 The tool that allows seal editions created with SigiDoc to become a proper digital publication is EFES. This was created as a customization platform for EpiDoc files, and was adopted by SigiDoc in 2017. EFES has since been adapted to the requirements of SigiDoc and of Byzantine sigillography and is the main way to present SigiDoc editions, with a web interface, indexes, and a search form. The platform thus allows extensive control over the editorial process at an individual level. Because EFES is the means by which the encoding work done with SigiDoc is presented to the final user, a great deal of attention has been paid to this platform from the beginning of DigiByzSeal. The list of indexes and the search form of the original EFES version created for EpiDoc had already been changed during the development process that led to SigiDoc 1.0, and these changes can be seen in the test corpus published then (now considered deprecated; see SigiDoc 2024a). These features have been changed again while preparing the version 2.0 of SigiDoc, as listed in **Table 1**. The major change that has been made so far to indexes for version 1.1 relates the first index, which is the prosopographical repertoire of the sigillographic edition: it has been renamed more appropriately as “Persons” instead of “Personal names” because it is not an onomastic list. In addition, there is only one index for “Offices,” which groups together the three types of functions of the Byzantine administration found on seals: the type of office—civil, military, ecclesiastical—is recorded for each entry.

List of indexes in EFES-SigiDoc	List of search filters in EFES-SigiDoc
Persons	Institutions
Place names	Collection
Dignities	Persons
Offices	Personal names
Marian terms	Family names
Christ-related terms	Milieu

(Contd.)

List of indexes in EFES-SigiDoc	List of search filters in EFES-SigiDoc
Saint-related terms	Place names
Iconography	Dignities
Monograms	Civil offices
Lemmata	Ecclesiastical offices
Legends' cases	Military offices
Metrical legends	Marian terms
Invocations	Christ-related terms
	Saint-related terms
	Iconography
	Monograms
	Legends' cases
	Metrical legends

**Table 1:** List of indexes and search filters in the EFES instance for SigiDoc.

§ 21 Perhaps more importantly, these indexes have been made interoperable with external data: columns have been added to the persons and place names indexes to link the data to external prosopographies, such as the *Prosopography of the Byzantine World* (PBW) (Jeffreys et al. 2017a), and to the *Pleiades gazetteer* (Stoa Consortium 2024); many other implementations of the same kind are planned during the next months.

§ 22 The decision was taken to limit the indexes to a short list covering the major features coming from the seals. Conversely, a more detailed list of search filters has been created in order to allow for a more granular search. Besides the topics already in the indexes, it is also possible to perform queries for repositories and collections, for onomastics, and for specific types of offices. On the other hand, topics such as invocations or lemmata have not been included because they offer limited possibilities of interaction with the other filters in order to refine the search.

## 2.2 Authority lists

§ 23 In order to create indexes and search filters, EFES uses authority lists in XML files. These lists contain all the data that a specific edition considers to be of importance: they are the essence of what legends offer to sigillographers and to historians. These lists contain a regularized version of names and terms and act as a set of controlled vocabularies to ensure consistency and interoperability within the same project and across projects: this is the reason that they cannot be the “possession” of a single project, but must be shared by all SigiDoc projects, each one enriching them for the use of all.



§ 24 All the changes to indexes and search filters mentioned above lead to changes to the corresponding authority lists. If one takes the index of “Offices,” for example, all the offices are listed in one single XML authority file, organized according to three `<list/>` elements, each with a different `@type` attribute (civil, ecclesiastical, military). The terms are nested inside the `<list/>` with `<item/>` elements, in which all the variants (excluding formal variants, i.e., different ways to write the same term) of a function are grouped together. This can be seen in the following example, in which the office of judge is recorded as such—under `<list type=“civil“/>`—but also together with the different types of judge and different geographical areas over which a judge might exercise their duties, each one having its own identification number:

#### START CODE

```
<item ana="κριτής" xml:lang="grc">
  <term xml:id="krites_001">κριτής</term>
  <term xml:id="krites_002">κριτής τοῦ βήλου</term>
  <term xml:id="krites_003">κριτής ἐπὶ τοῦ Ἴπποδρόμου</term>
  <term xml:id="krites_004">κριτής τῶν Βουκελλαρίων</term>
  <term xml:id="krites_005">κριτής τῶν Θρακησίων</term>
  <term xml:id="krites_006">κριτής τῆς Μακεδονίας</term>
</item>
```

#### END CODE

§ 25 The management of prosopographical data is one of the most challenging aspects of Byzantine sigillography, both for “analog” as well as digital data sets. And, of course, information on persons is traditionally what specialists, and more broadly researchers, are most likely to be looking for in a sigillographic edition. Thus this has been an area of considerable work in the development of SigiDoc by the DigiByzSeal team.

§ 26 The first solution that was adopted by the project for the problem of processing this type of data was to enhance the authority list, in accordance with the use of EFES and with what has been done for the other types of data. Consequently, the prosopographical authority list of SigiDoc 1.0 has undergone a deep restructuring to make it responsible for the “Persons” index and for the search filters “Persons,” “Personal names,” and “Family names.” Let’s consider the example of Romanos Kourkouas, a military officer of the 11th century. In version 1.0 of the prosopographical authority list, there was nothing more than a regularized form of the forename and of the surname, encoded together, with the corresponding `xml:id`. In version 1.1, on the other hand, the DigiByzSeal team has encoded a lot more information: “forename” and “surname” together and separated, their original and regularized forms, the edition from which the data comes, gender, dates of birth and death or

period of activity, and bibliographic references (in particular, links to external prosopographies):

#### START CODE

```
<person sex="m" xml:id="romanos-kourkouas" resp="#AS">
  <persName>
    <forename>
      <reg xml:lang="grc">Ρωμανός</reg>
      <orig source="TM_23">Ρωμανῶ</orig>
    </forename>
    <surname>
      <reg xml:lang="grc">Κουρκούας</reg>
      <orig source="TM_23">Κουρκούα</orig>
    </surname>
  </persName>
  <floruit notBefore="1000" notAfter="1020">Early 11th century</floruit>
  <bibl>
    <ptr target="PBW" />
    <citedRange>PBW Romanos 102</citedRange>
    <link target="http://pbw2016.kdl.kcl.ac.uk/person/Romanos/102/" />
  </bibl>
</person>
```

#### END CODE

With only one instance of markup and a single authority list, three different search filters and an index are generated. In addition, this allows two problems—of opposite types—to be avoided. On one hand, it avoids multiple duplicates in the onomastic list (the “Personal names” and “Family names” filters), where people with identical forenames or identical surname need to be grouped: for example, all the “George(s)” in the corpus need to be collected together by the “Personal names” filter as “George,” even if they are not the same person. On the other hand, in the prosopographical (i.e., real persons) list, there is the opposite problem, in that the homonyms that are different persons have to be separated: in this list, different “George(s)” must be separated if they are different persons just bearing the same forename but, conversely, grouped together when there are multiple mentions of the same person bearing the forename “George.”

§ 27 This solution solves several problems. However, the prosopographical list is far and away the most demanding list to maintain over the medium to long term, especially for a project that aims at sharing its foundation files with other endeavours using SigiDoc and EFES. This list is an ever-growing file, with an unpredictable input of data as the number of seals published and/or encoded increases: there are always new persons or new mentions of already known people to be encoded. In contrast, the other authority lists tend to be relatively stable, and, importantly, they do not present any challenge related to homonymy, disambiguation, or the identification of the same

individual across multiple seals, whether those seals are parallel or span different phases of a person’s career. For this reason, the DigiByzSeal team decided to adopt a completely different approach: instead of bringing the seal to the authority file, the idea is to embed the authority information into the seal’s XML edition file. A list of the persons cited by the seal (i.e., the issuers or owners of the seal) is to be created in the TEI header, inside the `<sourceDesc/>` element together with the metadata of the edition—for example:

#### START CODE

```
<listPerson type="issuer">
  <person gender="M" role="civil" sameAs="#s-a3v87b">
    <persName xml:lang="grc">
      <forename>Βασίλειος</forename>
      <surname>Μαυρός</surname>
    </persName>
    <persName xml:lang="en">
      <forename>Basil</forename>
      <surname>Mauros</surname>
    </persName>
    <persName xml:lang="fr">
      <forename>Basile</forename>
      <surname>Mauros</surname>
    </persName>
    <idno type="PBW" ana="http://pbw2016.kdl.kcl.ac.uk/person/Basileios/20221/">PBW
      Basileios 20221</idno>
    <idno type="PLP" />
    <idno type="PLRE" />
    <idno type="PmbZ" />
    <idno type="PcBE" />
    <idno type="PBE" />
  </person>
</listPerson>
```

#### END CODE

The `<listPerson/>` allows for more than a single issuer (i.e., there may be several `<person/>` elements within it); this is often the case with imperial seals or seals issued by fiscal agents known as *kommerkiarioi*. In the `<person/>` element, a `@sameAs` attribute is used to refer to the same person if they are already attested on other seals in the corpus; its value is the SigiDoc ID of these other seals (on which, see Section 2.3), and there can be more than one value.

§ 28 Note that within the XML edition file of the seal, the Greek spelling of the full name is presented only in its regularized form, followed by a translation (or transliteration) into modern languages. The prosopographical index will present the full name, in Greek, while among the search criteria there will be only onomastic lists—one for the first names and the other for the family names—to allow for a broader or, alternatively, for a more specific search as needed, since these filters can be applied

individually or in combination. For accessibility purposes, these onomastic lists will be populated with the English versions of the names (this is the default translation into a modern language), while the translations into other modern languages will be available through the string search, allowing searches for, for example, Ἰωάννης, John, Jean, Giovanni, and so on. For example, to find Romanos Kourkouas, one would perform two queries, one for “Romanos” and the other for “Kourkouas”; otherwise, selecting “Romanos” only, the result will give all the persons with this forename, including the Romanos with “Kourkouas” as a family name (and similarly if one selected “Kourkouas” only). The “original” spelling of the name, as it appears on the seal’s inscription, will be taken directly from the edition. Finally, `<listPerson/>` being a prosopographical element, it is not only appropriate but also necessary to associate this kind of data to already existing prosopographies, referring to their own ID, and explicitly linking to them if they are online, or simply encoding their IDs if they are non-digital resources.

§ 29 Before going any further, it is important to point out that physical persons were not the only ones to use seals: many were struck by institutions or groups, in different milieus, such as the Church (e.g., the priests of Hagia Sophia, the patriarchal see of the empire, or individual monasteries), the civil administration (e.g., the *kommerkia*, i.e., the fiscal services), and so on. These issuers cannot be encoded with elements designed for persons, and so a `<listOrg/>` element—structured in a way mostly parallel to `<listPerson/>`—has been added to the schema and to the template:

#### START CODE

```
<listOrg type="issuer">
  <org role="corporate" type="civil">
    <orgName xml:lang="grc">βασιλικά κομμέρκια</orgName>
    <orgName xml:lang="en">Imperial kommerkia</orgName>
  </org>
</listOrg>
```

#### END CODE

§ 30 The other major decision recently taken about the prosopographical data is not to assign an identifier to persons in the XML edition files, for a number of reasons. First, despite the importance of this kind of data for Byzantine studies, we are not preparing a prosopographical repertoire; our starting point is the seal, not the person. Second, the choice made by the project to have shared and common authority lists raises a major problem of coordination, and this is much more important for the prosopographical data than for other authority lists: it is particularly difficult to ensure in the long term that there is consistency in the attribution of person identifiers and to avoid duplication of entries for the same person. In addition, even if this means

that the prosopographical information must be encoded in every XML edition file, this solution is not a burden, and it does not really introduce redundancies across the different edition files. In fact, it makes the encoding simpler and quicker compared to the earlier solution. Previously, as explained above, a prosopographical authority list laid behind indexing and searching and, in the edition template, a now deprecated element `<author/>` was used to encode data on the issuer of a seal, specifically for the purposes of visualization on the web page after transformation. This introduced a coding duplication, in addition to not being an entirely satisfactory solution, as it had several limitations. In the new solution, both these tasks are dealt with by `<listPerson/>`. Now the issuers mentioned in the inscription are identified with their seals, which have IDs (see Section 2.3).

§ 31 However, discarding the prosopographical authority list and person IDs does not mean there is no persons list at all. Instead, the raw data encoded through the `<listPerson/>` elements now provides the foundation of the list. This data can be regularly extracted—at intervals based either on a fixed number of newly encoded seals (e.g., every 200 seals) or on a period of time (e.g., quarterly)—and a list compiled that associates the forename and surname of individuals from `<listPerson/>` with the SigiDoc ID of each seal on which they appear. The background work for this list is done by inserting a `@sameAs` (or alternatively a `@corresp`) attribute, as mentioned above, when there is a confirmed match with a person on another seal, whether that is within the same corpus or across different corpora—for example, when we know that Emperor Justinian I, who has seals in the Feind, Seyrig, and Zacos collections, is one and the same person. The list created in this process will be a versioned and public list, published as an ongoing data paper on a platform such as Zenodo, which automatically assigns a DOI and versioning, and so can be specifically referenced.

§ 32 As will be explained below (see Section 2.6), all the seals encoded with SigiDoc by different research teams will be made available through a federated search portal allowing the simultaneous perusal of all this material. This will enable an editor working on collection B to find—if available—a seal edited in collection A and issued by a person the editor considers to be the same as the one who issued the seal they are editing, and to point to it from that seal in collection B using the `@sameAs` attribute (whose value will be the SigiDoc ID of the seal from collection A). However, this solution must be considered merely as the first step, because the results to which it may lead are incomplete, for several reasons. First, the data encoded in the `@sameAs` attribute in collection B is one-way only, and cannot be applied—at least not automatically—to the relevant seal in collection A. Second, while collection B is being edited (or encoded), collection C may also be in preparation; thus neither collection B nor collection C is available through the common search portal, and so a

possible match between them would be totally overlooked. Third, the possibility of a mere oversight cannot be excluded, of course, especially if names appear in different forms on different seals. One consequence, among others, is that the same person may appear several times associated with different IDs (the SigiDoc IDs of the seals on which the name is inscribed), and with no link between them, because the seals come from different collections that did not communicate with each other.

§ 33 These issues lead to an inevitable conclusion: data reconciliation and quality control of the extracted persons list need to be done manually. A community effort is necessary: the researchers involved in DigiByzSeal can be considered as the core from which a dedicated community needs to develop—and this is one of the objectives of the project itself—but under no circumstances should these researchers consider themselves, or be considered by others, as the arbiters of the work of their colleagues, imposing their own choices on them. As soon as new data is extracted that reveals a need for a reconciliation, the editors and encoders responsible for the new seal editions will be alerted and asked to give their opinion on whether, for example, two seals bearing the same name were actually produced by the same person. In this way, the freedom of editors to autonomously interpret their own material will be preserved, while at the same time, the examination of similar material that had not been previously taken into consideration will retrospectively help improve sigillographic editions.

§ 34 After this human effort of reconciling data, all the SigiDoc IDs belonging to a single person will be gathered together under this person's name, and a meaningless and automatically generated identifier will finally be attributed to this name. This identifier then serves as the aggregating factor between identical individuals. Once the expert decisions have been made and any unique identifiers have been generated, so that the SigiDoc IDs of a recently edited collection are linked with any corresponding IDs in other collections, inside the EFES infrastructure, an index can be instantly generated that groups together any seals that refer to the same individual (whether there is only one such seal or many). There are cross-references between individuals and editions, enabling seamless navigation between the two. The prosopographical list and the index are constantly evolving, but they are versioned, allowing reference to the current version or any preceding version, and so can act as a prosopographical gazetteer.

§ 35 This is a more manageable solution than an authority list and an ID list that are left to individual projects or even to individual encoders. This would still require data reconciliation, but with a much greater degree of complexity and “messiness.” The community effort that will be demanded by the management of this list is no heavier than what already happens in comparable digital communities such as EpiDoc or Digital Classicist.

### 2.3 IDs

§ 36 In the previous section, we mentioned SigiDoc IDs. In order to have a proper digital existence, Byzantine seals need unique identifiers, which they certainly do not have today. Such identifiers, better than any other solution, would allow them to be integrated into the mass of data available online and, equally, to benefit from this same data; in short, they become linked open data. But IDs for seals are also an important tool for performing the most traditional and basic scientific activity during the edition of a seal: to find and gather together any parallels, that is, any seals struck with the same matrix. The SigiDoc edition of a seal includes a SigiDoc ID for the whole seal, considered as both the support and the imprint struck on it. The way that the DigiByzSeal team is currently implementing this is through a function that—when an XML edition record is created—randomly generates an alphanumeric string of six elements preceded by a prefix “s” (for seal), such as s-o5zje4 (or, once encoded, `<idno type="SigiDocID">s-o5zje4</idno>`). This “ID generator” will also be integrated into the tagless editor discussed in Section 1.5. Using this solution, the chance of identical IDs being created in different projects is virtually nil (as there are  $62^6$  or over 56 billion possible combinations), and the need—the heavy burden—for a centralized list of available and “to be distributed” IDs is obviated. These IDs could also be converted into URNs.

§ 37 However, as well as the seals, the matrices used to create them also need IDs. These matrices—iron plier-like instruments known as *boulloteria* (sg. *boulloterion*) in Byzantium—were used to strike the inscriptions and images of the finished seal onto a blank mostly made of lead. The inner surface of the two ends was engraved with the negative imprints of the two faces of the seal, as can be seen in **Figure 5**.



**Figure 5:** *Boulloterion*. Obverse: Bust of Saint Nicholas. Reverse: Inscription with invocation and owner's name and title. Object no. 1951.31.6, Harvard Art Museums collections online, <https://hvard.art/o/291105> (accessed December 13, 2023).



While Byzantine seals have never been uniquely identified, some of their *boulloteria* have received IDs within the *Prosopography of the Byzantine World*, with the concurrent creation of a permalink (see Jeffreys et al. 2017b). SigiDoc IDs for matrices will be generated following the same procedure adopted for seal IDs, with the difference that in the case of matrices the random sequence will be preceded by the prefix “m” (for matrix); so, for example, m-12rbX7. When a matrix is already in the PBW, a simple equivalence between the two IDs is established and shown on the web page of the edition. This information is encoded with the following elements:

#### START CODE

```
<rs type="matrix" subtype="not-surviving">
  <idno type="SigiDoc" />
  <idno type="PBW" corresp="URI" />
</rs>
```

#### END CODE

§ 38 Earlier in this section, we mentioned parallel seals. Thanks to the IDs assigned to both seals and the matrices that left their imprint on them, it is possible to gather together all the seals produced by the same *boulloterion*. This means that all the SigiDoc IDs of the parallel seals struck with a specific matrix will be brought together under the matrix’s unique ID. Seal IDs are not a problem, because each specimen has its own; in contrast, each matrix ID is “reusable,” in the sense that it should be applied to all the parallel seals struck with the same tool. Once the federated search portal is ready, an editor (or encoder) will be able to check for parallel seals already encoded in other collections and, once found, apply the same identifier to the matrix of their own seal(s). However, this solution raises issues similar to those already seen with the prosopographical list, and that is why the proposed solution is the same: periodically, matrix IDs will be extracted together with the SigiDoc IDs of the seals struck by them, and after a data reconciliation undertaken by a community of experts, the IDs of the parallel seals will be grouped under the same matrix ID (if this is necessary, of course, since seals not always have known parallels, so one matrix ID will correspond to only a single seal ID).

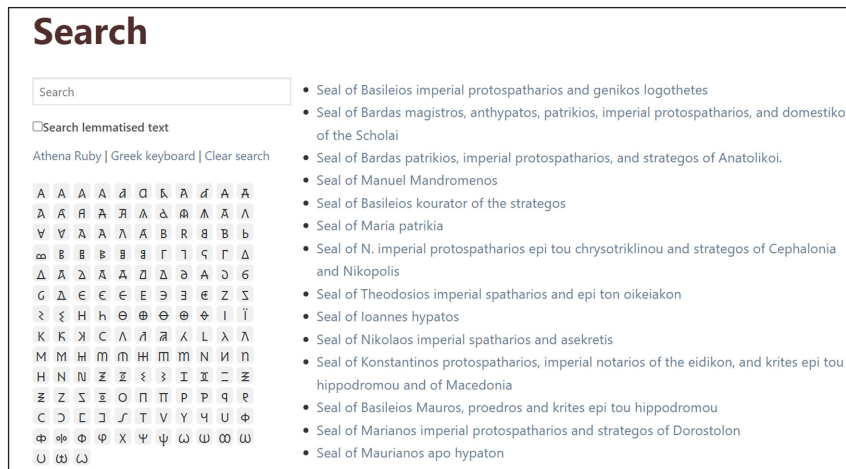
§ 39 As for the prosopography, a general index of matrices can be generated, thus fulfilling one of the ultimate aspirations of Byzantine sigillography: the virtual reconstitution of the matrices that produced those seals that have come down to us. An extremely small number of *boulloteria* (less than ten) still exist, which means that in the code cited above, the value of the @subtype attribute will almost always be “not-surviving.” Matrices are the best way to organize and aggregate seals, and to highlight their seriality, which is less obviously self-evident than it is for coins. They

can also become a virtual object: with the juxtaposition of several parallel seals, it will be possible to develop a clearer idea of the appearance of a matrix, especially in case of seals in a bad state of preservation, for which a “virtual collation” can return the text and image in their entirety. The Sigilla project on seals from medieval France uses the close concept of “sceau-type,” defined as “the digital avatar of the matrix used to produce the impression,” the “ultimate and absolute virtual impression, as imagined, defined, and validated by the owner of the seal” (Baudin et al. 2024; our translation).

#### 2.4 *Special fonts and the epigraphy of the seals*

§ 40 The legend inscribed on Byzantine seals is important as a meaning-bearing text, but equally there is a visual aspect that needs to be handled: the letters composing the inscribed text have specific shapes (a major dating criterion for sigillographers). This forces the researcher to choose between a typological and a visual rendering of the lettering, with specific fonts being required for the latter. The choice made for SigiDoc is visual rendering, through the adoption of an OpenType and Unicode-compliant font called Athena Ruby, which has been integrated into SigiDoc’s EFES version (for a discussion of why this choice was made, see Sopracasa, Filosa, and Stoyanova 2020, 113–118). This font, available at Dumbarton Oaks (Dumbarton Oaks Athena Ruby 2024), is used for the diplomatic transcription of the legend, but a definitive choice on how to encode it has not yet been made by the DigiByzSeal team. The alternatives are an extremely detailed encoding that includes all the characters or an encoding that is limited to characters considered to be “special”; the first solution promises the best results for searching within legends but is very time-consuming, while the second is more practical but raises the question of the criteria for making the selection. Thus far, Athena Ruby is being used primarily as a display font, with limited search possibilities: EFES now allows a basic search for specific characters in Athena Ruby thanks to a recently added keyboard (which needs to be expanded to include all the glyphs in the current Athena Ruby palette), as seen in **Figure 6**—in addition to the keyboard for polytonic Greek—but all character variants are mapped by default to the single corresponding letter (e.g., different variants of *mu* will give the standard  $\mu$  as a result).

§ 41 Athena Ruby’s palette is constantly growing: at time of writing, the current version is version 021, released in March 2021. This font can be freely modified; however, the developers of Athena Ruby, the research centre at Dumbarton Oaks (Washington, DC), have made themselves available to add any variants reported to them. Rather than creating local versions of the font, it is certainly preferable to adopt this second, more centralized solution, since in this way one can be certain that all



**Figure 6:** The Athena Ruby keyboard in the EFES search form.

potential users will have the same version available to them, accessible from a single place. For this reason, during the editorial process in the DigiByzSeal project, the sigillographers involved set themselves the task of listing any possible variants not found in the Athena Ruby palette. It turns out that the types of glyph that are most lacking are those for abbreviation signs and ligatures, plus some specific forms of letters such as *omega* or *mu*.

## 2.5 Visualization of geographical data: A Leaflet map

§ 42 Recently, the DigiByzSeal team started to address one of the major topics of the project: the management of geographical data. From a sigillographic point of view, this data is mainly of two kinds: places mentioned in the legend of a seal, and the place where the seal was found. The indexes and the search form take into account the first kind of place, while the second has a specific `<provenance type="found"/>` element in the template. A major shortcoming of Byzantine sigillography is the lack of information concerning the circumstances (where, when, who, how) of the discovery of a seal. For this reason, among others, a collaboration between the DigiByzSeal project, on the one hand, and the Regional History Museum of Shumen and the National Archaeological Institute with Museum of Sofia in Bulgaria, on the other, is of particular interest for the development of the SigiDoc infrastructure. Contrary to what has just been said about seals in general, there is a lot of information about finds on Bulgarian soil, and this is a test bed of the utmost importance for implementing the parts of the template that are dedicated to these topics (several `<provenance/>` elements with different `@type` attributes) and which are currently underdeveloped. One of the first endeavours is the development, currently underway by the Bulgarian team (in particular Vasko Nikolov),

of a Leaflet map gathering together findspots and places mentioned by the seals: this map is now part of the EFES version of the Bulgarian team and will be soon transferred to the standard EFES version available to all SigiDoc edition projects. This is certainly one of the best ways to add value to geographical data and to make it available as linked open data.

## 2.6 Federated search portal

§ 43 The ultimate goal of the DigiByzSeal project is to create a centralized hub for Byzantine sigillography, whose core will be represented by a federated search portal for all seals encoded with SigiDoc (on this portal, see also Neufeind et al. 2024). It is not a matter of gathering all Byzantine seals in one hyper-corpus. Quite the contrary: the use of SigiDoc as a common and shared tool allows, on the one hand, the totally independent creation of several collections or corpora, and, on the other hand, the interoperability between each of them, as all the editions made with SigiDoc will be unified virtually by the use of this tool, and their content will be available for consultation through the common search interface.

§ 44 A major commitment during the middle half of 2023 was to lay the foundations of a common ground of materials, in order to make them available for all the projects interested in editing Byzantine seals with SigiDoc: these are the basis on which the upcoming search portal will be built. This has involved the structuration of Git repositories, in order for them to be centralized when they contain shared material, but also expandable in order to host new projects; the overall structure can be seen in Table 2.

1. Central SigiDoc repository		2. Collection repositories	
1a. Unique repository for unchanging code	1b. Forkable code	2a. Forked code	2b. Raw XML data
schema	central EFES-SigiDoc instance	individual EFES-SigiDoc instances	XML files containing the edition of the seals
full template	authority files	authority files	
EpiDoc stylesheets			

**Table 2:** The structure of the Git repositories.

§ 45 One centralized repository, called “SigiDoc” (a GitHub organization; SigiDoc 2024b) contains materials designed to be used in two different ways. On one hand, there is the foundation code (the schema, the template, and the EpiDoc stylesheets),

which should not be altered by those working on individual SigiDoc-based projects and for which SigiDoc is the unique repository (column 1a in Table 2). Developments and changes to that code will be introduced by the owners of the repository, on their own initiative or based on suggestions and requests coming from individual projects, and they will be immediately available to each SigiDoc project without any delay and without any action required from the users.

§ 46 On the other hand, other parts of the code (column 1b in Table 2) are still centrally managed, but, for an effective workflow, individual projects should make this material their own. This means that SigiDoc makes available a centralized instance of an EFES customized version offering only the EFES software without any data or schemas: this is a “blank” version design-wise, from which several EFES instances, dedicated to the different collections, can fork for altering the design or adding specific features to meet the needs of individual projects. The other code in this category consists of the authority files and, among the centralized materials, the management of this is more delicate: what the DigiByzSeal team is currently doing is to “pre-prepare” long lists, full of data—taken from the major sigillographic publications and expanded even further—before the encoding of other sigillographic collections. These lists will then be available in the central SigiDoc repository, and can be forked together with the generic EFES-SigiDoc version by individual projects, which will then use their own forks of these materials (column 2a in Table 2).

§ 47 In order to create their own individual sigillographic edition projects, the members of the DigiByzSeal team created two GitHub organizations: the members based in Germany called theirs “byzantinistik-koeln” (Byzantinistik–Universität zu Köln 2024) while their colleagues based in France established “sceaux-byzantins-paris” (Sceaux-Byzantins-Paris 2024). Each of these sub-teams forked the central authority lists folder, and each one forked its own EFES-SigiDoc instances, one for each of the collections currently being worked on (for example, “EFES-SigiDoc-feind” or “EFES-SigiDoc-ifeb”).

§ 48 The crucial question is that of coordination. When a project team needs to change the authority files, they should first change their own authority files and then make pull requests to the central repository. At the same time, each project should make sure that they regularly update their files, that is, they pull from the central authority files repository, which potentially will have received new data from other projects. By using forked individual EFES-SigiDoc instances, each project will benefit from all the changes and improvements made to both the centralized EFES-SigiDoc instance and the “original” EFES instance for EpiDoc (EFES 2024), from which the SigiDoc version has in turn been forked.

§ 49 The last part of this structure is represented by the seals themselves; more precisely by the XML edition files containing all the sigillographic data to be published (column 2b in Table 2). Thus, for example, the previously mentioned GitHub organizations that contain the local authority lists and forked EFES-SigiDoc instances—“byzantinistik-koeln” and “sceaux-byzantins-paris”—also contain repositories bearing the XML files (such as “feind-collection” and “ifeb-collection”), which are the clearest expression of the work of an individual project. Unlike the code in column 1a, of course, the data in these XML file repositories can be independently modified by the individual project managers. Even though they show up as separate repositories on GitHub, each folder containing XML files has been integrated with the corresponding EFES-SigiDoc instance: for example, the “ifeb-collection” folder has the path EFES-SigiDoc-ifeb/webapps/ROOT/content/xml/epidoc.

§ 50 Having a structured working environment is of paramount importance for a project one of whose aims is to involve as many institutions, research centres, museums, and researchers as possible. As has been suggested, the SigiDoc and DigiByzSeal approach is not “Napoleonic”: in giving our colleagues tools and knowledge to independently develop their own projects, we also receive invaluable feedback, essential to further develop this knowledge, in a common effort to create a digital sigillographic community. Many researchers have shown interest in managing their own sigillographic material with SigiDoc, in collaboration with DigiByzSeal, as the map in Figure 7 shows.



**Figure 7:** SigiDoc in the world (Google My Maps). 🏠 = DigiByzSeal members (Institut d’Histoire et Civilisation de Byzance, Paris, France; Department of Byzantine and Modern Greek Philology and Cologne Center for eHumanities, University of Cologne, Germany); ◆ = Active collaborators (Musée d’Art et d’Histoire, Geneva, Switzerland; Regional History Museum, Shumen, Bulgaria; National Archaeological Institute with Museum, Sofia, Bulgaria; University of Bologna, Italy); ● = Interested partners (Numismatic Museum, Athens, Greece; Carthage National Museum, Tunisia; Dumbarton Oaks, Washington, DC, USA).



§ 51 Keeping in mind that this infrastructure has not yet been implemented, we can give some indications of how the portal will work since, like the individual project pages, the federated search portal will also be based on EFES. Individual sigillographic projects that want to integrate their own material into the portal for a federated search must make their data available on a public Git repository and set up a webhook. Webhooks are functions that can be set up in the common Git systems and that are triggered when something is committed to a Git repository (i.e., when a new XML file has recently been added). This allows a newly created endpoint, living on the federated search portal server, to run a specific function, that is, to pull the last commits from the repository. Then, another endpoint, which already exists in EFES, will be called, and a function for the indexing of the data in EFES will run. This latter action will take place on the server hosting the EFES instance dedicated to the federated search portal, and thanks to it, any new XML editions of seals will be indexed and made available for search. In this way, all the data will be directly integrated into the portal as and when a seal is edited.

§ 52 For example, if someone committed the seal “feind\_123.xml” into the “feind-collection” GitHub repository mentioned above, the webhook in this repository would make a request to the URL configured for the webhook, in our case the EFES federated search portal: this URL might perhaps be `DigiByzSeals.search.uni-koeln.de/EFES/newpull/feind` (note that this is a potential URL, not a real one, since the project team still has to decide where the federated search portal will be located). This would pull the new commit coming from the “feind-collection” repository onto the server hosting the EFES federated search portal. Once this action was done, the new seal would still need to be indexed: in order to do that, the URL defined in the EFES endpoint for indexing would run (this URL might perhaps be `.../admin/solr/index/tei/feind-collection/feind_123.xml`, for example).

§ 53 While the DigiByzSeal project does make extensive use of EFES, it should be remembered that the XML edition files prepared through the project can also be used outside this platform. This has a consequence for the centralized portal: projects can use EFES or some other software to present their collections, or they can simply use raw XML files. This choice has no impact on the effectiveness of the search performed by the portal, because the federated search operates on the basis of a list of online Git repositories (GitHub, GitLab, or something else), which it harvests. In the search results, a link is provided to the collection website (or to the source file, if the files are not displayed on such a website).

§ 54 With these upgrades to SigiDoc, we hope to put an end to the current dispersion of sigillographic editions and lay the foundations for a virtual *Corpus Sigillorum Ævi Byzantini*.

---



## Acknowledgments

The research presented in this paper has been supported by the Agence Nationale de la Recherche (ANR, National Agency for Research), within the framework of the French-German project “DigiByzSeal—Unlocking the Hidden Value of Seals: New Methodologies for Historical Research in Byzantine Studies” (<https://anr.fr/Project-ANR-21-FRAL-0008>) jointly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). We would like to thank Timothy Jowan Curnow for proofreading our paper and the two reviewers for their valuable suggestions.

## Competing interests

The authors have no competing interests to declare.

## Contributions

### Authorial

Authorship in the byline is given in alphabetical order after the corresponding author. Author contributions, described using the NISO (National Information Standards Organization) Credit taxonomy, are as follows:

Author names and initials:

Alessio Sopracasa (AS)  
Jan Bigalke (JB)  
Numa Buchs (NB)  
Sima Meziridou (SM)

Authors are listed in descending order by significance of contribution. The corresponding author is AS.

Conceptualization: AS  
Data Curation: AS, JB  
Funding Acquisition: AS  
Investigation: AS, NB, SM  
Methodology: AS, JB  
Project Administration: AS  
Resources: JB  
Software: JB  
Supervision: AS  
Validation: AS, JB  
Visualization: AS  
Writing – Original Draft: AS  
Writing – Review & Editing: AS, JB

### Editorial

#### Special Collection Editors

Martina Filosa, Universität zu Köln, Germany  
Claes Neuefeind, Universität zu Köln, Germany  
Claudia Sode, Universität zu Köln, Germany

### Recommending Referees

Antonello Vilella, Università degli Studi G. D'Annunzio di Chieti-Pescara, Italy  
Dimitar Illiev, Sofia University St. Kliment Ohridski, Bulgaria

### Section Editor

Morgan Pearce, The Journal Incubator, University of Lethbridge, Canada

### Chief Copy Editor

Christa Avram, The Journal Incubator, University of Lethbridge, Canada

### Layout Editor

A K M Iftekhar Khalid, The Journal Incubator, University of Lethbridge, Canada

## References

- Baudin, Arnaud, Clément Blanc-Riehl, Jean-Christophe Blanchard, Laurent Hablot, Philippe Jacquet, and Ambre Vilain. 2024. "The Sigilla Project." *Sigilla*. Accessed June 3. <http://www.sigilla.org/>.
- Bodard, Gabriel, Tom Elliott, Marion Lamé, Pietro Liuzzo, Emmanuelle Morlock, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, Scott Vanderbilt, Polina Yordanova, Scott DiGiulio, Hugh Cayless, Irene Vagionakis, and Martina Filosa. 2023. "EpiDoc Schema (Version 9.5)." *EpiDoc*. Accessed June 3, 2024. <https://epidoc.stoa.org/schema/9.5/tei-epidoc.xml>.
- Byzantinistik-Universität zu Köln. 2024. "Overview." GitHub. Accessed June 3. <https://github.com/byzantinistik-koeln>.
- Cheyne, Jean-Claude. 2019. *Les sceaux byzantins de la collection Yavuz Tatış*. İzmir: Metro Matbaacılık.
- Cheyne, Jean-Claude, Cécile Morrisson, and Werner Seibt. 1991. *Les sceaux byzantins de la collection Henri Seyrig*. Paris: Bibliothèque nationale de France.
- Dumbarton Oaks Athena Ruby. 2024. "Athena Ruby Inscription Font." *Dumbarton Oaks Research Library and Collection*. Accessed June 3. <https://www.doaks.org/resources/athena-ruby>.
- Dumbarton Oaks Seals Bibliography. 2024. "Dumbarton Oaks Byzantine Seals Catalogue, Bibliography." *Zotero*. Accessed June 3. [https://www.zotero.org/groups/4423194/dumbarton\\_oaks\\_byzantine\\_seals\\_catalogue/library](https://www.zotero.org/groups/4423194/dumbarton_oaks_byzantine_seals_catalogue/library).
- EFES (EpiDoc Front End Services). 2024. "EpiDoc Front End Services." GitHub. Accessed June 3. <https://github.com/EpiDoc/EFES>.
- Elliott, Tom, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, Scott Vanderbilt, et al. 2024. "EpiDoc Guidelines: Ancient Documents in TEI XML (Version 9)." *EpiDoc*. Accessed May 28. <https://epidoc.stoa.org/gl/latest/>.
- Jeffreys, Michael, et al. 2017a. *Prosopography of the Byzantine World, 2016*. King's College London. Accessed June 14, 2024. <http://pbw2016.kdl.kcl.ac.uk>.

———. 2017b. “Basileios Mauros.” *Prosopography of the Byzantine World*, 2016. King’s College London. Accessed June 14, 2024. <https://pbw2016.kdl.kcl.ac.uk/boulloterion/2534/>.

———. 2017c. “PBW Seal Editions.” *Prosopography of the Byzantine World* 2016. King’s College London. Accessed June 14, 2024. <https://pbw2016.kdl.kcl.ac.uk/ref/seal-editions/>.

Neuefeind, Claes, Jan Bigalke, Maria Teresa Catalano, Sviatoslav Drach, Sofia Efthymoglou, Pia Evening, Martina Filosa, Christos Malatras, Marcel Schaeben, and Claudia Sode. 2024. “Signed, Sealed, Delivered – Digital Approaches to Byzantine Sigillography.” *it – Information Technology*. Accessed June 17. <https://doi.org/10.1515/itit-2023-0030>.

Sceaux-Byzantins-Paris. 2024. “Overview.” GitHub. Accessed June 3. <https://github.com/sceaux-byzantins-paris>.

SigiDoc. 2024a. “SigiDoc Indices.” *SigiDoc* (Version 1.0). Accessed June 3. <https://sigidoc.raketadesign.com/en/indices/epidoc/>.

———. 2024b. “Overview.” GitHub. Accessed June 3. <https://github.com/SigiDoc>.

SigiDoc Editor. 2024. “SigiDoc Editor.” Cologne Center for eHumanities (Universität zu Köln). Accessed June 3. <https://digibyzseal.cceh.uni-koeln.de/>.

Sopracasa, Alessio, Martina Filosa, et al. 2023. “2. List of All Transcription Guidelines: Leidenisation.” *SigiDoc Guidelines: Byzantine Seals in TEI-XML* (Version 1.1). Accessed June 14, 2024. [http://sigidoc.huma-num.fr/Guidelines\\_XML/xml-html/SigiDoc\\_alltext.html](http://sigidoc.huma-num.fr/Guidelines_XML/xml-html/SigiDoc_alltext.html).

Sopracasa, Alessio, Martina Filosa, and Simona Stoyanova. 2020. “The Digital Enhancement of a Discipline: Byzantine Sigillography and Digital Humanities.” *magazén: International Journal for Digital and Public Humanities* 1(1): 101–128. Accessed May 28, 2024. <https://doi.org/10.30687/mag//2020/01/006>.

Sopracasa, Alessio, and Vivien Prigent. 2017. “Sceaux byzantins de la collection Sopracasa.” In *Oὐ δῶρὸν εἶμι τὰς γραφὰς βλέπων νόει: Mélanges Jean-Claude Cheynet*, edited by Béatrice Caseau, Vivien Prigent, and Alessio Sopracasa, 691–758. Paris: ACHCByz.

Stoa Consortium, eds. 2024. “What Is *Pleiades*?” *Pleiades*. Accessed June 3. <https://pleiades.stoa.org/>.

TEI Consortium, eds. 2024a. “Getting Started with P5 ODDs.” *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Accessed June 3. <https://tei-c.org/guidelines/customization/getting-started-with-p5-odds/>.

———. 2024b. “Customization.” *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Accessed June 3. <https://tei-c.org/guidelines/customization/>.

