Epigraphy in 2017

Hugh Cayless <hcayless_at_email_dot_unc_dot_edu>, University of North Carolina
Charlotte Roueché <charlotte_dot_roueche_at_kcl_dot_ac_dot_uk>, King's College London
Tom Elliott <tom_dot_elliott_at_nyu_dot_edu>, New York University
Gabriel Bodard <gabriel_dot_bodard_at_kcl_dot_ac_dot_uk>, King's College London

Abstract

Epigraphy as a discipline has evolved greatly over the centuries. Many epigraphists in the last 20 years have been investigating how to use digital technology to advance their research, but until the present decade, these undertakings were restrained by aspects of the technology. The late 1990s will be seen as a watershed moment in the transition from print-based to born-digital epigraphic publication. At present, the majority of new editions are still published solely in print, but by 2017 we believe this circumstance will change drastically. The history of epigraphy makes it quite clear that such transitions are natural to the discipline.

1. Background

Epigraphy is an area of study of the Graeco-Roman past which has been through centuries of change. Scholars in 8th century Constantinople tried to use inscriptions to make sense of their history:

> We have written this after reading it from inscriptions on marble tablets or after making enquiries of those who have read it. [Cameron and Herrin 1984, 24]

> To such viewers, in common with Pausanias in the second century, the main function of inscriptions was as labels, to explain monuments and statues, although the medieval viewers came to believe that there were also occult messages, to be deciphered by the learned. [Dagron 1984]

[1]

This basic tradition of reading an inscription as a label has never really been eliminated. But the awakening of scholarship in the Renaissance led western scholars — most strikingly, Cyriacus of Ancona — to start recording inscribed texts more fully. In western Europe, the relevance of Latin inscribed texts to local history was evident, and the style of such texts came increasingly to influence the funerary epigraphy of the period. The development of a standard education for gentlemen in both Latin and Greek meant that, as travel in the eastern Mediterranean increased in the 18th and 19th century, inscriptions started to be transcribed by visitors who were not necessarily classical scholars: William Sherard, who transcribed many important texts while he was British Consul at Smyrna in the early 1700s was a botanist who later endowed a Chair of Botany at Oxford (http://www.oxforddnb.com/view/article/25355?docPos=2&_fromAuth=1).

[2]

These travelers were working in difficult circumstances: Charles Fellows reports that there was an outbreak of plague in the area when he visited Aphrodisias; they even had to be sparing of paper (Wood copied some texts into the margins of the volume of Homer that he was carrying). It is astonishing that their copies were so often excellent. They tended to copy complete texts — rather than fragments — and did not spend much time recording details beyond the text itself. For examples, see the notebooks of John Gandy-Deering, from a visit to Aphrodisias in 1812: http://insaph.kcl.ac.uk/notebooks/deering/index.html.

[3]

Only gradually did a more "scientific" approach develop, as scholars began to travel with the specific purpose of

recording inscriptions: the Austrian Academy had special notebooks printed for recording such material, pre-printed with the headings for metadata: place, material, dimensions, etc. (For an example see http://insaph.kcl.ac.uk/ala2004/inscription/eAla016.html). Traveling scholars could make paper impressions of stones (squeezes), but photography was only possible at a large and well-established excavation, until the 1920s.

The publishing of inscriptions had similarly evolved, in a random and haphazard way; it was only in the 1930s that publishing protocols were agreed (the so-called Leiden conventions, [Van Groningen 1932]). Discussions and refinements have continued ever since. There is therefore a long tradition of change and development in the recording and publication of inscriptions. Principal developments have been an increasing emphasis on recording the support, and the archaeological context, of a text, and increasing discipline in presentation. This focus has tended to emphasize the role of epigraphy within archaeology and history, and to distance it from the study of text and language.

New approaches to the recording of inscriptions must, firstly, record the fullest possible account — any stone may disappear or be damaged. Such a record, nowadays, requires a very full photographic record. Secondly, that account must be published in a fashion which is sufficiently normative to be comprehensible, and that makes probable its long-term conservation: any publishing medium must aim at the longevity of manuscript. This is especially true for digital publications.

> ...it seems clear that those who create intellectual property in digital form need to be more informed about what is at risk if they ignore longevity issues at the time of creation. This means that scholars should be attending to the information resources crucial to their fields by developing and adopting the document standards vital for their research and teaching, with the advice of preservationists and computer scientists where appropriate. [Smith 2004]

Many epigraphists in the last 20 years have been investigating how to use digital capacities to serve their science; but until about 2000, these undertakings were restrained by aspects of the technology. The late 1990s will be seen as a watershed moment in the transition from print-based to born-digital epigraphic publication. At present, the majority of new editions are still published solely in print, but by 2017 we believe this circumstance will change drastically. It is our aim to ensure that such publication is not just driven by considerations of economy or space, but is developed to meet the academic requirements set out above. The history of epigraphy makes it quite clear that such transitions are natural to the discipline.

In trying to imagine where the discipline of epigraphy may go in the next ten years, we will use as a framework John Unsworth's thoughts on "scholarly primitives" [Unsworth 2000]. In some programming languages, data types that are built into the language, rather than being defined using the language, are called primitives. So by "primitives" Unsworth means basic building blocks: research methods common to all humanities research. He lists seven: discovery, annotation, comparing, referring, sampling, illustrating, and representing. The most obvious one of these that is enhanced by digitizing the research sources is discovery, or search. Enabling search across many texts has been a driving force behind digital initiatives in epigraphy. Digital texts may of course be searched with great ease. The prominence of database-driven projects in the outline of digital epigraphy projects below (section 1.2) underscores this motivation. Search is provided by many commercial applications, which means in addition that any project published online as static HTML may also be searched using Google or one of its competitors. We will argue, however, that while the ability to easily discover useful information is enhanced by the standard approaches to digitizing epigraphical documents, the other scholarly primitives are less well served, and that different digital representations may do a better job. That said, the importance of discovery in modern epigraphic projects makes a good jumping-off point for a discussion of the Leiden System as it has been applied to the electronic medium and some of its limitations and difficulties.

1.1 Leiden

The creation of large corpora of inscriptions, such as the *Corpus Inscriptionum Latinarum*, made clear the importance of consistent, universal, agreed standards for epigraphic conventions. The Leiden Convention, an international meeting of scholars in 1931, intended to draw up just such a universal standard, and to a large extent they have been successful.

Leiden specifies how features of an inscription besides the text itself should be represented in print. The system uses certain symbols and/or text decorations to convey the state of the original document and the editor's interpretation of that document.

A brief description of the Leiden conventions will be helpful here to the reader who is unfamiliar with them. Ancient inscriptions are obviously subject to all kinds of damage and wear, and therefore it is quite common for pieces of any given inscription to have broken off, or for sections of writing to have become hard to read. In addition, there are a number of features of the inscription itself that may bear notice by an editor publishing it: the text may contain abbreviations that warrant expansion, or a name may have been erased in antiquity and perhaps a new name inscribed on top of the erasure. In addition, there may be aspects of the editorial history of the inscription that are worthy of note. For example, a previous editor may have recorded letters that are no longer readable. These and other features of inscriptions (and papyri) are noted in the published versions using inline marks.

For example, missing text will be marked off with square brackets: "[...]". If the missing text can be supplemented, it will go inside the brackets. Letters that are worn or damaged and cannot be read with 100% accuracy will be marked with a dot underneath each letter. The expansion of abbreviations is marked with parentheses, such as "co(n)s(ul)" for "COS". Text deleted in antiquity will be marked with double square brackets, "[[Getae Caesari]]". Words read by a previous editor but unreadable by the present one are underlined, "Imperator <u>Caesar</u> ". As an illustrative example of some of these techniques, the following is an inscription from the city of Aphrodisias marked according to the Leiden conventions, along with an image of the stone. Note the brackets marking lost letters, the dots underneath damaged letters, and the parenthetic expansion of the name Flavius (*Φλάουιος*):

Figure 1.

τὸν εὐσεβέστ[α]-
τον καὶ καλλί -
νικον ἡμῶν
[δεσπό]την Φλ(άουιον)
5 [Ἰουστινια]νόν
[...]

[(e.g.) The city honours] our most pious and splendidly victorious master, Flavius
[?Justinia]nus. ala2004 (81)

1.2 Digital Epigraphy Projects

There are a great many extant digital epigraphy projects, some of which have been around now for decades. Notable among the early projects are the Packard Humanities Institute's Greek Epigraphy project and the Epigraphische

Datenbank Heidelberg (EDH). The PHI project has been publishing digitized inscriptions since the 1980s, initially on CD-ROM and more recently on the web. The Cornell Greek Epigraphy Project was started, on David Packard's initiative, in 1985. The first "Demonstration" CD (#1) was released in 1987; the last ("Greek Documentary Texts", #7) in 1996; the PHI epigraphical website was officially released in 2006. The EDH began in 1986 and the Epigraphic text database went online on the web in 1997. EDH consists of three databases, the first containing texts, the second bibliography, and the third over 20,000 images. EDH was an inspiration for and is one of the three original databases to be federated into EAGLE, the Electronic Archive of Greek and Latin Epigraphy, of which the other two constituents are the Epigraphic Database Bari (EDB), and the Epigraphic Database Roma (EDR). These are by no means the only digital epigraphy projects, but they do represent standard approaches to the problem of treating inscriptions and related data in digital form.

A more recent development comes in the form of competing initiatives to digitize and make available online the books in a growing number of libraries. Dessau's *Inscriptiones Latinae Selectae*, for example, is available (for U.S. readers only) on the Google Books site. Portions of the *Corpus Inscriptionum Latinarum* and other corpora have also been digitized, but are not made available in full, despite their being out of copyright. A competitor to the Google Books project, the Open Content Alliance, does not at last check have any of these works yet, but the OCA, unlike Google, is committed to making available out-of-copyright works, and therefore it may only be a matter of time before the great 19th-century inscriptional corpora are available and searchable on the World Wide Web. The utility of these works will be somewhat limited at first, since what is searchable is the raw output of optical character recognition, but these are significant developments nevertheless.

[13]

1.3 Epidoc

The majority of the projects mentioned so far take a straightforward approach to digitizing inscriptions: the Leiden-encoded inscription is transferred directly to digital form (with a few adjustments). A different approach, and one upon which we will focus in this paper is the EpiDoc initiative. EpiDoc was started in the late 1990s by Tom Elliott, then a graduate student in Ancient History at the University of North Carolina at Chapel Hill (U.S.A.). Elliott made public his initial work on epigraphic encoding in XML in response to the promulgation, by Prof. Silvio Panciera and colleagues, of a manifesto recommending the establishment of an online, free and unrestricted "database...of all surviving Greek and Latin epigraphical texts produced down to the end of Antiquity."

[14]

The manifesto itself had emerged from a round-table meeting on the subject of "Epigraphy and Information Technology" in Rome, convened by Prof. Panciera in May 1999 in his capacity as President of the "Commission for Epigraphy and Information Technology" of the Association Internationale d'Épigraphie Grecque et Latine. A second round-table meeting, held at Aquileia and Trieste in 2003, refined the initiative as a federation of epigraphic databases, EAGLE.

[15]

Among the salutary recommendations embodied in the report of the Rome meeting was a provision recognizing the importance of a platform-independent format suitable for backup, archiving and data interchange (i.e., XML):

[16]

> *È importante che siano utilizzati programmi che consentano l'esportazione dei documenti in "Document Type Definition (DTD) format".*

> It is important that the software used for the database support the export of data in a structured format definable by a Document Type Definition (DTD).

It was this provision that encouraged Elliott to post to the web, in the summer of 1999, a description of his work on XML for epigraphy. He hoped that this step might facilitate the realization of the Rome meeting's recommendations. Even at this early stage, EpiDoc was already a collaborative undertaking, having benefited from collegial help in Chapel Hill from Amy Hawkins, George Houston, Hugh Cayless, Kathryn McDonnell and Noel Fiser. The collaborators were seeking a digital encoding method that preserved the time-tested combination of flexibility and rigor in editorial expression to which classical epigraphers were accustomed in print, while bringing to both the creator and the reader of epigraphic editions the power and reusability of XML. By summer 1999, this search had led to the adoption of the Text Encoding Initiative as a foundation for specific epigraphic work. Although the TEI Guidelines did not specifically address epigraphic

materials — indeed, if they had, there would have been no need for the development of EpiDoc — many of their provisions for text criticism and transcription were readily adaptable to the needs of epigraphists.

The first draft of a set of guidelines for the application of TEI to epigraphic texts (a.k.a. The EpiDoc Guidelines) was promulgated in January 2001 with the assistance of Ross Scaife and Anne Mahoney (of the Stoa Consortium), and of John Bodel (then at Rutgers, now in the Classics Department at Brown University) and Charles Crowther (at Oxford's Centre for the Study of Ancient Documents). Crowther was a member of the AIEGL IT Commission and was planning with Alan Bowman and John Pearce the on-line publication of the Vindolanda Writing Tablets. Bodel's collaboration with Stephen Tracy on Greek and Latin Inscriptions in the USA. A Checklist (Rome 1997) was soon to spawn the U.S. Epigraphy Project. Both Bodel and Crowther had participated in the Rome meeting. The promulgation of an incomplete, draft set of guidelines drew the attention of a third epigraphist, Charlotte Roueché (in the Department of Byzantine and Modern Greek at King's College, London), who was exploring new technologies to apply to the publication of inscriptions from Aphrodisias. [17]

Bodel, Cayless, Crowther, Elliott and Roueché have remained heavily involved in subsequent developments. In particular, Bodel and Crowther both wrote in enthusiastic support of Roueché's April 2001 proposal for collaboration support and prototype funding from the Leverhulme Trust. This successful application, under the rubric of the EpiDoc-Aphrodisias Pilot Project, accelerated revision of the Guidelines and other EpiDoc resources through a series of workshops, later continued with funding from the Arts and Humanities Research Council. Through these projects and workshops, EpiDoc has grown and matured. Its scope has expanded beyond (though not abandoned) the original vision for a common interchange format. EpiDoc now aims also to be a mechanism for the creation of complete digital epigraphic editions and corpora. We will argue that EpiDoc represents a better digital abstraction of the Leiden conventions than is achievable by a simple mapping Leiden's syntax for printed publication into digital form. [18]

A full EpiDoc document may contain, in addition to the text itself, information about the history of the inscription, a description of the text and its support, commentary, findspot and current locations, links to photographs, translations, etc. For purposes of comparison with the example text above, below we see how it could be represented in EpiDoc: [19]

```
<lb/><supplied reason="lost">τ</supplied>ὸν
        εὐσεβέσ<unclear reason="damage">τ</unclear>
        <supplied reason="lost">α</supplied>
<lb type="worddiv"/>τον καὶ καλλί
<lb type="worddiv"/>νικον ἡμῶν
<lb/><supplied reason="lost">δεσπό</supplied><unclear reason="damage"
>τ</unclear>ην
        <expan>Φλ<supplied reason="abbreviation">άουιον</supplied></expan>
<lb/><gap reason="lost" extent="3" unit="characters" precision="circa"/>
        <supplied reason="lost" cert="low">Ἰουστινια</supplied>νόν
<lb/><gap reason="lost" unit="line" extent="unknown" dim="bottom" />
```

The first point the reader will notice is how much more verbose the XML version is. Compared to the Leiden version, the EpiDoc text contains many more characters. This version may easily be displayed in a Leiden form, however, if an XSLT stylesheet is used to transform it. A diplomatic version, where all editorial supplements have been removed would be equally easy to render from the same source, using the same method. Moreover, the EpiDoc project has developed tools that will handle the conversion of the Leiden-formatted version of this inscription (shown above) into the XML version automatically. The XML version therefore should not be viewed as a replacement for Leiden, which is easier for scholars to produce and to read, but as an interchange format to be used when Leiden needs to be read, manipulated, or transmitted by a computer. We will look in depth at the reasons why such a format is necessary in the next section.

2. Digital Leiden

Leiden is strictly a typographic standard. It specifies how an edited inscription should look in print. The "data dictionary" for Leiden consists of a number of published articles and the substantial body of practice. So there are places you can go to find out what a particular epigraphic siglum means. This typographic practice doesn't always carry over well into the digital realm. In Leiden, for example, a piece of underlined text means that a previous editor read it, but the current editor cannot see it. If one is storing information in a database, and is therefore restricted to undecorated text, underlining becomes a real problem. To convey this information when text decoration is inoperative, a workaround must be devised that modifies the Leiden conventions, for example, adding an underscore character "_" immediately before and after the text to be underlined.

Before proceeding with a deeper analysis of the distinctions between the Leiden System and EpiDoc, we must digress a little and explore the meaning of "plain" versus "decorated" text in the paragraph above. In print, there is no ambiguity between underlined and non-underlined text, nor between italicized and normal text. The former has below it an underscore and the latter is visibly different from a normal character. Their meaning is thus indicated by physical differences in the characters. In the digital realm, however, things are less clear. In a word processing format, for example, underlining is typically signaled by an instruction to the program displaying the text to begin underlining immediately before and to cease immediately after the string of text so marked. In RTF, for example,

```
\uUnderlined\unone
```

would produce <u>Underlined</u>. In HTML, there is more than one way to do this, but the underlying mechanisms are the same: the text to be underlined will be enclosed by a tag (perhaps `<span>`) which will have a style applied to it indicating that a browser displaying the text should underline it, so to produce the text Underlined we might see this:

```
<span style="text-decoration:underline">Underlined</span>.
```

As we noted, however, there are many ways to produce this effect, and more commonly in HTML one would use a Cascading Style Sheet, either contained within the HTML document or linked to it. In neither case is the underlining any longer a characteristic of the text itself, but of the surrounding markup and of the capabilities of the rendering system that will be applied to it. Therefore in this case the digital data is at a second remove from the meaning of the underlining: it depends upon formatting codes that surround the text read by a previous editor, but are not attached to the actual characters, and upon the ability of a rendering system correctly to interpret those codes. This separation can lead to data loss when the information is transmitted. For example, the stylesheet that creates the underlining in HTML might not travel along with the document. Or the text could be transmitted over email and the reader on the other end might not choose to display HTML email. The potential for this sort of information loss is not the most significant problem with these features of Leiden translated into the digital realm, however.

Some other features of Leiden are easier to encode. Expanded abbreviations, for example, are indicated by parentheses thus: D(is) M(anibus) S(acris). The digital representation of these parentheses corresponds directly to the typographic representation, character for character. This difference in the manner of digital representation (between previously seen text and expanded abbreviations in this case) highlights a central problem with mapping Leiden directly into an electronic form: such a mapping of necessity involves either a hybridized format, in which meaning is sometimes contained within characters and sometimes within formatting codes, or a conversion of the typographic forms in Leiden which would normally be represented as formatting codes to new characters. The latter approach is the typically the only one feasible in environments like a text column in a database.

Even with such a feature-to-character mapping, however, some problems remain in the rendering of some characters. Unclear letters, for example, are represented by a dot under the letter in Leiden. Combining sublinear dots are available in Unicode, but they require both a font that contains that particular code point and a rendering system smart enough to properly place the dot underneath the preceding character. Display of these features remains very inconsistent. A similar problem occurs with the typographic representation of letters incorrectly substituted or executed in the text, which are corrected by the editor. Panciera uses a pair of half square brackets to enclose such letters [Panciera 1991]. Unfortunately, while the left (opening) bracket is well represented in most popular Unicode fonts, the closing bracket is

absent from all but specialized fonts designed for use by Classicists. The digital encoding of these features is quite possible, therefore, but their display, at the present time, is often another matter.

To summarize: directly mapping Leiden print syntax to digital form involves a number of problems. Some conventions are represented differently from others, some may be inconsistently rendered when printed to the screen, some are not widely available. Moreover, if we examine the case of expanded abbreviations more closely, we will see that their treatment in print, which is perfectly clear and unambiguous, does not work well in a digital environment. In our example, D(is) M(anibus) S(acrum), the expansion of each word is enclosed by a pair of symbols, parentheses in this case. But what might we want to do with a digital abbreviation? One imagines, for example, being able to generate a list of abbreviations used in a corpus, or searching for instances of a word only when it is abbreviated. The parentheses are deficient in this case because they do not provide a simple way to pick up the entire abbreviated word. Only the expansions are marked. The problem becomes more obvious if we consider words where the abbreviated word contains characters not in the expanded word, e.g. AUGG for Augusti. In such cases we might wish to be able to get the abbreviation or the full term (and we will likely want both for the purposes of searching). Leiden in digital form makes this kind of operation more difficult than is perhaps necessary.

To return to our discussion of the scholarly primitive discovery, we must examine how this operation works. Searching across text involves pattern matching: an input query, such as "liberti" (freedmen), is matched against the documents in a digital corpus. This pattern matching may be subject to varying degrees of indirection, but the net result is that each document containing the string "liberti" is returned. Making this work when there is a system like Leiden imposed upon the text usually means ignoring the marks signifying editorial expansions or observations. "Liberti", for example, is commonly abbreviated, so the inscription might actually read "LIB", which the editor has expanded as lib(erti). In order for a word search to match on such abbreviated terms, the parentheses and other possible sigla would need to be ignored. This is in fact the behavior exhibited by online corpora such as EDH. It does make for problems, however, when more sophisticated searching is wanted — for example, when someone wants to search for "lib" when it begins an abbreviation. In such a case, just ignoring sigla for the purposes of search indexing doesn't help, but selectively paying attention to them is problematic also: a search for "lib(" might fail to pick up instances where other sigla are in the mix, such as "[l]ib(erti)". These kinds of problem arise because the digital treatment of Leiden relies on encoding it as a simple text stream.

Being able to digitally represent the features of Leiden is only a first step. In order to be able to take full advantage of the capabilities of the digital environment, that is, to make the texts fully queryable and manipulable, a further step is required. By the term "queryable", we do not simply mean that the text may be scanned for particular patterns of characters; we mean that features of the text indicated by Leiden should be able to be investigated also. So, for example, a corpus of inscriptions should be able to be queried for the full list of abbreviations used within it, or for the number of occurrences of a word in its full form, neither abbreviated nor supplemented. One can imagine many uses for a search engine able to do these kinds of queries on texts. An example of manipulation would be to display a diplomatic version of the Leiden text, with editorial supplements and expanded abbreviations removed. These kinds of operations can certainly be performed on digital texts in Leiden form, but some intervening steps are necessary.

In order for searches to leverage the structures imposed on a text by an editor using Leiden, the marked-up text would have to be parsed and converted into data structures amenable to the kinds of operations described above. The first step of such a parsing operation is to perform lexical analysis on the text, which is transformed into a stream of tokens to be fed into a parser, which in turn produces a hierarchical data structure (a parse tree) which can be acted upon and queried in various ways. The production of a diplomatic text, for example, would simply involve dropping those "branches" of the tree marked as abbreviations and supplements and serializing the data structure back to text. The types of query envisioned in the previous paragraph could be facilitated by the production of indexes using the parsed data. Here, however, we must note that this type of structure closely resembles that embodied by EpiDoc. XML is a hierarchical structure, easily parsed, with sophisticated tools for querying, linking, and transformation and with a well defined and standardized object model. In effect, EpiDoc represents a possible serialization of the Leiden data structure described above, with the additional benefit of having lots of tools available for working with it. Moreover, since the semantics of the XML document are clear, both from the elements themselves and from the formal guidelines for the

syntax upon which EpiDoc is based (the TEI), there is little opportunity for confusion over the meaning of this mapping from Leiden to digital form (expanded abbreviations are enclosed within `<expan>` tags, editorial supplements within `<supplied>`, etc.).

3. Epigraphical Databases and Digital Publication

Inscriptions have traditionally been published in the print world in three ways: as individual articles, in books, and in corpora which collect large numbers of inscriptions (usually focusing on a particular geographic area or type of inscription). These corpora have long been a major research tool for scholars using inscriptional evidence. In the digital realm, the clear inheritors of the mantle of the print corpora are the online epigraphic databases. There are a number of advantages to dealing with a corpus of inscriptions as a database, including obviously the query facilities that come with the use of a Relational Database Management System. We have already mentioned some of the well-established and growing database-based epigraphic corpora, including the Epigraphische Datenbank Heidelberg and the Epigraphic Database Roma (the PHI disks — and presumably the online version also — use a different approach). The typical format of these databases includes fields for metadata, such as find location, date, etc., and a large text field containing the text of the inscription in Leiden form. This kind of setup permits various kinds of fielded and full-text searches and is very easy to connect to a web-based front end with forms for generating the queries. The data is easy to extract using Structured Query Language (SQL), which is a broadly supported standard (although there are some differences in "dialects" used by different systems). Data is also easily added to these systems, so new inscriptions can be inserted as they become available. There are a number of very solid and well-established Open Source RDBMSs, as well as many proprietary ones, and interfaces to them from the kinds of programming languages used to run web sites are easily available, so there is a low barrier to entry for their creation.

The database-driven website as a venue for web publication is clearly an important and powerful tool, but there are some important features of print publication that are often thereby lost. Most important, from the perspective of digital preservation, is that copies of the corpus are not distributed the way that copies of print corpora are. "Web publishing" in this form results in a single centralized copy and only temporary and partial copies are distributed. Another important feature missing from many sites of this type is the ability to customize queries and to see how result sets are being constructed. The website developers will have made many decisions, such as how to turn information entered into a web form into a query and how to display the results of that query which while they are necessary in order to construct a functioning and usable interface, also obscure from the searchers how their requests on the system are being dealt with. There is a real risk that the form and capabilities of the interface will shape the kinds of scholarship that can be done using it. In other words, these technical decisions are also editorial and scholarly decisions (whether their nature is clear to those making the decisions or not). Giving access to the raw data provides others with the means to check on these decisions and correct them when necessary.

In the future, we hope that the ability to export parts of or all of the data in such systems into standard formats will be recognized as a basic requirement for any scholarly digital corpus of inscriptions. This requirement implies a need for persistent identifiers, such as DOIs, at the level of the digital object, which can be used to cite items retrieved from a digital corpus independent of the actual location from which the object was retrieved. In the kind of system we envision, in which published digital inscriptions may be stored in many locations (for example published on a website, retrieved by an aggregator, harvested by a local repository, then discovered and used by a scholar) the ability to cite accurately will be vital.

The EpiDoc project has spent considerable time and effort on tools to make EpiDoc documents interoperable with databases like EDH and EDR. In this scenario, EpiDoc would function as an interchange format to which part or all of a database could be dumped and saved or distributed and from which data could be extracted into such a database. A system like this is a great deal more "future proof" than reliance on a centralized website because it would allow the distribution of copies and subsets of the data, in a format that is able to be queried and manipulated digitally. In the event of a catastrophe, the central database could be reconstituted from a collection of these documents. This sort of distribution system would not remove the need for a central, online-accessible version of the data, which would be vital as a registry of published inscriptions, but it would allow scholars the freedom to search, analyze, re-edit, annotate, and

re-publish inscriptions in the digital environment.

4. The Scholar and Digital Texts

All of the arguments we have been making hinge in some sense on one central point: inscriptions are texts, situated in complex environments. This fact argues for treating them from the start as complex digital packages with their own deep structure, history, and associated data (such as images), rather than as simple elements in a standardized collection of data. Rather than engineering applications on top of a data structure that does not correspond well to the nature of the source material, we would do better to construct ways of closely representing the physical and intellectual aspects of the source in digital form, and then find ways to build upon that foundation. The former approach is a choice based on the (powerful) functionality offered by a particular set of tools, but XML and the technologies that have emerged around it offer the opportunity to choose a format that better matches the intent of the standard print publication and research methods for inscriptions and at the same time benefit from the same advantages as the older toolset.

<div style="float:right">32</div>

A common pattern in scholarly research is to start by compiling a dataset (this may be as simple as a collection of references on index cards), then analysing that dataset for patterns, forming hypotheses about the nature of those patterns, and then using the evidence gathered to make a formal argument for the truth of the hypotheses. Unsworth's scholarly primitives form the building blocks of this kind of process, and it seems unlikely that the fundamental nature of scholarly inquiry and output will change significantly over the next ten years. What we hope to see happen during the next ten years is that the tools for doing scholarship in the digital environment will become better fitted to the nature of the source material and to the needs of scholars. And that in turn, the types of inquiry that are enabled by these sources will expand. A dataset for scholarly inquiry will consist not simply of notes on flashcards, bookmarks, or even printouts or digital versions of search results, but will be a true dataset, able to be queried, mined, and transformed.

<div style="float:right">33</div>

We may imagine an epigrapher in ten years searching across multiple online repositories of inscriptional evidence to create a new, local, digital corpus. This corpus may then be loaded into tools designed and customized for the research needs at hand, analyzed in a variety of ways, and when the time comes to publish the results, the dataset used can be published digitally alongside the book or article, which may itself be either in print, or digital form, or both. What we hope will change and improve by 2017 are the tools available for gathering information and analyzing it as well as the forms and venues for publishing the results of scholarly inquiry. None of this will be possible unless information is published in such a way that it is not concealed behind an interface, but is in addition retrievable in bulk. EpiDoc is designed to be a format for precisely this type of publishing. We see EpiDoc as a small first step on the road to truly digitally-enabled epigraphy scholarship, in which not only will it be possible more efficiently to answer questions, but in addition it will be possible to ask and answer new types of question, and even to discover new questions to ask.

<div style="float:right">34</div>

Works Cited

Bodard 2009 Bodard, Gabriel. "EpiDoc: Epigraphic documents in XML for publication and interchange". In Francisca Feraudi-Gruénais, ed., *Latin on Stone: Epigraphic Research and Electronic Archives*. Rowan and Littlefield, 2009.

Bos 2007 BosBert, Tantek Çelik, Ian Hickson and Håkon Wium Lie, eds. *Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification*. 2007. http://www.w3.org/TR/2007/CR-CSS21-20070719/.

Bray 2006 Bray, Tim, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler and François Yergeau, eds. *Extensible Markup Language (XML) 1.0*. World Wide Web Consortium, 2006. http://www.w3.org/TR/2006/REC-xml-20060816/.

Cameron and Herrin 1984 Cameron, Averil, and Judith Herrin, eds. *Constantinople in the Early Eighth Century: The Parastaseis Syntomoi Chronikai*. New York: Columbia Studies in the Classical Tradition, 1984.

Clark 1999a Clark, James, and Steve DeRose, eds. *XML Path Language (XPath)*. *W3C Recommendation 16 November 1999*. Word Wide Web Consortium, 1999. http://www.w3.org/TR/1999/REC-xpath-19991116/.

Clark 1999b Clark, James, ed. *XSL Transformations (XSLT), Version 1.0*. World Wide Web Consortium, 1999. http://www.w3.org/tr/1999/rec-xslt-19991116.

Colin 2007 Colin, Matthew, Brian Harrison and Lawrence Goldman, eds. *Oxford Dictionary of National Biography*. Oxford: Oxford University Press, 2007. http://www.oxforddnb.com.

Dagron 1984 Dagron, Gilbert. *Constantinople Imaginaire: Études sur le Recueil des Patria*. Paris: Presses Universitaires de France, 1984.

Elliott 2007a Elliott, Tom, Lisa Anderson, Zaneta Au, Gabriel Bodard, John Bodel, Hugh Cayless, Charles Crowther, Julia Flanders, Ilaria Marchesi, Eli Mylonas and Charlotte Roueché. *EpiDoc: Guidelines for Structured Markup of Epigraphic Texts in TEI*. EpiDoc, 2007. http://www.stoa.org/epidoc/gl/5/.

Panciera 1991 Panciera, Silvio. "Struttura dei supplementi e segni diacritici dieci anni dopo". *Supplementa Italica* 8 (1991).

Reynolds 2007 ReynoldsJoyce, Charlotte Roueché and Gabriel Bodard. *Inscriptions of Aphrodisias*. 2007. http://insaph.kcl.ac.uk/iaph2007/index.html.

Roueché 2004 RouechéCharlotte. *Aphrodisias in Late Antiquity: The Late Roman and Byzantine Inscriptions*. 2004. http://insaph.kcl.ac.uk/ala2004/.

Smith 2004 Smith, Abby. "Preservation". In Susan Schreibman Ray Siemens and John Unsworth, eds., *A Companion to Digital Humanities*. Oxford: Blackwell Publishing, 2004.

Unsworth 2000 Unsworth, J. *Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?* Symposium on Humanities Computing: formal methods, experimental practice sponsored by King's College, London, 13 May 2000. http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html.

Van Groningen 1932 van Groningen, B.A. "Projet d'unification des systèmes de signes critiques". *Chronique d'Égypte* 7 (1932), pp. 262-269.