

This research was funded in part by the UK AHRC, grant number AH/W010682/1. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Open scholarship: epigraphic corpora in the digital age

John Bodel (Brown University), Jonathan Prag (University of Oxford), Charlotte Roueché (King's College London).¹

This paper began life as two sessions entitled ‘the future of corpora’ and ‘digital epigraphy’. For practical reasons, this became two linked papers co-authored by three speakers. In composing those papers it became increasingly clear to us that the two topics are intricately linked: not because we believe that any future corpus must necessarily be digital, but rather because the possibilities created by the rapid advances in digital technologies expose an increasing lack of clarity about the nature and purpose of both the ‘corpus’ and the individual ‘edition’. At the same time, the possibilities and the challenges of digital methods bring to the fore a wide set of issues concerning not only research methods but also research culture (it is no coincidence that open scholarship has emerged as a contemporary issue alongside digital scholarship).

The difficulties encountered by the earliest digital epigraphic projects have been well rehearsed: we do not attempt a history of the development of digital epigraphy, which is well documented.² The *technical* incapacities and incompatibilities that initially hindered full cooperation and sharing of information among even willing partners have been almost entirely overcome, although the scars from that experimental period of rapid development may well persist. Rather, it is our shared view that the largest obstacle still facing us as a discipline today is not technical but cultural: our academic culture has not caught up to advancements in technology that make possible the dissemination and sharing of information with unprecedented ease. Ever-advancing technological capabilities have opened up uncharted territory in the field of epigraphic research. If we are to embrace these developments - because they enable progress in our research, which is our *raison d'être* - we will need to adapt. Retaining to some extent the structure of our original presentations, the first part of this paper considers the interrelated questions of what purpose a corpus serves, what type of publication is best suited to achieving it, and what, increasingly problematised in the rapidly evolving world of digital publication, we

¹ Authors are listed in alphabetical order and share equal responsibility for the text. This work was supported by the UK AHRC grant number AH/W010682/1 (the FAIR Epigraphy project).

² See e.g. Rossi 2021, Elliott 2014, Bodel 2012, the papers in Feraudi-Gruénais 2010, Cayless *et al.* 2009; and for recent work and challenges, the collections of De Santis and Rossi 2018 or Velásquez and Espinosa 2021.

mean by an edition. We attempt to ground our ideas and concerns in a discussion of the evolution of the *Inscriptions of Roman Tripolitania*: throughout our writing of this paper we have had the remarkable figure of the late Joyce Reynolds in our thoughts: she was deeply committed to collaboration in the epigraphic community, hosting her own International Congress in 1967 before welcoming the development of AIEGL to take on the task. It was this commitment that led her to embrace the possibilities brought to epigraphic publication by digital developments (with a healthy scepticism of technology for its own sake). In the second part of the paper we develop some of these ideas, with a particular focus on some key challenges that are highlighted by the application of digital methods: sustainability, the fundamental importance of agreed standards, and the need for open and accessible data.

Epigraphic corpora

When August Boeckh in 1815 presented to the board of the Berlin Academy the proposal he had drafted on behalf of the Philological-Historical Class of that year for the publication of a *Thesaurus Inscriptionum* of all the inscriptions of antiquity “up to the beginning of the medieval period”, beginning with Greek and Latin (a practical constraint we should surely now move beyond), the core elements of how a geographically arranged corpus might be constituted and on what principles editions should be based had been largely worked out.³ Three hundred years earlier Cyriac of Ancona compiled the first large-scale epigraphic corpus by accurately copying and drawing the monumental inscriptions he encountered in his travels in Italy, Sicily, Dalmatia and throughout the eastern Mediterranean. By the end of the fifteenth century the principle that editions should be based on autopsy and, where possible, published with drawings or at least diplomatic transcriptions had been well established, and other basic subsidiary tools, such as indices of names, became common with the advancement of printed editions over the following century.⁴ But an enterprise on the scale that Boeckh and his colleagues proposed required something more: “The principal aim of a Royal Academy of Sciences,” he wrote, “must be to undertake projects and to complete work which no one person could achieve, in part because his abilities alone would not be sufficient, in part because it would require an expense which a private individual would not dare to incur”.⁵ The plan was to begin with Greek inscriptions, since Latin texts were more fully covered in existing

³ Boeckh’s handwritten proposal is reproduced and transcribed in Hallof 2012, 2-7.

⁴ Buonocore 2014 provides a masterful, if cursory, overview of the rich manuscript tradition. The three editions of Greek and Latin inscriptions published by Fra Giocondo of Verona in the last quarter of the 15th century were influential and set a high bar: *praeter quae vidi quaeque accurate exscripsi in hoc volumen nihil congesi* (“I have gathered into this volume nothing except what I have seen and what I have carefully transcribed”, quoted at Buonocore 2014, 30).

⁵ Hallof 2012, 2-3, “Der Hauptzweck einer königl. Akademie der Wissenschaften muß dieser seyn, Unternehmungen zu machen und Arbeiten zu liefern, welche kein Einzelner leisten kann, theils weil seine Kräfte denselben nicht gewachsen sind, theils weil ein Aufwand dazu erfordert wird, welchen kein Privatmann zu machen wagen wird.” Translation from Hallof 2009, 29, adapted. Boeckh went on to point out that only ‘a great prejudice’ (ein großes Vorurtheil) would deny to the fields of philology and history the same funding already bestowed upon the fields of mathematics and physics.

publications, a task that could be completed, it was thought, in four volumes to be produced over four years. When the fourth and final volume of the *Corpus Inscriptionum Graecarum* (*CIG*) did finally appear forty-four years later, a year before Boeckh's death in 1860 (the indices would not appear until 1877), it was clear already that a second edition and supplement were needed.

Epigraphists today, whether working in traditional ways or toiling in the digital trenches, can take comfort in three lessons that Boeckh's experience with *CIG* teaches us: in order to achieve success in any large-scale epigraphic project, long-term funding and collaboration are essential; such projects, if undertaken conscientiously, take more time than anticipated; and, no matter how professionally executed, they will inevitably require revision and updating. The successor project to *CIG*, the current *Inscriptiones Graecae* (*IG*), like the *Corpus Inscriptionum Latinarum* (*CIL*) initiated by Mommsen in 1853, was conceived through the recognised need in the nineteenth century to establish and maintain "a comprehensive and critical source-collection, presented geographically and by inscription-type, and one which is continually extended and updated".⁶ Two core points might be highlighted: one is the ambition to be comprehensive, combined with the assumption that there neither was nor would be any other comparably foolhardy project (whether conceived of as rival or fellow traveller); the other is the form and content of such a corpus.

Naturally, these corpora were constructed according to the best methods of the nineteenth century. The challenge of staying up to date has been constantly recognised, but hardly solved, at least outside of the internal archives of individual projects, the ephemeral *Ephemeris epigraphica: Corporis inscriptionum Latinarum supplementum*, or the early creation of *L'Année Épigraphique* (*AE*) and the *Supplementum Epigraphicum Graecum* (*SEG*). Other than the increased inclusion of photographs or drawings, more standardised basic metadata, and ever richer indices, the corpus volumes have not significantly changed over time.⁷ As Bodet put it when addressing the Oxford congress in 2007: "Innovations introduced into the entries in *CIL* ... have taken the medium of the codex ... nearly as far as it can go."⁸ A concordance is needed to navigate between the volumes of *CIL*,⁹ while *IG*, always intended to cover only the Greek inscriptions of Europe, now by design is destined to be incomplete. The basic question then, must be, is this still the best way to produce

⁶ Quoting from the ambitions of *CIL*: "It is the only systematic and critical edition which unites all the epigraphic sources from the former Roman empire (though only a selection of Christian inscriptions are included). As a comprehensive and critical source-collection, presented geographically and by inscription-type, and one which is continually extended and updated, it is an indispensable tool in the study of the ancient world." (Schmidt 2007, 19).

⁷ Certain editorial practices have changed: e.g., *IG* and *CIL* no longer attempt to reproduce the appearance of the carved letters, as they originally did; some practices, such as (in *IG*) the recording of 'fasti' of literary sources related to particular communities to accompany the epigraphic texts, have come and gone; the use of Latin persists; see Schmidt 2007, 25-27 with 16 and Hallof 2009, 32-45.

⁸ Bodet 2012, 280; the critique of *IG* by Merkelbach 1998 is noteworthy.

⁹ Fassbender 2003.

corpora, according to the methods of the nineteenth century, and not the methods now available to us in the twenty-first?

Günther Klaffenbach, director of *IG* from 1953 to 1972, recognized this limitation but saw it as inevitable. In response to the question “what are the users of a corpus looking for?”, he listed criteria we would all recognize as essential for a critical edition: information on the inscribed object (for *IG*, stone, itself a limitation that should perhaps be more open to question) and script, findspot and current location, a reliable text, a date, and “the minimum information necessary to understand the inscription”. None of this has changed; users of a corpus are looking for the same things today. But Klaffenbach went on: “For more than that, however, there is really no space, unless the acceptable length is to be exceeded, which would reduce the utility of the Corpus”.¹⁰ By acceptable length, of course, he meant for a folio volume, and ‘utility’ is here defined with reference to practicality. There was little else to imagine in Klaffenbach’s day. In 1967, he predicted “with confidence” that “the more the field diversifies over time, the greater the need will be ... for a few, large, comprehensive editions, to which every scholar of the ancient world can turn without lengthy searches and inquiries”.¹¹ In 1967 that was a reasonable expectation. Two years later the western world looked very different culturally, and digital technologies had begun to make their way into the world of epigraphy.¹²

Nowadays there is little reason to limit our idea of what a corpus should do by the technological capabilities of the nineteenth century, and no one today would argue that the best way to make inscriptions accessible to “every scholar of the ancient world” is through the medium of large folio volumes housed in major research libraries (and priced beyond the reach of individuals). The structure of an entry in the standard corpora is a beautiful thing—orderly, meticulous, concise, clean (in the sense that it is laid out visually in a way that facilitates comprehension)—and there is no reason to lose any of this utility or appeal; indeed all these features could be reproduced.¹³ But militant epigraphists have always pushed the technological capabilities of their day in order to acquire more accurate readings and to make inscriptions more legible to those not able to view them directly: drawings, squeezes,

¹⁰ Hallof 2009, 42-43, “Für mehr ist aber auch wirklich kein Platz da, wenn nicht der zulässige Umfang überstiegen und damit die Übersichtlichkeit des Corpus beeinträchtigt werden soll - und wenn der Bearbeiter auch wirklich fertig werden soll.” English transl. adapted.

¹¹ Hallof 2009, 24, “Je größer aber im Laufe der Zeit die Zersplitterung werden wird, desto größer wird (das glaube ich getrost behaupten zu können) wieder das Bedürfnis nach wenigen großen, zusammenfassenden Editionen werden, an die sich jeder Vertreter der Altertumswissenschaft ohne langes Suchen und Fragen wenden kann.” English translation on p. 27.

¹² Leaving aside the social and political upheavals of the day, Packard 1968 offered the first computer-based concordance of Livy, and work had begun on Jory and Moore 1974–1975: the IT revolution had begun (cf. Bodel 2012)..

¹³ Although it is worth bearing in mind the degree to which the *CIL* and *IG* entries, written still in Latin, are increasingly inaccessible, such that handbooks to epigraphy provide guidance on accessing them, e.g. Cooley 2012, 346-50 (cf. Merkelbach 1998, 295). For corpora entries, see Hallof 2009, 22-23; Schmidt 2007, 16, 18-23.

print photographs, text databases, digital imaging, photogrammetry and 3-D modeling, physical and chemical analysis have each, in their day, advanced the science of epigraphy and the accessibility of the material. None has fallen out of use with the arrival of each new technology, and the gains to knowledge have been cumulative. Is there any reason to anticipate lesser gains from transferring our accumulated knowledge from one medium of communication to another, one that enables additional capabilities and provides added value? Digital methods, notably but not only the possibilities of exploiting Linked Open Data (LOD), today enable us, and should require us, to do more.

To take one example: the so-called Leiden conventions were developed by papyrologists to meet their particular editorial needs. Subsequently adapted by epigraphists to the editing of texts incised in stone, the system enabled the consistent (i.e. widely comprehensible) and efficient presentation in the medium of print of core features of an inscribed text.¹⁴ The EpiDoc TEI standard was developed to encode the same features in machine-readable fashion, but such encoding is capable of capturing far more than Leiden or even an enhanced Leiden+ system ever could, and there is no reason artificially to limit the application of such encoding in line with the earlier limitations imposed by print.¹⁵ To what extent should the medium dictate the way we study and present our material?

These ever increasing possibilities beg the question, what is the purpose of the corpus? Any answer to that question will reveal an emerging tension, between the possibilities and desire for richer editions and the desire for comprehensive corpora to enable research. In the world of the codex, the two were seen to be mutually exclusive; and, in the world of the codex, the type of research that could be conducted over a collection of many thousands of texts was limited. We return to this theme below.

First, we examine a set of practices in which we have all probably engaged at some point, but which none of us consciously condones. These practices have emerged out of the existing model, based as it is upon the technology of the book, notwithstanding its many virtues. We refer to the habit of citing inscriptions by their entry numbers in standard corpora, or indeed in any of the supplementary series, anthologies or thematic and specialist corpora, as if the number were a mere descriptive identifier and did not in fact represent a specific edition made by a particular scholar on a particular basis at a particular time. This is not simply a question of potentially depriving an editor of credit for authorship (although this is itself an important point). Rather it is a matter of obscuring or even confusing

¹⁴ Panciera 1984 (repr. in Panciera 2006) provides a brief assessment of why the Leiden conventions never fully met the needs of epigraphists.

¹⁵ For EpiDoc, see Bodard 2010, Cayless *et al.* 2009, and <https://epidoc.stoa.org/> (= Elliott *et al.* 2006-2022).

authorship, and consequently the specific reading of the text adopted, when multiple editions of an inscription are cited as if equivalent. For example, when we cite a well-known epitaph of the late Republican period found outside Rome as *CIL* 1² 1221 = VI 9499 = *ILS* 7472 = *CLE* 959 = *ILLRP* 793, we leave entirely unclear whose edition we have consulted and on what basis the text that we have used in our research is constituted.¹⁶ The reasons for listing so many references are several (including the authorial claim to knowledge and authority), but above all the practice reflects the fundamental challenge of a world in which the original corpus edition of an inscription in *CIL* or *IG*, while still a reliable authoritative edition of record, is rarely the definitive edition (*CIL* 1² was published in 1918), given the multiplication of subsidiary corpora, further studies and new discoveries, with which the original corpus cannot hope to keep up to date; hence the value of incorporating at least one citation, where possible, of a reference in *AE* or *SEG*, particularly in the case of a publication that provides further material, contextual information or a photograph.¹⁷ Nonetheless, this citation habit combining abbreviation with feigned equivalence is deeply flawed: at best it is lazy, and fails to give credit where due; at worst, it makes it fundamentally difficult to engage critically with the research being conducted on the text in question. This practice was regularly criticised by Louis Robert, with his profound understanding of the importance of identifying and assessing every editor: “Quand l’éditeur n’est pas l’auteur de la copie, n’a pas vu la pierre ou l’estampage - c’est régulièrement le cas dans le *CIG* de Boeckh et de Franz -, il faut indiquer avec soin de qui est la copie; c’est capital pour la critique.”¹⁸ Furthermore, such practice increasingly conceals a crucial piece of information, namely which of these references point to genuine ‘editions’, contributing new or differing information and interpretation, and which ones merely reproduce a previously published text.

This situation has only deteriorated over time, firstly with the constant increase in paper publications; and secondly, with the rise of digital collections.¹⁹ The previous reference could, for example, be extended to include “= EDCS-19200211 = EDR167214 = TM574526”, and increasingly (and quite rightly), such digital identifiers are being cited in publications alongside paper ones. In the first place, we should note the current widespread failure to cite digital resources fully and formally in a

¹⁶ The text was transcribed and edited in *CIL* by Mommsen on the basis of autopsy (“*descripti*”) in the British Museum; the other editions listed here all leave implicit the source of their text, although all note that Mommsen transcribed it from autopsy.

¹⁷ The *EDR* entry for this particular inscription is especially instructive (http://www.edr-edr.it/edr_programmi/res_complex_comune.php?do=book&id_nr=EDR167214 = Butini 2022): the primary corpora references noted by us in the main text are supplemented by nine print and two digital editions / resources (some of the former are reported generally as works in *AE*, but this inscription is essentially not picked up for re-editions by *AE*, so finding more up to date information on this particular inscription would be a challenge without Butini in *EDR*!).

¹⁸ Robert & Robert 1954, 13-14 in a statement of their own principles for an edition; more trenchantly in critique of others, Robert 1959, 10-12 and 1965, 17-19.

¹⁹ We use this neutral term as a ‘catch-all’, precisely because the classification of digital resources, as part of a revised typology of editions, remains a desideratum.

comparable fashion to traditional printed resources.²⁰ Each of these identifiers, if you know where to look, will resolve as a webpage with further information about the inscription. Most of those citing these numbers do not bother to reference either the hosting website or the specific URL, and digital resources rarely appear in bibliographies with full details of publication date, authorship, etc.²¹ It is worth recognising, at the same time, that most of those citing *CIL* or *IG* leave many elements implicit, not least the date of the edition in question, as well as the identity of the actual editor responsible for that edition.²² However, at this point, comparison of these three specific digital examples highlights why current epigraphic citation practice ever more urgently needs revision.

The Epigraphik-Datenbank Clauss / Slaby (*EDCS*) boasts by far the largest number of texts of Greco-Roman inscriptions of any site on the internet (well over half a million at this point), and is an invaluable resource that is widely used. Texts are presented as search results via html, in a way similar to that of the Packard Humanities Institute (*PHI*) resource for Greek inscriptions, an aspect to which we return below. If we consider this specific case (*EDCS*-19200211), certain features can be observed:²³ the entry provides a transcription with supplements and editorial decisions such as corrected letters marked with the editorial conventions (an adapted Leiden system) peculiar to *EDCS*; a sequence of references to previous publications; an approximate date; and basic metadata about the original location and material of the support, the type of inscription, and persons mentioned in the text. There is no indication as to the source of the text and metadata (as there is in *PHI*) other than a listing of seven editions (or notices of editions, since *AE* 2006, 125 merely notes a volume which contains a new edition), nor is it immediately clear who prepared, published, or last revised the text and when. The *EDCS* ‘advice’ page notes that “The texts are based on the decision of the *EDCS* editors”, but this leaves many questions unanswered.²⁴ Currently there is no way to check the reliability of a text in *EDCS* or even to understand how it was created and where it might diverge from previous editions without going back to an earlier edition, and there is

²⁰ We attempt to model good practice in this paper, but doubtless with limited success.

²¹ Conventions for citing digital resources are well established: see e.g. <https://apastyle.apa.org/style-grammar-guidelines/references/examples/webpage-website-references> (= American Psychological Association 2021). The three sites in question are: <http://www.manfredclauss.de/> (= Clauss *et al.* n.d.); <http://www.edr-edr.it/default/index.php> (= Panciera *et al.* 1999-2023); and <https://www.trismegistos.org/> (= Depauw n.d.). In fairness to the user, many sites do not provide clear guidance for their own citation.

²² We note that there is a clear divergence of opinion in the community at present, expressed during the conference, as to whether an edition cited from a corpus such as *CIL* should in fact be attributed to the responsible individual, or whether the individual scholar should be subordinated to the collective enterprise (and so anonymised). It will be apparent that we subscribe to the former view.

²³ https://db.edcs.eu/epigr/epi_url.php?s_sprache=en&p_edcs_id=EDCS-19200211 (= Clauss *et al.* n.d.).

²⁴ This information can be found at: <https://db.edcs.eu/epigr/hinweise/hinweis-en.html>. Note that due to the use of frames on the site, following the link ‘Advice / Hinweise’ from the homepage to this particular page will not actually expose this URL, which has to be extracted by other means, making it even more difficult to provide accurate links to such information.

seemingly no way of knowing which edition was used to produce the text presented, in the many cases such as this where more than one edition is listed. Consequently, the increasing numbers of researchers who are citing inscriptions in their published research simply via an *EDCS* number are making arguments on the basis of a text of uncertain provenance and form - and which without a date of publication or revision may change not only without warning, but without ability to trace that fact - and thus falling short of the standards that we might expect of the scholarly community.

To be clear, this is not a problem with 'digital', or with the unique and invaluable *EDCS* database alone. Quite the opposite: the problem is widespread. It is a failure of academic practice, firstly on the part of those compiling the databases that do not provide users with critical information about how they are constituted, and secondly on the part of all those uncritical users (a majority, no doubt) who do not confirm the texts they report on the basis of more reliable, or at least, creditable editions. The fault lies on both sides, with both the producers and the users. On the side of the producers, clear and transparent documentation of how the data is produced and how it should be cited is essential; on the side of the users, it is imperative that we become more critical and responsible in our use of digital tools (but also, it emerges, in our use of published resources in general). Within the field of epigraphy, as already suggested, we need to become more critical in our appreciation of what constitutes an 'edition'. If the status of a text in *EDCS* is unclear (is it a copy of an existing edition? Or is it modified, and if so on what basis?), the same is not the case when we turn to the same inscription in the Epigraphic Database Roma (*EDR*).²⁵ In this case, a sequence of editions is reported, and a text is presented that is explicitly attributed to certain of those editions, with the further information that it has been checked against an image by the editor and with a limited apparatus; the author of this digital edition and the date of publication (and revision) is noted. The project's 'presentazione' page makes explicit the status of these texts as critical editions in their own right and provides a suggested form of citation for each record, which includes recognition of the authorship and date.²⁶

Lastly, we turn to the Trismegistos reference, the TM number cited above.²⁷ The TM number goes to the heart of the multiple editions 'problem', since the TM reference does not itself resolve as an edition of the text. Rather, the very purpose of the TM

²⁵ http://www.edr-edr.it/edr_programmi/res_complex_comune.php?do=book&id_nr=EDR167214 (= Butini 2022).

²⁶ http://www.edr-edr.it/it/present_it.php. It remains the case that the vast majority of citations of *EDR* in publications that we have observed do no more than quote the number, highlighting the fact that it is we, the users, who are above all at fault.

²⁷ www.trismegistos.org/text/574526. As a stable identifier, this is a different sort of resource when it comes to citation (see main text). The Trismegistos project simply suggests that in discussion of the resource one cites Depauw & Gheldof 2014 (see https://www.trismegistos.org/about/how_to_cite.php), on the grounds that reference to web addresses for bibliometric purposes "are often problematic" (!). For a more recent presentation and discussion, Depauw 2018.

text number is to provide a stable identifier to identify the inscription under discussion: consequently, what is provided is enough information, in the form of some basic metadata and core bibliographic references (including other digital resources), to refer unambiguously to this inscription and no other. In other words, the TM text number constitutes a reference to the inscription in the abstract, not a reference to a specific edition. Theoretically, therefore, in those situations where one is citing an inscription and the specific text/edition is not important, but one wishes to enable the reader to identify the inscription and locate editions of it, the TM number could serve in place of the entire list of references previously cited (or, one might imagine citing only the text edition actually employed together with the TM number, in order to enable the user to find a wider range of other editions if they wish to do so).²⁸ The TM reference thus makes particularly clear the need to be self-conscious and explicit as to what in fact one is citing, when referencing a text. But even this is not the whole story. Almost none of these editions make reference to the text-carrying stone's inventory number in the British Museum.²⁹ While it may not be the intention of an editor that the edition reference equates to an identifier for the text-carrier (i.e. physical object), this conflation of text with object is commonplace, at least in the practice of users. This confusion may even have been increased by the increased attention to a text's material context in recent decades (a repeated theme at recent AIEGL congresses). Although the TM number functions as an abstract reference to a text, the exact relationship between text and object in this case too remains opaque, because the TM number originated within the world of papyrological conventions, where commonly all texts on a single papyrus are treated as a single document.³⁰ This does not make the TM number less useful (although it does mean that there will be occasions where a single TM number may relate to multiple texts on a single object), but only reinforces the point that clarity of reference is essential, and there remains a further level of distinction, between text and text-

²⁸ Currently access to the concordance of identifiers provided by Trismegistos can only be accessed either by subscription or through direct use of the TexRelations Matcher API, which sets a high technical bar; consequently it remains inaccessible to most users and not fully 'FAIR' (on which see below at nn.72 and 79 especially).

²⁹ GR 1867.508.55 (the inventory number is recorded in *EDR* alone of the publications we have referenced). In fact there is a full online presentation of this record by the museum (including images, text transcription and translation), which is currently not cross-referenced by other resources: https://www.britishmuseum.org/collection/object/G_1867-0508-55 (= The Trustees of the British Museum 2023). The categorisation (and so title) of the inscription in the BM records as "tomb relief; block" is typically unhelpful to anyone searching for the inscribed text. Museum databases and catalogues seldom mention ancient inscriptions in their short titles and regularly fail to record them altogether.

³⁰ https://www.trismegistos.org/about/how_to_cite.php, and in particular "To determine what constitutes a document or book or inscription (and thus should become a separate record), we have given priority to material aspects: in principle all texts written on what was in antiquity a single writing surface belong together and form one document receiving a single Trismegistos number, unless there are good reasons to believe that the only (and unintended) relation between the two texts is the writing surface itself."

carrier, that has yet to be fully considered and integrated into many epigraphers' conceptual models.³¹

It is clear, therefore, that there needs to be an ongoing discussion about how best to express to a reader the precise source of the words they are seeing on the page, or upon which the argument that they are reading is based (i.e. an edition typology and best practice in the citation of editions). An example of such a categorisation can already be seen in the terminology employed in *CIL*, which includes:

- *descripsi*: original edition based upon autopsy
- *contuli*: revised edition (based upon autopsy?)
- *contuli quae supersunt*: revised edition based upon autopsy where only part of the original now survives compared to at the original moment of edition.
- *recognovi*: minor revisions of a previous edition (without autopsy)
- *recognovi et emendavi*: major revisions of a previous edition (without autopsy)
- *ex ectypo*: revision or edition based upon a squeeze

These could obviously be extended to include, for example, revision or edition based upon photograph (or other category of reproduction), but should also be extended to include categories such as the simple reproduction without emendation of an existing edition. The proposed guidance for recording observations of an inscription in the EpiDoc guidelines offers a basis for further development.³² Such an exercise in categorisation applies to editions of any kind, and is essential to the sort of scholarly rigour discussed above; but its necessity becomes ever more apparent as we start to encode texts in digital form.

At the heart of any such discussion is the identification of the necessary categories. How such categories should be expressed (in Latin, or any other language) is a secondary issue. The same applies to the presentation of text (it is not the shape of the brackets employed in Leiden or Krummrey-Pancierera that matters, but the varieties of text survival, restoration, and emendation that need to be represented), or to bibliography (it is not the abbreviation itself that matters, but the bibliographic edition that it represents). Digital encoding, in fact, enables such debates to be bypassed, since the encoding can be represented in any way one chooses. As the example of the Leiden conventions reminds us, it is in this crucial area of deciding on the appropriate categories that the community, we, the members of AIEGL, can play a valuable role by sharing our experience and pooling our expertise. We return below to the question of standards, which sits at the heart of Linked Open Data approaches.

³¹ This sort of conflation is accounted for within the CIDOC Conceptual Reference Model (CRM) developed by the cultural heritage sector to enable information integration particularly within the semantic web, which now has a proposed extension to incorporate ancient texts, called CRMt_{ext}, developed by Francesca Murano and Achille Felicetti: see <https://cidoc-crm.org/> and <https://cidoc-crm.org/crmtext/ModelVersion/version-1.0-0>, discussed in Felicetti & Murano 2017.

³² <https://epidoc.stoa.org/gl/latest/supp-histlocations.html> (= Bodard *et al.* 2022).

At this point in our discussion, it will be useful to consider the example of one regional corpus project that has successfully crossed the bridge from analog (that is, print) to digital publication: this may illustrate how the move to online dissemination can result naturally, even inevitably, from a consistent pursuit of the same primary responsibility of any epigraphist who is confronted with editing an ancient monument, which is to record it as fully as possible, as if she or he may be the only (or the last) witness who will ever do so. This recognition of our task as witnesses may be dulled in times of peace and easy travel, but it becomes clearer in times of war. When John Ward-Perkins, an archaeologist who had served with British troops in North Africa and Italy, was appointed Director of the British School at Rome in 1946, he understood the critical importance of recording archaeological evidence carefully and fully; from 1947 he visited Libya regularly, accompanied by an energetic graduate student, Joyce Reynolds. Ward-Perkins photographed, and Reynolds transcribed as many inscriptions as possible. Our choice of case-study is of course deliberate, not least as a tribute to Joyce. As Roueché has already observed elsewhere, “When the possibility of online publication arose, she embraced it – not because she had any affection for computers, but because she grasped that this could make her work accessible to so many more people all over the world.”³³

Reynolds and Ward-Perkins realised that their records would be imperfect, and that a full account would depend on collaboration with other scholars: it is particularly striking that, even in the sour aftermath of war, they worked closely with Italian scholars who had worked in Libya before them. They also saw the need to make their materials available as fully as possible, as soon as possible: *The Inscriptions of Roman Tripolitania* was published in 1952, within the constraints of post-war publication.³⁴ In 1953 Louis Robert observed the usefulness of the volume, while also regretting its limitations.³⁵ This was not the only time that Robert was to criticise corpora for being too limited: but the costs of printing and publishing would only increase. This is a situation where the best could so easily be the enemy of the good: the demand for ever richer commentaries, encouraged by Robert, has made publishing inscriptions ever more expensive. In particular, he commented that there were only 38 photographs for some 1000 texts: Robert had come to set great store by photography. These could be consulted at the British School at Rome: how much more secure they would be, he pointed out, if they were printed in a volume which could then exist in multiple copies. It was the need to disseminate the photographs which led the British School, fifty years later, to investigate the possibility of issuing an online edition of *IRT* with full illustration. Who can doubt that Robert would have welcomed this?

³³ Roueché & Thompson 2023.

³⁴ Reynolds & Ward-Perkins 1952.

³⁵ *Bulletin Épigraphique* 1953, no.25. “Il nous paraît cependant trop peu développé; on a fait un effort intense pour ne pas dépasser les limites de grosseur du volume et de prix de revient; il y a cependant un point à ne pas franchir.”

In 2009 a collaboration between the British School at Rome, the Institute for the Study of the Ancient World (New York) and King's College London led to the publication of *IRT2009* online.³⁶ This reproduced the 1952 volume in an enhanced, digitally encoded edition, with a few items that had been published immediately afterwards. The 2009 edition retained the *IRT* numbering: this reflects a reality, that people want simple ways of referring to a text in discussion. *CIL* VIII was published in 1881, so scholars have been dependent upon regional studies such as *IRT* to define the corpus in the subsequent 140 years.³⁷ Traditionally, one major contribution that a corpus can make, especially when providing the first edition of a text, is to give an inscribed text an identifying number, a name that can be used to refer to it, and that almost inevitably will be so used. One major challenge for the community (discussed above) is to revisit how this process works, since the fact that the number represents an actual edition tends to get lost over time. In the truly integrated multilingual environment to which the collective *corpus inscriptionum* of the ancient Mediterranean should aspire, a system of unique universal identifiers will be essential: at the same time it remains the case that abstract identifiers such as those generated by the Trismegistos project are more legible to machines than humans, and there will be a balance to be struck between interoperability (among computers, mostly taking place out of sight) and practicality (for researchers accustomed to deducing basic contextual information, such as language or location of an inscription, from the abbreviated title of the corpus in which it is published).

The 2009 edition did not attempt to provide new or improved commentary, except in one important regard: Joyce Reynolds did provide English translations for almost all the texts. In the digital format there were no constraints on space to prevent this; and translation is an essential form of commentary, particularly in a publication which is suddenly accessible to anyone anywhere in the world, not just to a few well-prepared scholars. *IG*, under Klaus Hallof's direction, has made freely available online translations (mostly in German) of all the inscriptions published in the corpus volumes since 1945, and the *Attic Inscriptions Online* Project (*AIO*) is providing enhanced English language translations and annotations of all inscriptions concerning ancient Athens and Attica; meanwhile, the hugely ambitious EAGLE MediaWiki project unites translations from multiple projects and individual contributors.³⁸ The other aim of the 2009 publication was to provide better location data. This had for many years been difficult in Libya, where geographic data were controlled for security reasons; but the arrival of freely available satellite maps, such as Google Earth, changed the situation. The editors therefore collected the

³⁶ Reynolds *et al.* 2009, available at: <https://inslib.kcl.ac.uk/irt2009/>.

³⁷ <https://irt2021.inslib.kcl.ac.uk/en/concordance/bibliography/cil.html> (accessed 2023-02-09) for the handful of inscriptions already in *CIL*.

³⁸ *IG* translations: <http://telota.bbaw.de/ig/> (= Berlin-Brandenburgischen Akademie der Wissenschaften n.d.). *AIO*: <https://www.atticinscriptions.com/> (= Lambert *et al.* n.d.). The EAGLE MediaWiki online at https://wiki.eagle-network.eu/wiki/Main_Page (accessed 2023-02-09), with Liuzzo *et al.* 2014 and Almas *et al.* 2017.

geographic data, with links to Google Earth. Over time the accessibility of Google Earth changed, and maps became less usable; the geodata were then used to build a Gazetteer of heritage locations in Libya.³⁹

The 2009 edition was widely welcomed and used, not least in Libya; colleagues appreciated the instant accessibility, and the many search tools. It was, however, clear that it could be improved, and linked to the Gazetteer. The core information had all been encoded according to agreed standards: it was therefore simple, and inexpensive to reuse it in a new edition. The aim of *IRT2021* was to add all known Greek or Latin inscriptions that had been published since 1952.⁴⁰ This was made possible by the diligent work of collection enterprises such as *AE*, *SEG*, and the Epigraphic Database Heidelberg (*EDH*).⁴¹ But even more important, in the spirit of the 1952 publication, was collaboration with expert colleagues, something made much easier in a digital environment. A first draft was made available to selected scholars online; the editors could then add comments, improvements, translations and above all much fuller illustration. Furthermore, it was now possible to use new standards for connecting information: the indices provide links for standard terms to Wikidata, or other resources, whether fully or partially digital, such as the *Lexicon of Greek Personal Names (LGPN)* or the *Prosopographia Imperii Romani (PIR)*.⁴² Consequently, the process of indexing not only connects data within the project, but links it out to other data sets, potentially without limits.

Publishing *IRT2021*, funded by the British Institute for Libyan and North African Studies (formerly the Society for Libyan Studies), was relatively inexpensive, and took less than twelve months. An immediate consequence is that it will lend itself easily to further editions, with new readings, new images, new information, multi-lingual translations, and more: an international team is already discussing how to develop the online collections of inscriptions from Libya, forming, as an initial step, the Libyan Epigraphy Research network.⁴³

Publication online, therefore, has been driven by very practical considerations. A modern witness to an inscription is required to record with care, and to publish a considerable body of information, particularly photographs. Even without the provision of a rich commentary to the standards of Louis Robert, the resultant publication presents demands that a printed book cannot meet. Modern archaeological activity in particular means that such collections need to be regularly updated, creating further expense. But the story of *IRT* demonstrates that the publication of an epigraphic corpus in digital form is an absolutely natural and

³⁹ <https://slsgazetteer.org/> (= British Institute for Libyan and North African Studies 2015)

⁴⁰ Reynolds *et al.* 2021, available at: <https://irt2021.inslib.kcl.ac.uk/en/>.

⁴¹ *EDH* online at <https://edh.ub.uni-heidelberg.de/> (= Heidelberg Academy of Sciences 1993-2021).

⁴² *LGPN* is mostly online and in evolution towards being a truly Linked Data resource

(<https://www.lgpn.ox.ac.uk/> accessed 2023-02-09); *PIR* is only very partially online

(<https://pir.bbaw.de/#/overview> (accessed 2023-02-09).

⁴³ <https://libyanepigraphy.org/> (accessed 2023-02-09).

perhaps inevitable evolution. Moreover, because digital publication offers new possibilities and has the potential to be transformative, it is challenging us to rethink our basic methods and approaches, our communal disciplinary culture, whether that is the simple nature of an edition, or the function of the corpus. *IRT*, like many other local and regional corpora, has also been strongly driven by the hope that this will make epigraphy accessible and relevant to more people than ever before, particularly in the communities where many of our inscriptions have been recorded and which bear the burden of preserving and curating them.

Open scholarship

It will be useful at this point to recall the words of Boeckh and Klaffenbach quoted at the outset: that corpora attempt to achieve what no one person could do alone intellectually or materially; but also that the greater the diversification of the field, the greater the need to enable the scholar to access core and reasonably comprehensive data without lengthy searches and inquiries. Digital work in epigraphy, it would be fair to suggest, has for some time pushed in two seemingly opposing directions, which themselves are reflected in the story of *IRT*: the possibilities and desire for richer editions; and the desire for comprehensive corpora to enable research.

A project such as *EDCS*, *PHI* or *EAGLE* exists, we would suggest, for exactly the same reasons that the original corpora were compiled: to make as many texts as possible accessible - and searchable - by researchers; and digital capacities make that possible in ways that were not conceivable in the nineteenth century. The digital projects of recent decades, however, were (and are) very much projects of their time, not only technically but intellectually, reflecting a world in which digital corpora have been conceived of as purely derivative of published editions, ancillary tools, rather than curated data and original editions in their own right. It is worth noting the secondary definition of 'corpus' that emerged in the twentieth century: 'a body of spoken or written material on which a linguistic analysis is based', which is to say, a text corpus for analysis.⁴⁴ Researchers increasingly want to do more with such collections of texts (and associated information) than simply use them as tools for searching text strings - a desire that goes to the heart of both how and why these texts and monuments and the information about them (collectively, our 'data') are published to begin with, while going beyond the original conception of these projects. In the sciences the need for transparency and reproducibility in data-based research is manifest, and a variety of international protocols has been developed to support it; there is, however, nothing peculiar to the domain of the natural or life sciences that makes these principles discipline specific. Accurate and careful

⁴⁴ The definition is taken from *The New Shorter Oxford English Dictionary* (1993), s.v. 'corpus', noted as mid-twentieth-century, and presented as a specific sub-meaning of the general concept of a corpus as "a body or collection of writing, knowledge, etc; the whole body of a particular category of literature, etc." (itself going back to early eighteenth-century (i.e. enlightenment) usage).

citation and referencing of data is as important to us as to the physicists: this was why classical scholars designed and developed the *apparatus criticus* in the first place. We can and should make these protocols our own. A text database can be recognized as a product of its time, a pioneering effort of the late 1980s that performs an invaluable service but has not kept up with technological developments. At the same time, as the case of *IRT* illustrates (and it is far from being alone), digital corpora are now being published that, rather than being derivations of published corpora, in fact sit parallel to them, or even supercede them.

Aggregator projects, as we might call them, such as *EDCS* or more recently, and complexly, The Europeana network of Ancient Greek and Latin Epigraphy (EAGLE),⁴⁵ reflect one response to the essential problem of the future of the corpus, itself reflecting the concerns of Boeckh and Klaffenbach already noted: the desire of the community to have a readily accessible, comprehensive set of texts on the one hand; and the need to navigate a proliferation of ever richer editions on the other. This proliferation of editions was a feature of the analog print world long before the digital turn, whence the increasingly confused modes of reference discussed above. It has been exacerbated as a problem for a variety of reasons, some of which are digital, but some of which are more general, not least the very question of the viability of the grand corpora themselves, as well as the changing national and institutional pressures on scholars to publish in certain ways. Several of the major attempts of the 1990s to compose large-scale corpora reflect in their evolving division of the Latin epigraphic world among *EDH*, Hispania Epigraphica Online (*HEpOnI*), Epigraphic Database Bari (*EDB*), and *EDR* the familiar challenge of any grand corpus project: large scale coverage was only possible through collaboration and division of labour, resulting in a number of separate but collaborating corpora projects.⁴⁶ Subsequently, with the development of standards such as EpiDoc XML, the development of increasingly rich, but smaller, more localised or thematic corpora, exemplified on our own parts by *IRT*, *USEP* and *I.Sicily*, has further complicated the digital picture, with ever greater overlap.⁴⁷ Such projects in turn collaborate to a greater or lesser degree with the larger scale projects, but all such collaboration is essentially ad hoc; the EAGLE project revealed, as Panciera saw, the challenges of

⁴⁵ See <https://www.eagle-network.eu/> (= Orlandi *et al.* 2013-2023) and the concluding conference, Orlandi *et al.* 2017.

⁴⁶ See Bodel 2012, 285-7 for a summary of this move to federation. We crudely distinguish here between 'aggregator' projects, which draw texts from existing corpora or publications, and primary 'corpora' which 'edit' the texts to a greater or lesser degree, generating new 'editions'. *HEpOnI* is at <http://www.eda-bea.es/> (Gómez-Pantoja n.d.), *EDB* is at <https://www.edb.uniba.it/> (Carletti *et al.* 1988-2023).

⁴⁷ Reynolds *et al.* 2021, online at <http://irt2021.inslib.kcl.ac.uk/>; Bodel 2003-2023, online at <https://usepigraphy.brown.edu/projects/usep/collections/>; Prag 2017-2023, online at <http://sicily.classics.ox.ac.uk/>. Any selection of projects is invidious but, acknowledging our hosts at the conference, we highlight the long-lived PETRAE project (<https://petrae.huma-num.fr/fr/projet> = Devillers *et al.* 2012-2018) as a valuable and evolving model at this regional level.

agreeing standards sufficient to enable interoperability.⁴⁸ In the rest of this paper, we focus on two key areas which are fundamental to a digital future, and which have been highlighted by constructive sceptics: questions of sustainability and preservation on the one hand, and of the increasing demand for common standards to enable interoperability on the other.

Sustainability is not a uniquely digital problem - every project has a limited life and limited funds - but digital outputs do present distinctive problems of sustainability. When Boeckh or Mommsen published a fascicle of *CIG* or *CIL*, they did not share our current perspective that such volumes might be thought increasingly inaccessible because existing in only a few expensive copies. More than this, however, they did so because they had confidence those copies would still exist, in multiple copies and be curated in libraries in fifty or one hundred years, preserving knowledge that in some cases only existed otherwise in a single manuscript. Digital epigraphists today can reach anyone with internet access but cannot be confident their work will be accessible at all in fifty or even twenty-five years: and this dilemma is doubtless one reason why digital publications, whatever their quality, continue to be treated with caution. Digital projects need both durable storage and ongoing infrastructure support - in other words, the equivalent of a library. Even running a static site (i.e. one which is no longer being updated, and which does not rely on software to generate content) on a server requires upkeep and expense, since both hardware and basic operating software still require regular maintenance and updates (just as a library needs maintenance, security, and staffing).⁴⁹ This is a real problem, not yet solved; but some key points should be noted.

Firstly, there is a difference between a static publication, i.e. a fixed edition (e.g. *IRT*) on the model of paper publication, and a dynamic publication, i.e. one subject to continuously published revision and updating (e.g. *I. Sicily* or *USEP*). This dynamic publication (i.e. one which is intended to allow for continuous revision and additions to individual text editions within a corpus, with no fixed end-point) clearly requires the ability to document, archive and reliably cite each distinct version (i.e. edition, however classified). Moreover, it almost certainly requires the ability to assign intellectual responsibility to very granular elements of a publication, since a new version is potentially constructed with a single correction to a text or its metadata, perhaps by a different individual (the sorts of changes traditionally united, retrospectively, at a single point of time in the future, in the *apparatus criticus* of a second edition).⁵⁰ This model, of course, reflects a very real digital solution to the

⁴⁸ The observations of Panciera 2012 remain extremely pertinent; Liuzzo 2015 is a sobering overview of the multiple challenges.

⁴⁹ *IRT2009* is maintained in this basic form, at <https://inslib.kcl.ac.uk/irt2009/> (“*IRT2009* has been archived, and thus search and map functionality is no longer available.”)

⁵⁰ The principles of version control, which extend far beyond the world of digitised inscriptions, are useful for designating minor changes that might not rise to the level of constituting a new ‘edition’: see the rich and wide-ranging discussion of Broyles 2020. A partial solution, whereby substantial revisions are archived to Zenodo.org with a distinct DOI, is currently in use for *I. Sicily* (see Prag 2021, 185-6

fundamental challenge presented to any corpus by the need to revise and update. However, it also highlights a further area for future community discussion - what constitutes a major or a minor revision? The Sicilian corpus of Greek and Latin inscriptions in *IG* and *CIL* is now well over one hundred years out of date; a new edition once every hundred years or more should not be something we have to live with any longer; and withholding individual revisions until a complete corpus is ready for publication is also no longer necessary.⁵¹

Secondly, there is a difference between data and software (this is hardly a new point in the digital world, but it deserves emphasising): the *I.Sicily* and *USEP* websites are never going to be sustainable indefinitely; each will require periodic upgrades in order to remain functional (indeed the current *I.Sicily* site, developed in 2015/16, is already beyond life expectancy and will need to be rebuilt shortly). It is the website and supporting software that generate the interfaces that make visualisation, searching, and analysis easier; but the actual data (notably the XML files containing the editions, and the images) can be readily stored and made available much more indefinitely and for very little expense. Existing examples include the model data repository of the currently static Heidelberg database (*EDH*), or the data deposits from multiple projects collected by the International Digital Epigraphy Association (IDEA) on Zenodo, or the data deposit of the constituent files of *IRT2021* in the King's College London digital repository.⁵² Any digital project should have a clear strategy for this sort of basic preservation of its data, but this is something that requires planning at the start, and until relatively recently was not something that was widely recognised.⁵³ It is true that this sort of publication is not so easily human readable, but to assume that the only value of publication is a traditionally human readable publication, even in digital form, is to miss a very large part of the value of digital publication: data that is available digitally can not only be searched and cited by humans, it can, potentially, be searched by machines, and reused to undertake further research.

At this point our different threads begin to come together. Multi-authored corpora came into being in order to make as many texts as possible accessible in reliable editions (ever more important in the face of potential loss); but also in recognition of the limits to what individuals could realistically achieve working alone. But corpora continue to multiply and expand in both quantity and richness of data (as the

and for an example, compare <http://sicily.classics.ox.ac.uk/inscription/ISic000537> with <https://doi.org/10.5281/zenodo.4384888> and <https://doi.org/10.5281/zenodo.6421667>); more granular versions could also be recovered from <https://github.com/ISicily/ISicily>.

⁵¹ The PETRAE project illustrates another approach, which is the parallel publication in paper and digital format (see <https://petrae.huma-num.fr/fr/publications> accessed 2023-02-09) with volumes produced from the digital corpus.

⁵² <https://edh.ub.uni-heidelberg.de/data> (= Heidelberg Academy of Sciences 1993-2021); <https://zenodo.org/communities/eagle-idea/> (accessed 2023-02-09); <https://doi.org/10.18742/c.6002275.v1> = Reynolds *et al.* 2022.

⁵³ See e.g. Rios 2018 discussing the general focus of research data management plans on data but not (until relatively recently) software.

inscriptions themselves continue to multiply). Digital tools first emerged as a means to analyse the already unmanageable mass of data in the printed corpora,⁵⁴ highlighting an essential purpose and value of corpora, the ability to study and analyse a large body of texts in carrying out our research. On the one hand researchers want to ask more questions of such data; and on the other, as the data itself becomes richer, those questions and research possibilities only increase; and digital recording and publication make possible ever richer data and analysis. Furthermore, if the purpose of a corpus is to enable research, it is not for the compiler of that corpus to second-guess future research questions, only to make the material available. Lastly, we are increasingly required by funders to make our data openly accessible, which is an admirable ambition (if one which funders currently do not fully resource). To return to one of our earlier ‘aggregator’ examples, *EDCS* enables an enhanced level of textual searching compared to paper indexes, but the data is not available in any readily reusable form (only as HTML in response to a query),⁵⁵ its legal status (whether one is allowed to re-use it in any way, and, if so, who should be credited) is unclear,⁵⁶ and the quality of the data is unknowable (for the reasons discussed above).

It is worth considering, therefore, what becomes possible when we make the data, even just the basic texts, available for use in digital form by our fellow researchers. Early, highly labour-intensive examples of this, such as Richard Saller’s and Brent Shaw’s 1984 study of “Tombstones and Roman Family Relations in the Principate”,⁵⁷ remain rare, because the data remains extremely inaccessible. More recent examples of such papers, whether focused on linguistics or epigraphic culture, consistently describe the basic challenges (and considerable effort) of trying to build a suitable dataset, notwithstanding existing digital resources, and the resulting limitations on the results.⁵⁸ It is particularly striking that linguistic analysis of epigraphic texts lags behind parallel work on literary or papyrological texts, a regrettable situation undoubtedly due in part to the limited availability of usable and reliable editions.⁵⁹ A couple of contemporary examples of digital projects will help to

⁵⁴ The principal example is Jory & Moore 1974-1975, who manually entered the texts of *CIL VI* in order to create a simple Key Word In Context index.

⁵⁵ Consequently, export or re-use of the data is only possible through the use of web-scraping (the extraction of data from a website, often using software to extract and standardise the data for further use: see e.g. Black 2016 for a discussion of some of the associated issues).

⁵⁶ The only licencing information is the generic internet disclaimer at <https://www.disclaimer.de/disclaimer.htm?farbe=FFFFFF/000000/000000/000000> (accessed 2023-09-02) and the disclaimer statement at <https://db.edcs.eu/epigr/hinweise/disclaimer-en.html> (accessed 2023-09-02). The *PHI* site presents similar problems for re-use, even if the data sources are clear (<https://inscriptions.packhum.org/> = Packard Humanities Institute n.d.).

⁵⁷ Saller and Shaw 1984.

⁵⁸ Adamik 2021 (linguistics); Heřmánková *et al.* 2021 (epigraphic culture).

⁵⁹ See CLaSSES (<http://classes-latin-linguistics.fileli.unipi.it/en> = Marotta 2023) for a smallscale example, or more recently and openly, CEIPoM (<https://reubenjipitts.github.io/Corpus-of-the-Epigraphy-of-the-Italian-Peninsula-in-the-1st-Millennium-BCE/> = Pitts 2022); contrast PapyGreek (<https://papygreek.hum.helsinki.fi/> = Henriksson and Vierros 2022) on papyrological data.

illustrate the challenges – and the potential – of both richer and more accessible data encoding.

The LatinNow project, directed by Alex Mullen, set out to employ sociolinguistics, epigraphy, and archaeology to write the social history of the Latinization of the north-western provinces of the Roman Empire. In the end, a huge part of the five-year project was devoted simply to preparing a dataset for analysis. The work would have been impossible without the massive works of data creation and aggregation already undertaken by a number of previous and ongoing projects, and the initial work of unification by the EAGLE project; but the limitations still present in that data and the work required to make it actually usable revealed the challenges that researchers still face. The remarkable GIS dataset that has resulted was released in beta form to the community at the time of the Bordeaux congress (August 2022).⁶⁰ The Ithaca project, led by Yannis Assael and Thea Sommerschild, has developed a machine-learning platform to assist epigraphers with the reconstruction and geographical and chronological attribution of Greek epigraphic texts.⁶¹ However, this project required several months of simple cleaning of the texts available in *PHI* in order to have a basic dataset necessary to make machine-learning possible; similar work is now underway to establish a comparable Latin text corpus from existing resources. It is sobering to realise, three conferences and fifteen years after the optimistic hope was expressed that we might be on the cusp of realising the possibilities of digital analysis based upon methods such as semantic mark-up, that we still sit on that cusp.⁶² There are two sides to this continuing problem: how digital datasets are made available (or not), and, looking back to considerations of academic practice discussed above, how they are both created and consumed. Projects such as these provide the merest hint of what any corpus project could, indeed should, now be enabling, far beyond the basic searching of metadata or texts, and far beyond the original questions or intentions of those who first created the data.

Re-use of data for further research brings to the fore the very problem with responsibility and citation that underlies some of the concerns discussed above (as well as the necessity of ‘clean’ or standardised data, to which we return below). An aggregator may bring together data in a way that makes it possible for others to re-use it; but if the aggregator does not record the source(s), not only is the validity of this data no longer (easily) verifiable, but the basic principles of intellectual property are violated. This is not, however, a problem with aggregators alone: plenty of existing projects make data available without providing sufficient documentation of intellectual responsibility or methods of data creation; and in turn, any re-user,

⁶⁰ LatinNow at <https://latinnow.eu/> (= Mullen 2017-2023); for the GIS data, see <https://gis.latinnow.eu/> and <https://latinnow.eu/online-resources-2/> (accessed 2023-02-09).

⁶¹ Assael *et al.* 2022, with the tool available online at <https://ithaca.deepmind.com/> (accessed 2023-02-09). The project essentially applies machine-learning to the traditional tasks that epigraphers regularly undertake on a daily basis, using paper corpora and now often also text-string searches in *EDCS* and *PHI*, in searching out parallels for reconstructing and contextualizing new inscriptions.

⁶² Bodel 2012, 282.

aggregator or direct research exercise, has a responsibility to document the source of their data, not always observed. As we noted earlier, this is not a specifically digital issue (and it is one that the *apparatus criticus* in part was intended to address) - but, it is made more urgent by the very nature and potential value of digital data. Established mechanisms exist for attributing and recognising authorship (whether original authorship, or other forms of data curatorship or data encoding), starting with the existence of digital identifiers for scholars that can be embedded in the data, such as ORCIDs, but it will take an effort of the community to embrace them and to do so in standardised ways.⁶³

These issues do not apply only to data publication for reuse, however (and let us be clear, when we talk about 'data publication' we are talking primarily about the publication of digitally encoded editions of inscriptions). They are the same issues as make collaboration a much more realistic possibility. In recent years, not least through the COVID-19 pandemic, we have all learned how much more interaction is possible digitally. Digital tools of varying levels of complexity, from Dropbox to Github, make highly collaborative corpus-building a reality. But such methods make it even more important, especially considering the need for early career researchers to be able to demonstrate their actual outputs and activity, that individual contributions in a collaborative environment be properly recorded and credited. Whether one believes that a particular edition of an inscription should be cited by its corpus number alone, or with reference to the one or more authors of that edition, it is essential that the question of who prepared the text, who recorded the metadata, who compiled the bibliography, who took the photograph, should be documented: this information has often been obscured to some extent by the nature of print editions; but it is easier to incorporate such information in a digital environment. This is one of several areas where it becomes apparent that using digital methods in fact requires that we hold ourselves to higher standards than those to which we have become accustomed; it is also notable that multiple authorship (common in the sciences) is becoming more commonplace. The purpose of documenting the construction of an edition in appropriate digital form is to ensure not only that the information can be retrieved but also that the work of individuals can be identified and catalogued;⁶⁴ and, as already noted, that information is important when it comes not just to citation, but to data re-use.

⁶³ ORCID stands for Open Researcher and Contributor ID, a community-built, not-for-profit, global organisation providing a unique, persistent identifier free of charge to researchers (with now in excess of 10 million registered users): <https://orcid.org/> (accessed 2023-02-09).

⁶⁴ See <https://github.com/ISicily/ISicily/graphs/contributors> for a basic example, from which it is possible to drill down to each change each individual has ever made. For a different sort of example, see https://en.wikipedia.org/wiki/Praeneste_fibula, together with https://en.wikipedia.org/w/index.php?title=Praeneste_fibula&action=history (accessed 2023-02-09). It is interesting to note the absence of epigraphers from this document; and the rather telling comment from a contributor at: https://en.wikipedia.org/wiki/Talk:Praeneste_fibula#Expert_still_needed.

These themes of collaboration, sustainability, proliferation and aggregation are not new, but digital methods make them ever more complex and important, and challenge some of our long-standing conventions of practice, which have evolved over the decades and centuries. The EAGLE project (2013-2016) confronted some of these challenges by attempting openly to aggregate the data from a large number of existing digital projects, exemplifying the principles of collaborative corpus-building and open data, in order to make real the ambition of federation that had earlier been embraced by a number of projects, with the encouragement of AIEGL (see above, nn. 45-46). The fundamental challenge that the EAGLE project faced (from our perspective) is that of trying to unify data from diverse projects retrospectively.⁶⁵ Even if data is openly available, if it is not prepared to the same standards (whether in inscription typologies or the treatment of past editions), it can only be combined with immense difficulty, as illustrated by the efforts required by the projects noted above, such as LatinNow and Ithaca. The EAGLE project valiantly attempted to align the 'metadata' labels (i.e. descriptive classificatory information, such as type of inscription, type of object, type of material) used across multiple projects, not only those involved in the EAGLE federation.⁶⁶ It thus instantiated one of the fundamental principles of Linked Open Data, the ability to align data through standardised terminology (and so to achieve any sort of integration or useful aggregation). It is worth recalling that the Greek and Latin epigraphic community has a strong tradition of evolving community standards, as being essential to communication. The primary example, already discussed, is the general adoption and development of the Leiden Conventions, albeit with some diversification between the Greek and Latin epigraphic communities. Those divergences, however, highlight a fundamental point, which is that the categories that the conventions capture are widely recognised, but there can be autonomy in their representation (and not all are needed by all, since some only apply in specific domains).⁶⁷ In this specific case, encoding the universally agreed categories in EpiDoc leaves entirely open the question of how the categories are subsequently displayed, so there is no loss of independence.

A second recent example is offered by the AIEGL proposal for standardisation of the abbreviations for Greek epigraphic publications, to which other projects can in turn

⁶⁵ The projects that joined the federation were born at different stages in the development of digital tools and methods from the 1980s to the 2000s; consequently, the variations in format and content among these projects are the result of circumstance, not design. We have not discussed here the further challenge of changing formats over time, positively exemplified by the PETRAE project's evolution between 1986 and the present day: outlined at <https://petrae.humanum.fr/fr/projet/historique> (accessed 2023-02-09).

⁶⁶ <https://www.eagle-network.eu/resources/vocabularies/> (accessed 2023-02-09), with an overview of the issues in Liuzzo 2015, and profound reflections on the project in Liuzzo 2018.

⁶⁷ Some of the principle divergences can easily be observed on the comparative table in the EpiDoc quick reference guide (Bodard 2021-02-01), available at: <https://github.com/EpiDoc/Source/blob/6041fccd12a309b689ec4223d52f33df7a1a8278/guidelines/msword/quickref.pdf>.

offer alignment, as recently demonstrated by *SEG*.⁶⁸ The next step in that particular case is to offer stable digital identifiers for each bibliography item, so that projects can maintain autonomy in their presentation of the item, while still ensuring the alignment that is the ambition behind the standardised abbreviation.⁶⁹ Both the AIEGL abbreviation list and the community development of the EpiDoc standard exemplify how well this can work. But at present the community lacks the detailed standards that would enable the easy comparison of data from one project to another: the EAGLE vocabularies are explicitly only a first draft, and it has so far been left to small voluntary groups such as the epigraphy.info community to make some first efforts in this direction.⁷⁰ This need for community agreed standards extends very widely, far beyond agreement on, e.g, how to classify a funerary inscription, and includes, as discussed above, the need to clarify categories of edition and principles of citation. None of these issues is new or unique to the digital world; but the possibilities opened up by digital methods make precision in these areas ever more important if we are to advance in our research.

EAGLE also came up against one of the other fundamental challenges noted above, sustainability: by aggregating the data from individual projects into a new database, EAGLE immediately faced the same sustainability challenge as any other project. In this case the problems are twofold. On the one hand, there is the need to maintain a database and its interface.⁷¹ On the other, there is the need for regular future intervention in order to keep such a dataset ('corpus') up-to-date, a problem noted repeatedly throughout this paper. The ability to revise and update remains one of the essential challenges of any corpus project, and one which digital capabilities in principle make easier, not harder. A serious challenge for any model of collaboration or aggregation, however, is how to capture and process the many revisions undertaken across many projects. Many mechanisms are needed to achieve this, most of which have been touched on in the preceding discussion, from version control, to authorial documentation, to agreed data standards, as well as open data and interoperability: from a technical perspective all of these things are possible. Some of the most fundamental elements required are already firmly established, with the EpiDoc standard for text-encoding, or the existence of unique identifiers for every inscription in the abstract (rather than denoting any particular edition), via the Trismegistos project, which creates the possibility for machines to connect every

⁶⁸ AIEGL Greek abbreviations list, Chaniotis *et al.* 2022, online at <https://aiegl.org/grepiabbr.html>; *SEG* concordance, Tsolakakis 2021, online at <https://scholarlyeditions.brill.com/sego/abbreviations/>.

⁶⁹ A pilot is under development at <https://github.com/FAIR-epigraphy> (Asif *et al.* 2022-2023), with the support of AIEGL.

⁷⁰ Specifically the working groups of epigraphy.info: https://epigraphy.info/working_groups/ (accessed 2023-02-09).

⁷¹ The EAGLE database can be searched at <https://www.eagle-network.eu/resources/search-inscriptions/> (accessed 2023-02-09). In the case of EAGLE, an initial solution to fund future work was the subscription model of the IDEA association (<https://www.eagle-network.eu/about/who-we-are/>); Trismegistos has adopted a similar approach (<https://www.trismegistos.org/registration.php>, accessed 2023-02-09), although in this case access to tools and resources is in turn limited for those who do not subscribe.

edition of a particular text unambiguously. Arguably, our greatest challenges are intellectual, academic, and cultural: firstly, the necessary cross-community collaboration on standards, which enables, for example, two epigraphists to agree, however loosely, on what might constitute a funerary inscription, or a meaningful change (i.e. new version) to an epigraphic text; and secondly the recognition that 'data' should be openly accessible, ideally in line with the now widely advocated FAIR principles: data should be Findable, Accessible, Interoperable, and Reusable.⁷²

We conclude with what we hope is an optimistic, rather than a pessimistic vision for the future of the corpus. Given the ever increasing challenges of scale, proliferation, funding and even sustainable travel, together with the fundamental opportunities for collaboration, more detailed analysis and more rapid revision that digital methods present, we would argue for an inversion of the existing corpus model, one that moves away from the single monumental corpus, and instead builds directly upon the proliferation of local and thematic corpora.⁷³ Fundamental to such an approach, however, are two basic sets of principles and approaches, alluded to and implicit in much of what has been said so far: the FAIR principles, for publication of data in ways that are Findable, Accessible, Interoperable and Reuseable; and the overlapping but distinct principles of Linked Open Data.⁷⁴ Essential for the success of any of this, however, is the ability of the community to agree on standards, whether for data categories, bibliographic and citation conventions, levels of versioning (editorial revision), or authorial responsibility. This challenge is hardly unique to the field of epigraphy: the frameworks for much of this are common to many disciplines, and do not need to be redeveloped for our own;⁷⁵ but some are discipline-specific to epigraphy, and need to be at least broadly agreed among us. To repeat the earlier point about standards and representation, it deserves emphasising that this does not require that any particular project actually calls a funerary inscription a *titulus sepulcralis*, but it does imply that all projects need to state somewhere in their data what categories they are using, and how those might align with an agreed list of categories - remembering also that in a digital

⁷² <https://www.go-fair.org/fair-principles/> (accessed 2023-02-09, and see Wilkinson *et al.* 2016). For a recent experiment with archiving a TEI-EpiDoc project employing FAIR principles, see Creamer *et al.* 2022.

⁷³ It is perhaps not inappropriate to quote Silvio Panciera: "One might claim that irresolvable technical divergences no longer exist these days. More important, in my view, is the need to join together in identifying a collective goal, which takes account of the necessity for different threads of research and of the best route to follow in order to reach it: this with overt protocols, which will subsequently allow the goal once reached to become a new point of departure. But on the one hand that calls for a spirit of unity and collaboration which so far frankly has been absent, and on the other for attaining a new point of balance between ambitions and possibilities, between excessive and over-modest expectations, between short-term projects and future opportunities." (Panciera 2012, 273.)

⁷⁴ <https://www.w3.org/DesignIssues/LinkedData.html> (= Berners-Lee 2009) .

⁷⁵ See in particular the examples of the Getty Art & Architecture Thesaurus (<https://www.getty.edu/research/tools/vocabularies/aat/>), the Ariadne portal for archaeology (<https://portal.ariadne-infrastructure.eu/>), or the Pactols thesaurus for archaeology (<https://pactols.frantiq.fr/opentheso/>) (all accessed 2023-02-09).

environment texts can be presented in multiple categories simultaneously, which is both a blessing and a challenge. That might just seem like good practice, and is no different from stating at the outset that one is, for example, using the Leiden conventions, or the AIEGL Greek epigraphic abbreviations. And indeed, what we propose is essentially that, a community of best practice, equipped with the community documentation and tools to make it (relatively) easy to do.

Adopting such an approach in fact already makes it possible, using existing and very standard technology, to aggregate live data from multiple projects in the simplest possible database. The FAIR Epigraphy project, directed by Marietta Horster and Jonathan Prag, has already created a simple pilot to demonstrate this, which pulls together a sample set of material from the available XML files of the Roman Inscriptions of Britain (*RIB*), the *Inscriptions of Greek Cyrenaica (IGCyr)* and *I. Sicily*, transforms them (a process described as ‘serialisation’) into a universal format called Resource Description Framework (RDF) (a standard format for the description and exchange of data on the web), and then makes this publicly available through a simple web interface at <https://inscriptions.org/>.⁷⁶ The framework (known as an ‘ontology’) for doing this was already devised by a collaborative group of epigraphers several years ago, building in part upon the work of the EAGLE project and almost entirely using existing resources.⁷⁷ This took a software engineer (Imran Asif) with no epigraphic background a very few hours to create, and could rapidly and easily be expanded to cover any project that can align its data to existing standards, such as the EAGLE or Getty vocabularies.⁷⁸ As the use of the Trismegistos number illustrates, such a method also directly opens the door to linking multiple corpora simultaneously for the same inscription and so would enable aggregating complementary data from multiple sources for individual inscriptions.⁷⁹ The same principle, inserting e.g. Pleiades geographic identifiers, or *LGPN* person identifiers, as already demonstrated by *IRT* (see above), opens the door to connecting inscriptions to larger datasets about the ancient world, and searches such as “show me all data from x place or about y person”.⁸⁰ This is the potential of

⁷⁶ The FAIR Epigraphy project, funded by the AHRC and the DFG, is intended to support the development and use of standards and best practices to realise the possibilities of ‘FAIR’ epigraphy. The RDF pilot and more information can be found at <https://github.com/FAIR-epigraphy> (Asif *et al.* 2022-2023). On RDF see <https://www.w3.org/RDF/> and, e.g. https://www.w3schools.com/xml/xml_rdf.asp (accessed 2023-02-09).

⁷⁷ See Bodard *et al.* 2021, also online at: https://epigraphy.info/ontologies_wg/ (the current draft in fact created only 3 new terms, otherwise using entirely existing resources: see <https://w3id.org/epont/ontology> (accessed 2023-02-09).

⁷⁸ We would note that this approach has already been adopted, with enormous success, by the numismatic community: see <https://nomisma.org/> (accessed 2023-02-09).

⁷⁹ Such linkage is already enabled by the Trismegistos TexRelations Matcher API ([TexRelations Matcher API Documentation](#), accessed 2023-02-09), for which a simple user interface developed by the FAIR Epigraphy project is now available in beta form at <https://id-resolver.inscriptions.org/> (= Asif & Prag 2023).

⁸⁰ See <https://peripleo.pelagios.org/about>, employing the Pleiades geographic references <https://pleiades.stoa.org> (= Bagnall *et al.* 2023), for an example of compilation of data from multiple resources with reference to ancient places.

Linked Open Data. An RDF data store of this sort can on the one hand act as a long-term and static (stable) data repository of core data from multiple projects; but on the other it can also operate dynamically, incorporating new and revised data from those projects, either through automatic or manual updates, through a very standardised process. Such an approach of course brings its own overheads, but they are much lower than those of a single comprehensive corpus or a traditional database, more responsive to change, and in the true spirit of linked data not only make core data readily accessible and searchable, but provide access for researchers to the individual projects that have studied specific inscriptions in much greater depth than the aggregator can hope to capture.

Any such approach, whether that of the EAGLE project, or that of the FAIR Epigraphy project, is, however, dependent upon community collaboration and respect, and therein lies the challenge - not a technical challenge, but a cultural one. Our concerns parallel those of the first epigraphic congress of 1938:

Le Congrès, rendant hommage aux fondateurs et aux directeurs actuels des Inscriptions Graecae et du Corpus Inscriptionum Latinarum; déplorant qu'il n'y a à présent uniformité ni dans le format ni dans les caractères ni dans la présentation générale des grands recueils d'inscriptions grecques et latines; émet le vœu de voir déterminer pour l'avenir par un statut international les conditions dans lesquelles seront publiés les futurs volumes.⁸¹

This request illustrates how collaboration and interoperability have been core concerns of AIEGL and a focus for congresses, from the beginning; the Association facilitates communication between larger and smaller enterprises, allowing the wider community to benefit from the work of both the subcommunities (e.g. epigraphy.info) and projects (e.g. FAIR Epigraphy) currently dedicated to this challenge. The future of the corpus, and our discipline is surely bright.

⁸¹ This printed statement was circulated (on what basis we do not know) at the 1938 congress in Amsterdam; a copy (our source) is preserved in the papers of W.H. Buckler who attended the conference.

Bibliography

Adamik, B. (2021) : “Romanization and Latinization of the Roman Empire in the light of data in the Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age”, *Journal of Latin Linguistics*, 20, 1, 1-19.

Almas, B., R. Baumann, M.-C. Beaulieu, H. Cayless, J. Cowey, P.M. Liuzzo, F. McCourt and J. Sosin (2017) : “TIGLIO. Translations and Images of Greek and Latin Inscriptions Online”, in : Orlandi *et al.*, ed. 2017, 221-232.

American Psychological Association (2021) : *Webpage on a website references* (<https://apastyle.apa.org/style-grammar-guidelines/references/examples/webpage-website-references>, accessed 09.02.2023).

Asif, I., P. Hermankova, M. Horster and J.R.W. Prag (2022) : *The FAIR Epigraphy Project* (<https://github.com/FAIR-epigraphy>, accessed 09.02.2023).

Asif, I. and J.R.W. Prag (2023) : *Inscriptiones Identifier Resolver (IDR)* (<https://id-resolver.inscriptiones.org/>, accessed 09.02.2023).

Assael, Y., T. Sommerschild, B. Shillingford, M. Bordbar, J. Pavlopoulos, M. Chatzipanagiotou, I. Androutsopoulos, J. Prag and N. de Freitas (2022) : “Restoring and attributing ancient texts using deep neural networks”, *Nature*, 603, 7900, 280-283.

Bagnall, R., R. Talbert and T. Elliott, ed. (2023) : *Pleiades* (<https://pleiades.stoa.org/>, accessed 09.02.2023).

Berlin-Brandenburgischen Akademie der Wissenschaften (n.d.) : *Inscriptiones Graecae* (<http://telota.bbaw.de/ig/>, accessed 09.02.2023).

Berners-Lee, T. (2009) : *Linked Data - Design Issues* (<https://www.w3.org/DesignIssues/LinkedData.html>, accessed 09.02.2023).

Black, M.L. (2016) : “The World Wide Web as Complex Data Set: Expanding the Digital Humanities into the Twentieth Century and Beyond through Internet Research”, *IJHAC*, 10, 1, 95-109.

Bodard, G. (2010) : “EpiDoc: Epigraphic documents in XML for publication and interchange”, in : Feraudi-Gruénais, éd. 2010, 101-118.

Bodard, G. (2021) : *Leiden and EpiDoc Quick Reference (shorter)* (<https://github.com/EpiDoc/Source/blob/6041fccd12a309b689ec4223d52f33df7a1a8278/guidelines/msword/quickref.pdf>, accessed 09.02.2023).

Bodard, G., H. Cayless, C. Cenati, A. Cooley, T. Elliott, S. Evangelisti, A. Felicetti, P. Granados, F. Grieshaber, E. Gruber, A. Hershkowitz, T. Hill, H. Kiiskinen, T. Kollatz, A. Levivier, P. Liuzzo, F. Luciani, A. Mannocci, E. Mataix, F. Murano, O. Murphy, E. Mylonas, J. Prag, V. Razanajao, S. Stoyanova, G. Tsolakis, C. Tupman, I. Vagionakis, V. Vitale and F. Weise (2021) : *Modeling Epigraphy with an Ontology* (<https://zenodo.org/record/4639508>, accessed 09.02.2023).

Bodard, G., S. Stoyanova, T. Elliott, J. Cowey and S. Vanderbilt (2022) : *Dates and locations of modern observations* (<https://epidoc.stoa.org/gl/latest/sup-histlocations.html>, accessed 09.02.2023).

Bodel, J., ed. (2003-2023) : *U.S. Epigraphy Project (USEP)* (<https://usepigraphy.brown.edu/projects/usep/collections/>, accessed 09.02.2023).

— (2012) : “Latin Epigraphy and the IT Revolution”, in : Davies & Wilkes, ed. 2012, 275-296.

Bolikowski, Ł., V. Casarosa, P. Goodale, N. Houssos, P. Manghi and J. Schirrwagen, ed. (2014) : *Theory and Practice of Digital Libraries -- TPDL 2013 Selected Workshops: LCPD 2013, SUEDL 2013, DataCur 2013, Held in Valletta, Malta, September 22-26, 2013. Revised Selected Papers*, Communications in Computer and Information Science, Cham.

British Institute for Libyan and North African Studies (2015) : *Heritage Gazetteer of Libya* (<https://slsgazetteer.org/>, accessed 09.02.2023).

Broyles, P.A. (2020) : “Digital Editions and Version Numbering”, *Digital Humanities Quarterly*, 14, 2 (<http://www.digitalhumanities.org/dhq/vol/14/2/000455/000455.html>).

Bruun, C. and J. Edmondson, ed. (2014) : *The Oxford Handbook of Roman Epigraphy*, Oxford Handbooks, Oxford.

Buonocore, M. (2014) : “Epigraphic Research from its Inception: The Contribution of Manuscripts”, in : Bruun & Edmondson, ed. 2014, 21-41.

Butini, M. (2022) : *EDR167214* (http://www.edr-edr.it/edr_programmi/res_complex_comune.php?do=book&id_nr=EDR167214, accessed 09.02.2023).

Carletti, C., A.E. Felle, V. Ambriola, F. Pallocca and A. Vilella (1988) : *Epigraphic Database Bari. Inscriptions by Christians in Rome (3rd-8th cent. CE)* (<https://www.edb.uniba.it/>, accessed 09.02.2023).

Cayless, H., C. Roueché, T. Elliott and G. Bodard (2009) : “Epigraphy in 2017”, *Digital Humanities Quarterly*, 3, 1 (<http://www.digitalhumanities.org/dhq/vol/3/1/000030/000030.html>).

Chaniotis, A., T. Corsten, D. Feissel, P.-L. Gatier, K. Hallof, M. Hatzopoulos, S. Orlandi, R. Parker, D. Rousset and C. Schuler (2022) : *Liste des abréviations des éditions et ouvrages de référence pour l'épigraphie grecque alphabétique* (<https://aieql.org/grepiabbr.html>, accessed 09.02.2023).

Clauss, M., A. Klob, W.A. Slaby and B. Woitas (n.d.) : *Epigraphische Datenbank Clauss - Slaby (EDCS)* (<http://www.manfredclauss.de/>, accessed 09.02.2023).

Clauss, M., A. Kolb, W.A. Slaby and B. Woitas (n.d.) : *EDCS-19200211* (https://db.edcs.eu/epigr/epi_url.php?s_sprache=en&p_edcs_id=EDCS-19200211, accessed 09.02.2023).

Cooley, A.E. (2012) : *The Cambridge Manual of Latin Epigraphy*, Cambridge.

- Corrado, E.M. (2019) : “Software Preservation: An Introduction to Issues and Challenges”, *Technical Services Quarterly*, 36, 2, 177-189.
- Creamer, A., G. Lembi, E. Mylonas and M. Satlow (2022) : “Archiving a TEI Project FAIRly”, *Journal of the Text Encoding Initiative*, Issue 14 (2021). Online since 08.12.2022 at: <https://doi.org/10.4000/jtei.4324>.
- Davies, J.K. and J.J. Wilkes, ed. (2012) : *Epigraphy and the historical sciences*, Proceedings of the British Academy 177, Oxford.
- De Santis, A. and I. Rossi, ed. (2018) : *Crossing experiences in digital epigraphy: From practice to discipline*, Warsaw, Berlin.
- Depauw, M. (2018) : “Trismegistos: Optimizing Interoperability for Texts from the Ancient World”, in : De Santis & Rossi, ed. 2018, 193-201.
- Depauw, M., ed. (n.d.) : *Trismegistos* (www.trismegistos.org, accessed 09.02.2023).
- Depauw, M. and T. Gheldof (2014) : “Trismegistos. An interdisciplinary Platform for Ancient World Texts and Related Information”, in : Bolikowski *et al.*, ed. 2014, 40-52.
- Devillers, O., M. Navarro Caballero, N. Prévôt and C. Bassion (2012-2018) : *PETRAE. Programme d'Enregistrement, Traitement et Reconnaissance Automatique en Épigraphie* (<https://petrae.huma-num.fr/fr/projet>, accessed 09.02.2023).
- Elliott, T. (2014) : “Epigraphy and Digital Resources”, in : Bruun & Edmondson, ed. 2014, 78-86.
- Elliott, T., G. Bodard and H. Cayless (2006) : *EpiDoc: Epigraphic Documents in TEI XML* (<https://epidoc.stoa.org/>, accessed 09.02.2023).
- Etienne, R., ed. (1984) : *Épigraphie hispanique: problèmes de méthode et d'édition.*, Publications du Centre Pierre Paris 10, Talence, France : Paris.
- Fassbender, A. (2003) : *Index numerorum: ein Findbuch zum Corpus inscriptionum Latinarum*, Corpus inscriptionum Latinarum. Auctarium ; series nova, vol. primum, Berlin.
- Felicetti, A. and F. Murano (2017) : “Scripta manent: a CIDOC CRM semiotic reading of ancient texts”, *International Journal on Digital Libraries*, 18, 4, 263-270.
- Feraudi-Gruénais, F., ed. (2010) : *Latin on stone: epigraphic research and electronic archives*, Roman studies, Lanham.
- Gómez-Pantoja, J. ed. (n.d.) : *Hispania Epigraphica Online (HEpOnl)* (<http://eda-bea.es/>, accessed 30.08.22).
- Hallof, K. (2009) : *Inscriptiones Graecae, editio altera*, Berlin.
- (2012) : *Inscriptiones graecae: imagines epigraphicorum = Epigraphikerbildnisse*, Berlin.

Heidelberg Academy of Sciences (1993-2021) : *Epigraphic Database Heidelberg (EDH)* (<https://edh.ub.uni-heidelberg.de/>, accessed 09.02.2023).

Henriksson, E. and M. Vierros (2022) : *PapyGreek* (<https://papygreek.hum.helsinki.fi/>, accessed 09.02.2023).

Heřmánková, P., V. Kaše and A. Sobotková (2021) : “Inscriptions as data: digital epigraphy in macro-historical perspective”, *Journal of Digital Humanities*, 1 (<http://doi.org/10.1515/JDH-2021-1004?locatt=label:IDHFULL>).

Jory, E.J. and D.W. Moore (1974) : *Corpus Inscriptionum Latinarum. Vol. VII, pars vii. Indices vocabulorum, fascs. 1-6*, Berlin.

Lambert, S., P. Low, P. Liddel and C. De Lisle, ed. (n.d.) : *Attic Inscriptions Online (AIO)* (<https://www.atticinscriptions.com/>, accessed 09.02.2023).

Liuzzo, P.M. (2015) : “EAGLE and EUROPEANA: Architecture Problems for Aggregation and Harmonization”, *Proceedings of the Symposium on Cultural Heritage Markup. Balisage Series on Markup Technologies*, 16 (<http://doi.org/10.4242/BalisageVol16.Liuzzo01>).

Liuzzo, P.M. (2018) : “EAGLE Continued: IDEA. The International Digital Epigraphy Association”, in : De Santis & Rossi, ed. 2018, 216-230.

Liuzzo, P.M., A. Zanni, L. Martinelli, L. Losa et P. De Nicolao (2014) : “The EAGLE Mediawiki. A fully collaborative database for academics, data engineers and the general public”, in : Orlandi *et al.*, ed. 2014, 187-200.

Lynch, C. (2014) : “The Next Generation of Challenges in the Curation of Scholarly Data”, in : J.M. Ray, ed. 2014, 395-408.

Marotta, G. (2023) : *CLaSSES. Corpus for Latin Sociolinguistic Studies on Epigraphic textS* (<http://classes-latin-linguistics.fileli.unipi.it/en>, accessed 09.02.2023).

Merkelbach, R. (1998) : “Nochmals Inscriptiones Graecae”, *Zeitschrift für Papyrologie und Epigraphik*, 122, 293-299.

Mullen, A., ed. (2017-2023) : *LatinNow. Latinization of the north-western Roman provinces: sociolinguistics, epigraphy and archaeology* (<https://latinnow.eu/>, accessed 09.02.2023).

Orlandi, S., ed. (2013-2023) : *The Europeana network of Ancient Greek and Latin Epigraphy (EAGLE)* (<https://www.eagle-network.eu/>, accessed 09.02.2023).

Orlandi, S., V. Casarosa, R. Santucci and P.M. Liuzzo, ed. (2014) : *Information Technologies for Epigraphy and Cultural Heritage*, Rome. (<http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf>).

Orlandi, S., R. Santucci, F. Mambrini and P.M. Liuzzo, ed. (2017) : *Digital and Traditional Epigraphy in Context. Proceedings of the EAGLE 2016 International Conference*, Rome. (<https://doi.org/10.13133/978-88-9377-021-7>).

Packard, D.W. (1968) : *A concordance to Livy*, Cambridge, Mass.

Packard Humanities Institute (n.d.) : *Searchable Greek Inscriptions* (<https://inscriptions.packhum.org/>, accessed 09.02.2023).

Pancierera, S. (1984) : “I Segni diacritici: riflessioni e proposte”, in : Etienne, ed. 1984, 372-379.

— (2006) : *Epigrafi, epigrafia, epigrafisti: scritti vari editi e inediti (1956-2005) con note complementari e indici*, Vetera 16, Roma.

— (2012) : “Epigraphy and Informatics”, in : Davies & Wilkes, ed. 2012, 271-273.

Pancierera, S., G. Camodeca, G. Cecconi, S. Orlandi, L. Fabriani and S. Evangelisti, ed. (1999-2023) : *Epigraphic Database Roma (EDR)* (<http://www.edr-edr.it/>, accessed 09-02-2023).

Pitts, R. (2022) : *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE (CEIPoM)* (<https://reubenjpitts.github.io/Corpus-of-the-Epigraphy-of-the-Italian-Peninsula-in-the-1st-Millennium-BCE/>, accessed 09-02-2023).

Prag, J.R.W., ed. (2017-2023) : *Inscriptions of Sicily (I. Sicily)* (<http://sicily.classics.ox.ac.uk/>, accessed 09.02.2023).

Prag, J.R.W. (2021) : “I. Sicily and Crossreads: a digital epigraphic corpus for ancient Sicily”, in : Prescott *et al.*, ed. 2021, 181-192.

Prescott, C., A. Karivieri, P.B. Campbell, K. Göransson et S. Tusa, ed. (2021) : *Trinacria, “an island outside time”: international archaeology in Sicily*, Oxford.

Ray, J.M. ed. (2014) : *Research data management: practical strategies for information professionals*, Charleston insights in library, archival, and information sciences, West Lafayette, Indiana.

Reynolds, J.M. and J.B. Ward-Perkins (1952) : *The Inscriptions of Roman Tripolitania*, Roma ; London.

Reynolds, J.M., J.B. Ward-Perkins, G. Bodard and C. Roueché (2009) : *Inscriptions of Roman Tripolitania (IRT2009)* (<https://inslib.kcl.ac.uk/irt2009/>, accessed 09.02.2023).

Reynolds, J.M., C. Roueché, G. Bodard and C. Barron (2021) : *Inscriptions of Roman Tripolitania (IRT2021)* (<https://irt2021.inslib.kcl.ac.uk/en/>, accessed 09.02.2023).

Reynolds, J.M., C. Roueché, G. Bodard, I. Vagionakis, C. Barron, F. Bigi, C. Dobias-Lalou, P. Kenrick, R.M. Kerr, M. Mackensen, and I. Tantillo (2022) : *Inscriptions of Roman Tripolitania 2021*. (<https://doi.org/10.18742/C.6002275.V1>).

Rios, F. (2018) : “Incorporating Software Curation into Research Data Management Services: Lessons Learned”, *International Journal of Digital Curation*, 13, 1, 235-247.

Robert, L. (1959) : “Griechische Versinschriften hrsg. von Peek. 1. Grabepigramme”, *Gnomon*, 31, 1-30.

Robert, L. (1965) : “Compte rendu du volume VIII des Monumenta Asiae Minoris Antiqua”, *Hellenica: recueil d'épigraphie de numismatique et d'antiquités grecques*, 13, 9-23.

Robert, L. and J. Robert (1954) : *La Carie: histoire et géographie historique avec le recueil des inscriptions antiques*, Paris.

Rossi, I. (2020) : “Qualitative and quantitative approaches in digital epigraphy”, *Archeologia e Calcolatori*, 31, 2, 145-156.

Roueché, C. and D. Thompson (2023) : “Joyce Reynolds (1918-2022)”, *CUCD Bulletin*, 52.

Saller, R.P. and B.D. Shaw (1984) : “Tombstones and Roman Family Relations in the Principate: Civilians, Soldiers and Slaves”, *Journal of Roman Studies*, 74, 124-156.

Schmidt, M.G. (2007) : *Corpus inscriptionum Latinarum*, 2. durchgesehene und aktualisierte Auflage. Berlin.

Sverkos, E. and Tsolakis, G. (2021) : *Supplementum Epigraphicum Graecum: consolidated concordances for volumes XLVI-LX (1996-2010)*, Leiden.

The Trustees of the British Museum (2023) : *Tomb relief; block* (https://www.britishmuseum.org/collection/object/G_1867-0508-55, accessed 09-02-2023).

Tsolakis, G. (2021) : “Epigraphic abbreviations”, in : Sverkos & Tsolakis, ed. 2021, XV-CXXVI.

Velázquez, I. and D. Espinosa, ed. (2021) : *Epigraphy in the digital age: opportunities and challenges in the recording, analysis and dissemination of inscriptions*, Archaeopress archaeology, Oxford.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, B. Mons, *et al.* (2016) : “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, 3, 1, 1-9.