Conclusions

The contributions collected in this volume, in particular those regarding the “marginal” epigraphies, bring to attention a considerable variety of themes, even beyond those initially envisaged while conceiving the volume. Even though a systematic vision and approach in digital epigraphy is still very far off, a summary of the different issues and positions highlights the common trends.

Albeit text and text-bearing object cannot be separated in the study of epigraphy, it is apparent that the text is the main focus in digitization. The recurring objective of epigraphic projects is making strings of inscriptional characters searchable, which may or may not include lacunae, integrations, variants and corrections. Presently, the most common practice is the transcription/transliteration and the XML encoding of different kinds of phenomena (structural, concerning the relationship with support, textual portions, transcription phenomena, editorial interventions, in-line *apparatus criticus*, PoS and morphological analysis, onomastics, etc.), depending on the topics and objectives of the projects. EpiDoc, a subset of elements of the TEI standard, is widely used both as archiving and exchange format. The agreement of the scientific community regarding well-defined best practices is an important achievement. Nevertheless, this method is not suitable for all the epigraphic materials, depending on the writing systems used and on the current degree of knowledge of the scripts and languages attested.

The contributions on the Maya, Linear B and cuneiform scripts witness alternative solutions with respect to the XML encoding of the texts, in relation to logo-syllabic writing systems. The *Sinleqiunnini* project (Di Filippo) shows the application of the relational model to the texts themselves, according to the “ordered hierarchy of content objects” theoretical framework. The limits of XML encoding in the annotations of not-contiguous portions of text, and in the management of overlapping hierarchies, are thus overcome. As texts are the sum of the instances of hierarchical entities, conflicting interpretations of the phonetic, morphological and semantic values can be recorded.

The *Text Database and Dictionary of Classic Mayan* (Prager et al.), on the other hand, is implementing a sign catalogue, in addition to the corpus of TEI encoded texts, which identifies each graph and relates it to its allographs using a propositional logic. The semantic modelling (CIDOC CMR and RDF encoding) allows for different, duly argued, readings.

The *RuneS* project (Zimmermann, Kezzazi, & Bahr) has a further approach. The different systems of Runic script are the focus of the project: the philological study of the graphic variants of signs is carried out through the annotation of the visual documentation, photographs in particular, which leads to the creation of a catalogue of signs. Such an approach, even though it does not allow the textual search, broadens the possibility of the palaeographic research, and is currently supported by the spread of standards for the interoperability of images, such as IIIF.

Even though the Palaeohispanic inscriptions of the *Hesperia* project (Estarán et al.) are transliterated in Latin characters, several phonetic interpretations, deriving from different editions of the epigraph, can be attributed to each graph so that the search engine retrieves all variants.

Those experiences attest to the diverse approaches to the structuring and archiving of textual data, depending on the characteristics of the epigraphic material and the objectives of the digitization project.

The practice of storing textual and extra-textual data in relational databases is indeed more widespread than the establishment of repositories of XML files for the same purpose. This is the case even for collections of texts written in fully deciphered scripts, whose transliteration and interpretation is plain. The modelling of discrete entities entails the independent descriptions of the various aspects of the inscriptional document, leading to the enrichment of information on the physical carriers – thus stimulating studies in the fields of material culture, iconography and history of art – or on their places of origin and provenance (Xella & Zamora). The implementation of contextual information, as discussed further on, is the foundation for establishing connections with archaeological datasets, and developing tools that support interoperability, such as chronological and geographical gazetteers.

Epigraphs are often the only sources for the study of languages whose fragmentary attestation currently hinders their complete comprehension. The digitization of textual corpora provides a host of data to be processed with the support of technology. However, lexicographic developments vary considerably, depending on languages and digitizing methods. In general, an automatic approach is preferred when very large digitized corpora are available.

NLP techniques are fruitful when applied to modern, living languages, which benefit from the mass of data collected from the web, as well as to Greek and Latin among the languages of antiquity. Greek and Latin have a considerable volume of digitized texts, which also include non-epigraphic documents such as papyri and manuscripts. Moreover, the availability of reference tools such as dictionaries and grammars, allows performing the automatic parsing of texts, PoS and morphological analysis, by matching each word with a fixed set of possible lemmata and word-formation rules.

On the contrary, a “manual” approach is preferred by projects focusing on languages, which are fragmentarily attested and not provided with those tools (the so-called “under-resourced languages”), or languages whose signs’ value or signs’ sequence interpretation is multiple. Some of them encode grammatical phenomena in the texts. Morphological analysis is carried out for each lexeme, which is then connected to a lemma and, in the case of Semitic languages, to a root as exemplified by the papers on the OIMEA (Novotny & Radner), OCIANA (Burt, Al-Jallad, & Macdonald) and Sabäisches Wörterbuch (Multhoff) projects. The strong repetitiveness of texts and, at the same time, the morphological ambiguity of the languages that make the direct encoding burdensome but still uncertain, may otherwise suggest the creation

of independent, lexical entries to be linked to occurrences in the texts, as in the DASI lexicon (Avanzini, De Santis, & Rossi).

KALAM (Ruzicka) is the first attempt at conducting an automatic detection of morphological attributes in pre-Islamic Arabian texts, on the basis of the generative grammar and with a synthetical approach. In any case, the lexical tools that are being developed will increasingly improve the linguistic knowledge and the available dataset, thereby enabling an automatic approach for these languages.

Given the specific needs and the differences so far described, we must underline a patent, yet still somehow underestimated issue: the boundaries of disciplines and methods should not affect the study of the societies of the past. A thorough knowledge is reached via a comprehensive approach, in which interoperability and open access to data play an essential role.

The contribution by Liuzzo clearly exemplifies the challenges to be faced when aggregating a huge amount of epigraphic records, although quite homogenous from the historical and the linguistic points of view. A complex harmonization of the tributary vocabularies has been necessary in the EAGLE project. If EpiDoc remains the most suitable schema for the interchanging of epigraphic records (also by virtue of its mapping to RDF), the exploitation of established thesauri and ontologies is a key to fostering interoperability – as exemplified by the case of the research platform EPIDAT for Jewish epigraphy (Kollatz). These tools force us to face the issue of matching the taxonomies that make sense in each domain. As for the encoding aspect, even if the EAGLE schema is based on the EpiDoc standard, the automatic up-conversion of the records provided by partners, has sometimes produced unsatisfactory results due to the personal and varied use of the EpiDoc elements by different providers.

It is right to ask ourselves whether the best practice is to conform to strict common operational guidelines, or to maintain diversity within a shared general framework, enacting mapping strategies when needed. In fact, the discrepancy in the choice of the phenomena to annotate, and the encoding solutions (though in the frame of the same guidelines), are usually objective-driven and therefore answer to different scientific questions. This is especially apparent when dealing with several editions for the same text across distinct repositories.

To preserve this wealth of information, and at the same time offering a single point of access to all the digital editions of the texts while disambiguating them, comprehensive indexes of texts – harvesting the existing repositories – are desirable. The project Trismegistos (Depaw), moves towards this objective by providing a unique ID for each textual document, and is now expanding its scope beyond the indexing of texts to the indexing of people and places mentioned in those sources, from a LOD perspective.

Encoding of onomastics is carried out by almost all of the projects represented in this volume, in some projects in addition to named entity recognition. The first successful attempts at connecting information and producing historical knowledge, through prosopographical authority files and geographical gazetteers, are very

promising. The identification of persons mentioned in the texts, the reconstruction of lineages, and the network analysis will enhance cross-referencing.

The general cooperation with initiatives like Pelagios and PeriodO (see Rabinowitz, Shaw, & Golden), and the references to their gazetteers of places and periods, would reduce the obstacles posed by the naming of geographical entities, which are continuously changing in relation to the historical context, and the dating systems used by different socio-political entities in the sources, as well as by different traditions of study.

As regards the interoperability of multilingual textual corpora, epigraphs witness, perhaps better than other sources, the cohabitation of different languages and scripts in the same periods and in the same regions (see the example of the same inscription attested in different languages in Bausi & Liuzzo), or the use of a language or script in very distant geographic, socio-political or historical contexts. Is it possible and worthwhile to imagine a simultaneous search on texts in different languages, and on concepts, across different corpora? The exploitation of lexical semantics and translations deserves deeper attention by the community of epigraphers in order to enjoy their full potential in terms of linguistic and cross-cultural research. Translations are not envisaged in many digitization projects. Indeed, their recording requires us to choose whether to provide a literal translation, or to include periphrases and metaphors, and then require conformity to this choice across the overall corpus. The standardization of translations within each project should be granted, and a strict relation between segments of texts and their translations should be envisaged, if not a proper encoding of the translations. Moreover, attempts at identifying and mapping common concepts in different corpora, at least those linguistically and culturally close, might produce unexpected results.

Operating at a translation level would allow scholars not familiar with a particular language to consider sources potentially useful for historical and cultural studies. Furthermore, as well as valorising the physicality of the inscriptions and their iconographic apparatus, translations would also make epigraphy more interesting for, and accessible by, a wider public. Projects as *I.Sicily* (Prag & Chartrand) are stimulating sensitivity towards the preservation and appreciation of the peculiar epigraphic historical witness, by operating within local cultural or educational institutions and by envisaging digital tools for the cross-cultural fruition of the sources within a defined geographic context.

The commitment to the digital preservation of the epigraphic heritage through a thorough, systematic collection of photographs is at the centre of many projects, acknowledging the importance of autopsy in the study of primary material sources, as exemplified by the *Karnak* project (Biston-Moulin & Thiers). Free access to and reuse of images becomes, thus, a key issue for the progress in epigraphic research, feeding the ongoing debate on the need of softening – if not removing – copyright obstacles by museums, archives, libraries, and even archaeological sites. This is

especially needed in contexts at risk of deterioration or destruction by human and/or environmental factors.

Last but not least, one major challenge emerging from the contributions in this volume is the sustainability of projects and their results.

This is probably one of the main factors that most discourages the creation of critical editions in digital format and the commitment to their implementation. Many scholars continue to prefer paper editions, as the editorial criteria of many projects do not allow the attribution of contributions and interventions to each individual researcher, thus impeding the proper citation of digital resources, and therefore the evaluation of his/her scientific activity. Moreover, traditional publications are still the preferred medium for a detailed communication to, and easier fruition by, the scientific community of the research's results (as stated by Cannata, and Biston-Moulin & Thiers), thus receiving a better appreciation by the specialized audience.

Furthermore the general fragility and volatility of digital knowledge threatens the total disappearance of scholars' scientific work in the medium term. The earliest, pioneering projects have not always had the opportunity to upgrade their technical infrastructure. The rapid, technical obsolescence, despite the rigorous scientific methodology, endangers access to their data.

Sooner or later, every digital epigraphy "venture" has to face this problem, due to the lack of permanent funding. Some projects are included in larger portals, like the one hosting and supporting Hittitology projects described in the paper by Müller & Schwemer. Some take the legal form of associations and foundations, and benefit from the support provided by their infrastructures: the IDEA initiative described by Liuzzo is the first attempt at exploring this path in the digital epigraphy domain. More often, engaging with new initiatives is the only viable solution (when achieved) to maintain previous (well established, but no more attractive) ones. Openness and reuse is the suggested best practice to circulate, multiply and save, at least, data.

Indeed, the definition of a common, theoretical and methodological framework that will make this research effective is only possible if we take into account the variety of questions, problems, approaches and solutions dealt with by the widest community of epigraphers. This volume was an attempt at gathering, virtually, at a "round table" some of those who practice digital epigraphy in very different cultural domains and with different scopes, and who have been engaged in this kind of research activity over different periods of time. They have shared their experiences to stimulate discussion of some of the main themes that are driving digital epigraphy forward in the future, while also shaping our approach to traditional epigraphy.