Le livre [sc. an electronic edition produced from a textual database] n'est plus qu'un produit secondaire. Ce qui compte vraiment, c'est le fichier, dont n'importe quel chercheur peut extraire le renseignement précis dont il a besoin.

Étienne Evrard

Pruned of the thick crown of metadata, which we have dealt with so far, at the core of Papyrology there are the texts: it is therefore an easy equation to state that at the core of Digital Papyrology there is the textual data. As we saw in the previous chapter, the first steps of the application of electronic calculators to papyrological research were moved in the sphere of the textual databanks, and the very first developments of a dawning Digital Papyrology were envisaged in terms of digital treatment of the texts. Subsequently came the rest. And since texts are now again in the spotlight, after fruitful seasons of flourishing metadata platforms (now fully functional and run) and imaging techniques, it is appropriate to conclude our survey precisely with a focus on this complex and fascinating universe, which recovers and develops what is probably the oldest dream of Papyrology: the creation of a complete directory of papyrus documents for reference and search, that is – essentially – for comparison.

8.1 Digital Encoding of Papyrus Texts: Theory and Practice

“Computational systems depend on resolving ‘real world’ situations into exact numerical strings”. This means that when digitising texts (namely Greek papyrus texts, in our case) one must take care of designing a code that translates the textual information into machine-readable conventions, i.e. “input conventions”. As already defined by R.J. Glickman, these must be designed for two types of data: the text proper and “identification labels”, i.e. all paratextual information, which differ slightly from what we called metadata in that the latter bear information about the papyrus as a material object and written artefact, while paratext bears information about the very essence of the text itself and of its linguistic / communicative con-

---

1 Evrard 1970, 125.
3 Terras 2010, 50.
4 On data input and text encoding see in general Laue 2004 and Rener 2004.
tents. Paratext\textsuperscript{6} is essentially the set of critical, diacritical, and punctuation signs that comes along with the text itself, whether in ancient or in modern times, being indispensable for its correct understanding\textsuperscript{7}, i.e. human decoding.

Encoding text poses some issues when dealing with non-Latin characters, especially if the electronic environment is designed to support a limited set of alphabetical signs and some other common symbols, as it was the case at the beginning of the digital era. Encoding paratext poses the same order of issues when dealing with specific conventional marks. Such issues – “the contemporary tower of Babel” in communication between men and machines, as it has been defined\textsuperscript{8} – are usually faced with the establishment of conventions, which can be more or less universally adopted but must be strictly followed in order to comply with the technological requirements: basically, they must be as standardized as possible. This is not that different from what happens in a printed transcription or edition, where typographical conventions are used to represent papyrological texts and paratexts, and one must follow the Leiden editorial conventions as well as the standard guidelines of journals, series, or publishers. Once data is encoded, it has to be converted again in human-readable format, so that the researcher can benefit of the output obtained from the computer. Operations of reversal encoding, or decoding, are therefore necessary as well.

Encoding is normally effected through input from keyboard. If the computer does not support Greek characters, a ‘transliteration’ is required: the input is encoded in plain ASCII characters, and then the computer renders it in (polytonic) Greek glyphs. This happened with the earliest work (Alpha and Beta Code, see below, § 8.3) as well as through more recent times, until Unicode was introduced (see above, § 1.2). LaserGreek, SuperGreek, and similar custom typefaces allowed indeed for encoding Greek with particular combinations of ASCII alphanumerical keys; the resulting text was displayed in Greek characters thanks to specific typefaces. Codes were different from font to font, and from platform to platform, generating a great mess: the text was in fact encoded and processed in ASCII, so that if one did not possess the exact typeface with which it had been encoded, (s)he was not able to display it correctly – not to speak of display issues in many web browsers. The introduction of Unicode (see above, §1.2) solved most of the problems: input was now effected directly in the Unicode Greek subset codes through appropriate keyboard layouts\textsuperscript{9}, ensuring cross-

\textsuperscript{6} Terminology refers to G. Genette’s textual theory (cf. Genette 1992, 83–4, as later developed in Genette 1997, 1–7).

\textsuperscript{7} Actually, with “identification labels” Glickman referred to the contextual metadata; I retain a narrower meaning, which suits my argument.

\textsuperscript{8} Melagakis 1996.

\textsuperscript{9} One of the most common is perhaps GreekKeys, owned and distributed by the American Philological Association (editor Donald Mastronarde). Its latest release (GreekKeys 2015) is free for APA members (https://classicalstudies.org/publications-and-research/about-greekkeys-2015; previous
platform compatibility. Unicode is in fact a great means of standardization, and its use is highly recommended, though unfortunately not universally adopted even today. Since some peculiar papyrological symbols were not included in the official Unicode release, some special Unicode font faces have been designed to support those missing glyphs in their own “private user area” (a special code range reserved for third-party personalizations) as well as nicer display of combining diacritics like the underdot. The most used are likely New Athena Unicode, developed by Donald Mastronarde and distributed by the American Philological Association (https://apagreekkeys.org/NAUdownload.html), and IFAO Grec Unicode, developed by the Institut Français d’Archéologie Orientale (http://www.ifao.egnet.net/publications/publier/outsils-ed/polices). Various utilities to convert legacy non-Unicode Greek text into Unicode have been released, and they prove very useful to recover old files, especially for the sake of entering papyrus texts in the textual databank.

---

10 On the architecture and the mechanics of character coding systems see Melagrakis 1996 and Macrakis 1996 (with an excursus on the encoding issues of ancient Greek alphabets, including diacriticals).

11 Though most of the Unicode diacritics for ancient Greek come as precombined characters (i.e. the letter + its diacritics form one single character), some of them are just appended to the letter as separated, combined (or “floating”) characters. The underdot belongs to the latter group. On this issue cf. Macrakis 1996, 278–9.

12 See below. For example, among the newest tools we can mention the open-source Theuth developed by Daniele Fusi, a Word addin that helps typing in Unicode and converting any custom-encoded polytonic Greek text into Unicode (http://fusisoft.azurewebsites.net/software/theuth), and Greek Transcoder, by David-Artur Daix, a Visual Basic routine (Macro) embedded in a Word template that offers a complete range of transcoding options (http://www.greektranscoder.org). One has just to open the template, paste the polytonic Greek text, run the Macro, choose the appropriate typefaces, and launch the routine.

---

7-bit ASCII chart & SuperGreek ASCII chart.

releases: http://apagreekkeys.org). Other options are described at https://wiki.digitalclassicist.org/Greek_Keyboard_) (Unicode).

---

See below. For example, among the newest tools we can mention the open-source Theuth developed by Daniele Fusi, a Word addin that helps typing in Unicode and converting any custom-encoded polytonic Greek text into Unicode (http://fusisoft.azurewebsites.net/software/theuth), and Greek Transcoder, by David-Artur Daix, a Visual Basic routine (Macro) embedded in a Word template that offers a complete range of transcoding options (http://www.greektranscoder.org). One has just to open the template, paste the polytonic Greek text, run the Macro, choose the appropriate typefaces, and launch the routine.
But we would see only half of the heaven (or of the hell…) if we did not take into consideration the encoding supports. As papyrologists are accustomed to deal with writing supports, their transformation over time and their intimate relation with the text they bear, digital papyrologists must cope with digital supports and with their intimate relation with the digital text they host. Digital texts are highly support- and device-dependent, and their ultimate usability lays on the usability of supports and devices. The first supports were punched cards, pieces of stiff paper in which the information was represented, in a binary way, by the presence or absence of small holes in predefined positions, which calculators could interpret and process (see above, § 4.4). These were huge machines and the intermediation of specialized technicians was indispensable (see above, § 7.1): we may define them the ‘cuneiform tablets’ of Digital Papyrology. It is probably for this reason that for the first systematic and comprehensive databanks we had to wait for the introduction of magnetic tapes, which made things slightly easier – they could be distributed quite easily, but still required tape drives to be transferred to large computers –, and the development of micro-computers, from the mid Seventies, which enhanced the personal use of the electronic resources (see above and below, §§ 7.1 and 8.3). At first, the textual databanks (namely the Duke Databank of Documentary Papyri, see below, § 8.3) were mostly relegated to a specific computer system, the Ibycus designed by David Packard to store and manage the Thesaurus Linguae Graecae directly in Greek characters (see below, § 8.3) – we can compare this phase to ‘hieroglyphic rolls’, slightly handier than punched cards (our metaphorical incised tablets) but still reserved to a narrow elite of experts. The quick diffusion of many different computer systems urged the developers to grant the conversion of proprietary Ibycus code into a universal code that could be read also by machines that could not manage Greek characters. Betacode, developed by Packard himself in the late Seventies as a way of representing Greek characters and diacriticals by means of Latin characters and special marks (see
below, § 8.3), became the first step towards universalism and standardization\(^{13}\) – a modern parallel to ancient ‘international Aramaic’, which became a sort of Mediterranean interlanguage also thanks to the ease of use of its alphabetic writing. The launch of the CD-ROM – a direct access memory device, sensibly smaller, more capacious, and faster than magnetic tapes, which were sequential access memory devices – facilitated the diffusion of the database, and at the same time the flourishing of special software programs designed to support the database on non-Ibycus systems\(^{14}\). The situation might now be compared, approximately, to classical antiquity: a widespread diffusion of Greek as ‘universal’ language declined in several handwritings, and a variety of writing supports with different characteristics and uses. Then, in 1991 Internet, the hypertextual revolution that completely changed the way of structuring and displaying data, saw daylight. This brought new ways of thinking the text in relation to the support, within the framework concept of tabularity that can be easily compared with ancient codex-format books\(^{15}\).

The conversion of Digital Papyrology to the World Wide Web was not exactly straightforward: many digital projects of those years started, or planned to start, on CD-ROM, but soon after turned to the Internet (see above, § 1.2). Anyway, within less than ten years (1997) the Duke Databank was fully working online in the Perseus hypertextual framework (see below, § 8.3). It took ten more years (2006/7), with the IDP project and Papyri.info (see below, § 8.4), to fully understand and exploit the real potentials of online databanking: open access, integration of resources, worldwide collaboration – after some smaller, pioneering projects already exploited some or all of these potentials. It must be clear that what we are now dealing with is not a new encoding support: it is a completely different way of structuring and conceiving data, that is texts. But we will face this last challenge in the next chapter.

\(^{13}\) Cf. W ILLIS 1988, 16. A strong claim to device-independency (quite interesting also because of its early appearance), especially from Ibycus systems, was made by O RE 1988, 28–9 (who also announced a translation of Kleve’s applications for Lacunology and Literalogy in such portable languages as Pascal and Modula-2: ORE 1988, 27–8).


\(^{15}\) See above, § 1.1. It is not for fun that I ceded to the folkloric and fashionable comparison between ancient and modern media. The relationship between medium and text is a constant factor in human cultural development (cf. B OLTER 1993, 160), and text transmission issues tend to reproduce themselves: “[i]t seems therefore reasonable to argue that we have returned to a situation somewhat like the one that existed in the ancient world and furthermore that perhaps some of the processes that governed the survival of ancient works might pertain to digital media” (C AYLESS 2010, 147). Also, computer power and storage cannot be underestimated as factors influencing the digital treatment of ancient texts (cf. C RANE 2004, 50 ff.) just as ancient roll and codex capacity determined many features of analogue text processing.
8.2 The Earliest Textual Databases

As anticipated above (§ 7.1), the first attempts to store Greek papyrus texts digitally were made in the Sixties at the LASLA in Liège under the guidance of Alfred Tomsin16, and at the Centro di Documentazione Automatica in Milan founded by Enrico Maretti, under the guidance of Mariangela Vandoni17. Both teams used punched cards to store the texts in electronic calculators (see above, § 4.4). However, though the support was the same, each team chose to use a different type of encoding for the Greek texts. The choices were made on the ground of the different research purposes.

The Belgians were eventually interested, alongside research features, in the indexing functions18 and in the printed output of the data19, and therefore decided to encode directly Greek characters along with spirits, accents20, iota subscript, diaeresis, underdots, capital letters21, so that they needed to modify existing IBM machines22, since they normally did not support Greek direct input. They encoded one word per card, along with reference codes (end of sentence, end of line)23. What is striking, in the text processing, is the high flexibility of the method: later emendations were added at a second stage, so that it was possible to reproduce both the emended version and the original editio princeps, alongside the diplomatic transcription24; moreover, as already noted earlier, a strong linguistic focus produced a deep ‘annotation’ of lemmatized, morpho-syntactic and semantic information for each word. Such a system was able to accept most of the editorial symbols as in the Leiden conventions, but some had to be changed for technical reasons (see the summary table below)25. In fact, when they presented their method at the 12th Inter-
national Congress of Papyrology (Ann Arbor 1968), they distributed a print sample of their work, consisting of a significant selection of papyrus editions printed from their database in the different possible formats, with indices and concordances generated automatically\textsuperscript{26}. This Choix de papyrus documentaires – Essai de traitement automatique opens with a prophetic introduction by Louis Delatte, the founder and director of LASLA, the concluding paragraph of which is perhaps worth to be quoted entirely for its historical and methodological significance:

> si chaque centre de papyrologie disposait d’un système 870 [i.e. the IBM machine used at LASLA] et si un centre général doté d’un petit ordinateur regroupait tous les fichiers, il ne serait pas ridicule d’imaginer qu’en quelques années, un corpus général de papyrologie pourrait être constitué et exploité\textsuperscript{27}.

It must be stressed that Delatte did not limit himself to prophesize a possible future database of all papyrological texts, but envisaged it in the same terms of worldwide collaboration that mostly inform Digital Papyrology today.

The Italians, on their side, aimed at providing the researchers with a work instrument easy and convenient, able to perform any possible search and based on a text that should be as close to the original as possible. For this reason, they encoded the texts in Latin transliteration, in capital letters, without diacriticals, and without any modern emendation/correction\textsuperscript{28}. Special codes were assigned to proper names and numerals (whether in figures, spelled out or fractional), as well as text sections (paragraph beginning, initial or final break, line end with or without word break, column beginning, verso, end of the papyrus)\textsuperscript{29}. A peculiar solution was adopted to indicate scribal corrections: \( * C p = q * C \), where \( p \) is the original word and \( q \) the corrected one\textsuperscript{30}. For other conventions see the summary table below. The experiment comprised a sample corpus of 50 Milan papyri and a program called ARSINOE, which was divided into five subroutines: ARSINOE 1 converted the machine-readable codes into human-readable format (e.g. adding line numbers, resolving abbreviations...) and checked possible ‘syntax’ mistakes in the perforation of the

\textsuperscript{26} BINGEN – TOMSIN – BODSON – DENOOZ – DUPONT – EVRARD 1968; cf. BINGEN 1968b, 379–80. Some conventions are still provisional in this volume, because it was printed quickly before the congress, as explained by BINGEN 1968b, 379. It is the case, in particular, with the numbers indicating illegible/lost characters.

\textsuperscript{27} DELATTE 1968. TOMSIN 1970a, 476 wished a progress from single test-cases to all texts as well.

\textsuperscript{28} Cf. ZARRI 1967, 57–9. The numerals koppa and sampi were transliterated with Q and V, stigma with the Arabic numeral 6.

\textsuperscript{29} Cf. ZARRI 1967, 71–2.

\textsuperscript{30} This, by the way, is a nice antecedent of modern ‘regularization’/’correction’ tags in Leiden+ (see below, § 8.5).
cards (e.g., a lacuna opened but lacking its closing code\textsuperscript{31}); ARSINOE 2 aimed at searching and extracting words for building indexes/concordances; ARSINOE 3 was devoted to the search for spelling variants; ARSINOE 4 and 5 performed searches for complex formulas\textsuperscript{32}.

Subsequently, Tomsin decided to apply the same automated method as used for Choix to the study of a dossier of papyrological documents related to the imperial estates in Egypt (\textit{ousiai}), envisaging a true quantitative analysis of the digital texts through partially automated lemmatization and semantic annotation, which produced special thematic indices (geographical, prosopographical, topographical, etc.)\textsuperscript{33}. It is a clear example of the potentials of a deeply annotated corpus: essentially, text and metadata integrated in a single, powerful database.

Of course, as Delatte himself noticed, punching cards to encode the basic data of all published papyri would have required an enormous work by several papyrological teams worldwide; but it is a real pity that when input methods became less complicated papyrus texts were digitized without annotations. This was still a priority when a project for digitizing all published documentary papyri from Oxyrhynchus was launched at Oxford in the late Seventies. Presented at the 16\textsuperscript{th} International Congress of Papyrology (New York 1980), the \textit{Oxyrhynchus Computer Project}\textsuperscript{34}, led by Peter Parsons, intended to encode texts along with quite detailed metadata: volume and edition number, of course, but also reference to photographs, bibliographical information, type of document, relevant subjects, date, current location, information on the other side, physical information. Texts were encoded in continuous transliteration (with possibility of converting to formatted Greek) and without editorial corrections. Further deep annotation stages were planned:

\begin{quote}
\begin{itemize}
  \item [n]umerals will need to be marked off in these texts to avoid their useless incorporation in word lists, as will words within restorations, affecting these same word lists and statistics based on them. Personal and geographical names, and the names of months, require distinction. Other specialized matters remain for the future, the largest being comprehensive lemmatization; marking off of homographs; analysis of syntactic function; and incorporation of scholarly conjectures and corrected orthography\textsuperscript{35}.
\end{itemize}
\end{quote}

Unfortunately, this remained a wish, as the next generation of textual database would be slightly different in shape and scope.

\textsuperscript{31} Another nice antecedent of a modern tool: the automatic validating check of the \textit{Papyrological Editor} (see below, § 8.5).


\textsuperscript{33} TOMSIN – DENOOZ 1974. The technique is described in details by TOMSIN 1970c; cf. also TOMSIN 1970b, 63.

\textsuperscript{34} Cf. KEFFE 1981.

\textsuperscript{35} KEFFE 1981, 684.
8.3 The *Thesaurus Linguae Graecae* and the *Duke Databank of Documentary Papyri*

1972 is the year of the *Thesaurus Linguae Graecae* (TLG), the tool that revolutionized the entire field of classical studies. The goal of the project, started at the University of California-Irvine under the direction of Theodore F. Brunner and funding by Marianne McDonald, was essentially to create a digital corpus of Greek literary texts from Homer to AD 600, later expanded to a selection of Byzantine works. Much has been written on TLG and it would be odd repeating everything here: but the event has been so paramount to deserve some notes. In particular, I would like to focus on some key features of the databank, and on the question whether it can be considered a ‘papyrological’ tool. Since the latter is a significant methodological and epistemological point, I will start with it.

In fact, the question has already been asked by Brunner himself, in the context of the 20th International Congress of Papyrology at Copenhagen (1992), and his authoritative answer was positive:

Ἐν ἀρχῇ ἦν ὁ λόγος – without an understanding of the words that it carries, any piece of papyrus would be merely a meaningless scrap of desiccated organic matter. One of the primary purposes of the *Thesaurus Linguae Graecae* is to provide us with a better understanding of text – of the use of words, and the relationship between words. Electronic data banks are superbly suited to facilitate the attainment of such understanding. Philologists, historians, linguists, theologians, and other non-papyrologists have been using the TLG’s resources for quite some time as a means to analyze and understand their raw material – their texts. In fact, the past few years have witnessed almost a quantum jump in the quantity and quality of scholarly publication in direct consequence of the availability of the TLG resource. Moreover, the availability of large text corpora in electronic (and thus easily accessible) form has stimulated research in areas long neglected simply because the raw materials were too extensive to be dealt with by means of traditional methodologies: 2.5 million words of Galen, or 4.5 million words of John Chrysostom suddenly no longer seem quite as formidable, now that they are accessible via a computer search consuming but a few minutes. Papyrologists can reap equal benefit from using the TLG as one of their basic research tools.

He gave some practical examples of unidentified literary fragments assigned to known authors thanks to the comparison with TLG texts, and of previously mismatched pieces then correctly reassigned. It is therefore in such “computer-
Früchte”39 – a neologism created by the same Brunner after the typical German expression *Lesefrüchte* – that we must see the main reason for which TLG “can be used as a papyrological tool”40: we have already presented the enthusiastic report by William Willis on its fruitful papyrological exploitation, and other similar cases can easily be evoked as well41. And since TLG is basically a philological resource (the encoded texts belong to literary material), it is really to be seen – again in Brunner’s own words – as a “unifying force”, a meeting point that can enable scholars from different fields to work together42, in view of that cross-disciplinary utopia that probably only digital tools, in their almost infinite potentials, can ensure (see above, § 3.3, the case of *Trismegistos*, expanding from Papyrology to more universal horizons). Moreover, we should not forget that TLG does in fact include papyrological texts, if they are the sole testimonies of literary authors43.

and enhanced by William Johnson to add features suitable to papyrological phenomena. Moreover, in BRUNNER 1988 he presented the case of P.Köln I 25 fr. k-o, previously assigned to *Iliad* II on the ground of the matches of the other fragments of the same papyrus, but actually not containing text strings significantly compatible with that book. See also BRUNNER 1986 and 1993b.

40 BRUNNER 1994, 606.
41 Cf. WILLIS 1984a (see above); FORTUNA – BINDI – BOZZI 1987 (at pp. 198–203, discussion of a software designed to perform automatic comparisons between fragmentary papyri and the literary database); BOUQUIAUX-SIMON 1991; 1992 (for the TLG at Liège see also MARGANNE 2007a, 16); HANSON 1997, 300–4 (on medical fragments); HANSON 2002, 196 (apropos of Louise C. Youtie’s work on the *Michigan Medical Codex*); RENNER 2009, 290–1.
43 “In 1976 [...] a growing concern with papyrus texts preoccupied the advisory committee in New York. Earlier, under the guidance of past APA president William Willis, who had been added to the membership in 1974, the committee had concurred ‘that literary papyri should be treated differently from documentary papyri, that they should be accepted as texts of ancient authors and handled together with other texts of Greek authors, and that the work of entering such texts into the data bank should be taken independently of any arrangements for documentary papyri.’ [“Minutes of the Irvine meeting of the APA Advisory Committee on the TLG (March 29–30, 1974)""] Following this advice, the TLG added numerous Greek papyrus texts to the data bank, although only edited texts would be represented, with diplomatic texts ignored altogether. [...] In order to avoid excessive duplication in the contents of the data bank, we would have to refine our thinking about papyrus texts: a distinction would have to be drawn between authors whose writings are preserved, either completely or in part, by codices, and authors whose known fragments derive entirely from papyrus and from quotations. Papyrus fragments of text supported by a manuscript tradition should be regarded as ‘alternative manuscripts,’ [quoted from “a statement by Pearson entitled ‘Fragments of Greek Authors in the Word Bank’, distributed to members of the advisory committee”, p. 1] to be treated as all other manuscripts in an apparatus criticus. Inasmuch as solutions for successful data entry of an apparatus criticus had not yet been developed, such papyrus texts were removed from the TLG’s immediate consideration. But the numerous papyrus texts that had no separate manuscript tradition would have to be represented in the data bank, and ‘generally the text to be followed will not be that of the original publication in a papyrus collection, but a critical edition of the author’s fragments.’ [“Pearson, ‘Fragments’, 2”]" (BERKOWITZ 1993, 45–6).
The technical details of TLG are of the utmost importance to understand the subsequent development of the biggest papyrological textual databank. The papyrological outcome of TLG is essentially literary, since it collects literary texts, while all papyrological digital resources before and for several years after it where devoted to ‘documentary’ data. It has been probably a matter of perspective that made the literary side of the texts prevail on the papyrological one for technical reasons, though Brunner himself expressed his wish that the gap could be filled and that “one of the Ds in DDBDP will disappear, and that a Duke Data Bank of Papyri will ultimately contain literary fragments and documents alike”.

The issues of the literary texts stored in TLG are well known and it is sufficient to enumerate them very quickly: the selection of one canonical edition and the absence of an apparatus criticus make it a powerful lexicographical tool, but not exactly an exhaustive scholarly reference. It can therefore be utilized with success, but being aware of the unavoidable risks and shortcomings; and this is even truer with Papyrology, since papyri very often attest to philological or linguistic variants that may differ from the accepted restitutions, or can even be unattested in the manuscript tradition. These circumstances of course affect the comparison of papyrus texts with the TLG databank, making Brunner’s wish for a database of literary papyri even more striking. Fortunately, this dream is going to be fulfilled, as we will see below (§§ 8.6–7).

From the merely technical viewpoint, the early TLG system used a Varian 620L minicomputer tied into Irvine’s mainframe. The texts were entered in the so-called Alpha Code, which conventionally used ASCII Latin characters and common symbols to represent Greek letters and diacritics. Both hardware and software had been designed by the visionary classicist David W. Packard, founder of the Packard Humanities Institute and son of the co-founder of Hewlett-Packard computer industry, who specifically implemented them to store, process, search, and display ancient Greek texts. Between the late Seventies and the early Eighties (1981) Packard developed new hardware and software and launched the Ibycus System, an IBM-

44 I am not speaking of metadata, of course. The first papyrological experiments, as well as the Oxyrhynchus Computer Project and the Duke Databank itself selected only ‘documentary’ texts. The first real examples of papyrological literary textual database are – significantly – the Catalogues of Mythographic and Paraliterary Papyri (CMP, CPP), launched in 2001 and 2003, respectively (see above and below, §§ 3.5 and 8.6).

45 Brunner 1994, 607.

46 The exact scope of the project, as defined in 1972, was “a lexicographical work which will collect, sort, and identify every single word extant in ancient Greek literary and non-literary documents” (http://stephanus.tlg.uci.edu/history.php). Then it was then decided to shape it as a literary textual databank. The lack of apparatus criticus is not to blame on the TLG team: “the field of Classics, asked by us to develop the principles that should underlie data entry of app. crit. materials, was not ready to provide the needed guidance” (Brunner 1994, 605).

47 Compare with the use of customized IBM machines at the LASLA (see above, § 8.2).
modified self-standing mainframe adapted to philological work, being specifically designed “to store, read, edit and search texts in Greek and Latin”\(^48\), as well as Beta Code, an enhanced release of the Greek encoding conventions. Beta Code uses upper-case Latin characters to represent Greek letters (a capital Greek letter is marked with an asterisk before it) and non-alphabetical common signs (parentheses, slashes, equals, pipe) to indicate spirits, accents, iota subscript (these must be added after the vowel, in the order: breathing-accent-iota subscript)\(^49\). *Ibycus System* used magnetic tapes to store information.

---

\(^48\) *Willis* 1984a, 163.

Magnetic tapes and the *Ibycus System*

It was in fact around an *Ibycus System* that the project of creating a textual databank of all documentary papyri was developed at the Duke University (Durham, NC), under suggestion of Packard and direction of John Oates and William Willis. Ten years earlier, TLG had abandoned the idea of including documentary papyri in its databank due to the technical difficulties of handling fragmentary texts; but the *Oxyrhynchus Computer Project* digitizing documentary papyri (see above, § 8.2) had also demonstrated that such a task was feasible. Therefore, the project officially started in 1982 and all published documentary papyri (i.e. Greek and Latin papyri, parchments, ostraka, and tablets from the III century BC to the VIII AD) started to be entered in Beta Code on magnetic tapes by means of the *Ibycus* machine acquired by

Duke. There, texts were typed directly in Greek51, and automatically converted in Beta Code by the system; subsequently they were proofread against the printed editions and the photographs at the University of Michigan under the direction of Ludwig Koenen52. The more recent editions were entered first, followed by the others backwards. Magnetic tapes were then released by Duke and delivered on request: the Beta Code could be easily displayed or printed in Greek characters by any conversion algorithm.


However, encoding papyri was not an easy task53, and it became necessary to apply some enhancements to the Beta Code. Indeed, TLG came without apparatus, but when dealing with papyri it is vital to consider the spelling variants and their regularizations in conventional Koine Greek, as well as scribal mistakes and their corrections: quite a different concern than TLG, which stands as a further example of the

51 OCR scans, as already evoked by KLEVE – FONNES 1981, 159, proved unfeasible: “the nature of papyrological data entry is infinitely more complicated than data entry from the standard printed volume of literary texts. Finally, the almost endless varieties of typeface used in presenting papyrological volumes, ranging from the different hands of the early BGU volumes to more recent typewritten and photographed editions, precluded the use of any kind of scanner” (OATES 1993, 64). Today, discussion of OCR in relation to ancient Greek texts is mostly focused on scanning modern editions: cf. BABEU 2011, 13–14.

52 Cf. OATES 1993, 63–4; GAGOS 1996, 15; 2001, 516 n. 10 and 525–6. Proofreading was twofold: it aimed at ensuring both accuracy (by checking any possible improvements to the texts) and user-friendliness (by standardizing variants, expanding abbreviations and symbols, etc.).

53 Cf. OATES 1993, 64–5.
papyrological care for the actual text in its precise features, as already seen in the earlier LASLA and Milan attempts at encoding emendations and editorial annotations. A new method was developed at Duke to encode such misspellings in the digitized texts, so that they could be easily retrieved with queries:

the conventional koine form is given first, followed by numbered braces enclosing the scribe’s form or the edition’s misprint: e.g., ὄνομα {4ωνομα}4 shows that the scribe has misspelled ὄνομα, ὑπὲρ {5υπαρ}5 that he wrote epsilon over alpha, αὐτοῦ {6αυτω}6 that he miswrote dative for genitive, Ἁθὺρ {7Ἁθὺς}7 that the edition has a misprint for Ἁθὺρ.

Special codes were introduced to encode original diaeresis (#80) and apostrophe (#81), as well as to represent other papyrological features like illegible characters or expanded abbreviations (see summary table below) and numerals, transcribed with the Arabic digit preceded by a grave accent. A manual was appended to each magnetic tape for ease of use. Thus, though the databank was admittedly designed to be a lexical and concordancing tools similar to TLG and not a substitute for printed critical editions, some articulated linguistic annotation was inserted somehow. Basic metadata (provenance, date, edition reference) were also added to each text in order to catalogue it and facilitate browsing; a more articulated set of metadata, including inventories, categorization, physical and palaeographical description, reference to photographs, was planned, though never developed. This was the picture of the Duke Data Bank of Documentary Papyri (DDBDP; later Databank, thence the slightly different acronym DDbDP) as portrayed by Willis at the 17th International Congress of Papyrology (Naples, 1983).

54 Willis 1984a, 169–70; cf. Oates 1993, 65 (the label {9} was also used, to indicate an alternative reading from a duplicate of the same text). It is worth noting, at this stage, a sort of ‘philological’ primacy of the ‘regularized’ text, accepted in the ‘main’ text, over the actual form attested on the papyrus, relegated in the parentheses. A change in this trend has been accomplished only recently, with the development of Papyri.info: see below, § 8.4.
55 Cf. Oates 1993, 64–5, for full discussion of the conventions used.
56 Cf. Willis 1988, 16.
57 “It is not our intention [...] to duplicate or replace the printed text of published papyri but solely to record them in such a way as to be instantly searchable in whole or by category as the scholar may wish. For consultation of the texts themselves, scholars will of course continue to repair to the published editions” (Willis 1984a, 169); see also Willis 1988, 16. Again, Oates 1993, 63: “The purpose of the Duke Data Bank is to make instantly accessible through search programs the total corpus of published Greek and Latin papyri. It is not intended to substitute for printed editions but rather to serve as a means of searching such volumes and of making concordances”.
58 Cf. Oates 1993, 68.
59 Willis 1984a.
A sample from the early DDbDP (P.Coll.Youtie I 33; screen output and Beta Code), from WILLIS 1984a, 170–1 (the ‘+’ marks a continuous line that is broken because does not fit the screen length).

In the following years, DDbDP absorbed the Oxyrhynchus papyri from the Oxford project, conveniently adapted to the new conventions\(^60\), and took great advantage of technological innovation coming from David Packard and the TLG project. Indeed, in 1985 the former launched *Ibycus Personal Scholarly Computer* (PSC), a machine equipped with the same features as *Ibycus System* but of a smaller size, which could read the new optical memory support called CD-ROM. The conversion of TLG to the CD-ROM was completed in the same year, and the TLG CD-ROM A was released by the Packard Humanities Institute (PHI). It is also known for having been the first published compact disk that did not contain music\(^61\). Both *Ibycus* PSC and the CD-ROM favoured the diffusion among scholars not only of the tools themselves, but also of a general technological know-how\(^62\). DDbDP converted itself to optical technology soon after TLG. In 1986, at the 18th International Congress of Athens, Willis announced the forthcoming news\(^63\), and presented them at the next meeting in Cairo (1989)\(^64\); the PHI-produced CD-ROM 2 (no. 1 being devoted to Latin and biblical texts), containing all documentary papyri published in 275 volumes between 1936 and 1988, totalling 19,500 texts containing 2.41 million Greek and Latin words, and an updated version of the Checklist, the conventions of which had been chosen to cite the papyri\(^65\).

---

\(^{60}\) Cf. WILLIS 1988, 15; OATES 1993, 63.


\(^{63}\) Cf. WILLIS 1988, 17–18.

\(^{64}\) Cf. WILLIS 1992. PHI CD-ROM 2 was released in December 1988 (in the same year TLG C, the second version of the literary database, was published too).

\(^{65}\) Cf. OATES 1993, 71; see above, § 2.3.
The impact of both resources on the papyrological community was immense; about twenty years later, Isabella Andorlini recalled that occasion with the following enthusiastic words:

Già nel 1989, in un disadorno corridoio dell’Università egiziana del Cairo che ospitava il XIX Congresso Internazionale di Papirologia, il collega Robert Kraft (University of Pennsylvania) si presentò attrezzato con computer, banca dati e sistema d’indagine Ibycus e fu in grado di fornire in tempo reale a chi lo domandasse le ricorrenze delle parole dei papiri nei testi della letteratura greca compattata su CD-ROM66.

Similarly, DDbDP on CD-ROM offered the possibility of wide-range speedy searches for words, phrases, strings, with or without blank spaces, but also for spelling variants, by searching for the labels instead of the words themselves67: “The value for editing texts justifies the creation of the data bank, but it also opens doors for historical, economic, and sociological studies as well as linguistic and stylistic analyses”68. This is an interesting statement: the original goals of a papyrological textual database remained unchanged, though many searching options were not embedded in the very texts any more69, and required the development of further tools.

The launch of DDbDP on CD-ROM, in the middle Eighties, coincided with the worldwide spread of personal computers: IBM PC dates to 1981, Apple Macintosh with its innovative graphical interface to 1984, and both operated with their own hardware and software architecture70. Since the database was specifically designed for Ibycus systems, third-party software was developed to allow processing the databank on different operating systems71. For example, Searcher was produced at the University of California-Santa Barbara for IBM machines; but the most famous and widespread were the Macintosh programs, namely SNS Greek & Latin developed at the Scuola Normale Superiore in Pisa72 and above all Pandora, designed by the Perseus Project at Harvard and based on HyperCard stacks73.

67 Cf. WILLIS 1984a, 170.
68 OATES 1993, 67.
69 Cf. QUENOUILLE 2016, 12.
70 Cf. HOCKEY 2004, 10–11.
71 Cf. WILLIS 1992, 126; OATES 1993, 65 with n. 6; WILLIS 1994, 628. A thorough comparison between SNS Greek & Latin and Pandora was developed by BÉGUIN 1995.
73 Cf. QUENOUILLE 2016, 7–8: “Die Suche konnten die NutzerInnen der DDBDP über „Pandora“ sowohl als genaue Suche als auch als „Wildcard“-Suche gestalten und dabei bis zu drei Begriffe, die nicht aufeinander folgten, gleichzeitig eingeben (Complex Search). Dabei konnten sie entscheiden, in welcher Reihenfolge die Wörter in den Texten vorkommen mußten bzw. ob sie überhaupt vorkommen durften. Für ganze Sätze reichte hingegen die einfache Suchmaske, in der der Satz(teil) eingegeben wurde (Simple Search). Durch Anklicken der entsprechend in einer Liste erscheinenden
In the Nineties, in particular, the software production for consulting the classical databases on CD-ROM – a second edition of the Duke Databank is dated to 1991, within PHI #6; a third one to 1997, within PHI #7 – flourished in plenty of different tools, for different platforms, with partially different functions, working and producing output in different character sets (font faces), almost not compatible with the other operating systems: one may refer to a useful list provided by the TLG website itself for a full overview updated to 2009. In a more or less sophisticated way, all these programs were able to process several different automated operations on the textual databank – word/phrase/string searches at various level and in personalized sub-corpora; index lists; sorting; browsing; displaying/exporting; etc.

The confusion was quite great, and one had to choose carefully according to needs and preferences because fees had to be paid for most of these programs, and sometimes also for some proprietary fonts used by them for output. Fonts that, in turn, often created severe compatibility problems across platforms. These issues are well known and it is not worth dealing with them in details. Suffice it to recall that though most of these tools are now discontinued – Pandora, for example, in its latest release 3.0, developed by Daniel Riaño, does not run on the newer Intel-based Macintosh operating systems –, if one is in need for consulting the Duke Databank on PHI #7 CD-ROM (or the TLG CD-ROM) can still rely on a simple but powerful open-source, cross-platform and Unicode-compatible software: Diogenes, written by P.J. Heslin (Durham, UK) in Perl script and XULRunner runtime environment (the same as Mozilla browsers), which can be installed also in server mode and runs on Mac OS X, Windows, Linux. This might still be an option for the unavoidable moments of Internet blackouts and it provides a rather helpful embedded engine for morphological analysis and links to a stand-alone electronic version of Liddell-Scott, being therefore a sort of offline version of Perseus (see below), but be aware that the latest DDdbDP CD-ROM release covers publications up to 1996 only.

Editionshinweise gelangten sie auf den Volltext, in welchem das Gesuchte vorkam und markiert ausgegeben wurde. Die Ergebnisse konnten schließlich inklusive Volltext in ein Worddokument exportiert werden”. HyperCard was a sort of primitive hypertext application: see HOCKEY 2004, 10–11.

74 Cf. WILLIS 1994.
75 Both TLG and PHI databases on CD-ROM are largely dealt with by SCHÄFER – MEIER 2003, 8–67.
77 https://www.tlg.uci.edu/about/cd_soft.php. Add former tools like L-Base and View & Find (cf. SCHÄFER 1993, 156–65), and Diogenes (see below). SCHÄFER – MEIER 2003, 67–250 (an updated version of SCHÄFER 1993), provide a detailed survey of the different programs, with a particular focus on the indexing and concordancing functions; this state of the art, of course, dates back to nearly 15 years ago.
78 https://community.dur.ac.uk/pjheslin/Software/Diogenes; cf. QUENOUILLE 2016, 9. Its latest version (3.2.0) dates to 2007 but still works well even on Windows 10.
A couple of screenshots of the latest release of Pandora (from QUENOUILLE 2016).
A couple of screenshots of the latest release of *Diogenes* (note, in the second one, the morphology and dictionary tool).
Indeed, the development of the Internet, as from 1991, opened new ways of conceiving the management of information (see above, § 1.2). Following the trend of other resources, also DDbDP moved on line, and in 1996/7 migrated from PHI #7 CD-ROM to the open, web-based Perseus Project, the renowned hypertextual online digital library hosted at Tufts University and directed by Gregory Crane (http://www.perseus.tufts.edu)\textsuperscript{79}. Perseus still offers a powerful platform: not only does it store information in an open universal format, but also develops automated routines of search and analysis. It is particularly remarkable the Greek Word Study Tool (http://www.perseus.tufts.edu/hopper/morph), which allows performing morphological analysis of ancient Greek words and cross-referencing each to the corresponding entry in an online version of Liddell-Scott-Jones’ Greek-English Lexicon, as well as in the abridged version called Middle Liddell. Each word in the classical corpus, including Greek and Latin literature – original texts and English translations, can be analysed with this tool, and search functions for exact forms or for all inflected forms of a lemma.

Perseus texts were marked with SGML (recently moved to XML) compliant with TEI\textsuperscript{80}, and therefore it was necessary to apply this markup to the plain Beta Code of DDbDP. The online search options proved very powerful: “l’impresa risulta tanto più straordinaria in quanto è la prima volta che su Internet appare un corpus complessivo di documentazione sul mondo antico in lingua originale”\textsuperscript{81}. The Duke Databank of Documentary Papyri, regularly updated, has been hosted by the Perseus Digital Library until 2010, when a completely different online platform was eventually released.

### 8.4 The Papyrological Navigator

In 2004/5 DDbDP (led by Oates, then passed away in 2006, and by Josh Sosin) started collaborating with HGV (led by James Cowey) to map the datasets of both to each other. But the sustainability of the entire Duke Databank was at stake: the increasing amount of primary data, favoured without doubts, in turn, by the spread of the digital techniques, was making update increasingly difficult and economically unaffordable for the small team at Duke, especially with the end of income from the CD-ROMs. In the meantime, Roger Bagnall had gone on with the concept of resource

---

\textsuperscript{79} Perseus is online since 1995; it was formerly released in CD-ROM (Perseus 1.0 and 2.0) but in a slight different shape (mostly didactical resources). It might be nice to stress the papyrological primacy in this case: TLG (under the new direction by Maria Pantelia, as of 1996) migrated to the web with a subscribed-access version only in 2001, after producing a last CD-ROM (TLG E, 2000). For Perseus as hypertext see Bolter 1991, 544.

\textsuperscript{80} See above, § 3.6, and cf. Rinear 2014, 225 ff.

\textsuperscript{81} Cristofori 2000.
integration envisaged at least since 1992 and developed with APIS (see above, § 3.6). In 2006, at Columbia University, he promoted the prototype of a new database based on portlet technology (different web modules contained by a portal) and a powerful image display platform: the idea was to aggregate digital pictures (mainly via APIS), metadata (via HGV and APIS), and text (via DDbDP) in a single hub. The project was called Papyrological Navigator (PN) and served as the starting point for a subsequent, wider project named Integrating Digital Papyrology (IDP), which started in 2007 with the goal of setting common standards in the papyrological resources and enhancing simultaneous access to them through a single interface, involving several different institutions (Columbia, King’s College London, New York University, Heidelberg Institute of Papyrology, University of Kentucky, Duke University) across the years.

Integration is based on the RDF (Resource Description Framework) model, which allows for connecting and merging the different sources in the framework of the so-called “Semantic Web”, and for possible further connections in the future.

The legacy encoding and markup of the Duke Databank via Perseus (Beta Code and SGML) were converted to modern, robust and shared standards: Unicode and the TEI/EpiDoc XML schema. Migration from Beta Code to Unicode was run through a Transcoder module, and then another applet (Chapel Hill Epigraphic Text-Converter = CHET-C, originally designed for digitizing epigraphical texts) was used to convert the Leiden editorial conventions of DDbDP, including legacy Beta Code escapes, into standard EpiDoc XML markup. A further effort was required by the encoding of digits, which had been entered as Arabic numerals in the earlier DDbDP: a Greek Number Converter was applied to transcode them to the corresponding Greek Unicode characters and to add XML numerical tags for future computational applications. HGV metadata underwent the same migration to XML. However,

82 Cf. Bagnall – Gagos 2007, 63–5: at the Copenhagen International Congress of Papyrology he described a “dream machine” fitting all main papyrological electronic resources (including a futuristic Berichtigungsliste, a digital Wörterbuch, prosopographical databases, even a Demotic dictionary) “into a comprehensive system of scholarly information”; Bagnall 2012a, 2.
83 Cf. Sosin 2010.
85 For technical details and further information cf. Cayless 2011 and 2013; in general on RDF cf. https://www.w3.org/RDF with further references. I am grateful to Hugh Cayless for advising on this.
86 Cf. Bodard – Sosin 2011; see above, § 1.2.
89 Some Leiden remainders were fixed during the following project phase.
the DDbDP conversion was run as a single and ultimate process, so that all future work would be conducted directly in Unicode and EpiDoc XML. On the contrary, HGV would continue to be maintained in its FileMaker database at Heidelberg, and its conversion would therefore have to be regularly repeated: this was obtained with a process called HGV Metadata Crosswalk, an XSLT that can convert the XML output of FileMaker into EpiDoc-compliant XML\(^90\); an HGV Translation Crosswalk was similarly applied to HGV German and English translations. The three separated XML outputs (DDbDP, HGV metadata, and HGV translations) were then merged into a unique XML file by an Aggregator, which was also able to process items provided with HGV metadata but lacking DDbDP text, allowing for their future addition. To merge DDbDP texts with HGV metadata, the unique numerical identifiers assigned by Trismegistos to the papyri (TM number: see above, § 3.3) played an essential bridging role\(^91\). Customized EpiDoc XSLT stylesheets were then applied to generate plain text (UTF-8) and HTML output from the XML files, in order to obtain a human-readable version of the content, featuring text, metadata, image, and translation juxtaposed in a synoptic view (see above, § 3.1).

The output has been indeed designed according to the typical papyrological editions, with Leiden conventions, metadata above text, a basic apparatus criticus below, etc.\(^92\) In this way, it has been designed a new concept of textual database, fully integrated with metadata (according to the earliest attempts) and featuring an apparatus criticus showing alternative readings, spelling variants, editorial corrections\(^93\). Below, a short text (SB XIV 11942, the early papyrus of Peukestas) with simple samples of Leiden critical marks (square brackets for ancient deletion) and of apparatus entries (regularization of iotacism), in both XML source and HTML output:

```xml
<TEI n="0239;14;11942" xml:id="sb.14.11942" xml:lang="en">
  <teiHeader>
  ...
  </teiHeader>
</TEI>
```

\(^{90}\) Cf. Bodard 2010, 105–6

\(^{91}\) On the issues of integration cf. Babeu 2011, 147, and see above, § 3.4.

\(^{92}\) Particularly remarkable is the change in the display of misspellings and “regularizations”, which initially followed the practice of placing normalized/corrected forms in the text and the ancient reading in the apparatus, as a consequence of the adaptations from the old DDbDP markup (see above, § 8.3). As of September 2011, the two elements have been swapped with each other: cf. http://digitalpapyrology.blogspot.it/2011/09/just-posted-to-papylist-dear-colleagues.html. This required a huge effort, because the ancient reading was originally transcribed diplomatically without spirits and accents, but its inclusion in the text made it necessary to add them. This was mainly driven automatically by a script handling a large table of equivalences and with the help of the TLG morphological engine; but some remaining mess has been fixed manually over the time.

<fileDesc>
  <titleStmt>
    <title>sb.14.11942</title>
  </titleStmt>
  <publicationStmt>
    <authority>Duke Collaboratory for Classics Computing (DC3)</authority>
    <idno type="filename">sb.14.11942</idno>
    <idno type="ddb-perseus-style">0239;14;11942</idno>
    <idno type="ddb-hybrid">sb;14;11942</idno>
    <idno type="HGV">4274</idno>
    <idno type="TM">4274</idno>
    <availability>
      © Duke Databank of Documentary Papyri. This work is licensed under a <ref type="license" target="http://creativecommons.org/licenses/by/3.0/">Creative Commons Attribution 3.0 License</ref>.</availability>
  </publicationStmt>
  <sourceDesc><p/></sourceDesc>
</fileDesc>

<profileDesc>
  <langUsage>
    <language ident="en">English</language>
    <language ident="grc">Greek</language>
  </langUsage>
</profileDesc>

<revisionDesc>
  <change when="2012-10-26T08:50:07.916-04:00" who="http://papyri.info/editor/users/james.cowey">Finalized - Ready.</change>
  <change when="2012-10-26T08:50:07.903-04:00" who="http://papyri.info/editor/users/james.cowey">Vote - Accept-Straight-to-Finalization - Good catch. Fine.</change>
  <change when="2012-10-23T10:45:17.621-04:00" who="http://papyri.info/editor/users/simoeis">Line 2: reg tag added and --></change>
  <change when="2011-12-14" who="http://papyri.info/editor/users/gabrielbodard">rationalized languages in langUsage</change>
  <change when="2011-12-14" who="http://papyri.info/editor/users/gabrielbodard">changed editor names to URIs</change>
  <change when="2011-10-31" who="http://papyri.info/editor/users/gabrielbodard">changed type=inWord to break=no</change>
  <change when="2010-05-05" who="http://papyri.info/editor/users/gabrielbodard">changed schema; added xml:space=preserve; indented; moved title/@n to idno</change>
  <change when="2009-11-12" who="http://papyri.info/editor/users/"/>
<body>
  <head n="4274" xml:lang="en">
    <date>331BC?</date>
    <placeName>Saqqara</placeName>
  </head>
  <div xml:lang="grc" type="edition" xml:space="preserve">
    <ab>
      <lb n="1"/><del rend="erasure"> ε</del>
      <lb n="2"/>
      <lb n="3" break="no"/> ε . <choice><reg> ε</reg><orig> Ι</orig></choice>
    </ab>
  </div>
</body>
</text>
</TEI>
Another striking innovation is that

the complete set of IDP XML files are published with a Creative Commons Attribution 3.0 License, explicitly permitting the typical varieties of scholarly reuse and citation anticipated for the data, in line with other recent calls for open access in the humanities\(^94\).

As in the traditional print editions, each item has a call number, which is this case is represented by the permanent URL assigned to it. This means partial standardization, since each URL is unique. However, since each resource (HGV, DDBDP, APIS) has its own URLs, following its own conventions, it happens very often that an item has three different URLs (e.g. http://papyri.info/ddbdp/p.got;101 = http://papyri.info/hgv/30696 = http://papyri.info/apis/gothenburg.apis.112), even more if it has been republished (e.g. http://papyri.info/ddbdp/p.got;20 = http://papyri.info/ddbdp/sb;20;14671 = http://papyri.info/hgv/38507 = http://papyri.info/apis/gothenburg.apis.14). Also in this case, a unifying factor is given by TM numbers, acting as unique identifiers (30696 and 38507, in the given examples).

On the access side, PN deploys a user-centred interface (http://papyri.info/search). The search functions are particularly articulated and represent a decisive improvement with respect to the Perseus platform, though no morphological analysis is possible any more\(^95\). They have been constantly enhanced during the past years\(^96\). Full-text word, phrase and substring queries in texts, metadata or translations support Betacode/Unicode input, proximity customizations (definition of proximity character or word ranges\(^97\)), Boolean operators (AND, OR, etc.) for combination of strings, regular expressions\(^98\), search for abbreviations\(^99\), lemmatized

94 Baumann 2013, 93. Cayless 2010, 146 contends that CC license, and in general open access, fosters the digital permanence of scientific publications; see below, § 9, for openness as a requisite for digital criticism.

95 Cf. Quenouille 2016, 12.


97 It must be stressed that word proximity searches for full words; if one wants to search for any other text string, then character proximity must be used. This affects significantly the searching.

98 Regular expressions (REGEX) are formalized patterns describing certain amounts of text. They use ‘literal characters’ and ‘metacharacters’ (characters with a special meaning) to represent the text searched for. For example, given that \b indicates a ‘word boundary’ (where \ is an escape code to indicate that what follows is not a literal character), . (dot) stands for ‘any character’, \{\} (braces) indicate a repetition range, one can write REGEX αὐτο\b.{1,20}\bκαι in the query box to search for “a string beginning αὐτο- within 20 characters of a string beginning και-.” (cf. http://
searching (via a term index built by mapping the original word forms to the morpho-
logical tables developed by the Perseus Project), case\textsuperscript{100} and diacritic-sensitivity. Documents can also be browsed and searched by inventory and edition reference, provenance, date, language. Any kind of query generates a single searching instance, which may be combined with others or subsequently narrowed by closing the appropriate instance, which is displayed as an autonomous box on the top of the resulting list of hits. The results are sorted by publication reference, and a quick preview of date, provenance, and the immediate context of the word(s) searched for are provided\textsuperscript{101}. Of the integration with the Bibliographie Papyrologique, active since 2011\textsuperscript{102}, we have already discussed above (§ 2.1); we can just add that, unlike the previous versions of DDbDP, Papyri.info features also Coptic texts\textsuperscript{103}, and some preliminary attempts to add Unicode Arabic texts worked fine, despite the different right-to-left direction of writing\textsuperscript{104}. A bigger challenge for the future might be taking into consideration Demotic documentary texts:

Demotic is a more difficult matter, as it seems impossible to define a standardized set of characters, and therefore a Unicode encoding standard, for a script with so many variations from one scribe to the other, not to mention the different methods of transcription used around the world\textsuperscript{105}.

digitalpapyrology.blogspot.it/2012/03/idp-updates.html). Complete explanations and tutorials about regular expressions, which can support even extremely complex combinations, can be found at http://www.regular-expressions.info.


\textsuperscript{100} This is useful if one has to search for proper names (persons, gods, places, months...), since by rule in the digital editions of papyri no word is capitalized but – indeed – proper names.

\textsuperscript{101} Sometimes, the highlight of the term(s) searched for does not work properly for some bugs not yet completely fixed.


\textsuperscript{103} Coptic characters do indeed have a dedicated Unicode subset. Previously, Coptic documentary texts were collected by the Banque de données des textes coptes documentaires, a.k.a. the Brussels Coptic Database (BCD) developed by Alain Delattre at the Université Libre de Bruxelles (http://dev.ulb.ac.be/philo/bad/copte; see above, § 3.3). This database was last updated in 2014. For further Coptic resources see DELATTRE – HEILPORN 2014, 324.

\textsuperscript{104} Cf. GAD 2016.

\textsuperscript{105} DELATTRE – HEILPORN 2014, 322. For now, one should refer to the Demotic texts stored in the Thesaurus Linguae Aegyptiae (http://aaew.bbaw.de/tla); see also MADERNA-SIEBEN – WESPI – KORTE 2016 (above, 5.4). A very short Demotic transcription (Ptwrs, a personal name written on the verso of P.Tebt. I 110) can be found only in http://papyri.info/ddbdp/p.tebt;1;110.
A typical Papyri.info integrated record (HGV + TM + APIS + DDbDP + image; Greek papyrus).
In the following pages, samples of Coptic, Arabic and Latin entries.
<table>
<thead>
<tr>
<th>Papyri.info</th>
<th>From Textual Databases to Digital Scholarship</th>
</tr>
</thead>
</table>

| **Title** | From Textual Databases to Digital Scholarship |
| **Publication** | Papyri.info |
| **Language** | English |
| **Availability** | Available online |

**Apparatus**

<table>
<thead>
<tr>
<th>p.worp.65 = HGV P.Worp 65 = Triamisipatos 115592</th>
</tr>
</thead>
</table>

| **Title** | p.worp.65 = HGV P.Worp 65 = Triamisipatos 115592 |
| **Publication** | P.L. Simas 2020 |
| **Language** | Greek |
| **Availability** | Available online |

**Apparatus**
Title: Resuviso of an Apisid-brood (presumably P. Bm). Inv. 406
Publications: P. Bagnall 3 (item in item list)
     Ostraca (item in item list)
Date: 226 BC
Language: Latin
Provenance: Ostracarea (direction unknown)
Place: Paphlagonia
Subject: Ostraca, Ostraca (Ostraca) (Babylon)
     Ostraca, Ostraca (Ostraca) (Hellenistic)

Catalog Record:

Title: Request to inherit [not before 120]
Author: Aurelius Dionysus (from Ostracarea, Egypt)
Summary: Request to inherit an ostracarea from Ostracarea (modern name Babylon), Egypt, from Dioscurides, one of the scribes of the ostracarea. The request is to be made in front of the scribes of the ostracarea, and the request is to be witnessed by a witness from the ostracarea.

Citations:

Support/Dimensions: 1. From: papyrus. Two fragments which do not join, in place, in the ostracarea.
Language: Greek
Note: 750 CE
Note: 750 CE
Note: 750 CE
Note: 750 CE
Note: 750 CE

Subjects: Archontes (list of Babylon from Egypt), the scribes, and the ostracarea from Egypt.

DO&DP transcription:

II.2.2
1. [ ... ]
2. [ ... ]

Apparatus:

Notes:

It is possible that one or both of these ostracareas are from the ostracarea from Egypt, as indicated by the record of the ostracarea from Egypt.

Editorial History:

This ostracarea is based on the ostracarea from Egypt, and it is therefore included in the DO&DP project.

Linked Data:

This ostracarea is linked to the DO&DP project and the Papyrological Navigator project.
8.5 The Papyrological Editor

The most revolutionary improvement came during the second stage of the Integrating Digital Papyrology project (IDP2, 2008–2010): alongside some residual fix from the preceding phase, an innovative method was deployed to ensure an appropriate update of the Duke Databank in the face of the increasing scholarship in the field and of the parallel increasing shortage of economic and human resources at Duke\textsuperscript{106}. The idea of a dynamic web-based editorial platform open to contribution from the entire papyrological community owed much to Ross Scaife’s pioneering project of making the translation of Suda (the X-century encyclopaedic lexicon) open to the users’ contributions, according to the web concept of wiki (Hawaiian for “quick”) i.e. collaborative editing of hypertextual pages, developed as of 1995\textsuperscript{107}. Scaife’s project, called Suda On Line (SOL, www.stoa.org/sol)\textsuperscript{108}, started in 1998 (three years before Wikipedia), and its legacy is reflected in the name chosen for the software environment conceived for the papyrological collaborative platform: Son of Suda On Line (SoSOL), which is admittedly a tribute to Scaife’s idea. The collaboration takes place in the so-called Papyrological Editor (PE, http://papyri.info/editor), a platform where any registered user can edit existing database entries or adding new items. Since IDP uses Git, a distributed version control system designed to handle multiple versions of project information, as its public data repository (via GitHub), since 2011\textsuperscript{109}, all changes to the dataset are recorded and tracked in a version history\textsuperscript{110} (see picture above):

[p]ermanent transparency is the guiding principle behind SoSOL. The system keeps track of everything. When you log in and submit a text, SoSOL records it; when you submit a text or propose an emendation SoSOL will not let you submit until you have written a message explaining what you propose. Similarly, SoSOL will not allow Editors to vote on a text without explaining why they vote the way they do. For every single text SoSOL keeps a permanent and comprehensive record of every single change. Users can see this, forever. The discipline of transparency and permanence has the virtue of requiring all of us to live up to the high standards of our field’s motto, and make that motto meaningful: amicitia papyrologorum. Collegiality is, in effect, a technical requirement of SoSOL. It also means that all proposals must be offered and scrutinized with utmost seriousness, since our comments are visible to all, forever. And, that under SoSOL accurate scholarly attribution is very easy to enforce. Moreover, we assume

\textsuperscript{106} Cf. Bagnall 2012a, 2–3; Babeu 2011, 148.
\textsuperscript{108} Cf. the history sections (http://www.stoa.org/sol/history.shtml and http://www.stoa.org/sol/about.shtml) as well as Mahoney 2009 and Baumann 2011.
\textsuperscript{109} Cf. http://digitalpapyrology.blogspot.it/2011/01/idp-data-available-on-github.html: “it represents an enabling of true community ownership of the data”. Note that, though it is possible to edit the source code via Git and GitHub, it is recommended to perform small text interventions via the Papyrological Editor interface: see the Readme.md at https://github.com/papyri/idp.data.
\textsuperscript{110} On both Git and the revision history cf. Baumann 2011. Cf. also Bagnall 2012a, 5.
that even suggestions judged by the Editors to be incorrect might one day be judged right, in the light of new finds, or might, though wrong, nevertheless inspire someone else to solve even an unrelated puzzle. So, SoSOL does not throw away rejected ideas; it simply stores them in the Comments page for every text, accurately attributing and time-stamping every single comment, for posterity, and for purposes of rigorous scholarly attribution\textsuperscript{111}.

One strength of this model is that rejected proposals are not deleted forever, and are instead retained in the digital record, in case new data or better arguments appear to support them. Additionally, all accepted proposals are attributed to their contributor so that proper scholarly credit can be given to them\textsuperscript{112}.

However, the editing process takes place in the PE (SoSOL) platform (“Advanced Create” in the user’s home page, after login), where one can choose between changing existing material, picking DDB, HGV or APIS entries up, or creating a completely new “publication”, as the editing instance is called\textsuperscript{113}. Each “publication” is divided into “identifiers”, i.e. the constituting resources (DDB text, HGV and APIS metadata); each “publication” and each “identifier” are provided with a unique number, which forms the URI of the editing instance (e.g. http://papyri.info/editor/publications/55694/ddb_identifiers/126674/edit points to a DDbDP text editing instance within publication no. 55694)\textsuperscript{114}. The process ends with the submission of the changes made by the users to an editorial board (currently composed of Rodney Ast, James Cowey, Paul Heilporn, Todd Hickey, Cisca Hoogendijk, and Josh Sosin\textsuperscript{115}), which conducts a thorough double peer review on the proposals and decides whether accepting, correcting, refusing, or sending them back to the user for more substantial changes. If the changes are accepted, the updated file will be eventually published in the database and publicly available, with all comments and remarks traced in the history log\textsuperscript{116}.

\textsuperscript{111} S OSIN 2010.
\textsuperscript{112} B ABEU 2011, 148.
\textsuperscript{113} An “Assignments” Google spreadsheet is available online to advise about papyrological editions still missing from the database: anyone can choose a text and put his or her name to claim it for digital entering (https://docs.google.com/spreadsheets/d/1DFnkrgqtcn4erxuP3_TkW-6LFQr2PW4WS0F-ys0oBNo/edit#gid=0).
\textsuperscript{114} Cf. S OSIN 2010; C AYLESS 2011; B AUMANN 2013. For a beginner’s guide to PE (updated to 2012) compiled by Paul Heilporn and others, see https://docs.google.com/document/edit?id=1w0TXTq5V ulzQxGYq9vOOCJRER33Ir6tKmwwaCzyGrs&authkey=CKnGk_ML&hl=en#.
\textsuperscript{115} Senior editors have also been appointed for advice on the most difficult or complex cases. They are Roger S. Bagnall, Willy Clarysse, Hélène Cuvigny, Nikolaos Gonis, Dieter Hagedorn, Ann E. Hanson, Andrea Jördens, James G. Keenan, Klaas A. Worp, and formerly the late Isabella Andorlini.
\textsuperscript{116} APIS and HGV metadata can be updated or entered following the very same pipelines.
Texts can be edited from their XML code, but in order to facilitate the work of papyrologists, a tag-free markup language has been developed so that it be the closest possible to the traditional Leiden editorial conventions\textsuperscript{117}. This language is therefore called Leiden+ because it is a digital enhancement of the Leiden system. It has a double advantage: it is comparatively easy to learn and use for non-XML experts, and it allows to copy and paste text from digital sources, and to adjust it with minimal changes (of course, the source text must be typed in Unicode characters). Many signs remain the same as their Leiden antecedents; others undergo little adjustments (see the summary table below) because Leiden+ markup has to be automatically transcoded into the corresponding XML tags\textsuperscript{118}.

\textsuperscript{117} Cf. \textsc{Baumann} 2013, 6–10.
\textsuperscript{118} Detailed guidelines are available at http://papyri.info/docs/leiden_plus. On transcoding see \textsc{Bodard} – \textsc{Sosin} 2011. It was effected by means of a parsing dual-syntax converter called \textsc{XSugar}, which supports conversion from both XML to Leiden+ and vice versa (http://www.brics.dk/xsugar).
For example, abbreviated words – traditionally resolved with the expansion in parentheses – must be enclosed into an extra pair of parentheses, because also the entire word is marked as an expanded abbreviation in XML, e.g. αὐτός → (αὐτός) → <expan>αὐ<ex>τός</ex></expan>. It must be stressed that while Leiden+ is a descriptive markup system, i.e. it tends to reproduce the papyrological features of a text, XML is a semantic markup language, i.e. tends to describe the meaning of those features. This leads to interesting theoretical conflicts with the traditional papyrological editorial practice.

For example, a lacuna is, papyrologically speaking, a physical gap of the papyrus, where some text is missing. This is marked, according to the Leiden conventions, with square brackets. Sometimes lacunas can be supplemented, either completely or partially, on the ground of parallels or conjectures or such, but from the papyrological viewpoint they are still lacunas. There is no papyrological difference between αὐ[. . . ], αὐ[το . ], and αὐ[τός]: square brackets always indicate the same circumstance. TEI XML, on the other hand, being a text-focused markup, distinguishes between textual portions and non-textual portions, so that an unsupplemented lacuna (non-textual portion) remains a lacuna and is labelled with the <gap> tag. Therefore, our first example will be encoded as αὐ[.3] and transcoded into αὐ<gap reason="lost" quantity="3" unit="character" />, i.e. “a 3–character long gap of lost text”. On the contrary, a supplemented lacuna is encoded as supplied text, with the <supplied> tag; therefore, our third example will be transcoded into αὐ<supplied reason="lost">τός</supplied>. This has consequences on the Leiden+ markup, because one must be careful in separating supplemented and unsupplemented portions when they occur within the same lacuna: our second example must be encoded as αὐ[το][.1] so that it be correctly transcoded into αὐ<supplied reason="lost">το</supplied><gap reason="lost" quantity="1" unit="character" />. On the other hand, what for a papyrologist represents illegible characters, expressed with dots (e.g. αὐ . . . ) but not in square brackets because actually visible on the papyrus, for XML is a non-textual portion, since it does not express any meaningful text, and is classified as a <gap>. The only difference from a proper lacuna is the “reason” attribute, “illegible” instead of “lost”: αὐ<gap reason="illegible" quantity="3" unit="character" />, i.e. “a 3–character long gap of illegible text”.

Another important caveat is that Leiden+, as all markup languages, though tag-free, is nonetheless a mathematical expression and its logical syntax must be respected. Therefore, if a papyrologist has to transcribe a lost line end, (s)he can print an opening square bracket followed by a blank; but in the digital encoding, (s)he must close the bracket, because any opened tag must be closed properly: αὐτός [ → αὐτός [?]. Syntax mistakes are always noticed by a validation checker, which will display a red banner when the editor attempts to save the work; on the contrary, the
platform cannot detect semantic errors, like the said lacuna case, and it will display a green banner anyway when saving\textsuperscript{119}.

This kind of markup causes elements of the apparatus criticus to be encoded directly within the text: the terms that need to be pinpointed in the apparatus are marked through special tags, and their display at the bottom of the text is just a matter of HTML visualization, admittedly to emulate a printed edition (see above, § 8.4). This is probably the best and clearest example of the fact that such a semantic markup as XML / Leiden+ is content-focused rather than display-focused, which means that what really matters is the correct encoding of the textual features, their semantic substance, and not their rendered appearance:

there is much emphasis in the modern study of digital preservation on preserving the appearance of documents [...]. But an overemphasis on appearance pushes one in the direction of technologies that I will argue are not the ideal vehicles for digital preservation\textsuperscript{120}.

A closer look at the apparatus cases considered by the papyrological XML / Leiden+ markup seems to be worthwhile, since it is the main milestone that differentiates purely textual databanks like TLG from proper digital editions\textsuperscript{121}. All tags work the same way in Leiden+: the opening mark <: is followed by the ‘correct’ or main instance of the text (the preferred alternative to be printed in the text; the newer reading; the regularized or corrected form), then by the appropriate tag, finally by the other instance or instances (multiple alternatives and editorial corrections are supported), before the closing mark :>. However, the HTML rendering can differ: usually the term on the left is displayed in the main text, but the |reg| tag works the other way around, as noted above (§ 8.4). Furthermore, the EpiDoc XML code behind Leiden+ points to slightly different concepts: alternatives and editorial corrections belong to the <app> type, expressing “one entry in a critical apparatus, with an optional lemma and usually one or more readings or notes on the relevant pas-

\footnote{119}{In the years, the editing syntax has been improved to better respond to the users’ and the scientific needs. Among the most remarkable enhancements, we can mention the possibility to encode multiple alternative readings and ‘regularizations’ with an easier markup (cf. \url{http://digitalpapyrology.blogspot.it/2011/09/just-posted-to-papylist-dear-colleagues.html}; \url{http://digitalpapyrology.blogspot.it/2011/12/papyriinfo-updates.html}), and above all the replacement of the tag initially used to indicate misspellings (which was |orth| for all cases) with a more nuanced distinction between ‘regularizations’ of linguistic variants (tag |reg|) and simple ‘corrections’ of outright scribal mistakes (tag |corr|) (cf. \url{http://digitalpapyrology.blogspot.it/2011/03/new-in-ddbdp.html}). An interesting enhancement has also been the addition of different types of editorial corrections (BL, proposals from printed publications, proposals via PE: cf. \url{http://digitalpapyrology.blogspot.it/2011/12/papyriinfo-updates.html}) and the possibility of nesting several different cases into one another.}

\footnote{120}{CAYLESS 2010, 145.}

\footnote{121}{On the issue of the apparatus criticus in the digital editions of ancient texts see BOSCHETTI 2007; AGNESINI 2008, 114; MAGNANI 2008, 132; BABEU 2011, 158; DAMON 2016.}
sage"\(^{122}\); regularizations and orthographic corrections belongs to the <choice> type, defining a number of alternative encodings of the same text portion due to editorial interventions\(^{123}\); the scribal correction, <subst>, points to ancient interventions, which belongs to a third different category because records actual text features and not modern editorial changes\(^{124}\).

The following examples are taken from the online guidelines, http://papyri.info/docs/leiden_plus:

<table>
<thead>
<tr>
<th>Case</th>
<th>Description</th>
<th>EpiDoc XML</th>
<th>Leiden+</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternate readings</td>
<td>different possible readings of uncertain words</td>
<td>&lt;app type=&quot;alternative&quot;&gt; &lt;lem&gt;Ὀχυρυγχίτου&lt;/lem&gt; &lt;rdg&gt;Ὀξυρυγχίτου νομοῦ&lt;/rdg&gt; &lt;/app&gt;</td>
<td>&lt;:Ὀχυρυγχίτου</td>
<td>alt</td>
</tr>
<tr>
<td>Modern editorial corrections</td>
<td>newer improvements in readings proposed by the previous editors(^{125})</td>
<td>&lt;app type=&quot;editorial&quot;&gt; &lt;lem resp=&quot;resp&quot;&gt;αἱ τοῦ&lt;/lem&gt; &lt;rdg&gt;Θίτου&lt;/rdg&gt; &lt;/app&gt;</td>
<td>&lt;:αἱ τοῦ=resp</td>
<td>ed</td>
</tr>
<tr>
<td>Spelling regularizations</td>
<td>phonetic or morphological deviations from the 'standard' Greek(^{126})</td>
<td>&lt;choice&gt; &lt;reg&gt;φρόντισον&lt;/reg&gt; &lt;orig&gt;φρόνδεισον&lt;/orig&gt; &lt;/choice&gt;</td>
<td>&lt;:φρόντισον</td>
<td>reg</td>
</tr>
</tbody>
</table>


\(^{125}\) The resp attribute can refer to modern authors, printed bibliography, BL corrections, or PN corrections directly suggested via the Editor.

\(^{126}\) This markup is suggested also for all cases of iota adscript.

\(^{127}\) When the mistake involves an extra or a missing character, the use of Leiden brackets is recommended (e.g. στρατ(τ)ηγός, στρατευ(τ)ηγός, corresponding to different EpiDoc XML codes), but a markup like <στρατ(τ)ηγός|corr|στρατέυς> is also suggested. It must be said that actually there is a certain degree of inconsistency, which may lead to some erroneous encoding. The main example provided, in fact, could also be encoded as <:τ<ι>μ{μ}ὴν|corr|τμμὴν:>, depending on the editor's opinion about the nature of the scribe's actual mistake.
As is apparent, the HTML display is just a temporary display of a user-friendly adaptation (Leiden+) of a deeply semantic markup, which requires a thorough understanding of the ancient text. The traditional editorial practice is unavoidably affected by instances of uncertainty and incoherence, which should be cleared during the digital encoding of the edition, without very little care for its final graphical appearance. It has been noted, for example,

that even with standard conventions such as [printed] Leiden, not all the conventions were applied evenly, as some scholars used ‘underdots’ to indicate partially preserved characters while others used them to demonstrate doubtful characters. The use of EpiDoc consequently addressed these types of issues with Leiden encoding as it was commonly practiced: “This example illustrates the primary advantage of encoding the editions in XML. If editors wish to differentiate uncertain characters and broken characters they can encode them with different tags. They can then transform both tags into under-dots if they still wish to present both instances as such or they can decide to visualize one instance, underlined and the other under-dotted to distinguish between them.”

The carefulness required by the digitization of a papyrus text according to a strict set of standard conventions leads us to make a fundamental observation. Digitizing a papyrus edition is itself an editorial work, a philological reconsideration of the printed edition(s). To properly encode the text in a formalized structure, the digital editor is compelled to analyse the reference edition thoroughly in order to understand what the original editor meant to express, and possibly also to check any reading against the original piece or, at least, a digital reproduction of it. Moreover, a global reconsideration of the papyrus may lead to corrections or reading improvements, which can be directly annotated in the digital framework. From this viewpoint, the digital edition is an edition of an edition, but not in the (Platonic) pejorative sense: on the contrary, it increases editorial akribeia exponentially, becoming an intimate part of the process of scholarship, and not a mere supporting tool:

---

128 Note that the original form is encoded without diacriticals.
129 BABEU 2011, 150. EpiDoc envisages the <damage> tag to indicate characters that are broken but legible (http://www.stoa.org/epidoc/gl/8.16/trans-damaged.html), which is different than <unclear> marking uncertain characters (http://www.stoa.org/epidoc/gl/8.16/trans-ambiguous.html). Such distinction is not retained in Leiden+, so that both cases tend to be encoded (as in the printed editions) with the underdot, which corresponds to the <unclear> XML tag only.
130 ROUED-CUNLIFFE 2009, [2].
il modello editoriale digitale [...] è rigido e flessibile al tempo stesso, e obbliga l’editore virtuale ad adeguare ai parametri condivisi del sistema le idiosincrasie sempre più marcate dei moderni editori di papirosi che trascurano le raccomandazioni del sistema codificato di Leida e costruiscono edizioni critiche sempre più personali; [...] lo sforzo che l’editore virtuale fa nel convertire in linguaggio elettronico i problemi di lettura e di comprensione non risolti o la molteplicità delle soluzioni alternative educa la comunità degli esperti ad una più consapevole applicazione di criteri editoriali condivisi e di rigore interpretativo131.

### Comparative table of different markup for papyrus texts.

<table>
<thead>
<tr>
<th>Case</th>
<th>Leiden</th>
<th>LASLA</th>
<th>ARSINOE</th>
<th>DDbDP Beta Code</th>
<th>Leiden+</th>
<th>TEI/EpiDoc XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclear / uncertain</td>
<td>α</td>
<td>Α</td>
<td>α</td>
<td>α?</td>
<td>α</td>
<td>&lt;unclear&gt;α &lt;/unclear&gt;</td>
</tr>
<tr>
<td>illegible (1)</td>
<td>.</td>
<td>1</td>
<td>!</td>
<td>.1</td>
<td></td>
<td>&lt;gap reason=&quot;illegible&quot; quantity=&quot;1&quot; unit=&quot;character&quot;/&gt;</td>
</tr>
<tr>
<td>illegible (15)</td>
<td>-15-</td>
<td>15</td>
<td>!15</td>
<td>.15</td>
<td></td>
<td>&lt;gap reason=&quot;illegible&quot; quantity=&quot;15&quot; unit=&quot;character&quot;/&gt;</td>
</tr>
<tr>
<td>lacuna</td>
<td>[ . . ] [ ] [3], [ . ] [3], [.?]</td>
<td>[13], [c%1] [.3], [.?]</td>
<td></td>
<td>&lt;gap reason=&quot;lost&quot; quantity=&quot;3&quot; unit=&quot;character&quot;/&gt; , &lt;gap reason=&quot;lost&quot; extent=&quot;unknown&quot; unit=&quot;character&quot;/&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>supplement</td>
<td>[α]</td>
<td>[α]</td>
<td>(α)</td>
<td>[α]</td>
<td>[α]</td>
<td>&lt;supplied reason=&quot;lost&quot;&gt;α&lt;/supplied&gt;</td>
</tr>
<tr>
<td>omission</td>
<td>(α)</td>
<td>&lt;α&gt;</td>
<td></td>
<td>&lt;α&gt;</td>
<td></td>
<td>&lt;supplied reason=&quot;omitted&quot;&gt;α&lt;/supplied&gt;</td>
</tr>
<tr>
<td>superfluous</td>
<td>{α}</td>
<td>&lt;&lt;α&gt;&gt;</td>
<td>{α}</td>
<td></td>
<td></td>
<td>&lt;surplus&gt;α&lt;/surplus&gt;</td>
</tr>
<tr>
<td>deletion</td>
<td>[a]</td>
<td>[a]</td>
<td>[4α]4</td>
<td>[a]</td>
<td></td>
<td>&lt;del rend=&quot;erasure&quot;&gt;α&lt;/del&gt;</td>
</tr>
<tr>
<td>interlinear</td>
<td>'α'</td>
<td>((α))</td>
<td>\α/</td>
<td></td>
<td></td>
<td>&lt;add place=&quot;above&quot;&gt;α&lt;/add&gt;</td>
</tr>
<tr>
<td>abbreviation</td>
<td>α(β)</td>
<td>α(β)</td>
<td>α(β)</td>
<td>α[1β]1</td>
<td>α(β)</td>
<td>&lt;expan&gt;&lt;ex&gt;β&lt;/ex&gt;&lt;/ex&gt;</td>
</tr>
<tr>
<td>symbol</td>
<td>(α)</td>
<td>*α</td>
<td>((α))</td>
<td>[1α]1</td>
<td>(α)</td>
<td></td>
</tr>
<tr>
<td>doubt</td>
<td>?</td>
<td>*</td>
<td></td>
<td>?, (?)</td>
<td>cert=&quot;low&quot;</td>
<td></td>
</tr>
<tr>
<td>misspelling</td>
<td>apparatus</td>
<td>* C α = β C (4 ) , etc. according to types</td>
<td>&lt;choice&gt;&lt;reg</td>
<td>β&gt;:&gt;; &lt;α</td>
<td>corr</td>
<td>β&gt;:&gt;</td>
</tr>
</tbody>
</table>
8.6 From Digital Editions to Digital Scholarship

The final, open access version of the integrated database (PN) and of the editorial SoSOL platform (PE) was released in 2010 under the name of Papyri.info (http://papyri.info). It is apparent that we are dealing with a completely different concept of papyrological database, where the instances of integration and collaboration have profoundly transformed what was originally a searching/indexing/concordancing tool\(^{132}\). The Duke Databank is no more a fixed collection of canonical reference texts: “it is a collection of conjectures, now easily capable of being revisited, revised, and improved”\(^{133}\): a dynamic workspace for a digital scholarship, the true representative of a “discipline in flux”\(^{134}\) such as Papyrology itself, and something else than the original DDbDP.

The texts, already provided with a basic critical apparatus, after the third phase of the IDP project (2010–12) are equipped with the possibility of adding an introduction and line-by-line commentary\(^{135}\): they are, therefore, potentially closer to the concept of digital critical edition than to that of textual databanks à la TLG. A recent experiment conducted at Heidelberg, during the Seminar of Digital Papyrology held by Rodney Ast, Lajos Berkes and James Cowey, led to the creation of born-digital critical editions of unpublished papyri. A group of *descripta* of the Gothenburg collection was studied and edited directly online via PE. The results – some of them are already available in the public database\(^{136}\) – showed that the potentials of PE go far beyond the collection of already published texts and their open update.

There exist some other online resources providing digital editions of papyrus texts. Several of them chronologically precede the IDP project, and clearly express the feeling of expanding the then existing digital papyrus corpus (DDbDP via Perseus) by taking into consideration different textual categories (namely, paraliterary and literary papyri\(^{137}\)) and/or a deeper level of information (articulated metadata, apparatus criticus, descriptions and commentaries\(^{138}\)).

---

132 See above, §§ 7.1 and 8.2. The 1968 AIP recommendations (above, § 1.1) also focused on indexing issues.
133 BAUMANN 2013, 105.
134 HANSON 2002; see above, §§ 1.1–2.
136 http://papyri.info/ddbdp/ddbdp;2015;1; http://papyri.info/ddbdp/ddbdp;2015;2; http://papyri.info/ddbdp/ddbdp;2015;3. Note the way they are recorded: since they are *descripta* and do not have any printed *editio princeps*, they have not been called after the official abbreviation P.Got., but with a progressive “ddbdp” number, which makes it clear that the papyrus has just this online edition. Cf. BERKES 2017.
137 See the project of Kathleen McNamee for creating a database of marginal annotations in literary and paraliterary papyri (MCNAMEE 1984). For Arabic papyri see already above, § 3.5
138 See e.g. the cases of collections catalogues that include also transcriptions or editions of some texts: the Spanish *Ductus* and the German *Papyrus Portal* (see above, § 3.6).
A groundbreaking effort came from the already mentioned *Catalogue of Paraliterary Papyri* (2003), which beside metadata chose to provide the full texts of the documents, both in plain Beta Code (transcriptions without accents, diacriticals, papyrological signs, only to facilitate search) and in Unicode Greek, encoded in TEI XML, converted in an HTML display, and provided with a critical apparatus. Yet CPP is not a mere reproduction of the existing editions:

Although the CPP collection does not have the ambition to produce new scholarly editions, the texts are never simple reproductions of one particular edition but they are based on our own representation of the most recent edition or simply of the one we considered the best. In many cases, it is the result of the comparison between two or more editions. When this is so, variants among the different editions are noted in the apparatus. For the purposes of scholarly research, however, consultation of the printed editions remains necessary139.

This statement is remarkable for two reasons: first, it demonstrates what we noticed above of the philological flavour of the digitizing task; second, the recurring observation that a digital resource, even detailed and almost complete, never replaces other traditional sources of information.

Another pioneer in the digital edition of papyrological texts has been the *Vindolanda Tablets Online* portal (VTO, http://vindolanda.csad.ox.ac.uk), directed by Alan Bowman, Charles Crowther, and John Pearce (Oxford/CSAD), offering a complete and updated online version of T.Vindol. II, which superseded T.Vidol. I. This is a

nice example of integration between printed and digital resources: introductory parts of the volumes are reproduced, useful concordances between printed and digital documents are provided, and a detailed section of *addenda* and *corrigenda* is maintained (http://vindolanda.csad.ox.ac.uk/tablets/TVaddenda.shtml). General introductions to the tablets and their context (“Exhibition”), a reference section with information about the documentary context (names, military terms, numerals, dates, currency, measures), and a complete guide to the database form a wide help tool to better use and understand the published material. All sections of the site are searchable, and the database itself can be browsed by several fields (publication number, subject, category i.e. chapter headings, document type, people mentioned, places mentioned, military terms, archaeological context…) or searched with various criteria (Latin text, metadata text, publication number). The texts are published alongside a zoomable digital picture, an extensive commentary, and an English translation. The texts themselves are encoded in a modified version of TEI XML\textsuperscript{140}, called “Vindolanda XSL Style Sheet”, with apparatus and notes that pop up in separate windows; the notes from T.Vindol. I and the *addenda* are also available when applicable. The VTO site was developed in 2001–2003, and since 2011 it is flanked by a second website, *Vindolanda Tablets Online II* (VTO2, http://vto2.classics.ox.ac.uk), developed by Henriette Roued-Cunliffe. It is not intended to be a replacement of VTO but an updated re-elaboration of the concept of digital edition. While VTO was shaped as a database, VTO2 is designed as a series of XML documents encoded in Creative Commons license according to EpiDoc standards, from which information is extracted through a web service (APPELLO) specifically developed for this site\textsuperscript{141}. The new collection, comprising T.Vindol. I-II as well as the more recent third vol-

\textsuperscript{140} Cf. http://vindolanda.csad.ox.ac.uk/tablets/TVdigital.shtml; BABEU 2011, 146.
\textsuperscript{141} Cf. http://vto2.classics.ox.ac.uk/index.php/about/appello-web-service; ROUED-CUNLIFFE 2009; BABEU 2011, 151–2 and 157. APPELLO also allows for automated reading suggestions (see above, § 7.1).
From Textual Databases to Digital Scholarship

volume T.Vindol. III, can be browsed by publication number, but is not searchable; thematic indices are provided instead. Each document exhibits image, inventory number, introduction, text, translation, and commentary. A remarkable characteristic is that the text is annotated: the words feature different colours according to their category (persons, military terms...: a “contextual encoding”\(^{142}\)) and are lemmatized, so that by clicking each one a pop-up window gives reference of the lemma, its occurrences within the *corpus* (concordance), dictionary entries from the *Perseus Project*, and a definition taken from Thomas Cooper’s *Thesaurus Linguae Romanae et Britannicae*.

The integration between text and images is even deeper in the online edition of the *Codex Sinaiticus*, the famous 4\(^{th}\)-century biblical codex (http://www.codexsinaiticus.org). Since the artefact had been dispersed among four different institutions (British Library, National Library of Russia, St. Catherines Monastery, and Leipzig University Library), an international project has been launched to reunite the entire manuscript in digital form and make it accessible to a global audience. The *Codex Sinaiticus Project* is therefore, first of all, a remarkable case of virtual reunification of scattered pieces of the same documents, performed through digital imaging techniques\(^{143}\). Then

\(^{142}\) BABEU 2011, 151.

\(^{143}\) Cf. VANNINI 2016 and see above, § 5.3. A digital reunified *Codex Sinaiticus* is available also through the *Turning the Pages* project of the British Library (http://www.bl.uk/turning-the-pages) for virtual leafing through (see above, § 5.3 as well).
these high-quality pictures\textsuperscript{144} are integrated, within a single interlinked interface, page by page, with: (a) a TEI-compliant transcription of the text, supporting either a view by physical page or by biblical verse\textsuperscript{145}, including all corrections; (b) modern translations in Russian, Greek, German, and English of selected passages; (c) detailed physical description of each page. The text is digitally aligned to the image (see above, § 7.1), so that clicking a word in the transcription highlights the corresponding word in the picture, and can be browsed by page or by biblical passage.

A particular corpus that deserved a special attention is the Herculaneum one, for the understandable peculiarities that justify the existence of a “Papirologia Ercolanese”\textsuperscript{146}. An automated indexing of the Herculaneum papyri was attempted first by Knut Kleve – after all, Lacunology and Literalogy were by-products of his work on the carbonized rolls (see above, §§ 5.4 and 7.1) – and Jan Songstad, who produced (in 1975) an \textit{Index to Works of Philodemus} in which the lines of the papyri are numbered in a continuous series, and a concordance called \textit{Works of Philodemus} that lists the words alphabetically giving the line number, the reference to the edition, and the context of the phrase. This was followed in 1987 by a similar \textit{Concordance to Philodemus} and by Daniel Delattre’s attempt, in the Nineties, to digitize the Philodemean texts in a Word file, to perform queries with its ‘search’ tool\textsuperscript{147}.

In 2002 Gianluca Del Mastro started a new enterprise: he noticed that many Epicurean texts were missing from TLG, which therefore could not be used as a valid search tool for the Herculaneum papyri; on the other hand, he also noted that the Herculaneum texts, unlike the literary works recorded in TLG, are by nature subject to constant update; it was also necessary that data from all various editions be available, in order to have the entire editorial history on hand\textsuperscript{148}. He therefore launched the \textit{Thesaurus Herculanensium Voluminum} (THV, \url{http://www.thvproject.it}), started in 2008 with the collaboration of Holger Essler (Würzburg University). This ongoing database (26 papyri uploaded so far) is searchable with various text combinations; each papyrus is encoded with basic metadata (catalogue number; author, work title, and volume number, with indication of the degree of certainty of the attribution; bibliographical reference to edition) and the text, with interpretations and notes to lines displayed in pop-up windows, in a hypertextual architecture resembling that of VTO2 (see above). Text is encoded, rather uncommonly, in SuperGreek (see above, § 8.1; the corresponding font can be downloaded from the site). The choice is explained by the fact that this font contains all symbols used to edit Herculaneum papyri and perfectly interfaces the programming language used to

\textsuperscript{144} A page of the website is devoted to their technical details.
\textsuperscript{145} Cf. Babeu 2011, 123.
\textsuperscript{146} We have already encountered the issues related to the digital imaging of the Herculaneum papyri: see above, § 5.3.
\textsuperscript{147} Cf. Del Mastro 2012, 176–7.
\textsuperscript{148} Cf. Del Mastro 2012, 177–8.
build the MySQL database. However, it goes without saying that a Unicode-compliant font would be much more universally integrated with other resources and other software: a future conversion, fortunately, is not excluded\(^{149}\), also in view of the ongoing project of linguistic annotation of Herculaneum papyri (see above, § 7.1); THV is indeed a major contributor to DCLP (see below, § 8.7). The most interesting feature is the possibility for the registered scholarly users\(^{150}\) to propose emendations to the texts, in a collaborative spirit that parallels the Papyrological Editor, “nella convinzione, che fu di Marcello Gigante, che solo grazie alla collaborazione internazionale la papirologia ercolanese potrà continuare a lungo il suo cammino”\(^{151}\). Future integration with the catalogue Chartes (see above, § 3.6), eventually in an aggregated portal, is under consideration\(^{152}\). On the side of Herculaneum Papyrology we shall mention also the forthcoming final outcome of PHerc project (see above, § 6.6), in that it envisages “a DVD including an interactive edition of the critical text with direct links to all the relevant papyrological documentation and a virtual reconstruction of the original papyrus roll”\(^{153}\): an integrated digital critical edition that raises particular expectations.

The Derveni Papyrus Online, developed in 2012 by the Center for Hellenic Studies (principal editor Ioanna Papadopoulou), follows a different format, and shows some noteworthy features. The text from Kouremenos, Parássoglou and Tsantsanoglou’s editio princeps\(^{154}\) is hosted on the iMouseion Project (http://dp.chs.harvard.edu/...
index.php?col=1&ed=KPT), a platform designed to allow annotations, indices, and collaborative work on digital editions of ancient texts\(^{155}\). The text is encoded in content-based, annotated XML and Unicode font; the apparatus can be toggled into a different window on the right. The platform offers also a reproduction of the newer edition of the papyrus by F. Ferrari\(^{156}\), with English translation and apparatus criticus. The two versions can be displayed in two parallel columns, generating therefore a “multiversion”. The same “multiversion” can be obtained with a third edition of the text, the more recent one established by Alberto Bernabé and Valeria Piano. This is a very interesting example of the evolving nature of a digital critical edition: in a scientific background of fluidity, the hyperspace is used to store more than one version of the text, so that the possible different solutions be compared and evaluated (see below, §§ 8.7 and 9). A further feature should be stressed: thanks to the technical peculiarities of the platform, textual supplements in lacuna can be displayed or hidden at alternate clicks\(^{157}\). As we remember, the possibility of having separate outputs for the diplomatic transcription and the emended edition is an old desideratum in Digital Papyrology, in order to gain as much as possible a representation close to the original fragment.

A fresh ongoing project is the database to be developed in the framework of the PLATINUM project (Papyri and Latin Texts: Insights and Updated Methodologies) led by Maria Chiara Scappaticcio at the University of Naples “Federico II”. The project aims at providing complete editions of all Latin texts on papyrus and ostraka; the texts will


\(^{157}\) One may have experienced the same feature in the Searchable Greek Inscriptions database by PHI, http://epigraphy.packhum.org.
be subsequently collected in a databank (https://platinum-erc.it/database, technical editor Andrea Bernini). Though most of them are already included in Papyri.info, the idea clearly attests to the need for some more advanced resources dealing with special corpora.

It is apparent that there is a generalized need for something more than a plain textual databank, even though with advanced search functions. A first trend is towards the creation of a collaborative workspace where papyrologists can offer their individual contribution, share knowledge, and even interact with each other in a virtual way. In short, a place for dynamic digital scholarship which seems to be the increasingly precise incarnation (or shall we say excarnation?) of the ideal amicitia papyrologorum, which could never find a real proper way of expression in paper format. The Internet undoubtedly favoured this concept: as its creator Tim Berners-Lee put it,

[the Web] is an information space through which people can communicate, but communicate in a special way: communicate by sharing their knowledge in a pool. The idea was not just that it should be a big browsing medium. The idea was that everybody would be putting their ideas in, as well as taking them out158.

A second trend can be described as the deployment of an integrated and interconnected network of data, metadata and images that goes beyond the traditional fixity of canonical critical editions. A third trend is the need for resources devoted to non-documentary papyri159. The Derveni Papyrus Online stems from a specific research interest, but CPP and THV are admittedly aimed at filling in a DDbDP gap that even TLG cannot cover because of its own nature.

From all the said trends stemmed Proteus, a forthcoming Oxford project announced as “a platform that digitally captures the evolving data of Greek and Latin literary and subliterary papyri as they are edited and re-edited over time, [...] a digital ecosystem for both creating next-generation born digital critical editions and generating the textual criticism that underwrites them” 160. Essentially, it is conceived as a philologically-oriented “Papyrological Editor”, where users will be able to create born online critical editions complete of diacriticals (in-browser keyboards and menus are planned to help in this) and apparatus, and to emend and update existing data. The platform is still under construction at http://www.proteusproject.uk, but is announced as articulated into an editorial section (DELFHI, i.e. “the Digital Editor for Classical Philology”) and the Proteus Search Interface.

---

158 From the transcript of his talk to the MIT LCS 35th Anniversary celebrations, Cambridge MA, April 1, 1999: https://www.w3.org/1999/04/13-tbl.html). See Introduction above (§§ 1.1–2).
159 In fact, as we saw, an exact categorization is often impossible – I should perhaps say ‘non-HGV/DDbDP papyri’.
Some of the official screenshots of Proteus (apparatus, XML source, Markdown markup, and – in the next page – a collation of two different editions).
8.7 New Standards for Digital Literary Papyri

The Proteus project is admittedly rooted in the statement that

[d]espite being a focal point for modern digital papyrology, the [Papyri.info] application targets only documentary papyri and consequently cannot be used to create born digital critical editions of literary papyri\textsuperscript{161}.

It therefore aims at providing an innovative tool to perform this task by developing the encoding standards already implemented by the IDP project. The new platform will be based on a redesigned TEI-compliant XML schema, cognate to but different than the EpiDoc one that informs Papyri.info, since the latter was originally designed for epigraphy, and the Proteus developers produced “a new XML standard for philological studies of papyrological material”, which they called Critical Syntax for Papyri (CSYN-P)\textsuperscript{162}. Consequently, the Leiden+ markup has been rethought too, and


\textsuperscript{162} “As the standard was designed for epigraphy, many of the standard’s XML tags and attributes provide little to no meaning in the context of papyrology and obfuscate the XML structure of a literary papyrological edition” (WILLIAMS – SANTARSIERO – MECCARIELLO – VERHASSELT – CARROLL – WALLIN – OBBINK – BRUSUELAS 2015, 2). One may note, however, that many epigraphic tags are not used at all for encoding papyri, and do not disturb the users’ work on the texts, if this is what the Proteus developers mean. On the other hand, the common EpiDoc ground can ensure high degrees
a new user-friendly annotating syntax has been developed under the name of CSYN Markdown, as a result of a combination between the Leiden editorial conventions “and the popular Markdown language”, which is actually a minimal transcription language for study material and. Moreover, “a custom XML parser” to render the XML file into HTML human-readable display will be used instead of the standard XSLT transformation schemas. The announced innovations are great, especially after almost ten years of EpiDoc/Leiden+ addiction. What may strike is that, at least for now, no reference is made to data circulation, open-source software or Creative Commons licenses, nor to the possible compatibility of the new standards with all existing resources, nor even to cross-resource integration. Even more striking is perhaps the assertion that “the current information model for Greek and Latin digital texts fails to include the vital components necessary to create complete born digital critical editions and facilitate the scholarly use and citation of such editions”, while, as we saw above (§ 8.6), the SoSOL Papyrological Editor is fully equipped for supporting both live emendations and born-digital critical editions of papyri, citation of which is made rather easy by the unique identifying URLs associated to the digital documents. They are papyri of documentary type, of course: but recently a new project have been launched to extend the experience of Papyri.info to literary and paraliterary material – a fact that is somehow acknowledged by the Proteus developers.

The project in question is the Digital Corpus of Literary Papyri (DCLP, http://litpap.info), and has been launched in 2013 by the Institute for the Study of the Ancient World (ISAW) at New York (Roger S. Bagnall, Tom Elliott) and the Heidelberg Institute of Papyrology (Rodney Ast, James Cowey), with technical collaboration of the Duke Collaboratory for Classics Computing (DC3), which manages Papyri.info (Ryan Baumann, Hugh Cayless, Josh Sosin), expressly to extend the PN/PE functionalities to the whole world of literary and paraliterary papyri. The main coordinates of this still ongoing effort are the very same as the documentary Papyri.info: a Papyrological Navigator with the same searching options (currently from

of compatibility between cognate documents (let us just think of Trismegistos planning to include inscriptions, see above § 3.3).

163 Cf. https://daringfireball.net/projects/markdown; VOEGLER – BORSCHEN – WEBER 2014. As to now, Markdown lacks true standardization, contrary to XML.


166 “Although plans have been announced to extend its functionality to literary papyri, we are unable to evaluate their proposed system as the application’s changes are still a work-in-progress” (WILLIAMS – SANTARSIERO – MECCARIELLO – VERHASSELT – CARROLL – WALLIN – OBBINK – BRUSUELAS 2015, 1).


From Textual Databases to Digital Scholarship

litpap.info it is possible to search in both corpora: documentary and literary; metadata (taken from TM and LDAB, of course, rather than from HGV); an Editor section; a TEI/EpiDOC-based XML source code publicly available via GitHub. The difference, and the main issue, lied in adapting the encoding parameters established for the documentary papyri to texts that present slightly different features.

As we have already cleared (see above, § 8.5), in the digital encoding of any text what really matters is the information stored in the computer, not the pure display output. Therefore, it is true that Papyri.info initially did not support the full set of paratextual symbols that one can find in non-documentary papyri (coronides, diplai, diplai obelismenai, stigmai, etc.), but is also true that EpiDoc XML is flexible enough to allow extending its capabilities far beyond the original design. For example, the EpiDoc tag <g> is used to label non-standard characters or glyphs (with a “type” attribute specifying the name of the symbol), and is extensively used in the Duke Databank to mark, e.g., original dots (<g type="dot"/>), S-shaped symbols for etous “year” (<g type="s-etous"/>), check marks (<g type="check"/>). These are all rendered in Leiden+ as the symbol name between two asterisks: *dot*, *s-etous*, *check*, etc. Of course, literary diacritical marks are not included in list of <g> types used for documentary papyri; nevertheless, the language is flexible enough to allow using the same syntax for virtually any glyph: this means that new tags <g type="coronis"/>, <g type="diple"/>, <g type="diple-obelismene"/> and so on can implemented, and they indeed are recognized and accepted by the system. This works also on the Leiden+ side: *coronis*, *diple*, *diple-obelismene*, etc. are accepted by the PE and correctly converted in the corresponding XML. The only “problem” is that they don’t have a specific graphical display in the HTML output – but this is an issue shared with many other ‘documentary’ <g> types, also due to the lack of specific Unicode characters, and a secondary one, since the important point, as we stated, is correct encoding.

Another example of the flexibility of EpiDoc/TEI XML is the treatment of layout features. In text of literary and paraliterary nature, even more than in ‘documents’, the mise en page is a fundamental part of the text itself, and quite often plays a primary role: the articulation of the content bears meaning and needs to be encoded properly. Ekthesis and eisthesis, for example (extension and indention of lines), are not only significant from the bibliological and palaeographical viewpoints, but are themselves parts of the work, contribute to its meaning in defining sections of text. Originally not conceived for the encoding of documentary papyri, such layout devices can now be marked through appropriate XML and Leiden+ tags.

169 A list of the <g> types currently featured by Papyri.info is available at http://147.142.225.252/paptrac/wiki/gtypes.
170 Detailed discussion of this issue will be offered in REGGIANI 2018c, 2018d, and 2018e.
Discussion of literary and paraliterary text encoding has been carried on by the DCLP developers together with the participating projects, in particular the Würzburg team directed by Holger Essler and the Parma team led by Isabella Andorlini. Indeed, both were dealing with very peculiar groups of texts – the philosophical treatises preserved in the Herculaneum carbonized rolls, with their fundamental and complex editorial history (see above, §§ 7.1 and 8.6), and the corpus of the Greek medical papyri, which comprises literary as well as documentary and paraliterary works, i.e. technical texts with a very peculiar textual scenario (reuse, annotations, abbreviations and symbols, heavy paratextual devices, idiosyncratic variants from the medical writers\textsuperscript{173}). Several joint meetings led to the definition of a complex stylesheet for the encoding of ancient punctuation, diacriticals, symbols, layout devices, editorial features, not least the variant readings, which express loci where the papyrus deviates from the manuscript tradition or other sources and are of course totally absent from ‘documents’. Some issues are still under evaluation and development, but the Papyrological Editor can indeed evolve to expand the capacity of the papyrological database to encompass all the types of written materials. Both from Würzburg and from Parma came a significant contribution to the rising DCLP: the former provided annotated texts (lemmaization layer), the latter built full critical editions complete of introduction, apparatus, line-by-line commentary, and translation\textsuperscript{172}. These are nice examples of the potentials of a versatile database that can become also a space of discussion and confrontation: DCLP will offer the same editing possibilities as Papyri.info.

The medical papyri, in particular, have been published with summarized information taken from the main reference editions, and therefore exhibit a very basic apparatus criticus and commentary, essentially reporting – beside the usual editorial corrections and ‘regularizations’ – relevant parallel passages in medical authors\textsuperscript{173}. Moreover, the Parma team is also planning to develop some experimental born digital critical editions of unpublished medical papyri, in order to envisage the most suitable way to deploy the editorial workspace of DCLP. Actually DCLP, like Papyri.info, is not “unrelated to the task of creating born digital critical editions”\textsuperscript{174} – it is just designed to be a workspace for digital scholarship, and the developments in the research will contribute to shape its nature\textsuperscript{175}.

\textsuperscript{171} Cf. REGGIANI 2017c, 2017d, 2018b, 2018d, 2018e.

\textsuperscript{172} The encoded texts are listed at https://goo.gl/ZBbHkp.

\textsuperscript{173} Cf. http://www.papirologia.unipr.it/ERC. The project is mentioned at https://wiki.digitalclassicist.org/Parma_Digital_Medical_Library. See below, Appendix 2.


\textsuperscript{175} An interesting claim to avoid project-specific markup is advanced by MONELLA 2008 (cf. BABEU 2011, 34).
A sample medical papyrus on DCLP