DIGITIZATION AND THE DICTIONARY OF CELTIC LATIN

At the 2006 Prague colloquium of editors of the medieval Latin dictionaries of Europe I gave a paper that, like many of the other contributions, was subsequently published by the Czech Academy of Sciences: it appeared under the title “From Full-Text Database to Electronic Lexicon and Beyond: The Role of Computers in the *Dictionary of Celtic Latin* Project”. The subject being primarily that indicated, the paper therefore covered an aspect of the Dublin project that was not addressed in respect of its sister enterprises until four years later, at the table ronde in León (June 2010). However, François Dolbeau has kindly offered me the present opportunity to summarize the approach taken by the *Dictionary of Medieval Latin from Celtic Sources* (DMLCS) to digitization, and to outline developments that have taken place in this area since my Prague paper was published.

The fact is that, throughout the history of DMLCS, we have accorded a much more central role to information technology than have at least some of the other Medieval Latin enterprises, and with the current trend towards digitization we happen to find ourselves in the lead. It was no particular virtue inherent in DMLCS that meant that this project had a head-start in the use of information technology over most of its sisters; it was just that the Dublin enterprise was embarked upon later than the others, and at a time when it was becoming clear that full computerization was the approach to adopt. For although valuable preparatory work had been carried out earlier, it was not until the year 1980 that the final terms of reference for the Celtic-Latin project were fixed, and by that time most of the other ventures had already appointed teams of readers and accumulated large quantities of slips. They were therefore committed to using the traditional methodology ( excerption – lemmatization – selection and classification – arrangement and composition), even if they are now busy transferring this methodology to electronic supports. DMLCS, on the other hand, could start with a clean slate.

The form this fresh approach took was the decision to capture, in their entirety, on computer, the corpus of relevant texts from approximately the period 400 to 1200 A.D. In the specially commissioned *Bibliography of Celtic-Latin Literature* (Dublin, 1985), which defined this corpus for the first time, Michael Lapidge and Richard Sharpe showed that the totality of the material consisted of about thirteen hundred distinct texts from the former Roman Britain, Ireland, Scotland, Brittany, and the Hiberno-Latin continent; we now reckon that, all told, the material adds up to six or seven million words of continuous

---

2. A semi-schematic outline of the plan for the whole project has already appeared as “Chro-

text. The main purpose of the resulting full-text database has been, of course, to provide the reservoir from which the Dictionary itself goes on to draw its examples; however, it was always also intended that, as a searchable repository, the Celtic-Latin corpus in digital form would come to constitute a unique and valuable resource for scholars working in many branches of the Humanities other than lexicography. The question was how to make it available to them.

That question is being answered with Brepols’ publication, from the DMLCS database, of a series of three cumulative editions of a full-text Archive of Celtic-Latin Literature (A CLL). Originally modelled upon Brepols’ existing CLCLT (now LLT, the Library of Latin Texts), this was designed to constitute with that corpus the beginnings of a pan-European electronic library of patristic and medieval Latin material. The initiative has been proceeding as follows:

- ACLL-1, the first, preliminary edition, was originally released on CD-rom in 1994 and, in a revised form, on line in 2007. This extensive collection consisted of over four hundred selected texts, representative of the authors, nationalities, periods and genres in the Celtic-Latin domain. As befitted a database designed to encapsulate what was most distinctive in Celtic-Latin literature, the texts in ACLL-1 were largely chosen for their embodiment of that difference.

- ACLL-2, the second, developed and much expanded edition, as completed and published on line earlier this year (2010) at Brepolis, shifts the emphasis to the fact that Celtic authors of Latin in the period from the fourth century to the fourteenth also contributed to mainstream European Latin culture in ways for which they have not hitherto been accorded sufficient credit. While continuing to present all of the material that was captured for ACLL-1, a principal intention of the second edition is therefore to provide searchable access also to Celtic-Latin material that went on to make its mark across western Europe. More than a hundred significant additional works have been included to this end.

- ACLL-3. The initiative is planned to culminate in ACLL-3, a third, definitive edition, once the database has reached the maximum feasible size – that is, when all accessible Celtic-Latin texts have been captured and processed to standard. In the meantime, the availability of Brepols’ Cross-Database Searchtool means that even the current release of the Archive is jointly searchable on line both with LLT and, now, with eMGH (the electronic Monumenta Germaniae Historica) and, for that matter, the Aristoteles Latinus Database (ALD). As we like to say in our publicity material, “methodology and content have thus already combined to integrate ACLL fully into the digital portfolio available to scholars of the written heritage of a thousand years of Western European history”.

Key theological writings appearing for the first time in this context are the complete Pauline commentaries of the heresiarch Pelagius, Eriugena’s weighty contribution to controversies concerning predestination, and much of the influential output of Abelard; while work that profoundly influenced British national historiography is represented by Geoffrey of Monmouth’s Historia regum Britanniae, and by further books from Gerald of Wales. Among additional legal and penitential material, an edition of the vastly influential early eighth-century Collectio canonum Hibernensis is now included. At the same time, other categories already opened in ACLL-1 have been filled out with further Saints’ lives, grammatical treatises, annals, inscriptions, poetry, charters, and the remaining two books of Eriugena’s Periphyseon.
As with LLT, a series of filters enables the Archive to be searched not just for Latin forms within the texts, but also by author, or by text-title, or by period, or indeed (in the online version) by century. Furthermore, the slightly richer mark-up in use in Dublin means that, in the case of ACLL, one can also search by geographical area, or by original language-family in the case of unlatinized words embedded in DMLCS material. As regards the nature of the Latin forms that can be searched for, these can (as in Professor Tombeur’s project) be whole individual words, or part-words, or collocations of words. The second of these three facilities, in particular, is helpful given the fluctuating orthography encountered between (and even within) individual works of Celtic-Latin literature as captured in the Archive, and given the need to anticipate the appearance of inflected forms of lexemes.

Having produced the first edition of the Archive in the mid 1990s, the DMLCS staff turned their attention to lexicography in the strict sense. As I pointed out in my Czech Academy paper, this was the point at which we encountered the “down side” of having adopted the full-text database methodology: it was immediately clear that, with the extremely modest resources that the project had or seemed likely to have, the production of a dictionary that dealt with all the words in that database would take us not just decades, but centuries! In any event, as was observed in a review of progress that took stock of the situation, “university library shelves were replete with fascicules of definitive dictionaries of various languages that were complete for the first few letters, but that then petered out, either abandoned ignominiously or else still in progress after decades; and scholars were as likely to wish to look up a word beginning with S or T as they were one commencing with A or B”. So the approach taken by DMLCS was to determine, in the light of our circumstances, what approximation to a full dictionary could reasonably be considered feasible, given that this would have to be done in a period of time that would also be defensible. The upshot was our strategic decision in the first instance not to attempt to proceed by dealing with all of A, then all of B, and so on.

“Instead, DMLCS would conduct a number of lexicographical sweeps through the entire alphabet, each restricted to certain categories of word and, as such, complete and constituting a worthwhile publication in itself, but each capable of forming part of an eventual whole that need not be very different from a traditional definitive dictionary... The first such sweep was designed to provide authoritative treatment of those words encountered in any part of the Celtic-Latin corpus that were not found in the Classical Oxford Latin Dictionary (even with a change of meaning).”

Such non-Classical items were considered to be in many ways the most interesting words, and by definition they were the most distinctive; and DMLCS already had at its disposal a skeleton collection of appropriate headwords. This initial sweep, then, is the phase of the project that, as far as lexicography is concerned, the project staff have engaged with first, ACLL’s valuable word-searching software being ideally suited to ferreting out further items for it. The component of the Dictionary that appeared from Brepols of Turnhout in 2005 accordingly took the form of Volume I of The Non-Classical

---

4 This is an alternation that has continued since then: as soon as the first constituent volume of the Dictionary had duly appeared ten years later, it was time to concentrate once more on the Archive, which is why work on ACLL-2 has represented our main effort since the Prague gathering.

5 A. Harvey, “Royal Irish Academy Activity in Celtic-Latin Studies”, in The Scriptures and Early Medieval Ireland, ed. T. O’Loughlin (Turnhout, 1999), pp. 117-124 (p. 120).

6 Ibid., pp. 120-121.
Lexicon of Celtic Latinity (thus NCLCL-1); covering the letters A to H, it contains something under 6000 headwords, and the plan is for the rest of the alphabet to be dealt with in Volume II. In the meantime, the whole of the collection of headwords has already been made available as an on-line Celtic-Latin Word-List on the DMLCS website; the List is updated letter by letter as each is treated for the Lexicon, and serves as a sort of advance catalogue to the major work.

As a work of primary lexicography, NCLCL has the objective of constituting a documented guide to the meaning and usage of the individual non-Classical items of vocabulary found in the corpus. It seeks to achieve this aim by analysing, in a detailed way, the specific examples of the words involved, as these are to be found in the full-text Archive; it is by reference to their location in that database that the examples are consequently cited, and this is done by means of codes (to which the key is of course provided). The intention here is to provide, along with the location, the maximum amount of metatextual information possible in a limited space. In addition to a number corresponding to that of the text in the Bibliography and Archive (these using the same system), and the number of the electronic pagina in the Archive where the example is to be found, each reference-code therefore also contains an indication of the author or, in the case of anonymous works, the genre to which a text is assigned, and a geographical marker. To save space, the examples are not actually quoted and, whereas it is recognized that the dictionary’s definitions may therefore occasionally appear somewhat cryptic or elliptical given that the examples themselves cannot be immediately viewed in their contexts, it is felt that this will be more than compensated for by the further (and more important) role that is to be played by the reference-codes in the future. For these will really come into their own with the planned publication of an electronic version of the Lexicon, the point being that these codes correspond to those already used in the internal indexing of ACLL. Therefore, without rekeyboarding, they provide a ready-made means by which the digitized version of the Lexicon can be grafted onto a future edition of the Archive: the codes will serve as hypertext links between the instances of the words in the archived texts and the Lexicon’s definitions of those words. With the codes providing immediate access from the full texts

---

7 It may be consulted by following the “Wordlist” link from <http://journals.eecs.qub.ac.uk/DMLCS>, while progress on other elements of DMLCS work, such as ACLL and our Scriptores Celtigenae medieval Latin text series, may likewise be tracked by following the appropriate links from that URL.

8 Thus, for example, the word exhonorare, when used in its pan-European, Late Latin sense of “to despise or treat with contempt”, is shown to occur with coordinates THL:B340@13 — that is, on pagina 13 of an anonymous theological work from geographical zone B (namely Ireland) having text-number 340 (identified in the key as a seventh-century Commentarius in epistolas catholicas). On the other hand, used in the context of Gaelic law with the meaning “to deprive of personal honour (Old Irish ‘enech’)”, the word is found at JCLN:E1018@227 — pagina 227 of a work by Jocelin of Furness; written in zone E (Scotland), this is text 1018, shown by the key to be a Vita S. Kentigerni. And when, influenced by (or perhaps with a play on) Classical Latin exonerare, an example of our word is found in the sense of “to deprive of specific honour(s)”, the coordinates GRLD:A60EPIST@221 tell us that the author is Gerald of Wales, writing, in zone A (the former Roman Britain), text 60EPIST (the Epistolae from his work Symbolum electorum), the form in question occurring on pagina 221.
Die lateinische Mediävistik wird in der Tschechischen Republik an den Universitäten, in Instituten der Akademie der Wissenschaften, so wie auch in Archiven, Museen und anderen wissenschaftlichen und kulturellen Institutionen betrieben; dieser Bericht will vor allem die Ergebnisse besprechen, die mit der mittellateinischen Philologie und Editionstätigkeit zusammenhängen und die auch den nicht-tschechischen Benutzern dienen können.


Die Mitarbeiterinnen des Wörterbuches beteiligen sich auch an der zweisprachigen Herausgabe (lateinisch und tschechisch) von Isidors *Etymologiae*. Im Jahr 2009 wurden zwei Bände veröffentlicht, und zwar die Bücher XIX-XX, übersetzt und kommentiert von Iva Adámková, Kateřina Vrševská, Barbora Kocánová und Hana Florianová, und Buch XI, an dem Barbora Kocánová, Hana Šedinová, Lenka Blechová-Čelebrić und František Šimon gearbeitet haben. Die Enzyklopädie des Bischofs von Sevilla erscheint im Prager Verlag Oikumene als Teil der Editionsreihe *Knihovna středověké tradice (Bibliothek der mittelalterlichen Tradition)* und das Projekt nähert sich schon seinem Abschluß, weil nur noch die Bücher V, X und XVIII auf Übersetzung und Kommentar warten. Es ist wahr, dass die Übersetzung und Einführungsstudien nur dem tschechischen Leser völlig verständlich sind, aber auch die andere Benutzer können von diesen Publi-

---

9 As a separate undertaking, however, and to facilitate access in the opposite direction, the forthcoming second volume of the *Lexicon* (non-Classical vocabulary, letters I to Z) will be accompanied with a digital version of all the entries (letters A to Z) in which, along with the codes identifying them, the passages containing at least the principal examples will actually be quoted from the texts.