

Pádraic Moran

Irish glossaries and other digital resources for early Irish studies

Abstract

Students of early Irish history, archaeology, language and literature are in many ways well served with digital resources. This paper outlines some of the key resources already available, and discusses a contribution currently under preparation by the *Early Irish Glossaries Project* at the *University of Cambridge*. It also makes some general remarks on potential future directions for electronic resources in the field.

1. Introduction: Digital resources for early Irish studies

Students of early Irish history, archaeology, language and literature are in many ways well served with digital resources.¹ The long-running *Corpus of Electronic Texts* (CELT) [9], based at *University College, Cork*, currently provides access to more than 1,000 historical and literary documents, many of which are directly relevant to early Irish studies. More recently this has been supplemented with the collections of texts available at *The-saurus Linguae Hibernicae* (TLH) [29], at *University College, Dublin*. Many manuscript witness are accessible through high-resolution images at the *Irish Script on Screen* (ISOS) [20] and the *Early Manuscripts at Oxford University* [14] websites.² Scanned texts are available from the *Internet Archive* [19], *Google Books* [17] and the *Celtic Digital Initiative* [6]. The earliest witnesses for the Irish language, the ogam inscriptions, are catalogued at the *TTTUS Ogamica* website [16] (partly) and the *Celtic Inscribed Stones Project* [7].

¹ The following list is an outline survey, and is not intended to be comprehensive.

² See especially the latter's Laud and Rawlinson manuscripts in the Bodleian Library collection. Images of some other important manuscripts are available elsewhere, for example St Gall, Stiftsbibliothek, MS 904 [10], or the Würzburg, Universitätsbibliothek, MS M.p.th.f. 12 [28], both containing large corpora of Old Irish glosses.

Reference works include a digital edition of the fundamental lexicographical resource for early Irish, *Contributions to a Dictionary of the Irish Language* [15], launched by the *University of Ulster* in 2007. The standard index on people and place names in early Ireland, *Onomasticon Goedelicum* (Hogan, 1910) [24], has been made available by the *Locus project* at *University College, Cork*, who are currently working on a modern replacement.

A smaller amount of secondary literature is becoming available through some of the resources already listed [6][17][19], in addition to broader collections such as JSTOR [21], and the websites of individual publishers. Bibliographical resources are also available, most notably the *Dublin Institute for Advanced Studies’ Bibliography of Irish Linguistics and Literature* [3] and the CSANA *Celtic Studies Bibliography* [8].³

Key resources for the study of the Hiberno-Latin tradition have been made available through the work of the *Dictionary of Medieval Latin from Celtic Sources* (DMLCS) [12], based at the *Royal Irish Academy*.⁴ The first edition of their *Archive of Celtic Latin Literature* (ACCL) was released on CD-ROM,⁵ and in 2007 became available on the Brepolis website [5]. The project promises two further, expanded editions of the archive, which will eventually be integrated with an electronic version of the finished *Non-Classical Lexicon of Celtic Latinity*, the first volume of which has already appeared in print.⁶ In 2008, the DMCLS undertook a promising new initiative, the St Patrick’s *Confessio* Hypertext Stack Project [26], which will integrate a range of existing resources for this important text.

This article discusses a contribution by the Early Irish Glossaries Project (EIGP) [2] at the *University of Cambridge*, which is working to establish new editions of three Irish-language texts. Part of the project’s scope is to create electronic resources that will make available transcriptions of all the relevant manuscripts, integrated with other existing resources as far as possible, and supplemented with custom-built search and concordance tools that will allow readers to locate material and better understand the relationships texts. In this way the project aims to exploit as far as possible the potential of the electronic medium, moving beyond the limitations of the printed format. The article first outlines the main fea-

³ The CELT website [9] also provides excellent bibliography on individual texts, and there are very useful resources for narrative literature in Irish at MsOmit [11], Scéla [22] and the Cycles of the Kings [32].

⁴ The DMLCS embraced the potential of electronic tools from its inception: Cf. Devine/Harvey/Smith (1987). CELT [9] also contains a small number of Hiberno-Latin texts.

⁵ Harvey 1994.

⁶ Harvey/Power 2005.

tures of the EIGP editions, and goes on to discuss the project in the context of other digital resources available for early Irish studies, offering some suggestions for future directions.

2. Early Irish glossaries

While this is not the place to discuss the value of early Irish glossaries for the study of medieval Irish culture, an outline of their general characteristics may be instructive.⁷ The principal texts are known as *Sanas Cormaic* (*Cormac's glossary*), *O'Mulconry's glossary*, and *Dúil Dromma Cetta* (the collection of Drumcett).⁸ The first of these is traditionally associated with the king and bishop of Cashel, Cormac mac Cuilennáin, who died in 908, although the reliability of this ascription may be called into question.⁹ The authorship of the other texts is not known.¹⁰

The glossaries are written primarily in Irish, albeit with a frequent admixture of Latin phrases, and many references to words in Latin, Greek, Hebrew, Brittonic, Norse, English and even Pictish. They each comprise between around 650 and 1,300 discrete entries, each with an Irish headword followed by discussion ranging in length from a single-word gloss to prose narrative extending across several pages of printed text. The following entries from *Cormac's glossary* are by way of illustration:¹¹

- Y 822 Lín a lino. (*Lín* [Irish: ›flax, linen‹] from *linum* [Latin: ›flax, linen‹].)
- Y 523 Enbroth .i. en usce ⁊ broth arbor; brot autem nortmanica est lingua. (*Enbroth* [Irish: ›gruek?‹], that is, *en-* ›water‹ and *-broth* ›grain‹; *brot* [Norse: *braud* ›bread‹] moreover is Norse language.)
- Y 751 Ídol ab idolo; idos isin gréic, forma isin laitin; unde idolum .i. delba ⁊ arrachta inna ndúlae do gnítis in geinti. (*Ídol* [Irish: ›idok‹, from *idolum* [Latin: ›form‹]; εἶδος in Greek, ›form‹ in Latin; from which *ido-*

⁷ For surveys of texts and manuscripts see Mahon (1987: ch. 1), Russell 1988 and Moran (2007: ch. 1). For an early survey of the genre see Stokes (1862: v-lxxv). A full bibliography for early Irish glossaries is available at [2].

⁸ The principal editions of these texts are currently Meyer 1913, Stokes 1900 and Binchy (1978: ii 604.39-622.12), respectively.

⁹ Cf. Moran (2007: 20-25).

¹⁰ O'Mulconry's glossary was named almost arbitrarily by its editor (cf. Stokes [1900: 232]), presumably confusing the author with the scribe of the only complete extant copy.

¹¹ Numbering follows the edition of Meyer 1913 based on manuscript Y.

Jahrbuch für Computerphilologie 10 (2008), Moran über “Irish Glossaries”

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

lum [Latin: ›idol], that is, the forms and representations of the idols which the heathens used to make.)

These few examples may give some flavour of the content: explication of Irish words, frequently beginning with some sort of etymological analysis, with notes touching on topics including native law, poetry and poetics, traditional history, and other aspects of the curriculum of medieval Irish schools.

Glossaries may be categorised as open recensions: texts that were subject to expansion, conflation and abridgement throughout the course of their transmission. Accordingly dating is problematic. Linguistically, a core appears to be Old Irish (pre-900), with a few elements dateable to the early eighth century. However, there is also a significant number of Middle Irish forms (circa 900-1200), and probably later elements. Our manuscripts are generally considerably later, for the most part written between the 14th and 16th centuries, and generations of scribes were liable to update older linguistic forms or to impose archaic spellings, producing a distinctly variegated picture. There is also considerable textual overlap between the glossaries, indicating that the extant recensions have been through an editorial mill, combining and reworking individual entries from a variety of pre-existing sources. This textual overlap makes it expedient to edit the three main glossaries together. However, given that each glossary has a distinct manuscript transmission, the project will produce a separate edition for each text, without attempting to reconstruct any hypothetical archetype not directly found in the manuscript tradition.

2.1 Rationale for digital edition

The characteristics outlined above are well suited to digital editing. The structure of the texts is well defined, and apt to being marked up. Individual glossary entries tend to be clearly distinguished in the manuscripts – having larger, sometimes coloured, initials, and frequently beginning on a new line.¹² Within an entry, the headword is generally clearly separated with a linking formula such as ›.i.‹ (*id est*) or similar. Entries are further

¹² See for example, the images of Oxford, Bodleian Library, MS Laud Misc 610, fols. 79r on, at [14].

Jahrbuch für Computerphilologie 10 (2008), Moran über “Irish Glossaries”

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

grouped into alphabetical blocks, each new block marked with a large letter initial.

As a category of text, glossaries work well in the digital medium. Glossaries may be counted with annals, genealogies, commentaries and other types of scholia that are effectively compilations or lists. Taken in their entirety, such texts have little literary merit, and can instead be rambling and indigestible. Very few scholars will read a glossary from start to finish, and most will prefer to dip in, to search for entries relevant to a topic of interest, or for early attestations of a particular word. Therefore, while search tools can be useful but not essential resources for some types of digital text, with compilations such as glossaries they can unlock the text and provide access in ways not possible in print.¹³

Perhaps the most significant advantage of publishing these glossaries in digital format, however, lies in the ease of cross-referencing. As noted above, the overlap between texts is such that the same entry often occurs in more than one glossary. In all versions, many entries have been corrupted or significantly reworked – sometimes only by exploring an entry in several different versions can an original sense be recovered. By providing concordance tools that cross-reference and collate different versions, readers have an opportunity to understand entries in a fuller context.

The concordance tools are also valuable for exploring the textual history of each individual text. In common with much of early Irish literature, our texts survive mainly in manuscripts much later than the time of their original composition. The accompanying stemmatic diagram illustrates the situation with Cormac’s glossary.

¹³ Printed glossaries rely heavily on indices, which are by their nature selective.

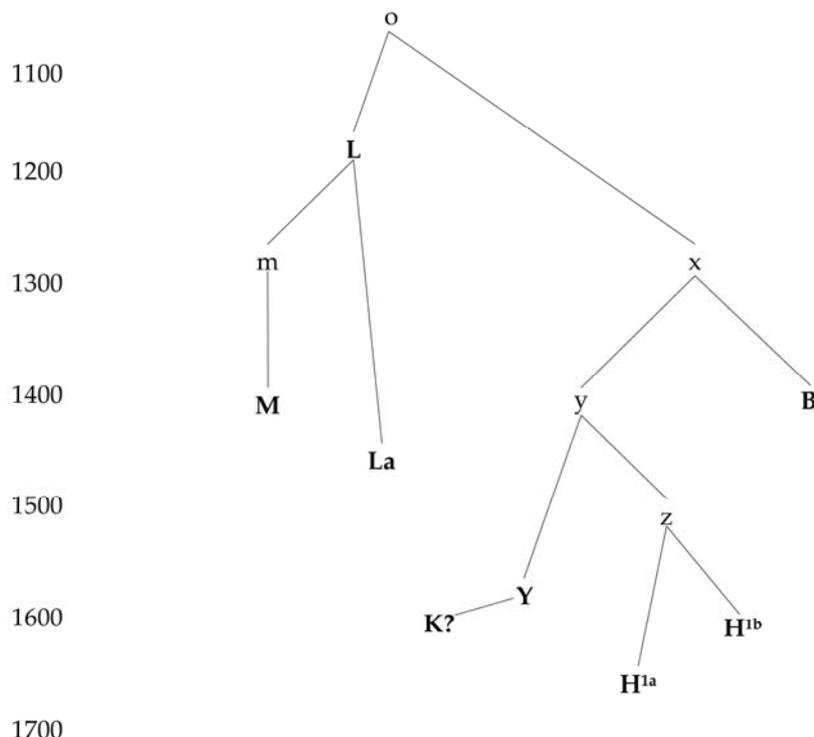


Figure 1: Manuscript stemma for *Cormac's glossary*¹⁴

The oldest manuscript is *L*, a section of Book of Leinster written around 1186. This would form a good basis for the edition, were it not a tiny fragment, containing just 21 entries out of around 750. The next oldest are *B* and *M*, both written around 1400: the language of the former has been significantly updated, and the latter is often so corrupt as to be completely nonsensical. *La* (circa 1450) preserves an early text, but it is missing its first half. We are left with *Y* and related manuscripts, the text of which is clearly modernised and reworked in some areas, but less so than *B*, and being the most complete version gives a representative picture of the overall textual tradition. Basing the edition on this group seems preferable to the patchwork effect of compositing other fragments. However, readers interested in the earliest versions of individual

¹⁴ Based on Russell (1996: 156), with a manuscript timeline, and omitting short witnesses for single entries.

Jahrbuch für Computerphilologie 10 (2008), Moran über “Irish Glossaries”

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

entries will very likely prefer to cite *L* where possible, or alternatively *La*, or perhaps *M* with caution. The digital edition can therefore present the evolution of the text and give readers access to specific branches of the stemma in a way far more accessible than a traditional manuscript apparatus.

2.2 Digital versus print

On the other hand, a digital edition has some drawbacks. The web may indeed be worldwide, but books arguably still retain a broader readership in more traditional fields of scholarship. A lack of engagement is manifest in the way digital resources are still entirely excluded from the peer review process. I am not aware of any written reviews of the resources outlined at the start of this paper.¹⁵ Given this state of affairs, should we publish our work only in electronic format, there is every likelihood that it would be completely overlooked in scholarly publications. And without peer review, would other scholars regard it as a serious contribution? Would they ever cite it? There are also practical implications. How would the resource be treated in Research Assessment Exercises and similar? Those of us on research contracts might wonder what standing it would have with hiring committees for academic jobs. Whatever the merits of electronic publication, I think it is fair to say that printed books, at least for the moment, are treated more seriously in an academic culture that is not generally known for being quick to embrace innovation.

With these considerations in mind, we opted for a dual (or hybrid) format, combining print and digital elements. The print edition will follow the traditional scholarly format: edited text with critical apparatuses, translation, commentary and introductory material. The digital edition will complement this with manuscript-related resources: full transcriptions and manuscript images, with search tools, customisable concordances, and integrated dictionary resources.¹⁶

¹⁵ My own review of the online Bibliography of Irish Linguistics and Literature is due to appear in a forthcoming edition of *Peritia: Journal of the Medieval Academy of Ireland*.

¹⁶ We must also take into account the commercial imperatives of the publisher, who is reluctant to publish material otherwise freely available on the web.

3. Features of digital edition

3.1 Manuscript transcriptions

The digital resource currently under development, and due for launch in July 2009, is an expansion of a pilot project launched online in August 2006 [2]. That resource comprised lists of headwords only for each manuscript version of each glossary, and was intended to provide an index for locating specific glossary entries, while concordance tools allowed users to explore the structural relationships between texts. The pilot project is now being expanded to include complete manuscript transcriptions.

An important aspect of this expansion is the aim to integrate with other resources as far as possible. Each glossary entry links to a manuscript image at the relevant page.¹⁷ Clicking on any word in the text will link to the online *Electronic Dictionary of the Irish Language* (eDIL) [15]. For each glossary headword, we have identified the corresponding lemma within eDIL, sometimes distinguishing between numbered homonyms (for example, between *dobur 1* ›dark, unclean‹ or *dobur 2* ›water‹). We have not undertaken to do this for every word in the glossaries. However, because eDIL very often lists a variety of grammatical forms and attested orthographies under each headword, and moreover supplies comprehensive citations (including from the glossaries themselves), the relevant dictionary entries can usually be readily identified by the reader.¹⁸

These manuscript transcriptions will be available in XML format encoded according the TEI P5 guidelines [27]. In deciding what aspects of the text to mark up, our choices were largely governed by considerations of time, available resources and greatest utility. Initial aspirations for a close and highly descriptive mark-up were scaled back after exploratory

¹⁷ Manuscript images are mostly available from ISOS [20] and the Oxford collection [14]. Where not otherwise available, we have made scans from microfilm for internal use only, to assist the work of editing.

¹⁸ For Latin words, the challenge of locating the relevant dictionary entry without individually associating each word is greater. Inflected forms may be identified (though sometimes with ambiguities) with resources such as the *Persens Morphology Tool* [25]. However, its effectiveness is reduced by the incidence of medieval Latin spellings and scribal corruptions. Links to dictionaries for other languages – such as Greek, Hebrew or Brittonic – are more problematic still and potentially misleading: these words are best treated on an individual basis in the textual commentary.

work showed that this was not achievable within the project schedule with the resources available. Moreover, the utility of this approach seemed uncertain: manuscript images were readily available for reference, there seemed to be no particular scholarly application for rich descriptive mark-up, and it seemed better to apply resources to the very basic requirement of establishing a reliable text and translation, with an appropriate commentary to guide the reader.

As a result, a light approach to mark-up has been adopted. For example, for manuscript references we have supplied only page and column details for each glossary entry. We have not recorded the precise location of every entry on the page, which readers must locate for themselves. With around 8,600 entries in total, and given that headwords are commonly marked with large initials, this seemed like the most practical approach.

A more unusual choice was not to mark up every manuscript suspension, contraction and abbreviation. These are certainly used liberally in all Irish manuscripts, and are invariably marked with italic type in printed diplomatic editions. However, their very ubiquity might itself diminish the importance of marking every occurrence. Suspensions and contractions such as *dt* with overstroke for Latin *dicit* ›says‹, *dr*-stroke for Latin *dicitur* ›is said‹, *di*-stroke for Irish *didiu* ›thus‹, *f*-stroke for Irish *for* ›on‹, *n*-stroke for Latin *non* ›not‹, straight and curved strokes for *n* and *m* contractions, respectively, and sigla such as the H symbol for Latin *autem* ›moreover‹, reversed *c* for *con*-, and 2-like siglum for Latin *est* ›is‹ are very commonplace and rarely contestable. Wherever there is any ambiguity it has been duly marked.

The transcriptions will encode all divisions of text: letter blocks, entry divisions, headwords, column and page breaks. Marginal and interlinear glosses are recorded as such, as are supplied text, corrections, deletions, manuscript gaps, irregular spaces, and unclear or damaged areas of writing.¹⁹ We intend also to mark personal and place names, as well as foreign (non-Irish) words in the text. Departures from conventional transcription may in part be justified by the ready availability of high-quality manuscript images. Manuscripts are ultimately the foundation of medieval textual scholarship, and there is arguably no substitute for getting to grips with the original context.

¹⁹ Other aspects of the transcription follow standard practices for early Irish texts, with modern word separation, use of hyphens (to distinguish initial mutations before vowels), and commas and semicolons for readability. The transcriptions retain the Tironian signs 7 (Latin *et* ›and‹) and 1 (Latin *uel*, Irish *nó* ›or‹).

3.2 Concordances

The second major feature of the Irish glossaries database is a tool for dynamic generation of glossary concordances. By selecting a base text, and any number of others to compare against it, we can explore the relationships of different versions to each other. The applications are two-fold. Firstly, a reader can compare readings for the same glossary entry across all the texts which contain it, and thus observe textual variations, sometimes substantial, from one manuscript to the next.²⁰ This might involve entries in one glossary text only (for example, *Cormac's glossary*), or entries found in more than one text (for example, *Cormac's glossary* and *O'Mulconry's glossary*). Secondly, the concordances allow readers to explore the broader structural relationships between different glossary texts, investigating the composition of texts by observing common blocks of entries.

3.3 Flexible searching

One of the most useful attributes of any electronic text is the ability to search for specific words or phrases. For early Irish and other vernaculars, however, this can be problematic owing to wide variations in orthography. To take one unremarkable example, the word cited in DIL s.v. *lóchar(n)* meaning ›lamp‹ (from Latin *lucerna*), is a headword in Cormac's glossary, spelt variously in manuscripts as follows: *locharn* (B), *luacarn* (H1a), *luacharn* (H1b), *luacharnn* (La), *lochbrann* (M), *luacharn* (Y). (To confuse the issue further, vowels are frequently ligatured, and the medieval Irish approach to accentuation might be regarded as erratic at best.) Many such variations are explainable by reference to sometimes conservative, sometimes hypercorrect scribal practice in the face of historical developments in Irish phonology. Conventional word-searching in electronic resources is bound to failure unless one has the knowledge with which to predict the forms that might typically exist for any given word. Even with the flexibility of Boolean operators or regular expressions, searches are at best time-consuming and repetitive; at worst, it is all too

²⁰ This feature will usefully supplement the critical apparatus destined for the print editions.

Jahrbuch für Computerphilologie 10 (2008), Moran über “Irish Glossaries”

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

easy to miss potentially important matches, and inadvertently make assumptions based on an incomplete picture.

To address this issue, we have developed a search tool that takes typical patterns of variation into account. At present any word-search takes into account 22 pattern-matching rules, and for longer words the patterns become increasingly complex. Thus, with reference to the example given above, a search for ›lócharn‹ generates the following search pattern:

`/l(f|fh|ph)?(o|ó|ua)+i?ch?(f|fh|ph)?(a|á)?i?rn(n|d)?/`

The search tool is not intended as a substitute for linguistic knowledge. In fact, it regularly returns matches that are linguistically impossible. This apparent imprecision jars somewhat with a training in historical linguistics, which rightly emphasises precision and solid methodology. However, as long as it is currently impossible to fine-tune any search mechanism to provide word-perfect results, the search feature is offered as a tool, an aid to research. Accordingly, it is calibrated to allow researchers the opportunity to discriminate genuine from irrelevant matches, without overlooking potentially important ones.

4. General desiderata

As outlined at the start of this paper, the digital resource created by the *Early Irish Glossaries Project* does not stand in isolation, but is one of a range of electronic resources available to researchers in early Irish studies. Most of these projects have developed more or less piecemeal over the past decade, and show some variety in approach. As long as there is no peer review for digital resources, there is very little opportunity for feedback and constructive criticism. Accordingly, there follow here some general observations that have emerged from challenges and discussions arising from the glossaries project. It should be emphasised that these remarks are in no way intended to diminish what has often been pioneering work in the field. Moreover, standards and expectations are much higher today than when some of these projects originated, and funding structures generally weigh in favour of expansion over consolidation. However, it is hoped that by encouraging discussion we might at least identify current opportunities and future aspirations.

4.1 Interface design and integration

With the extent and variety of resources currently available (textual, lexicographical, manuscript images, bibliographical, et cetera), there is excellent potential for better integration. Many of the texts on CELT [9], for example, are found in the manuscripts digitised on ISOS [20] and the Oxford collection [14], and some are in fact diplomatic transcriptions. Interlinking these resources more thoroughly would mean not only greater efficiency and productivity for researchers, but would also draw attention to supplementary resources where available, and encourage more frequent and perhaps more imaginative use of them.

Unfortunately, integration with some resources is made more difficult because of certain technical choices. Use of frames, for example, generally means that an individual page (within a frameset) cannot be accessed in its original context on a distinct URL (for example, [3][7][9][14][20]). Likewise, search results pages that rely on server-side session variables (as in [5][15]) do not have a permanent URL, and therefore cannot be directly referenced.²¹

The design of user interfaces may be improved in other ways. For example, where translations of texts are readily available, why not lay these out in parallel with the original text? Even such a simple adaptation would be significantly more helpful for readers. We might hope for more refined search tools that take into account not just common orthographical variations as discussed earlier, but also the distinctive characteristics of different types of texts, and accordingly the varying needs of researchers. For example, when searching the extremely copious body of Irish annals, why not restrict searches to within date ranges, say a specific century or the life span of a particular individual? This would be of immense utility in locating material in, for example, the *Annals of Ulster*, which run from AD 431 to 1541.²² Why not search across all versions of the annals and automatically collate the various entries by year, to allow for easy comparison? After all, this is what any historian will inevitably need to do. In general, the one size fits all approach currently prevalent could be reconsidered with a view to the fine-tuning the presentation and search interface to the distinct characteristics of different texts.

²¹ For similar reasons, the use of frames and sessions variables means that web pages are either entirely invisible to search engines or cannot be accessed satisfactorily.

²² Available on CELT [9]. I should acknowledge the immense value in these annals being available in any searchable format on the CELT website.

There are some general standards of best practice, in web design particularly, that may be followed more closely. If we unanimously endorse the TEI standard for mark-up, why not the more general standards of the *World Wide Web Consortium* (W3C) [31], in particular XHTML and CSS?

The term »accessibility« may refer to the unrestricted availability and ease-of-use of content on the web, or in a more specific sense to technical standards and other forms of best practice intended to ensure electronic resources can be used by people with disabilities such as visual or hearing impairment, or reduced motor skills. The W3C, for example, has published guidelines and discussion documents under its *Web Accessibility Initiative* [30].²³ Many of the W3C accessibility standards in fact conform to best practices in web design in general, and are consistent with universal design principles – such as simplicity, clarity and tolerance for error – which aim to make resources intuitive and easy to use for everyone.²⁴ These principles can be taken into account in constructing the interactive elements of a user interface. Navigation, for example, should be succinct, clearly presented, and displayed consistently on every page.²⁵ And users navigate in two directions: the back button on browsers, an excellent example of error tolerance, should not be obstructed through the use of session variables.

Good visual design does not merely reflect personal taste or passing trends, but follows certain universal principles, many of which have been established over centuries of typography and book production. On the other hand, dark background colours and intrusive background images, unrestrained use of colour, generic knotwork designs and other clip art, pseudo-historical typefaces and 3D baubles do not improve the reader’s experience. The poor quality of visual design prevalent in scholarly electronic resources in general may be attributable to the DIY ethic that has always informed web publishing, given the low technical barrier and freedom from traditional restraints of print. Responsibility for graphic design is therefore given to technical or content specialists, rather than trained graphic designers, no doubt partly also for budgetary reasons.²⁶ It

²³ Ireland’s *National Disability Authority* publishes similar guidelines on AccessIT [1].

²⁴ In this broader usage, the term »accessibility« is sometimes interchanged with »usability«. For an excellent introduction to user-centred design, see the Web Style Guide [23].

²⁵ Consistent navigation is all the more important given that many visitors will come directly to a page via search engine results.

²⁶ Web development agencies in the commercial sector typically have a range of specialists focused entirely on aspects of graphic design and interface development.

Jahrbuch für Computerphilologie 10 (2008), Moran über “Irish Glossaries”

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

may seem trivial to criticise a more cosmetic aspect given the serious scholarly endeavour that underlies many of these projects. However, if we aspire for electronic resources to have the broad readership and the credibility of mainstream academic publication, we ought to ensure that their standard of presentation is at least as high as for print publication.

4.2 Target content

The CELT project [9] above all has distinguished itself in pioneering the distribution of early Irish texts online, making an enormous number freely available to those without access to a research library. More recent projects such as the TLH [29] are adding to this growing corpus. However, many of these texts are also now becoming available through other enterprises, operating outside the field of early Irish studies. Book digitisation projects such as *Google Books* [17] and the *Internet Archive* [19] are making many scans of out-of-copyright books freely available to download,²⁷ including out-of-copyright nineteenth- or early twentieth-century editions of early Irish texts, often the same editions used for the creation of electronic editions. The scanned format carries advantages and disadvantages. Though searchable, the machine-readable text in scanned books is far less accurate than in electronic editions which have been painstakingly proof-read. On the other hand, the readable text of scanned books is arguably less likely to contain errors, as the images are extremely faithful to the original books. Moreover, they are more easily printable, and preserve the better standards of typography of the printed medium. The greatest benefit in using scanned books, however, is in the supplementary content they contain which is typically omitted from electronic editions: in particular, critical and other apparatuses, indices and textual commentaries, scholarly introductions. These resources are essential for any serious textual research, and without them, current electronic editions are often a first port of call before recourse to the original books. With the rapidly increasing availability of scanned books, many of these electronic editions, in their current format at least, may soon become obsolete.

²⁷ Microsoft’s Live Search Books project was terminated in May 2008, having by then scanned more than 750,000 books [4].

Jahrbuch für Computerphilologie 10 (2008), Moran über "Irish Glossaries"

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

To realise the full potential of electronic editions (text-based rather than scanned resources), it may be worth focusing attention on aspects which break from the traditional linearity of the printed medium. The suggestions already made in relation to integration with other resources, customised interfaces and search tools would enhance their value significantly. The scope of content in electronic editions could also be reconsidered, particularly with a view to incorporating more supplementary resources, such as diplomatic manuscript readings. This would give readers deeper insights into the textual history of a work than is possible with a traditional critical apparatus, and particularly benefit texts with a complex transmission or contaminated recensions.²⁸

Some of the most fundamental resources in the field have not yet been digitised, and would particularly benefit from new approaches such as those outlined above. Most notable are the *Thesaurus Palaeohibernicus*,²⁹ still the indispensable collection of reference material for Old Irish, and the *Corpus Iuris Hibernici* (Binchy 1978), a six-volume collection forming the basis of study for early Irish law.³⁰ Historians would profit from greater access to the extensive corpus of genealogies, much of which has not yet been published in any medium. An obvious starting point would be a searchable and cross-referenced edition of the *Corpus Genealogiarum Hiberniae*.³¹ Among reference literature, Thurneysen's standard reference grammar for Old Irish (1946) might also be considered.

4.3 Scholarly community

As previously noted, the lack of broader engagement with digital resources in early Irish studies seems to be reflected in their exclusion from the traditional peer review process. There may be several reasons for this. How do you review an electronic publication at all? What aspects should given consideration? Who would be best qualified to assess them? In the first place, the basic content of an electronic resource should presumably be subject to the same kind of rigorous assessment given to print publications, and would therefore require only a qualified

²⁸ This is the main rationale for the digital aspect of the Early Irish Glossaries Project.

²⁹ Stokes/Strachan 1903.

³⁰ Aaron Griffith has made his revised edition of part of the *Thesaurus* available online [18].

³¹ O'Brien 1962.

Jahrbuch für Computerphilologie 10 (2008), Moran über "Irish Glossaries"

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

specialist in a relevant academic field. For some projects, where a digital resource republishes material that has already been available in print and subjected to peer review, this aspect may not need reviewing at all. A second aspect may relate to its technical infrastructure: for example, compliance with various standards and best practices, functionality on different platforms and software, compatibility with other electronic resources, software bugs and other unexpected occurrences. This level of testing would ideally require the expertise of a software developer or digital humanities specialist. There may also be a third aspect to consider, lying somewhere between the previous two: does the scholarly content work effectively within the medium? In other words, would someone with average technical skills find the resource simple and intuitive to use? An adequate review may therefore fall outside the competency of a single reviewer, and it would be useful to bring together experts in various disciplines, both technical and non-technical, with a view to establishing some sort of guidelines.

There is certainly much to be gained through discussion and collaboration between the various projects outlined at the start of this article and more. Any effective integration of resources will require both imagination and technical collaboration. Part of this process would be to form a closer consensus on the application of standards and best practices to follow. This might best be realised through a network that could in turn provide a forum for new ideas, feedback on work in progress, and advice for new and prospective projects. A comprehensive bibliography of digital resources for early Irish studies, set out according to clear editorial criteria, would be a useful starting point.³²

5. Conclusion

In the light the absence of any review process or forum for discussion for resources in early Irish studies, this paper has hoped to stimulate some discussion about current standards and best practices, as well as appropriate directions for future development. It is hoped that the Early Irish Glossaries Project will make a useful contribution to the field, and that its experience may be instructive for other projects. Researchers in

³² It is hoped that the new Digital Humanities Observatory [13] will have a active role in drawing projects together.

Jahrbuch für Computerphilologie 10 (2008), Moran über "Irish Glossaries"

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

the field of medieval Irish studies are fortunate, not only because progress has been made, but because much important work remains to be done.

References

- Binchy, Daniel A. (ed.)
1978 *Corpus Iuris Hibernici. Ad fidem codicum mancriptorum recognovit.* 6 vols. Dublin: Dublin Institute for Advanced Studies.
- Devine, Kieran/ Anthony Harvey/ Francis J. Smith
1987 *The Database of Medieval Latin from Celtic Sources. A study in computer-assisted lexicography.* Dublin: Royal Irish Academy.
- Harvey, Anthony (ed.)
1994 *Archive of Celtic-Latin Literature.* Turnhout: Brepols. [CD-ROM]
- Harvey, Anthony/ Jane Power (ed.)
2005 *The Non-Classical Lexicon of Celtic Latinity. Volume 1: Letters A-H.* Turnhout: Brepols.
- Hogan, Edmund (ed.)
1910 *Onomasticon Goedelicum Locorum et Tribuum Hiberniae et Scotiae.* Dublin: Hodges, Figgis & Co.
- Moran, Pádraic
2007 *Sacred Languages and Irish Glossaries. Evidence for the study of Latin, Greek and Hebrew in early medieval Ireland.* Unpublished PhD thesis, National University of Ireland, Galway.
- Mahon, William J.
1987 *Contributions to the Study of Early Irish Lexicography.* Unpublished PhD thesis, Harvard University.
- Meyer, Kuno (ed.)
1913 *Anecdota from Irish Manuscripts 4.* Dublin: Hodges Figgis (facs. repr. Felin-fach: Llanerch, 1994).
- O'Brien, Michael A. (ed.)
1962 *Corpus Genealogiarum Hiberniae. Vol. 1.* Dublin: Dublin Institute for Advanced Studies.
- O'Sullivan, William
1966 Notes on the script and make-up of the Book of Leinster. In: *Celtica* 7, pp. 1-31.
- Russell, Paul
1988 The Sounds of a Silence: The Growth of Cormac's glossary. In: *Cambridge Medieval Celtic Studies* 15, pp. 1-30.
1996 *Dúil Dromma Cetta* and Cormac's Glossary. In: *Études Celtiques* 32, pp. 147-74.

Jahrbuch für Computerphilologie 10 (2008), Moran über "Irish Glossaries"

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

Stokes, Whitley (ed.)

1862 Three Irish Glossaries. London: Williams & Norgate (facs. repr. Felinfach: Llanerch, 2000).

1900 O'Mulconry's Glossary. In: Whitley Stokes and Kuno Meyer (eds.): Archiv für celtische Lexicographie 1. Halle: Niemeyer, pp. 232-324, 473-81, 629.

Stokes, Whitley/ John Strachan (eds.)

1903-1910 Thesaurus Palaeohibernicus. A Collection of Old Irish Glosses, Scholia, Prose and Verse. 2 vols & supplement. Cambridge: University Press (repr. Dublin: Dublin Institute for Advanced Studies, 1975, 1987).

Thurneysen, Rudolf

1946 A Grammar of Old Irish. Translated by D. A. Binchy and Osborn Bergin. Dublin: Dublin Institute for Advanced Studies. (Translated and revised edition of Handbuch des Altirischen. Grammatik, Texte und Wörterbuch. 2 vols. Heidelberg: C. Winter, 1909.)

Websites

- [1] AccessIT. IT Accessibility Guidelines <<http://accessit.nda.ie/>> (21.02.2008).
- [2] Sharon Arbuthnot/Pádraic Moran/Paul Russell: Early Irish Glossaries Project (EIGP) <<http://www.asnc.cam.ac.uk/irishglossaries/>> (21.02.2008).
- [3] Bibliography of Irish Linguistics and Literature <<http://bill.celt.dias.ie/>> (21.02.2008).
- [4] »Book search winding down«. At: Live Search. The official blog of the Live Search team at Microsoft <<http://blogs.msdn.com/livesearch/archive/2008/05/23/book-search-winding-down.aspx>> (21.02.2008).
- [5] Brepolis. Brepols Publishers Online <<http://www.brepolis.net/>> (21.02.2009).
- [6] Celtic Digital Initiative <<http://www.ucc.ie/academic/smg/CDI/>> (21.02.2009).
- [7] Celtic Inscribed Stones Project <<http://www.ucl.ac.uk/archaeology/cisp/>> (21.02.2009).
- [8] Celtic Studies Association of North America (CSANA): Celtic Studies Bibliography <<http://www.humnet.ucla.edu/humnet/celtic/csanabib.html>> (21.02.2009).
- [9] Corpus of Electronic Texts (CELT) <<http://www.ucc.ie/celt/>> (21.02.2009)
- [10] Codices Electronici Sangallenses <<http://www.cesg.unifr.ch/>> (21.02.2009).
- [11] Johan Corthals: Manuscript Sources to Old and Middle Irish Tales (MsOmit) <<http://www.ucc.ie/celt/MS-OMIT/index.htm>> (21.02.2009).
- [12] Dictionary of Medieval Latin from Celtic Sources <<http://journals.eecs.qub.ac.uk/dmlcs/>> (14.08.2008)
- [13] Digital Humanities Observatory <<http://www.dho.ie/>> (21.02.2009).
- [14] Early Manuscripts at Oxford University <<http://image.ox.ac.uk/>> (21.02.2009).
- [15] Electronic Dictionary of the Irish Language <<http://www.dil.ie/>> (21.02.2009).
- [16] Jost Gippert: Ogam Inscriptions <<http://titus.uni-frankfurt.de/ogam/>> (21.02.2009).
- [17] Google Books <<http://books.google.com/>> (21.02.2009).
- [18] Aaron Griffith: A Dictionary of the Old-Irish Glosses in the Milan Codex Ambrosianus C 301 <<http://www.univie.ac.at/indogermanistik/download/projekte/milan.htm>> (21.02.2009).
- [19] Internet Archive <<http://www.archive.org/>> (21.02.2009).

Jahrbuch für Computerphilologie 10 (2008), Moran über "Irish Glossaries"

<<http://computerphilologie.de/jg08/moran.pdf>> (30. April 2009)

[20] Irish Script on Screen (ISOS) <<http://www.isos.dias.ie/>> (21.02.2009).

[21] JSTOR <<http://www.jstor.org/>> (14.08.2008).

[22] Štěpán Kosík: Scéla. Catalogue of medieval Irish narratives & literary enumerations <<http://volny.cz/enelen/>> (21.02.2009).

[23] Patrick Lynch/Sarah Horton: Web Style Guide, 2nd edition <<http://www.web-styleguide.com/>> (21.02.2009).

[24] Onomasticon Goedelicum <<http://publish.ucc.ie/cocoon/doi/locus>> (17.03.2009).

[25] Perseus Digital Library: Morphology Tool <<http://www.perseus.tufts.edu/hopper/morph.jsp>> (21.02.2009).

[26] St Patrick's *Confessio* Hypertext Stack Project <<http://www.dho.ie/confessio.html>> (14.08.2008).

[27] Text Encoding Initiative: P5 Guidelines <<http://www.tei-c.org/Guidelines/P5/>> (21.02.2009).

[28] Thesaurus Indogermanischer Text- und Sprachmaterialien (TITUS). The Old Irish ›Würzburg‹ Glosses <<http://titus.fkidg1.uni-frankfurt.de/texte/celtica/wbgl/wbgl.htm>> (21.02.2009).

[29] Thesaurus Linguae Hibernicae (TLH) <<http://www.ucd.ie/tlh/>> (21.02.2009).

[30] W3C Web Accessibility Initiative <<http://www.w3.org/WAI/>> (21.02.2009).

[31] World Wide Web Consortium (W3C) <<http://www.w3.org/>> (21.02.2009).

[32] Dan Wiley: The Cycles of the Kings. <<http://www.hastings.edu/academic/english/Kings/Home.html>> (21.02.2009).

Pádraic Moran

Dept. of Anglo-Saxon, Norse and Celtic,

9 West Road, Cambridge CB3 9DP

pm394@cam.ac.uk

www.pmoran.ie