Gregory Toner, David Stifter,
Nina Cnockaert-Guillou, Beatrix Färber, Deborah Hayden,
Máire Ní Mhaonaigh, Joanna Tucker, Christopher Guy Yocum

Developing a Digital Framework for the Medieval Gaelic World

Report

Draft version – 3 March 2022

Table of Contents

Introduction

In 2020, the Research Network entitled 'Developing a Digital Framework for the Medieval Gaelic World' was established.[1] This project was funded by UKRI-AHRC and the Irish Research Council under the 'UK-Ireland Collaboration in the Digital Humanities Networking Call' (grant numbers AH/V00235X/1 and IRC/V00235X/1). The network aimed to to bring together scholars working across various aspects of medieval Celtic Studies in order to assess where we stand in terms of the digitisation of resources relating to medieval Ireland and Scotland, and to work towards a consensus on the way forward.

The network organised four workshops and a training event, which were all held virtually due to the ongoing COVID-19 pandemic. While the focus of these events remained on medieval Gaelic research, the workshops also drew on scholarly expertise from outside of the field of Celtic Studies, in the knowledge that there is still much to be learned from other disciplines in understanding the current 'state of the art' in digital humanities more broadly. Attendants often numbered close to a hundred, with many participants coming back to later workshops to take part again in our network's ongoing conversation. The online format chosen due to the pandemic opened up the workshop to a wide range of attendees who might not have been able to join us in person.

The first workshop, organised by the University of Cambridge, entitled 'The Changing Face of Research in the Digital Age', aimed to provide a starting point to reflect on progress in digitisation achieved thus far, identifying key areas for development. The second workshop, 'Representing Texts: from Material to Digital', organised by Maynooth University, focussed on the impact of different forms of textual representation on research. The third workshop, 'Exploring Texts: Revealing Hidden Heritage through Online Resources', organised by University College Cork, examined how digital resources have helped a wider audience engage with medieval Gaelic material and facilitated new forms of teaching. Finally, the fourth workshop, 'Conserving Electronic Resources: Sustainability in a Digital World', organised by the University of Glasgow, focussed on the key issue of sustainability, which was brought to the fore repeatedly at each of the first three workshops.

As many participants of each workshop voiced the need for more training opportunities, the network also decided to facilitate such an event to conclude its discussion, 'Digital Resources, Manuscripts and Texts: an Online Training Event focussing on London, British Library, Manuscript Harley 5280', which was organised by the University of Cambridge in September 2021.

The creation of an Inventory of Digital Projects[2] in advance of the first workshop helped to provide a first snapshot of progress, notwithstanding the fact that information provided was not uniform in all cases and that the projects are at very different stages in terms of active engagement, funding status and the like. An inventory of this kind updated annually would

---

[1] Project website: https://www.qub.ac.uk/schools/ael/Research/ResearchinLanguages/imdorus/ (accessed 27 October 2021).

[2] https://www.qub.ac.uk/schools/ael/Research/ResearchinLanguages/imdorus/DigitalProjectsfortheGaelicWorld/ (accessed 27 September 2021).

provide a useful resource and orientation map in terms of digitisation in the field.[3] As the issue of interoperability between the various projects emerged as a major theme, it was later considered by a consultant on behalf of the Research Network (for which see Appendix 2, Recommendations on Data Interoperability in Medieval Celtic Digital Projects).

The four workshops organised by the Network were praised by participants and attendees alike for creating the space to discuss important issues in the field of medieval Gaelic studies, and starting to propose ways of moving forward. While it was stressed that these conversations need to continue in order to affect real change and progress, we hope that this report provides a useful record of the current 'state of the art' of digital medieval Gaelic research.

---

[3] A document entitled 'Digital Tools for Medieval Gaelic Research', prepared by Pádraic Moran for the Training Event in Cambridge in September 2021, with input by other scholars, can be accessed and added to: https://docs.google.com/document/d/19z1w3DQO_eFKiDPFG7BEgDnJntzutlIDAZMJ1N0df9E/edit#heading=h.gy6a8xdj8x71.

I. Workshop 1, University of Cambridge, 'The Changing Face of Research in the Digital Age', 26–27 November 2020

Nina Cnockaert-Guillou and Máire Ní Mhaonaigh, July 2021

Introduction

The first workshop in a series of four, this meeting sought to explore how digitisation and digital humanities more generally have influenced how we carry out research in the field of medieval Gaelic studies, and have also shaped the underlying research questions we address in a more fundamental way. The starting point was to reflect on progress in digitisation achieved thus far, identifying key areas for development. The topic was addressed from the perspective of studying History and Texts, on the one hand, and Language and Linguistics, on the other. The methodologies of Artificial Intelligence and Data Analysis in particular were explored (see Appendix 1 for the workshop programme).

1. *The Nature of Research in the Digital Age*

a. The Relationship between Print and the Digital

The underlying truth that Digitisation and Digital Humanities are not one and the same was reiterated at various points throughout the workshop, an acknowledgement that many researchers, as well as those engaging with the field in other ways, continue to equate digital research with accessing resources online. The potential for utilising these online resources to structure and analyse data (of all kinds) in new ways and so ask different types of research questions is not being fully realised. Training events focussing on the possibilities of digital research have a part to play in changing this culture. Those creating digital resources can also assist in ensuring that the interface is as intuitive as possible and that there are clear instructions available directing users to the full searchability and so functionality of the resource in question.

Not all digital resources allow the same possibilities in this regard, some very useful online resources maintaining a stronger link than others to an underlying print format; the online version of Edmund Hogan's *Onomasticon Goedelicum* produced as part of the Locus project is an example of this type.[4] In very many cases, however, the online project has been augmented and developed in fundamental ways. This was illustrated clearly during the workshop by Mark McConville in relation to the *Digital Archive of Scottish Gaelic* (DASG).[5] A digitisation of archival material initially, further funding, as well as support from many organisations, has enabled its development into a multi-faceted, successful online resource which is regularly updated and continues to embrace new digital projects. Other online resources which were presented at subsequent workshops in Maynooth, Cork and Glasgow are

---

[4] https://research.ucc.ie/doi/locus (accessed 27 September 2021).
[5] https://dasg.ac.uk/en (accessed 27 September 2021).

similar in this regard; the *Electronic Dictionary of the Irish Language* (eDIL),[6] the *Corpus of Electronic Texts* (CELT)[7] and *Irish Sagas Online*,[8] for example, commenced life as digitisation projects of existing print resources, before becoming fully-fledged digital resources. User statistics in the case of eDIL, however, indicate that the advanced searches possible are not always utilised. It is hoped that a recent series of eight YouTube videos, 'A Guide to Using eDIL'[9] will aid researchers in understanding how they can use the resource to maximum effect. The complex relationship between print and the digital in our field underlies these issues, as the workshop highlighted at various points.

<p style="text-align:center"><u>b. Harnessing Accessible Resources</u></p>

The universal accessibility of digital resources has fundamentally changed the ways in which research is carried out, reducing the frequency with which recourse to a specialist library is required and 'democratising' research to a certain extent. This positive development was noted by many speakers in the workshop, as was the fact that true accessibility means that digital resources should be easily searchable (by means of an 'intuitive' interface, as noted above). Efficient and effective digital searching increases the potential for discovery of new material and allows for connections to be made across sources and texts. This calls for particular interpretative skills and in training upcoming researchers, university departments should take account of and respond to this changing nature of research (and fora by means of which best practice could be shared in this regard might be considered). The accessibility of these resources also provides opportunities to non-specialists to engage with them, be they members of the public, or researchers from different fields, as discussed in the third workshop (p. 22).

Dialogue must even be occasionally initiated across branches of a single field. In his contribution to the workshop, Theodorus Fransen highlighted the discrete nature of lexicographical resources pertaining to medieval Irish, on the one hand, and Modern Irish, on the other. In examining both modern and historical layers of language, the *Cardamon Project*[10] seeks to bridge this particular divide. Similar issues were explored by other contributors to later workshops, including Brendan Kane in presenting the *Léamh* project focussed on Early Modern Irish (third workshop, p. 23).[11] Pertinent to these issues also is interoperability (see Appendix 2); linking various projects would increase the value of the individual resources, as well as enabling the criss-crossing of inter-related corpora, stimulating new directions in research (see further below). On a practical level, this would involve greater co-operation between different areas in a research field; across institutions; among projects and the researchers connected with them. In this connection, the productive discussion facilitated by this AHRC-IRC Research Network should be maintained beyond the lifetime of the Network in less formal ways.

---

[6] http://www.dil.ie (accessed 27 September 2021).
[7] https://celt.ucc.ie/ (accessed 27 September 2021.
[8] https://iso.ucc.ie/ (accessed 27 September 2021).
[9] https://www.youtube.com/watch?v=WX0HrWqS0iA (accessed 27 September 2021).
[10] http://www.cardamom-project.org/about/ (accessed 27 September 2021).
[11] https://léamh.org (accessed 27 September 2021).

c. New Tools – New Methods – New Questions to Address

Accessing sources differently alters the way in which we undertake research and the positive aspects of this new methodology was highlighted in a number of areas. The digitisation of manuscripts and inscriptions, available via databases such as ISOS, has allowed much easier access to original manuscript sources on the part of the scholarly community, including those researchers who are geographically far removed from the institutional repositories in which those manuscripts are held. This has not only made it far easier for scholars to check readings and consult variant versions of a text, but has also led to a broader shift in the focus of research in the field of medieval Gaelic studies to incorporate more discussion of subjects that have in the past been comparatively neglected, such as scientific or technical works.

Another result of the increased availability of digital resources is the fact that more attention has been paid to manuscripts as whole artefacts, rather than simply as repositories of individual texts. Physical models and tools for the visualisation of the physical constructions of codex manuscripts such as the Schoenberg Institute for Manuscript Studies' *VisColl*[12] have underlined the importance of the materiality of sources; the capacity to analyse digitally medieval handwriting, with recourse to *DigiPal*,[13] for example, aids in the identification of specific scribal features and hands. Thus, a digital turn has brought the physical object, the form of the original source, greater to the fore. Inherent in such tools is a facility for comparison across manuscripts and inscriptions, as well as between hands. The advent of the *International Image Interoperability Framework* (IIIF)[14] will greatly assist such comparisons; its application to a digitised collection of Gaelic manuscripts, *Irish Scripts on Screen* (ISOS) was discussed at the second workshop by the director of that project, Anne Marie O'Brien (pp. 19–20). What was repeatedly underlined in this first workshop is the necessity of various digital platforms to be flexible to enable such comparisons. It is the interfaces and the search engines that are paramount in allowing users to read and interpret the data, as the perspective from other fields presented by María José Estarán Tolosa and Philip Durkin similarly underlined with reference to *Hesperia* and the *Oxford English Dictionary* respectively.[15]

Such tools for comparison and the effective harnessing of accessible resources more generally, as well as doing research in new ways to utilise the full potential of those sources raises the possibility of new research questions being addressed, as several contributors to the workshop explored. Digital tools bring previously unrecognised patterns to light, raising new questions and prompting novel approaches to source material. Manipulation and analysis of 'big data' allows strands to be identified, within a holistic framework. The usefulness of data visualisation tools for different types of sources was highlighted by Pádraic Moran and Dauvit Broun; the potential of quantitative studies in the fields of palaeography and manuscript studies was illuminated by Orietta da Rold and Elaine Treharne. Treharne's research as part of *Stanford Text Technologies*[16] involves enormous quantities of data; yet such tools and approaches can

---

[12] https://viscoll.org/ (accessed 27 September 2021).
[13] http://www.digipal.eu/ (accessed 27 September 2021).
[14] https://iiif.io (accessed 27 September 2021).
[15] http://hesperia.ucm.es/ and https://www.oed.com/ (accessed 27 September 2021).
[16] https://texttechnologies.stanford.edu/ (accessed 27 September 2021).

also be effective in a scaled-down context. Conceptualising data differently will allow other research questions to emerge.

2. *The Changing Face of Research: Moving Forward*

a. Technological and Research Challenges

Changing a research culture and embedding new technologies and methods brings challenges in its wake, a number of which were raised by contributors to the workshop. The most pressing issues include lack of standardisation and agreement of norms across websites in terms of layout and also citation practice; and inconsistency as far as tokenisation and use of unique identifiers across various Gaelic resources is concerned which can cause confusion and hinders linkage and interoperability.

Interoperability emerged as a key concern, and is the subject of a separate report (see Appendix 2). The importance of thinking about interoperability when planning a new digital project was underlined by a number of speakers, with an emphasis on the challenges brought by constantly evolving technology. A recurring message was to keep things simple and stick to standard technologies which can be migrated more easily as new software is created. The centrality of metadata was emphasised, since it is paramount if a project is to be comprehensible, citable and interoperable in the longer term. Even with foresight, challenges caused by developing technologies cannot always be avoided; the complexities of moving metadata to an Open Linked format and hosting it on a more sustainable Resource Description Framework (RDF) server were illustrated by Dauvit Broun and Mark McConville in relation to two projects, *People of Medieval Scotland* (PoMS)[17] and the Digital Archive of Scottish Gaelic (DASG) in turn.

The **sustainability** of digital resources emerged continually as a theme and was the focus of the fourth workshop (p. 29). Many websites, notwithstanding their outdated technologies, continue to be used, as they form indispensible resources for researchers; *Hesperia*, a database of resources pertaining to palaeohispanic languages, and CELT, a corpus of electronic texts in various languages relating to Ireland, are two such examples (CELT was discussed in more depth at the second workshop, see p. 14). Changing technologies might cause greater difficulties in the future, however, and smaller, extremely useful resources, such as the *Early Irish Glossaries Database* (EIGD)[18] can only be maintained with investment of expertise and time (and at some point funding). Kevin Scannell set out clear strategies to plan for longer term sustainability of data (defined as between fifty and one hundred years); these included putting material in the public domain, or under a Creative Commons licence (such as CC-BY);[19] incorporating data within a 'software pool'[20] and into an Open Linked Data domain.[21] The significance of documenting data throughout by means of a 'data statement' was also

---

[17] https://www.poms.ac.uk/ (accessed 27 September 2021).
[18] https://www.asnc.cam.ac.uk/irishglossaries/ (accessed 27 September 2021).
[19] https://creativecommons.org/licenses/ (accessed 27 September 2021).
[20] Streiter, O., Scannell, K.P. and Stuflesser, M. 'Implementing NLP projects for noncentral languages: instructions for funding bodies, strategies for developers'. *Machine Translation* 20, 267–289 (2006). https://doi.org/10.1007/s10590-007-9026-x (accessed 27 September 2021).
[21] An example is: https://lod-cloud.net/ (accessed 27 September 2021).

underlined,[22] as was the consistent use of permanent object identifiers (see fourth workshop, p. 31 especially). Site security was also identified as a major sustainability issue, since problems with security certificates often led to the removal of important digital research resources from host-servers. Ongoing technical and administrative support would alleviate this but this is rarely present in the case of projects whose period of funding lie in the past. Clear communication with institutional hosts (often universities) was essential, though this could not always solve an ongoing problem. Collaboration within the field and sharing of best practice was also identified as important. Matters concerned with sustainability were treated in depth at the fourth workshop of the series (p. 29).

b. Human and Digital Resources

Communication and collaboration illustrate a recurring theme – the primacy of people, in creating, maintaining and refining digital resources. In terms of sustainability, ongoing human technological and academic expertise throughout a project and well beyond the initial phase of a digital resource is essential to ensure its continued maintenance and development, as we have noted. Technological and research specialists must work together (they are rarely one and the same); the latter remain at the heart of digital projects no matter how advanced the technology. Indeed, the greater the input of academic experts in the development of the tool, the greater its use, the illustrative example of the intense human involvement in training an algorithm to recognise manuscript features being offered by Elaine Treharne. Moreover, the starting point for digital resources should be determined by specialist understanding of what we need to know to address a particular problem or clarify a concept, rather than driven by what might be technologically possible. Communication between the different types of collaborators is crucial at every stage.

At all stages interpretation is fundamental, commencing with interrogation of a dataset, often wrongly considered as hard evidence. As a human creation, databases have inherent biases and are neither neutral nor meaningful in their own right. Elaine Treharne and Pádraic Moran, among others, stressed the need for data analysis by specialists whose focussed interrogation can yield significant results. Patterns identified can then be skilfully evaluated; and the significance of non-emerging patterns will also be recognised and therefore addressed. But interpretation is also a matter of technical expertise; tagging and coding involves precise analysis of the meaning of, for example, a word, phrase, term. This necessitates the acquisition of a specific skills set, and so the matter of training comes again to the fore.

Digital training events of specific relevance to the field of medieval Gaelic studies could increase the level of technical expertise among academics and, at the very least, enhance the communication necessary for the creation of targeted, multi-functional digital resources. Such training should be offered to scholars at all stages; postgraduate students and early career researchers may be more likely to incorporate digital technologies in their research, but more established scholars may play a supervisory role in such research projects and their own

---

[22] Bender, E. M. and Friedman, B., 'Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science'. *Transactions of the Association for Computational Linguistics* 6, 587–604 (2018). https://doi.org/10.1162/tacl_a_00041 (accessed 27 September 2021).

research too will benefit from engagement with the possibilities offered by digital tools. This topic is addressed in more detail in the discussion about the roundtable at the third workshop (pp. 26–28).

Ultimately, this Research Network found that training is key to realising the potential of developing digital resources, and organised such an event in September 2021 (for which see below, p. 35). It aimed to help users understand what technological tools can achieve so that they can be utilised with all their functionality, as new research questions will be driven by increasing familiarity with the digital and greater confidence in its use. To embrace such possibilities, a shift in thinking is required away from a print format (even if represented on a computer screen) to digital mode. By way of illustration, Elaine Treharne aduced the possibility of exploring numerous pages of the same manuscript side by side on a digital screen (or screens), a feat not possible in the case of the same manuscript in physical form.

<u>c. Research Standards and Peer Review</u>

The limits of our known research horizons make venturing into the lesser known digital world more difficult in some cases but other aspects also contribute to a reluctance to rely on and cite digital resources (particularly when there is a print alternative, even if more out of date). Since databases can be frequently updated and revised, it is important to keep an accurate record of those revisions and make it clear to users precisely how the source should be cited in secondary scholarship. The *Electronic Dictionary of the Irish Language* (eDIL) employs permanent URLs for each lexical entry in the database, thereby enabling other online research projects to create permanent links to particular entries in perpetuity, and it provides information about citation on its homepage. There is also an 'archive' link via which readers can consult the supplement to the dictionary published in 2013; this 'versioning' process was discussed in the fourth workshop (p. 31). While the creation of a 'how to cite' button is the responsibility of those directing specific digital resources, the creation of recognised standards for such resources is a matter for the field as a whole. Their lack, especially when compared with long-established conventions in the case of print publications, create uncertainty in terms of the perceived importance and reliability of digital counterparts. This is true not just for the field of medieval Gaelic studies, but across humanities disciplines more generally, and addressing these issues should take into account practices in related subject areas.

Digital resources are expensive both to build and maintain. The establishment of recognised universal standards and peer review on a par with those in place for print publications will lead to greater involvement with digital resources. They may also encourage funding agencies and institutions to support digital tools which are increasingly central to our field, shaping both the questions we ask and the way we do research. Embracing them alongside more established print technologies allows a myriad of research possibilities not available with recourse to print resources alone.

Conclusion

Orietta da Rold highlighted the innovative ways in which vellum and paper were used side by side in the late medieval period, as the new technology – paper – was made to work alongside

older parchment in complex ways. Making the old and the new work together to support and advance our research is our challenge. Discussions initiated in this the first workshop of the Research Network, Developing a Digital Framework for the Medieval Gaelic World, and continued in three further workshops, culminating in a digital training event, is an important way forward.

II. Workshop 2, Maynooth University, 'Representing Texts: from Material to Digital', 21–22 January 2021

Deborah Hayden and David Stifter, July 2021

Introduction

The second workshop examined how different forms of textual representation, whether facsimile, photograph, diplomatic, critical edition or database, have changed scholars' perspectives on texts and their relationship to original sources. The availability of diplomatic editions of a small number of medieval Gaelic manuscripts has, in the past, focussed considerable attention on those particular sources, often to the neglect of others. This has sometimes resulted in a disproportionate amount of scholarly attention being paid to certain versions of medieval texts, or even to certain sub-disciplines within our field, with profound implications for how we understand and interpret the broader scope of history and literary culture in medieval Ireland and Scotland. The digitisation of manuscript images has more recently served an important role in redressing this imbalance by providing increased access to a wider range of primary sources. There is still progress to be made in many areas, however, such as how we represent the scribal ambiguities or idiosyncrasies of the original manuscripts in a digital format, or provide a better representation of textual apparatus showing original and variant readings in some derivative digital editions of printed texts. Issues relating to scholarly authority and citation are also a key point of concern.

Session 1 featured talks exploring the methods, problems and impact of digital editing in the field of medieval Gaelic studies, while in Session 2 speakers focussed more specifically on projects and strategies concerned with the capture and markup of texts. The presentations in Session 3 looked at techniques of data extraction, visualisation and analysis in relation to projects concerned with both medieval and early modern Gaelic studies and the wider sphere of cutting-edge work in the digital humanities. The roundtable discussion, which included five representatives from the British Library, Royal Irish Academy, National Library of Scotland, the *Irish Script on Screen* project (ISOS, hosted by the Dublin Institute for Advanced Studies) and the *Watermarks in Irish Documents* project (University College Cork, UCC)[23] considered progress, challenges and desiderata in the digitisation of Gaelic manuscript sources from an institutional perspective. Their discussion explored the state of the art in manuscript digitization, the opportunities offered by emerging technologies such as IIIF and multispectral imaging, issues surrounding the funding of library digitization projects, and the tensions inherent in collaborating with researchers and serving the general public. The final session of the workshop featured a preview of the CorPH database, a major new digital resource that has been developed by the ERC-funded ChronHib project based in the Department of Early Irish from 2015–2021.

Many of the key themes that emerged from this workshop both informed and expanded on those addressed in the other three meetings, including issues surrounding the interoperability, sustainability, and potential duplication of digital resources for medieval

---

[23] https://watermarks.ucc.ie (accessed 27 September 2021).

Gaelic studies, as well as their potential for use in a pedagogical or outreach context. Particular emphasis was placed, however, on the ways in which digital resources can give rise to new types of questions about textual sources in our field, while also still being subject to different types of bias, as already mentioned at the first workshop. This second workshop also fostered valuable dialogue concerning the challenges and desiderata relating to the availability of digitised manuscripts and other resources from an institutional perspective. In sum, the meeting allowed us to gain more insight not only into the state of digitisation in our field and the opportunities that may lie ahead, but also into the perceived impact of digitisation on our understanding of 'text' in particular.

1. *Editing of Gaelic-language sources: past, present and future*

Texts provide the foundation for our scholarly endeavours as historians, linguists and literary researchers, but our different disciplinary backgrounds and needs mean that we approach them, and therefore digitise them, in different ways. Printed editions of texts have long been recognised as an authoritative means of making manuscript material more accessible to a wide audience, usually one of an academic orientation. In this format, it is possible to present manuscript text in a more legible format than is often found in the manuscript source, as well as to supply facing translations and textual notes. However, diplomatic editions focus only on one specific manuscript witness of a text, while other editions often privilege the manuscript witness that is thought by the editor to provide the 'best reading' of a text, even if variant readings from copies in other manuscripts might be provided at the bottom of the page. The prevailing focus on reconstructing 'ur-texts' has often led to the neglect of the study of purposeful textual variations that might reflect the historical and social context in and for which manuscript copies were produced. A consequence of all this is that modern scholarship has often paid disproportionate attention to the contents of a small number of manuscripts, to the neglect of others, while the erroneous interpretation of ambiguous abbreviations or obscure terms on the part of an individual editor might be carried forward into subsequent scholarly discussion of a text. An additional problem with printed editions is the fact that many are only accessible to those scholars who have access to specialist research libraries, and cannot be purchased individually due to the fact that they are either out of print or prohibitively expensive. A side effect of digital editions (in any form) can be to shine a more prominent light on hitherto neglected texts, neglected because they were hard to access, or because they were considered 'minor', like, for instance, the so-called 'Minor Glosses' that Elliott Lash prepared for the CorPH database, or that had been 'forgotten' because they did not form part of a classical handbook such as the *Thesaurus Palaeohibernicus*.

As discussed in the first workshop, the current availability of images of a wider range of Gaelic manuscript sources has already radically changed how scholars within these disciplines work, and the ease and searchability of online texts have allowed researchers to ask new questions about the contents of medieval manuscripts. Future work in the area of manuscript digitization will increasingly engage with emerging technologies that will facilitate the accessibility of manuscript images, the interoperability of different databases, collaboration between individual academic projects and library repositories, and wider public engagement. A key development in this area is the implementation of IIIF standards by libraries and

academic institutions (discussed further below). However, significant concerns about the sustainability of digital resources in the field were raised by various speakers in the workshop. Some of these are addressed briefly here, but a more in-depth discussion of sustainability can be found in the report on the fourth workshop, which was devoted specifically to this theme (p. 29).

2. *Types of text digitization*

Medieval Irish studies have been at the vanguard of textual digitisation since the infancy of the World Wide Web. The *Corpus of Electronic Texts* (CELT) was initiated as early as 1994 and the standard of mark-up evolved over the years in tandem with the sophistication of textual representation in the World Wide Web. Texts captured in CELT do in fact represent editions that are different from the original printed texts. Information about the versioning is stored in the hypertext, but it is not technically possible to provide access to all accumulated versions. Its comparatively long history makes CELT a valuable model for how internet resources of this type are used, namely as a helpful tool in teaching and for research. Moreover, combined approaches where CELT is used secondarily through Google searches must not be underestimated. CELT also illustrates a common practical dilemma: at some point the size of corpora becomes too large to allow sweeping adaptations to the structure. The functioning and fate of CELT is currently intimately linked to the person of Beatrix Färber, which highlights a critical issue of many digital projects.

An important distinction should be made between 'databases' and 'digital editions of texts'. With regard to the latter questions that arise are: What is an original? What is an edition? What is an edition of an edition? Such issues can help raise the awareness of the fluidity of texts, against the static illusion created by traditional editions. Databases are a useful way of presenting material that is more lexical in nature, such as dictionaries or medieval glossary-texts, since they can present concordances of multiple manuscript witnesses in a more flexible way than printed concordances. For example, the *Early Irish Glossaries Database* (EIGD), of which two separate versions were published in 2006 and 2009, contains sophisticated search functions that allow researchers to identify patterns in textual variants across manuscript witnesses that would not be easily recognisable from a comparison of printed concordances. Yet while the EIGD contains transcriptions of texts from manuscripts, it does not provide a critical edition or textual notes on those transcriptions; instead, it was intended to serve as a useful search tool for scholars, with the long-term aim of creating printed editions of individual glossary texts with more extensive annotation.[24] This can provide an important 'intermediary step' in bringing neglected manuscript material to light, particularly given the long length of time often required to prepare and publish printed editions (a point that is often not adequately appreciated by funding bodies), as well as in highlighting potential new avenues of research. The same issue was also illustrated in a presentation of a subset of the lexicographic database *Corpus PalaeoHibernicum* (CorPH), namely the minor Old Irish glosses. Elliott Lash argued

---

[24] A printed edition of a glossary text that has resulted from this project is Pádraic Moran (ed. and trans.), *De origine scoticae linguae (O'Mulconry's glossary): an early Irish linguistic tract, edited with a related glossary, Irsan*, Lexica Latini Medii Aevi 7 (Turnhout: Brepols, 2019). It is expected that other printed editions will appear in due course.

that the inclusion of this group of diverse texts (taken from a diverse range of printed sources from across the 20th century, and often checked against the digitalised manuscripts) does not constitute an edition in its own right. Databases are tailored towards machine-searches for linguistic and statistical purposes, while editions aim at wider philological uses. Nevertheless, databases are steps towards non-traditional editions, and they can improve on errors in existing printed editions.

Other project websites aim to provide digital editions of texts that contain the same information that one would typically find in a printed edition, but present it in a more flexible format. An example of this discussed at the workshop was the collaborative project on medieval Welsh saints' lives based at the University of Wales Centre for Advanced Welsh and Celtic Studies.[25] This website presents parallel transcriptions and translations with annotations alongside manuscript images that readers can click on and examine in juxtaposition with the edited text.

3. *The capture and markup of texts*

The key advantage of digitalised texts over printed editions is the possibility of adding markup, i.e. annotating them with meta-information. Both with regard to the added contents and to the formal methods, the possibilities of markup are almost unlimited. A recurring theme, expressed succinctly by Chris Yocum (*IrishGen* project),[26] is therefore that the digital capturing of texts needs to be aimed at a goal, for instance increasing the speed of searching material and research on the material. It is essential that both the goal of the markup and the methodologies be well thought-out before the start of the practical work. The science of how to curate data is called 'data governance', and there are dedicated websites that provide examples of best practice.[27] There is no single format that is fit for all data; the format has to be guided by the goal. It is very advisable to write up, in a log, one's methodology and the annotation choices one makes as one goes along. Doing this retrospectively will lead to omissions. It must be kept in mind that beyond the narrow goals for which digital resources were originally conceived, corpora and databases can always have the secondary function of serving as training data for other applications, and can thus take on a life of their own.

Fangzhe Qiu presented the concept of Variation Tagging developed in the ChronHib project for *Corpus PalaeoHibernicum* (CorPH). The advantage of this approach is that annotations resulting from new research can be added back in to the corpus. This in turn creates the opportunity for academic crowd-sourcing of resources, in that new annotation schemes can be developed as new research questions are being asked of a database such as CorPH. The more diverse the tagging is, the more tags can be combined in searches and correlations between tags and their values can be studied. Chris Yocum highlighted the usefulness of Linked Open Data and *Web Ontology Language* 2 (OWL 2)[28] for the same purpose; this is applicable not only

---

[25] https://welshsaints.ac.uk/ (accessed 28 July 2021).
[26] https://github.com/cyocum/irish-gen (accessed 27 October 2021).
[27] E.g., W3C Data on the Web: https://www.w3.org/TR/dwbp/ (accessed 24 August 2021), Digital Curation Centre: https://www.dcc.ac.uk/ (accessed 24 August 2021).
[28] https://www.w3.org/TR/owl2-overview/ (accessed 24 August 2021).

within a single database, but also across different resources. As a query language, SPARQL facilitates federated queries across resources.

A reference point for many digital projects aimed at texts is TEI XML, as evidenced by Martina Maher and Eystein Thanisch (*Faclair na Gàidhlig*).[29] The central axiom for their work on a corpus of tagged manuscripts and texts is achieving a balance of quality vs. quantity of the material and bridging the gap between faithfulness to the text as written in the manuscript while also having a centralised search system. The fact that each token in their respective corpus receives a unique identifier is a commonality between *Faclair na Gàidhlig* and CorPH. This approach allows, for instance, the creation of diplomatic and semi-diplomatic editions at same time, or the visualisation of vocabulary in network analyses.

While Fangzhe Qiu and Martina Maher, coming from a philological background, stressed the manual aspects of the annotation work, Chris Yocum, with an IT background, recommended leaving as much work as possible to the computer. There is clearly a practical divide there depending on scholars' expertise in coding, but the gap may close as younger generations of scholars are more familiar with IT techniques.

4. *Techniques of data extraction, visualisation and analysis*

Digital network analysis is increasingly being used for research in the fields of medieval and early modern Celtic Studies and is proving to be a useful tool for shedding light on aspects of the historical record that have been sidelined or neglected in previous scholarship. For example, the MACMORRIS project (*Mapping Actors and Cultures: Modelling Research in Renaissance Ireland in the Sixteenth and Seventeenth Centuries*),[30] based at Maynooth University, has been aiming to create a digital map of the full range of cultural activity across languages and ethnic groups in Ireland from 1541 to 1691. Project members have drawn upon existing printed and digital sources, such as the *Dictionary of Irish Biography*[31] and the *Bardic Poetry Database*,[32] to compile a list of 'cultural actors' and their patrons active across Ireland during the period in question, and to conduct a 'deep mapping' of the province of Munster. Their network uses Python code to link poets, patrons and other available metadata, such as the time period in which a given poem was written and the degree of certainty with which it can be attributed to a given poet. This broad digital approach has been shown to complement the narrower, more focussed work of editing individual texts or collections of poems, which may only provide insight into one specific region or patron. It also sheds light on connections between lesser-known figures and regions where there might have been more extensive cultural activity, even if only a few poems have survived to the present-day. The network analysis element of the project thus acts as a catalyst for future research and editorial work, and has demonstrated how digital humanities can redress the tendency to privilege English-language sources over Irish ones in the writing of Irish history.

---

[29] http://www.faclair.ac.uk (accessed 27 September 2021).
[30] https://macmorris.maynoothuniversity.ie (accessed 27 September 2021).
[31] https://www.dib.ie (accessed 27 September 2021).
[32] https://bardic.celt.dias.ie/ (accessed 29 July 2021).

Peter Stokes similarly highlighted the importance in digital humanities research of using computers to facilitate new questions rather than to simply find answers to existing ones, thus underlining the idea that digital humanities programmes should not replace human expertise (a point also made by Oksana Dereza during the roundtable in workshop 3, see p. 27).[33] This was well illustrated by his demonstration of the open-source Archetype software, which has been used extensively for other types of medieval manuscripts, but not yet for Gaelic sources. Archetype was originally designed for analysis of palaeographical features, but can also be used to filter other kinds of images or metadata, such as dates or attributions, across a wide corpus of manuscript material. The *Transkribus* project based in Innsbruck, Austria has established models for Handwritten Text Recognition for Carolingian scripts, and Bernhard Bauer highlighted the potential for applying such techniques in a Gaelic-language context as a means of making transcription work less time-consuming for researchers.[34] Many issues surrounding the implementation of these tools remain, however, including the logistics of doing collaborative work.

5. *Digital resources and scholarly authority*

A recurring theme of the workshop was that of the perceived authority and trustworthiness of digital resources. Databases such as ISOS include catalogues for individual manuscripts based on printed catalogues from different sources, some of which are in need of updating in light of current scholarship. However, the database also includes other, more comprehensive or accurate, catalogues that either do not exist in printed form or supersede existing catalogues. As users of the database will trust those online sources as authoritative, it is important to supply the author and date associated with each catalogue entry to ensure that appropriate credit can be given.

There is a tension between the flexibility of online representations of texts, which can be updated, corrected or just changed at any time, and version control, i.e. the possibility to attribute specific variants of a text constitution to a specific date and person. It was emphasized that digital work is generally more collaborative than is often the case in humanities research, and clear and consistent acknowledgement of the input of all contributors to digital resources needs refinement. In this regard, Science disciplines provide a model which can be moulded to serve the type of team-work involved in building first-rate digital facilitators of research.

This acknowledgement is important not only in terms of the perceived trustworthiness and quality of the sources themselves on the part of the user, but also for ensuring that scholars receive due credit for their work for the purposes of career progression. Many early career scholars are increasingly being employed on short-term projects involving a substantial digital element, and such work may detract from the time they have to complete other, more 'traditional' types of scholarly outputs, such as articles or monographs. A closely related issue is the recognition of digital research and work in the CVs of scholars. As mentioned in the first

---

[33] Stokes quoted Donald Knuth, 'The Concept of a Metafont', *Visible Language* 16/2 (1982), 3–27, at 5: 'The best way to understand something is to know it so well that you can teach it to a computer'.
[34] https://transkribus.eu/lite/ (accessed 30 July 2021).

workshop (2.c., p. 10), a clearer consensus should be reached for the appropriate methods of peer-reviewing and citing digital content.

6. *Project management and sustainability*

As digital resources proliferate in our field, users become increasingly reliant on them. A consequence of this is that when databases suddenly become unavailable, researchers may not have adequate printed source-material to hand as a replacement. This is particularly problematic in the case of students, who may have been asked to use digital resources for their coursework and have limited financial means for purchasing expensive printed editions of texts. For example, it was observed that the *Early Irish Glossaries Database* has periodically been taken down by IT administrators at the University of Cambridge on account of external threats (e.g. bots seeking illicit access to server space) or weaknesses identified by testers aiming to identify those threats. Responsibility for bringing digital resources back online swiftly and efficiently often lies with the academics who created them, but this process may be adversely affected if the responsible individual is no longer available to advocate for the resource (e.g. on account of retirement). It was emphasised that researchers need to begin thinking at an early stage about the long-term sustainability of digital projects, as well as the importance of long-term elements such as storage, backup and the potential ability to migrate files to new technology in the future. These issues were treated in more depth in the fourth workshop (see p. 29).

7. *Libraries and digitization*

Four main themes came to the fore in the roundtable discussion on this theme:

Even short-term projects require long-term planning. People applying for short-term projects need to liaise with libraries at an early stage, as library representatives often have to go through internal processes before the digitization of a manuscript collection can begin (the importance of libraries in project sustainbility was also addressed in the fourth workshop, for which see p. 31). Libraries are also often working on multiple projects simultaneously while dealing with staff limitations. Another factor to consider is the potential need to carry out other kinds of work on a given manuscript before it can be digitized, e.g. conservation checks for more fragile manuscripts or improvement of existing catalogue descriptions. The time required to carry out this work is often not fully appreciated, and the limited 'manpower' that is typically available for such work means that difficult decisions sometime have to be made about what projects take priority at any given time.

Funding. A great deal of funding for library digitization projects is external. Although academics frequently recognise the necessity of obtaining funding for digitization in the context of individual applications, very often funding bodies will not accommodate such requests, and therefore the digitization element of the project falls back on internal library resources. While libraries are usually very happy to collaborate with academics on such projects, this method of funding can also mean that the work of digitization ends up being more

'funding-led' than 'research-led', and the resulting gaps in library collections can in turn have an impact on the focus of future research initiatives. On a more positive note, it was observed that some individual donors or other funders, such as the Polonsky Foundation,[35] have been very keen to finance digitization projects, as seen in the third workshop (p. 25). Such bodies have enabled the British Library to get funding for creating or revising catalogue records to bring them up to an acceptable standard – work that can itself be highly beneficial to researchers, since it is often even more difficult to get funding for cataloguing projects than for ones involving digitization of sources.

Tension between academic work and public engagement. The time and resources needed for libraries to engage with the public are also often underestimated, and library staff need to balance their commitment to serving the academic community with that of making their collections accessible to a wider audience. It can sometimes be challenging to get members of the public to engage with medieval manuscripts in particular due to barriers such as the script or language of texts, or unfamiliarity with specialist terminology (e.g. 'codex'). A result of this is that illuminated manuscripts, such as the Book of Kells, are more frequently the focus of such initiatives.[36] However, exhibitions such as the 'Anglo-Saxon Kingdoms' event at the British Library have been highly successful, and might serve as a model for future work in this area.[37] It is often helpful if a manuscript exhibition coincides with an anniversary, as illustrated by the recent online exhibition curated by the Royal Irish Academy on the *Cathach* (RIA MS 12 R 33), which was created to celebrate the 1500th anniversary of the birth of Colm Cille.[38] Providing interpretative webspaces entails another cost to libraries, but can make a big difference in terms of the extent to which the wider public can engage with medieval manuscript sources. Many libraries and research institutions increasingly use social media platforms to highlight aspects of their collections, as discussed in the third workshop (p. 25). This is also an area where IIIF technology (discussed in the following section) will be advantageous in future.

Technology. Libraries and academic research institutions such as DIAS are increasingly moving toward the use of IIIF (*International Image Interoperability Framework*) technology. IIIF standards are 'a set of shared application programming interface (API) specifications for interoperable functionality in digital asset repositories' that facilitate the parsing and sharing of digitized materials across technology systems.[39] This technology has been described as a 'game-changer' for how researchers plan and implement projects dealing with medieval manuscripts, in the sense that it allows libraries to act as places of permanent storage not only for the original manuscript artefacts, but also for their digital 'surrogates', thereby facilitating issues surrounding the sustainability of project websites. Another major advantage of this technology is that it enables more 'open' use of data: researchers who are planning projects

---

[35] https://polonskyfoundation.org (accessed 27 July 2021).
[36] An example of this is the exhibition on the 'Book of Kells' (TCD MS A I. 58) at Trinity College Library (https://www.tcd.ie/visitors/book-of-kells/; accessed 27 July 2021).
[37] https://blogs.bl.uk/digitisedmanuscripts/2018/10/anglo-saxon-kingdoms-a-once-in-a-generation-exhibition.html (accessed 27 July 2021).
[38] https://www.ria.ie/library/digital-collections/library-exhibitions (accessed 27 July 2021).
[39] https://iiif.io/community/faq/#what-is-iiif (accessed 27 July 2021).

that will draw on manuscripts from several different institutions will no longer have to go through the time-consuming process of negotiating with each of those institutions individually, often paying hefty prices for the use of their images. IIIF will bring significant improvement to existing resources in the field of medieval Celtic Studies. For example, while it is recognised that while the *Irish Script on Screen* (ISOS) website is easy to navigate in its current form, it does not have very extensive searching capabilities; IIIF technology will enable catalogues to be linked to manuscript images, and will allow for keyword or page searching through catalogues. Users will be able to import images from various repositories into their own 'virtual workspace' and easily add annotations to these, making the technology very useful for educational purposes. However, the process of adapting existing databases to these standards is not entirely straightforward and can be time-consuming: for example, the ISOS project must process 20 years' worth of legacy data and ensure that the catalogues that are linked to manuscript images are up-to-date and reliable.

Libraries also currently offer multispectral and 3D imaging services to researchers. Multispectral imaging is a useful way of getting 'under the skin' of a manuscript to see what is beneath the ink, as well as to analyse features such as stains.[40] The Royal Irish Academy (RIA) has been doing this kind of imaging for a small number of Irish manuscripts, partly in collaboration with an IRC Advanced Laureate Award project on the history and materiality of the Irish book led by Prof. Pádraig Ó Macháin (UCC).[41] The British Library has made a number of its collections items available for viewing in 3D format via its Sketchfab page,[42] and also conducts multispectral imaging; at present, the latter is usually carried out at the request of individual researchers and for manuscripts that are damaged or treated with reagents, but the library will be able to make such images more widely available using their new 'Universal Viewer'. There can be a tension between doing multispectral imaging on a small number of manuscripts vs. large-scale digitization, as libraries have limited numbers of curators, and these individuals are typically doing other things as well. It is also recognised that, although the digitization of manuscripts can reduce the time that researchers have to spend handling the original artefact, which can be beneficial in terms of conservation, it is important that researchers have an opportunity to view the actual manuscript as well in order to get a good sense of aspects such as size and collation; moreover students need to be trained in how to handle manuscripts properly.[43] Although a perception may exist among some researchers that libraries are increasingly using digitization as an excuse for restricting access to original manuscripts, it was emphasised that this is not at all the case: libraries still recognise the importance of seeing manuscripts in the original, and digitization projects are only intended to complement this.

---

[40] An example of the latter is the 'Library of Stains' project, funded by a Postdoctoral Fellowship Microgrant from the Council on Library and Information Resources (CLIR): https://zenodo.org/record/2528377#.YQBfPbqSnIX (accessed 27 July 2021).

[41] https://research.ie/2019/04/11/irish-research-council-to-invest-e12-million-in-world-class-frontier-research/ (accessed 27 July 2021).

[42] https://sketchfab.com/britishlibrary (accessed 27 July 2021).

[43] Digital advancements relating to the visualization of manuscript collation have, however, also been made recently by projects such as VisColl: https://viscoll.org/ (accessed 27 July 2021).

Recommendations

➤ Databases and other digital resources need to be built around a goal, as this predetermines the practical decisions about data structure and data format. Researchers planning digital projects need to begin thinking at an early stage about issues relating to the goal(s) of their resources, interoperability, data management and the long-term sustainability of their projects.

➤ Documentation of the choices, the procedures, and the steps taken is vital for the long-term operability of a resource. Likewise, with regard to specific digitised objects such as texts, version control needs to be carried out.

➤ Similarly, those intending to begin projects that involve digitization elements should engage with data curators and librarians at the earliest possible time in order to facilitate collaboration.

➤ Funding bodies should recognise the importance of digitization as an integral part of many humanities research applications, as well as the fact that libraries usually have limited internal resources for accommodating such work. Researchers should therefore be permitted to budget for digitization of manuscript sources in project applications.

➤ Libraries and other academic institutions that are engaged with the digitization of medieval Gaelic manuscripts should implement IIIF standards in order to make their digital collections as open, flexible and accessible as possible for researchers.

➤ A clearer consensus needs to be reached concerning appropriate methods of peer-review and citation of digital resources in order to ensure confidence in these resources on the part of users, and also to allow the researchers who have contributed them to receive appropriate recognition for their work.

III. Workshop 3, University College Cork, 'Exploring Texts: Revealing Hidden Heritage through Online Resources, 25–26 March 2021

Beatrix Färber, July 2021

Introduction

This mainly practically-oriented workshop contained three sessions, one keynote speech and a final roundtable discussion. Following on from workshops 1 (Cambridge) and 2 (Maynooth), this third workshop focussed on how digital resources have helped a wider audience engage with medieval Gaelic material and facilitated new forms of teaching language, literature and history within universities. The presentations discussed the creation, maintenance and updating of the digital infrastructures behind the showcased instances of our cultural heritage. This heritage has been appreciated mainly by scholars and other specialists so far, and much of it has remained 'hidden' from public view for a long time. Through public outreach, it is becoming increasingly visible, both by researchers' physical visits to schools and other relevant groups or places, and via social (mass) media, podcasts, tutorials and educational videos, exhibitions and printed books.

The various projects presented at this fourth workshop engage both with the general public and with specific communities. Emerging best practice from these experiences was demonstrated. Such engagement has at times been a project funding requirement, and across the board has resulted in excellent feedback and community-building at all levels.

Orla Murphy's talk rounded off the thematic sessions, showing how Digital Humanities (DH) in Ireland can benefit from a change in thinking across the European research landscape. She pointed participants to training courses on the *Digital Research Infrastructure for the Arts and Humanities* (DARIAH) website[44] in her reflections on 'Digital Humanities as Enabler', and encouraged scholars to stress the value of their work, and demonstrate its contribution to society. She showed that it is worth building bridges to other disciplines and organisations in order to facilitate each other's endeavours, for example with alliances supporting DH such as DARIAH and the *Common Language Resources and Technology Infrastructur*e (CLARIN)[45] on an EU level. Especially now that the Humanities are no longer 'excluded' from the Sciences, our disciplines would do well to band closer together to achieve more. Issues of interoperability and sustainability also came to the fore, and are addressed further in Appendix 2 (p. 43) and the fourth workshop (p. 29) respectively.

Another aspect treated in this workshop emerged from previous feedback and was incorporated as a roundtable discussion on the use and experiences of digital resources in teaching. It concluded the workshop with a lively discussion involving postgraduates, early career researchers, and academic teachers, who talked about their exposure to digital resources, their experiences of using digital content for study and (where applicable) for their own teaching, and their suggestions to improve these resources.

---

[44] https://www.dariah.eu (accessed 27 September 2021).
[45] https://www.clarin.eu (accessed 27 September 2021).

1. *Online teaching and its infrastructure*

Session 1 examined challenges in current research projects across research institutes and universities in Ireland and the US, which provide infrastructure, teaching tools, cross-linking and integration, particularly in third-level education, in the area of medieval Gaelic and cognate disciplines.

The presenters focussed on the underlying tasks and challenges of delivering, maintaining and updating the infrastructure for teaching, including self-directed study. They showcased individual projects of diverse backgrounds, financial and staff support, and technical features. Discussion contributions explored how we typically make use of this rich range of teaching tools. Evaluating different levels of technical functionality raised questions like: Which skill threshold does website navigation require? Does the functionality offered correspond to users' needs? Can it assist researchers and teachers in discovering parallels, and making comparisons? Does it help you see what you (think you) know in new contexts, or invite you to engage creatively with data hidden within corpora and databases? How can experts and learners pool their strengths to deliver better teaching tools? The *Léamh* project especially was commended for putting together a strong teaching resource for Early Modern Irish, covering all pedagogic aspects learners might see addressed.

2. *Visibility and impact: public engagement*

Session 2, 'Enhancing visibility and impact on public engagement/self-study' and session 3, 'Reaching out to communities: more public engagement' were closely related. The overlap between the topics addressed was intended, creating scope to explore different aspects of common themes and the challenges discussed. One central theme was exploring how the research community might learn from each other, and was assisted by the chat contributions. Another focus was on examples of projects engaging both with general public and with specific communities. Presenters gave examples of what helped to improve their projects' outreach and impact.

The impressive example of *Logainm*[46] showed a well-designed and well-resourced site first intended as digital infrastructure for a modest number of users. The project has been developed over fourteen years by a dedicated team, interlinking linguistic, toponomastic and technological expertise, and it invites non-experts to contribute as well. Adding tools created by related government projects has expanded its potential, and listening to its users has enhanced its functionality. Including users in transcribing documents has widened its appeal to a broad range of people.[47] In addition, expertise lent by collaborating institutions, such as the Ordnance Survey, was tapped to optimize the experiences for new users.

Beyond generous funding, small and under-resourced projects set up for collaboration such as the *Collaborative Online Database and e-Resources for Celtic Studies* (CODECS) have also delivered impressive results.[48] Dennis Groenewegen, a developer of great technical and

---

[46] https://www.logainm.ie/ (accessed 27 September 2021).
[47] Meitheal Dúchas (https://www.duchas.ie/en/meitheal/) is the website, as part of https://www.duchas.ie/ (accessed 27 September 2021).
[48] https://www.vanhamel.nl/codecs/Home (accessed 27 September 2021).

linguistic skill, showed how his project, set up using Wikimedia and for the most part maintained by himself, has survived for over a decade without any institutional funding. CODECS also hosts *Tionscadal na Nod*,[49] which lets users compare images of scribal abbreviations found in Irish manuscripts in digital images, an excellent resource for self-study of palaeography.

All these projects are great examples of constructing an infrastructure, involving communities and enabling them to contribute, and in some cases they have built their own community to engage with them.

3. *Reaching out to communities*

Session 3 focussed on connecting with existing communities and establishing new ones around the project theme, in order to achieve greater public impact, awareness, and public support for relevant academic work. Popularizing academic research without compromising accuracy or complexity is essential. It was pointed out that direct interaction with the public gets the scholarly message across undiluted, and that creativity and humour work at any level. The talks showcased projects using enjoyable educational activities relevant to everyday life, and suitable for teaching outside third level. They range from exploring local place-names and cultural artefacts such as Ogam stones, to monastic scribes, and to folk stories about origin, meaning, and history of words and myths. These activities and events have the potential to increase the impact of research in the field of medieval Gaelic studies on a wider audience, as well as encouraging multi- and inter-disciplinary research. For this potential to be realised, however, pedagogy is key: explaining what is complex material to an interested public so that they can interact more meaningfully with it, in the first place, while simultaneously enabling mutual exchange of specialist knowledge across disciplinary boundaries. Such knowledge-transfer needs particular skills to be effective; and there is often a substantial amount of preparatory work that goes into outreach efforts of this kind, such as creating resources to allow very different constituent audiences informed access to the world of digital research, or ensuring that accessible and up-to-date catalogue entries are available online to complement manuscript or artefact images being shared.

Nora White started to give talks about the background of the *Ogham in 3D* project[50] to local groups and historical societies, which then sparked a community project that is still active locally and on Twitter. The Heritage Council outreach programme for visits to primary schools, *Heritage in Schools*, is now online via Zoom and shows how scholars have adapted to restrictions due to the pandemic.[51] The ongoing maintenance and hosting by DIAS has proved vital to the *Ogham in 3D* project. ISOS secured more funding to update the visual display of the stone artefacts with a IIIF-compliant viewer. However, generally, funding for older projects remains a difficulty, since funders prefer new and innovating resources, but investing in maintaining and upgrading existing ones is just as important to ensure sustainability. Integrating current resources with other projects in new ways is essential, e.g. when new

---

[49] https://www.vanhamel.nl/codecs/Show:Tionscadal_na_Nod (accessed 27 September 2021).
[50] https://ogham.celt.dias.ie (accessed 27 September 2021).
[51] http://www.heritageinschools.ie (accessed 27 September 2021).

standards become available, otherwise the resources used confidently and easily now will be lost to science (see p. 29 for further discussion of sustainability).

Calum Cockburn highlighted the power of a project outreach backed by household names such as the British Library and the Bibliothèque Nationale, namely the *Polonsky Foundation England and France Project*, drawing on existing social media channels and a team of experts from their network adding content.[52] The artistic, intellectual and cultural connections between medieval England and France in the period 700–1200 come to life in manuscript images, enriched by videos and commissioned articles to bring both enjoyment and educational potential of medieval culture to a public familiar with similar projects and exhibitions. Statistics underline the impact of social media, as for example the British Library Medieval Manuscripts Blog and Twitter accounts (with 111,200 followers for the latter)[53] have been highly successful, with a fifth of website traffic originating from social media. They have demonstrated that there is a huge global appetite for information about this material that can be built on in order to help convince funders of the international impact of work on medieval manuscripts. Another example is the ISOS project at the Dublin Institute for Advanced Studies (DIAS), which uses Twitter to highlight some of the more accessible digitized manuscripts in their collections.[54]

Sharon Arbuthnot reported how eDIL has grown from a specialist resource for academics and researchers, building its website over various project cycles, adding functionality, and updating it twice. Similar to *Logainm*, eDIL started community engagement (a funding requirement) later, to reveal the Dictionary's enormous hidden potential. Engagement was channelled through the website, Twitter, local events, and finally a printed book, *A History of Ireland in 100 Words*, written in an accessible style for the general reader.

A follow-up project called *Spreading the Words* builds on its success. It is funded specifically to increase public impact and engagement, to reach new readers, listeners and viewers via various media, posters and events. The words chosen in the book reflect the materiality behind the eDIL entries, suited to exploring and explaining the hidden meaning. Marketing experts consulted recommended to do this by storytelling. Illustrations commissioned for the book turned out like woodcuts, striking and original. Of strong visual impact, they were reused in posters and in an exhibition, marking perhaps the beginning of branding. Regarding second-level outreach, *Spreading the Words* provides teaching and learning resources in Irish history, filling a perceived need. Users can avail of podcasts on Radio MoLi (Museum of Literature Ireland).[55] Yet another successful strategy adopted was to show presence at related cultural events, such as Heritage Week. Combining varied approaches has sparked scholarly creativity and opened up new audiences.

In the discussion, some contributing factors for project success were highlighted, such as the level of online engagement and creation of new content visitors can relate to, a strong

---

[52] https://www.bl.uk/projects/polonsky-foundation-england-and-france-digitisation-project (accessed 27 September 2021).
[53] Blog: https://blogs.bl.uk/digitisedmanuscripts/; Twitter account: https://twitter.com/BLMedieval (accessed 27 September 2021).
[54] https://twitter.com/dias_isos (accessed 27 July 2021).
[55] https://moli.ie/radio/ (accessed 27 September 2021).

connection with local communities or aspects, and the human element inherent in cultural resources.

Among the challenges mentioned (and echoed in other workshops) were how to deal with lack of funding creatively, and how to keep a project alive after the end of the funding period, especially if the people involved have left (see fourth workshop on sustainability, especially p. 33). This can result in a loss of research resources as brought up in the discussion regarding the *Thesaurus Linguae Hibernicae* (TLH, University College Dublin) having gone offline.[56] On the other hand, travel restrictions during the pandemic have been a constant factor for shifting public engagement even more to the digital realm. 'Research Infrastructures' have been named in Pillar I, 'Excellent Science' for *Horizon Europe*, the EU's research and innovation programme until 2027;[57] therefore Digital Humanities need to keep advocating to maintain, improve and reuse infrastructures, or else what was built may be lost.

4. *Use and experiences of digital resources in teaching*

The roundtable discussion brought together experienced teachers of medieval Gaelic and Indo-European languages, with one representative each of early career researchers and doctoral students. It revisited the theme of teaching tools, current practice, skills-pooling and training in greater depth.

A variety of issues were discussed. The problem that rich website functionality does not automatically translate into an intuitive interface was raised, as some website user interfaces are complex and may intimidate students. Inductions via video tutorials were suggested to encourage increasing use, and such an example can be found in the recent series of eight YouTube videos, 'A Guide to Using eDIL'. Specific digital projects might also incorporate user training into their overall programme, though this will be dependent on what a specific funding package might allow. Recent online events introducing *Corpus Palaeohibernicum* (CorPH)[58] to users, for example at a Celtic Studies postgraduate students training meeting (CLARSACH) provide one possible model, though no project will be able to provide this in a sustained way.

On the issue of how students and staff across the board could up-skill in Digital Humanities tools and techniques, including those outside project settings, the participants agreed that time is often scarce, and prioritized for research. To the question of how to integrate computer science and software skills into humanities courses, the US model of taking elective courses during undergraduate study was recommended as a suitable option. Still, more training opportunities including peer-training are required, as often scholars cannot attend a specialized course, and/or are not sure where to start for their specific needs. The time academics can spend is key as they often juggle many competing commitments, and this needs to be addressed in shaping the training of the future. This training should ideally incorporate the various strands of research, linguistic, literary, historical, palaeographical and others, that the field of medieval

---

[56] It is now online again at https://www.ucd.ie/tlh/ (accessed 27 September 2021).
[57] https://horizoneurope.ie/excellent-science/research-infrastructures (accessed 27 September 2021).
[58] https://chronhib.maynoothuniversity.ie/chronhibWebsite/home (accessed 27 September 2021).

Gaelic studies entails. The training event with which the activities of the Research Network has concluded is a response to this clearly identifiable need (for which see p. 35).

The question of whether digital projects were suitable for PhD theses was raised. Geraldine Parsons pointed out that not only did the traditional administrative guidelines still favour the printed monograph over web-based work, with questions arising about the perceived parity of a digital edition in relation to a printed book (discussed in the first and second workshops), but that the supervisor also needed special skills for supervision, or might face an ethical dilemma. On the other hand, as jobs in academia are scarce in the humanities, digital skills enhance graduates' broader employability.

Götz Keydana gave an example of how technology has changed the way we teach, especially by widening the scope of small areas and their practical limitations, e.g. subjects, time or curriculum. He explained that increased input from specialists had enhanced his own teaching, but also that viewing other academics on video made it easier for the students to overcome their shyness and contact them. Pooling expertise and resources in small departments or subject areas, via a funded, collaborative project thus resulted in win-win situations for students and staff alike, as Indo-Europeanists from many universities came together to record lectures in their area of special expertise. As the dangers of this collaboration leading to staff replacements was addressed, Oksana Dereza remarked that technology was designed to teach more in a more time-saving and efficient manner and to interact with students, but not to replace staff. However, institution administrations, usually keen to save expenditure, may not always keep this in mind.

The COVID-19 pandemic was mentioned again and again. It has increased our reliance on digital resources and skills. Nina Cnockaert-Guillou shared her experience that institutions, while generally helpful, may not be fully aware of the difficulties for Masters and PhD students with hybrid teaching and learning. Geraldine Parsons added that a very high number of students do not have a PC though they own a smartphone, so interfaces and resources must be compatible with and optimised for smartphones. The participants also noted the limitations of library scanning services where these were restricted to one article per journal issue for Intellectual Property (IP) reasons, and in fact may obstruct research. In contrast to the sciences, medieval Gaelic studies still rely on old journals whose content is long out of copyright, but that have not always been digitized. These should be exempt from such regulations.

In the discussion, topics relating to authority and reliability versus standardisation were brought up repeatedly, after having been addressed at the second workshop (p. 12). A certain tension emerged between the expectations of different user groups. When texts or data created by previous scholars are harvested for specific purposes, the texts are presented 'as is'. Encoding cannot anticipate needs of projects down the line, which intend to reuse the data and build on them for other ends. Such data reuse needs to be addressed at the point of harvesting. The issue of which kind and/or amount of standardisation might be required for a machine-readable text, and of how this may be feasible for large corpus linguistics and similar applications is best left to the ingenuity of programmers who know their requirements.

Conclusions and recommendations

➤ Providing infrastructures for research and teaching compliant with the **FAIR** principles for scientific data management and stewardship – **F**indability, **A**ccessibility, **I**nteroperability and **R**euse – can benefit from pooling teaching resources, and from reusing content on other platforms.[59]

➤ **Public engagement:** Funding requirements ought to recognize and adequately budget for the outreach benefits at all levels explored here, from post-primary and secondary school projects, to the general interested public, societies and the like, to third level students, and scholars across cognate and even unrelated disciplines. Community engagement and community building can work out as a win-win-situation over time, contributing to raising the project profile, getting more insights, creating two-way-conversations, recruiting new researchers, volunteers or interns, and may result in more project spin-off opportunities. A high level of engagement requires careful previous planning, including for staff, finances and time, but will pay dividends. Long-standing, or more generously funded projects may draw on a whole range of outreach activities, blend online and physical activities, link in to existing blogs etc., or consult marketing experts, whereas minor projects can creatively use social media, or piggy-back on popular events.

➤ **Peer training:** An existing DH network such as DARIAH could be explored, as it offers peer training in different formats. If joining such a network, regular sessions relevant to your area could be organized first, supplemented later with sessions to fill perceived knowledge gaps.

➤ **Peer training:** Try to establish a training culture in your department. Allow for one or two days each term where the department members (or neighbouring departments) start by training each other. These could be recurring events following or preceding conferences or seminars. In order to keep organising and input manageable, this could be in a familiar setting where presentations are only the first step, followed by direct conversations; and it may be best to focus on fewer choices, but more intense hands-on training. Follow on from this by connecting to interdepartmental peers. Recordings could be kept internally for refreshers and for new colleagues. Be inclusive.

➤ **Peer-review:** Joanna Tucker pointed out the review journal for digital editions and resources (RIDE) published by the Institut für Dokumentologie und Editorik.[60] It promotes peer-reviewing for digital editions and resources, with a particular focus on applying humanities methods to historical documents.

---

[59] https://www.go-fair.org/fair-principles/ (accessed 27 September 2021).
[60] https://ride.i-d-e.de (accessed 27 September 2021).

Workshop 4, University of Glasgow, 'Conserving electronic resources: Sustainability in a digital world', 10–11 June 2021

Joanna Tucker, July 2021

Introduction[61]

The fourth workshop explored the topic of sustainability in the digital world. The first session attempted to confront what sustainability really means from the perspective of digital humanities and the implications for digital humanities research. The second session built on this by looking at digital sustainability 'in practice'. Seven individual digital research projects were presented as 'lightening talks', including many tools familiar to those who research and teach on the medieval Gaelic world. This session provided an opportunity to look at the reality of sustainability for projects past, present and future. The third session turned to libraries and considered the important work that has been ongoing in relation to making digital resources and data sustainable, as well as some of the concerns about the future. In the final session, participants were divided into breakout rooms where they were given the opportunity to share their experiences and air their concerns about digital sustainability. This allowed the network to reflect on the future of digital resources, and what challenges and opportunities lie ahead for the community.

All of the network's workshops demonstrated that sustainability is now a widely shared concern. An initial question that the workshop faced, therefore, was how to define 'digital sustainability'. It was suggested that this can encompass technological decisions, financial sustainability, the environment, and human sustainability. Themes that therefore emerged during the workshop ranged from the more technical aspects of sustainability (including, for example, how to structure data in order to maximise its longevity), to the institutional aspects (especially the advocacy work that is required in order to reform funding and the long term hosting of digital research projects). The 'human' aspects of sustainability underpinned many of the discussions (especially how to train and educate people in relation to digital sustainability). It was acknowledged that sustainability will remain an ongoing concern, and as such workshops and networks such as this one are part of the 'solution' to addressing sustainability itself by offering a space to share experiences across institutions, disciplines and projects.

1. *Sustainability of resources for the digital Gaelic world: past, present and future*

A central aim of the network was to take stock of our achievements as a community to date, as well as look to the future of medieval Gaelic studies in the digital world. Seven speakers were asked to consider what 'sustainability' means in the context of their digital resource, and whether this impacted decisions relating to hosting or archiving, data format or standards, technical support and documentation, users and continued promotion of the resource, or further

---

[61] The themes from this report have subsequently been expanded into a longer article: Joanna Tucker, 'Facing the challenge of digital sustainability' (forthcoming).

funding. They also commented on how optimistic (or otherwise!) they felt about the future of their digital resource specifically, and whether steps could be taken in future to ensure greater sustainability.

Three long established projects shared their experiences of sustainability:

1. *People of Medieval Scotland* (PoMS). Dauvit Broun explained how PoMS resulted from a succession of funded projects from 2007 onwards. In the last few years, John Bradley has converted PoMS's data into Resource Description Framework (RDF) triples, allowing for more interoperable analysis in the future (as discussed in the first workshop, p. 8).

2. *Electronic Dictionary of the Irish Language* (eDIL). Greg Toner noted that eDIL, originally begun in 2003, is based on the well supported TEI XML framework and is now hosted continuously by Queen's University Belfast.

3. *Corpus of Electronic Texts* (CELT). Beatrix Färber described CELT's beginnings in 1997. Now, CELT is hosted by University College Cork and continues to make important texts from the medieval Gaelic world widely available to its users.

We also heard from newer projects which must now plan for sustainability within their early development. Those examples we heard from were:

4. Pádraic Moran's *The Gloss Engine* and *Manuscripts with Irish Associations* (MIrA).[62]

5. David Stifter's project *Chronologicon Hibernicum* (ChronHib), which resulted in the creation of the lexicographic database *Corpus PalaeoHibernicum* (CorPH).

For older projects, the path to sustainability is often less clear, and there are significant challenges for updating. This was the case for our final two examples:

6. *Thesaurus Linguae Hibernicae* (TLH). Fangzhe Dimurjan Qiu explained that TLH is only recently back online as a result of some security risks, which have now been addressed.

7. *Monasticon Hibernicum*.[63] Nora White outlined some practical steps that could soon be undertaken to enhance the sustainability of *Monasticon Hibernicum*, many of which could apply to other project websites as well: persistent URLs for individual records; enabling exports of records and search results; updating copyright restrictions; uploading data to GitHub; linking into other more recent online resources; refreshing the basic look of the website; and optimising for mobile devices (see Appendix 2).

2. *Technical aspects of digital sustainability*

Many sustainability issues are raised when thinking about a project's **'data'**. Datasets have the advantage that they can be made visible and available beyond a project's lifetime, which can partially help to meet the imperative from public funders for 'impact' (though impact is much

---

[62] http://www.glossing.org/glossengine and http://www.mira.ie (accessed 27 September 2021).
[63] https://monasticon.celt.dias.ie (accessed 27 September 2021).

more than just dissemination, of course; as discussed in the third workshop, p. 22). However, data is not automatically sustainable. For one thing, it is usually accessed via an **interface** which can pose its own issues in terms of maintenance. The **'quality'** of the data is also of vital importance, and can have a direct impact on sustainability and the potential for future interoperability. For example, it is vital that scholarly annotations are of a high standard and that there is appropriate metadata to allow future users to contextualise the production of the original data.

For websites and web resources, the use of **'Persistent Identifiers'** are a significant aspect of data management and reuse. These are a form of permanent identifier that establish a link to an object or information about it. As Rachael Kotarski explained, Persistent Identifiers are particularly important for repositories of digital materials such as libraries, but also for research projects or even commercial websites. The sustainability of Persistent Identifiers is not something technical but rather is achieved through active governance. Libraries are key here, given their expertise in data management. This is, as Rachael Kotarski noted, 'sustainability as a community effort'.

The use of **shared standards and frameworks** can be one way to enhance the potential for reuse of data in future. Notable examples of shared frameworks include the *Text Encoding Initiative* (TEI),[64] and the *International Image Interoperability Framework* (IIIF). Sustainability is likely therefore to be more feasible when using tried and tested methods and technologies. King's Digital Lab (KDL) at King's College, London provide a useful case study here, as discussed by Arianna Ciula. Because of its large portfolio of digital research tools (some dating back to the 1990s), the Lab's technical stack has increasingly been 'homogenised' to allow them to manage this range of legacy projects.

Research projects also face the challenge of how to **update, alter or extend their 'data'** beyond the lifetime of the original project. Indeed, Pádraic Moran spoke compellingly about 'survival' of digital resources as 'evolution'. An example of a project which has gone through successive phases of evolution is eDIL. Greg Toner described 'versioning' as part of the process of maintaining eDIL. These changes have been documented in a Supplement (2013) and Companion (2019). Another example is PoMS. Dauvit Broun noted that the database has changed with each successive project in order to address a new research question.

The workshop also brought to light the crucial role of **digital preservation** and what this can involve at an organisational level. As Lee Hibberd showed, significant staff and infrastructure costs associated with digital preservation mean this requires meticulous planning, including the use of 'digital preservation and access plans' for different types of digital record (discussed in the second workshop, see pp. 15–6). HEIs are also increasingly supporting and enabling the preservation of digital scholarship in the humanities. Cillian Joy discussed various examples of this approach at NUI Galway. As a result, such hosting and preservation can now be more readily quantified in terms of costs and can be more routinely included within business cases. KDL now even advertise the post-project management solutions they can offer.[65] Though digital preservation practices may be unique to an individual

---

[64] https://tei-c.org (accessed 27 September 2021).
[65] https://kdl.kcl.ac.uk/our-work/archiving-sustainability/ (accessed 27 September 2021).

institution, it was emphasised in the workshop that it is to the benefit of all to continue to share experiences and best practice.

Such technical concerns make it all the more imperative that projects have robust **data management plans** inbuilt from the outset. Arianna Ciula emphasised that, at KDL, planning a collaborative research project will include early discussions about the sustainability of the data that will be produced. This means that project leaders must take into consideration the questions of standards, hosting, maintenance, data preservation, data quality, and therefore overall sustainability.

3. *Designing for sustainability*

In the course of the workshop, a few points emerged that could be said to be starting principles for designing with sustainability firmly in mind. While the points below are specific, in general it was emphasised that digital resources should adhere to the FAIR principles.

1. Ensure the resource and its data are accessible and usable for a range of communities over a long period.
2. Ensure the resource is citable (not just overall but each page), preferably with permanent links achieved using Persistent Identifiers.
3. Ensure that all those individuals involved in production of the website and its data are appropriately acknowledged.
4. Consider simplicity of design, though without losing the richness of a digital resource where this might matter.
5. Design resources with 'modularity' in mind, especially separating out the interface, the code behind this, and the content itself.
6. Allow for flexibility, since there are many aspects of sustainability that cannot be controlled (especially the 'human' factors, but also technical ones).

4. *Environmental sustainability*

The environmental impact of digital sustainability is a particularly ominous aspect. Lee Hibberd reminded us that 'cloud' storage is only a metaphor, with servers often simply out of sight but still requiring enormous amounts of energy. Local advocacy can potentially drive change here, such as by requesting that cloud storage providers declare their greenhouse gas emissions or integrating environmental concerns into procurement processes. Funders might also add environmental sustainability as a criteria for funding.

5. *Funding sustainability*

It is generally recognised that a key obstacle for digital sustainability is funding. A particular issue is the **short-termism** of the funding models whereby funders look for transformative digital resources which will last but also three- to five-year projects that are based on a specific 'research question'. This model – alongside the fact that public funders do not currently offer pots of money for basic updates – has led to many research tools being abandoned and left to

the goodwill and enthusiasm of volunteers (usually those involved in the original funded project), as evidenced in the previous workshops. Thankfully, short-termism in funding models is now being challenged through advocacy work by the likes of Orla Murphy at a European level (for which see p. 22).

A number of actions were identified in the course of discussions that may help, or are currently helping, to move the issues with funding in a new direction. One concrete suggestion was the need for more **costed models** for sustaining digital resources to aid individuals and institutions in planning for a project's future. A helpful concept here is that of **'distributed responsibilities'** of sustainability, where many individuals within an institution share the benefits and the responsibilities of continuously supporting a project. There is also hope in the emerging role played by **national and international infrastructures** as 'data stewards'. For many smaller or more local institutions (including HEIs), supporting digital resources on an ongoing basis and in unpredictable conditions is a real challenge, especially given their inflexible budgets. Instead, it was recognised that the collective ambition means that solutions need to look beyond individual projects or local institutions. Notable current initiatives include the *Digital Research Infrastructure for Arts and Humanities* (DARIAH) and the *Digital Repository of Ireland* (DRI).[66] It is vital to understand the role of **researcher advocacy** for such infrastructure. Indeed, Orla Murphy reminded us that medievalists have a particular insight, given our long view on issues associated with survival. For any infrastructure solution to have universal impact, it is paramount that national funding bodies engage with these discussions explicitly.

6. *Training for sustainability*

Training is not always explicitly associated with sustainability. It was clear at the workshop, however, that this was a fundamental part of maintaining digital resources into the future. DH training is nearly always context specific. It depends on disciplinary background (whether humanities researchers, librarians and archivists, software developers) as well as career stage (whether postgraduate and undergraduate students, or also those at a later career stage). There are a range of settings within which training might happen: day-long workshops or summer schools; as part of a defined career path or a funded project; or as part of an internship or secondment.

One distinction that was highlighted was between training to build and training to use a digital resource, as mentioned during the roundtable in the third workshop (p. 26). Another crucial distinction that was discussed was between 'targeted' training in particular tools and 'holistic' training in DH generally. While both have their advantages, it was noted by many that more targeted, 'site specific' experiences of DH technologies are particularly valuable for the humanities. Brendan Kane of the University of Connecticut discussed this approach to DH teaching in particular. Such targeted training could also help to address the challenge of managing 'transitions' of digital projects, where new leadership is required.

---

[66] https://www.dri.ie (accessed 27 September 2021).

Recommendations

➢ Foster a broader understanding of digital sustainability in the humanities, one that recognises sustainability as involving a combination of technical, financial, environmental and human factors .

➢ Continue to run events that address digital sustainability concerns to allow individuals to hear about current and alternative approaches in other disciplines and institutions.

➢ Support the development of national/international infrastructures as professional 'data stewards' for future projects.

➢ Digital resource design should consider simplicity over bespoke designs, and with reference to the FAIR principles (Findability, Accessibility, Interoperability, Reuse). This encompasses the quality of the data and the potential use of shared standards such as TEI, IIIF and Persistent Identifiers.

➢ Project planning should involve early and detailed discussions with all relevant partners about the sustainability of the data produced: what needs to be preserved beyond the funded project and why/for whom; how it will be stored/maintained/accessed; and who is responsible for this. Discussions should result in a robust data management plan.

➢ Project planning should consider the environmental impact of digital projects where possible, e.g., by selecting a cloud storage provider on the basis of their stated greenhouse gas emissions and carefully planning any travel for research purposes.

➢ Project planning should build in targeted training to expand the community of users and to address succession planning for the digital resource.

V. Training event, University of Cambridge, 'Digital Resources, Manuscripts and Texts: an Online Training Event focussing on London, British Library, Manuscript Harley 5280', 7–9 September 2021

Máire Ní Mhaonaigh, October 2021

The final gathering hosted by the Digital Network was an online training event held virtually in Cambridge. 'Digital Resources, Manuscripts and Texts: An Online Training Event', was a response to the need raised for general digital training during all workshops of the Network, and also marked the digitisation of an important medieval Irish manuscript held at the British Library, Harley 5280. The digital photography required was funded by the UK Arts and Humanities Research Council, as part of the Network grant.

The first day focussed on the digital environment and the medieval Gaelic world, with a focus on the digital tools available and the creation of metadata. A session on funding applications directed at early career researchers in particular highlighted how to incorporate digital technologies into research projects. The remainder of the training event explored the use of digital resources for manuscript research, with a focus on image and format, on the one hand (day 2), and script and text, on the other (day 3). Sessions addressed IIIF and the visualisation of manuscript collation, an introduction being provided to the recently launched VCEditor by means of which collation diagrams can be produced.[67] The core principles of the Text Encoding Initiative (TEI) were outlined and examples relating to the medieval Gaelic world explained. The event also included a transcription session focussed on the new images in IIIF format, published online by the British Library of Harley 5280.

---

[67] https://vceditor.library.upenn.edu/ (accessed 27 October 2021).

Appendix 1: Workshop Programmes

Workshop 1: University of Cambridge, 26–27 November 2020
The Changing Face of Research in the Digital Age

Thursday 26th November

2pm: Greg Toner (Queen's University Belfast), 'Developing a Digital Framework for the Medieval Gaelic World: Aims and Opportunities'

Session 1, 2.15–3pm
The Impact of Digitisation on the Nature of Research: History and Texts
Moderator: Joanna Tucker (University of Glasgow)

Dauvit Broun, University of Glasgow
Pádraic Moran, National University of Ireland, Galway

Session 2, 3.15–4pm
The Impact of Digitisation on the Nature of Research: Language and Linguistics
Moderator: Sharon Arbuthnot (Sabhal Mòr Ostaig / University of Cambridge)

Theodorus Fransen, National University of Ireland, Galway
Mark McConville, University of Glasgow

4–4.30pm: Break

Session 3, 4.30–6pm
Artificial Intelligence and Analysing Data
Moderator: Máire Ní Mhaonaigh (University of Cambridge)

Orietta da Rold, University of Cambridge
Elaine Treharne, Stanford University

Friday, 27th November

Session 4, 11.30–12.30pm
Structuring Data
Moderator: Greg Toner (Queen's University Belfast)

María José Estarán Tolosa, University of Zaragoza
Philip Durkin, Oxford English Dictionary

12.30–2pm: Break

Session 5, 2–3.30pm
Digital Resources for the Medieval Gaelic World:
new ways of asking research questions and doing research
Discussion leader: David Stifter (Maynooth University)

Kevin Scannell, St Louis University
(Summary of information provided by researchers in existing projects in the field, followed by discussion informed by Prof. Scannell's contribution).

Representing Texts: from Material to Digital

Thursday 21st January

2pm: Opening remarks (Deborah Hayden, Maynooth University)

Session 1, 2.10–3.40 pm
Digital editing: methods, problems and impact
Chair: Deborah Hayden (Maynooth University)

Paul Russell (University of Cambridge): 'Reflections on the *Early Irish Glossaries Database* (and Other Online Materials) in its Second Decade'

Beatrix Färber (CELT project, University College Cork): 'Squaring the Circle? Electronic Editing and the Integrity of the Text'

Elliott Lash (Universität Göttingen): "Why the Minor Glosses Portion of *CorPH* is a Linguistic Database and not a Digital Edition"

3.40–4 pm: Break

Session 2, 4–5.30 pm
The capture and markup of texts
Chair: David Stifter (Maynooth University)

Chris Yocum (Independent Scholar): 'IrishGen: Capture Structure within Texts'

Fangzhe Qiu (University College Dublin): 'Capturing Linguistic Changes in Early Irish Texts: Variation Tagging in *CorPH'*

Martina Maher & Eystein Thanisch (Faclair na Gàidhlig, Sabhal Mòr Ostaig): 'Hand-crafted: the Manual Capture of a Corpus of Tagged Gaelic Manuscript Texts and its Potential'

Friday 22nd January

Session 3, 11 am–12.30 pm
Techniques of data extraction, visualisation and analysis
Chair: Joanna Tucker (University of Glasgow)

Bernhard Bauer (Maynooth University): '(Semi-)automatised Transcription of Medieval Manuscripts: Optical Character Recognition & Handwritten Text Recognition'

Pat Palmer, Deirdre Nic Chárthaigh & Evan Bourke (MACMORRIS project, Maynooth University): 'Visualising Literary Patronage Networks in Gaelic Ireland c. 1550–1650'

Peter Stokes (École pratique des hautes études, Paris): 'From Images to Information: Some Digital and Computational Approaches to Manuscript Studies'

12.30–1.30 pm: Lunch Break

Session 4, 1.30–2.30 pm
Roundtable discussion on the digitisation of Gaelic manuscript sources
Moderator: Greg Toner (Queen's University Belfast)

Participants: Claire Breay (British Library); Barbara McCormack (Royal Irish Academy); Ulrike Hogg (National Library of Scotland); Anne Marie O'Brien (ISOS project, Dublin Institute for Advanced Studies); Pádraig Ó Macháin (*Watermarks in Irish Documents* project, University College Cork)

2.30–2.40 pm: Break

Session 5, 2.40–3.30 pm
Presentation of the *Corpus PalaeoHibernicum* (*CorPH*) database by members of the *Chronologicon Hibernicum* project (Maynooth University)

Workshop 3: University College Cork, 25–26 March 2021
Exploring Text: Revealing Hidden Heritage through Online Resources

<u>**Thursday, 25 March**</u>

1.20pm: Opening remarks (Beatrix Färber, University College Cork)

Session 1, 1.30–3.30pm
<u>Online teaching and its infrastructure</u>
Chair: Beatrix Färber (University College Cork)

Anne Marie O'Brien (Dublin Institute for Advanced Studies), 'Exploring Collections through Irish Script On Screen'

Kevin Murray (University College Cork), 'Celtic Digital Initiative, Irish Sagas Online and the secondary digital environment'

Brendan Kane (University of Connecticut), with Deirdre Nic Chárthaigh (Maynooth University), Emmet de Barra (University of Connecticut), Christina Cleary (Dublin Institute for Advanced Studies), 'Exploring texts word-by-word: Léamh.org and networking the learning of Early Modern Irish'

3–3.30pm: Questions/discussion

3.30–4pm: Break

Session 2, 4–5.30pm
<u>Enhancing visibility and impact on public engagement/self-study</u>
Chair: Greg Toner (Queen's University Belfast)

Mairín MacCarron (University College Cork), 'Digital Approaches for a Divided Archive: the Edgeworth family archive in the National Library of Ireland and the Bodleian Library'

Brian Ó Raghallaigh (Dublin City University), 'Logainm – Enhancing the public impact of the Place-names Database of Ireland'

Dennis Groenewegen (CODECS, Stichting Van Hamel), 'Mapping the sources with CODECS: some challenges and opportunities of developing a web-based project platform'

5–5.30pm: Discussion and conclusion

Friday, 26 March

Session 3, 2–4pm
Reaching out to communities: More public engagement
Chair: David Stifter (Maynooth University)

Nora White (Maynooth University), 'Ogham in 3D and engaging digitally with our earliest writing in Irish'
Calum Cockburn (British Library), 'Promoting The Polonsky Foundation Eng-land and France Project: Public Engagement in Online Environments'
Sharon Arbuthnot (Sabhal Mòr Ostaig / University of Cambridge), 'eDIL and Opportunities for Spreading the Words'

3.30–4pm: Discussion

General reflections, 4.05–4.25pm
Chair: Beatrix Färber
Orla Murphy (University College Cork), 'Digital Humanities as Enabler'

4.25–4.40pm: Break

Session 4, 4.40–5.40pm
Roundtable discussion: Use and experiences of digital resources in teaching
Chair: Deborah Hayden (Maynooth University)

Geraldine Parsons (Lecturer, University of Glasgow)
Götz Keydana (Professor, Universität Göttingen)
Oksana Dereza (Doctoral student and Early Career Researcher, National University of Ireland, Galway)
Nina Cnockaert-Guillou (Doctoral student, University of Cambridge)

5.40pm: Concluding remarks

Workshop 4: University of Glasgow, 10–11 June 2021
Conserving Electronic Resources: Sustainability in a Digital World

<u>Thursday 10th June</u>

1–1.15pm: Welcome

Session 1, 1.15–2.30pm
<u>Sustainability from a Digital Humanities perspective</u>
Chair: Joanna Tucker (University of Glasgow)

Orla Murphy (University College Cork)
Arianna Ciula (King's Digital Lab, King's College London)

2.30–3pm: Break

Session 2, 3–4.30pm
<u>Sustainability of digital research projects past, present and future</u>
Chair: Máire Ní Mhaonaigh (University of Cambridge)

'Lightening talks' from DH projects relating to the medieval Gaelic world, featuring *People of Medieval Scotland, 1093–1371 (PoMS); Thesaurus Linguae Hibernicae (TLH); Corpus of Electronic Texts (CELT); The Gloss Engine; Electronic Dictionary of the Irish Language (eDIL); Monasticon Hibernicum and Chronologicon Hibernicum (ChronHib).*

4.30pm: End

<u>Friday 11th June</u>

11–11.10am: Welcome

Session 3, 11.10–12.30pm
<u>Digital sustainability from a library perspective</u>
Chair: Claire Breay (British Library)

Rachael Kotarski (British Library)
Lee Hibberd (National Library of Scotland)
Cillian Joy (National University of Ireland, Galway, Library)

12:30–1:30pm: Lunch break

Session 4, 1.30–3.10pm

<u>Looking to the future: challenges and opportunities for sustainability</u>

Chair: Greg Toner (Queen's University Belfast)

Explanation (10mins)

Breakout room discussions, led by facilitators (45mins)

General discussion (45mins)

3:15pm: End

Appendix 2: Data Interoperability Report

Recommendations on Data Interoperability
in Medieval Celtic Digital Projects

Christopher Guy Yocum[68]

Introduction

Data Interoperability on the Web is a core concern for the future of both the Web as a whole and Medieval Celtic Digital projects in particular. While user interfaces and database design are usually at the forefront of researchers' minds, the ability for digital projects to collaborate at a distance can:

- enhance the usefulness for researchers and the general public;
- strengthen the conclusions drawn from a particular digital resource;
- allow new kinds of research and connections which will create new research questions and new, possibly revolutionary, results.

This guide and set of recommendations draws its inspiration from the World Wide Web Consortium (W3C) *Recommendation on Data Best Practices on the Web* (https://www.w3.org/TR/dwbp/). While the W3C's recommendations are highly technical, this guide is meant to recommend a subset of these for researchers who are contemplating digital resources and help them interoperate with many of the best available research oriented digital resources in Celtic Studies. However, it is highly recommended to also consult the W3C's recommendations in detail when contemplating a digital project and its data interoperability. The question of interoperability arose during workshops hosted by the Digital Framework for the Medieval Gaelic World network, jointly funded by the Arts and Humanities Research Council and the Irish Research Council (2020-21).[69] As the workshops continued, there was an opportunity to evaluate and understand the needs of the various projects. Through this process, network members became aware that the ecosystem of digital projects could be made more robust by introducing the concept of data interoperability. There is a clear desire for the employment of simple methods for ensuring that a project's data set can be easily referenced and used by other websites without introducing any great technical or logistical burdens to the work.

It will be helpful to begin by defining what is meant by Data Interoperability and to give a few examples to illustrate its value both for Digital Humanities projects as a whole and

---

[68] ORCID: https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Forcid.org%2F0000-0002-7241-3264&data=04%7C01%7Cg.toner%40qub.ac.uk%7C32c0680116cb4ffd0c6e08d964c4c0ad%7Ceaab77eab4a549e3a1e8d6dd23a1f286%7C0%7C0%7C637651618635826387%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=sSpznN%2Fe%2BQ5W82NWrjbWK4ElYMuRaV5u72dpdYV9yj4%3D&reserved=0.

Medieval Celtic digital projects in particular. Data Interoperability is the ability of two separate computer programs to request, ingest, and act on data from each other. What this means in a practical sense is that the two programs use the same communication methods to share data with each other, and at the simplest level that two websites can refer to each other consistently and reliably.

Data Collection Methods

Before embarking on creating recommendations, the project met and interviewed over a number of weeks with many of the teams who are leading digital resources in Celtic Studies today. Their concern was mainly driven by the need for researchers to be able to accurately cite and communicate information that can now only be accessed via the Web. We met with Ogam in 3D (https://ogham.celt.dias.ie), CODECS (https://www.vanhamel.nl/codecs/Home), CELT (https://celt.ucc.ie/), BILL (https://bill.celt.dias.ie/), LOCUS (https://www.ucc.ie/en/locus/), ChronHib/CorPH (https://chronhib.maynoothuniversity.ie/chronhibWebsite/home), and eDIL (http://dil.ie). These interviews were conducted in a conversational style and many different aspects of data interoperability were explored. Notes were taken and these were distilled into the present set of guidelines and recommendations.

Recommendations and Guidelines

The recommendations and guidelines observed in this document are meant to help researchers as they create their research proposals and guide their projects. These are not meant to be rigid rules. Recommendations can be regarded as optional or not relevant to any particular digital project, depending on its needs and resources. However, researchers contemplating developing a digital project should consult with a professional Software Engineer or Computer Scientist before making a final decision.

Permanent URLs

All digital projects presently require a presence on the Web. In fact, with one exception, all the projects interviewed had websites and, in most cases, interactive Web Applications. However, it was determined that not all of these projects had Permanent Uniform Resource Locators (pURLs).[70] URLs (Uniform Resource Locators) are a fundamental function of the Web.[71] Researchers will be familiar with URLs such as http://www.google.com that appears in their web browsers. Permanent Uniform Resource Locators (pURLs) are a very simple key to making your data easily referenced by other projects. A pURL is a permanent URL which directs a user (either a web browser or some other piece of software) to a piece of data within the project. While an ordinary URL may change over time, making it difficult for users to find the resource at some future date, a pURL will always retain the same form, so that a pURL

---

[70] For much more detailed information see also https://www.w3.org/TR/dwbp/#DataIdentifiers
[71] For a technical description, see https://url.spec.whatwg.org/.

generated today will still be the same in 10 or 20 years. These pURLs have become indispensable for researchers communicating on the web. For example, eDIL uses pURLs to indicate headwords (see http://www.dil.ie/50141, for instance). Another pURL that researchers may be familiar with is the DOI (Digital Object Identifier, https://www.doi.org/) which are used to unambiguously and permanently locate a document online and are often used when citing research in many different realms of scholarly communication, particularly in the STEM disciplines. An example DOI is https://doi.org/10.1515/9783110680744-005. Another example that researchers may have encountered is ORCID (https://orcid.org/) which provides unique identifiers for individual researchers.

In terms of Data Interoperability, pURLs allow two systems which would not usually communicate to direct users to specific pieces of information or data without needing to know anything about the underlying resource. For humans, this is often their web browser but these standards allow any program to consume and use data in ways that the original designer may not have anticipated. In this way, Data Interoperability can contribute to innovation.

pURLs are meant to be permanent and, as such, the hosting organization or project must undertake to maintain the URL in perpetuity. Critically, this means that rights to the domain name, which is the google.com part of a URL, must be paid on time. Many embarrassing incidents have happened where the domain name registration was allowed to lapse and the domain name taken over by a malicious third party. If the project is hosted by a university under their domain name, it is critical to communicate the nature of pURLs to the hosting organization. For instance, CELT pURL for Táin Bó Cúailgne is now http://research.ucc.ie/celt/document/G301012. This will always point to the first recension of the Táin and will remain as long as the site remains operative.

Recommendation 1: Determine key areas of data that other users might want to reference (e.g. a dictionary entry, a bibliographical entry, or a manuscript) and consider creating pURLs for these.

Recommendation 2: You must not change pURLs once established and you should discuss the format with the domain name owner (e.g. your university) in advance.

Entities

Determining what should be a pURL and how to construct one is a topic that can be particularly difficult but is greatly facilitated by the consideration of entities. An entity may be defined as a thing capable of an independent existence that can be uniquely identified.[72] For instance, an author in a library catalogue or bibliography may be described as an Entity. How much detail may be held about an author may differ from system to system but there are some common themes, such as first name, last name, and year of birth. To take another example, a book could be described as an entity with attributes such as author, title, year of publication, publisher, etc.

---

[72] See *Beynon-Davies, Paul (2004). Database Systems. Basingstoke, UK: Palgrave: Houndmills and https://en.wikipedia.org/wiki/Entity–relationship_model*

The owners of the data set must decide which are the important entities that other users will want to use and that will, therefore, require a pURL.

How to bring Entities under the umbrella of a pURL is something that depends on how the data is encoded and structured and should be done in consultation with an experienced database designer who can guide scholars on which set of Entities should be referenced together as pURLs or which single Entities would be pURLs on their own. However, there are some basic principles that can be followed.

Returning to our previous example, we have seen that a bibliography contains various entities of significance to users including author and book. We might decide that the main thing that users will wish to reference is the work, which would lead us to want to create pURLs for books. In that case, the book's ISBN is a unique identifier and may form the basis of a pURL. In a different kind of bibliography, we may regard articles within books and journals as entities of key interest and so we would wish to create pURLs for them.

It is best practice to use numbers when creating Primary Keys and other unique identifiers. The main reason for this is that using something opaque, such as a number, removes semantic meaning which is based on language and allows one and only one meaning to both computer and human. Additionally, numbers also can be processed by a computer in a much more efficient manner than language or other more esoteric methods. Moreover, numbers are guaranteed to be unique and there is no chance that a pURL that uses numbers could resolve to two entities without a bug being discovered.

pURLs in Relational Database Management Systems

Within Database Systems theory, the most commonly used implementation is the Third Normal Form (3NF), which breaks up information into a series of interlinked tables. Tables should contain a primary key which is typically a number that uniquely identifies a record in a database. All data dependencies should be resolved by the query used to fetch the data so that all information relevant for the entity which may reside in different tables will be presented to the user. These keys are the ideal basis for creating a pURL. For instance, in http://www.example.com/123 , the number 123 would be the Primary Key for the entity that is being referenced.

pURLs in XML

Much of the data generated by Digital Humanities projects is stored in Text Encoding Initiative-conformant XML. While this data is semi-structured in contrast to the RDMS structured format, entities that can have pURLs are generally easy to determine. In this case, a researcher can annotate a tag with the xml:id attribute which can be used in a similar way to the use of the primary key described above. An entity that can be referenced should then either link to another reference within the same document for extraction and presentation to a user or have child tags which are then extracted and presented to the user. For instance, as we have seen, eDIL gives each headword an identifier (or key) that is unique to that headword. This allows the user and systems to uniquely reference an eDIL entity (the entry in this case).

A note of caution should be observed here. The xml:id has certain restrictions, such as it must be unique throughout the document and it must not contain the ":" character, and XML processing systems may reject a document that does not conform to these restrictions (see https://www.w3.org/TR/xml-id/#processing and https://tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CORS1). If a researcher does not wish to be restricted by the xml:id attribute, a different scheme will be needed but is outside the scope of this document.

Recommendation 3: pURLs should be based on a primary key or other identifier that is unique to the entity being referenced

Recommendation 4: it is best practice to write the unique identifier of a pURL in number form rather than in a form intelligible to humans.

pURLs aiding communication between computers

While on the one hand, pURLs as described here are very helpful for humans who wish to cite digital resources in their written or other work, another aspect of this is ingestion by machines. For this to occur, Data Interoperability relies heavily on the use of well-known and well-specified formats, such as XML, Resource Description Framework (RDF) for Linked Data, SQL dump file, which is a text file which contains a series of SQL commands, or a Comma Separated Value (CSV) file.[73] Using standard data formats is fundamental to sustainability but is also essential for data interoperability. The reason for this is that standard formats tend to have well-supported open source libraries meaning that they will remain usable in the longer term. Additionally, standard data formats have formal specifications so even if there are no implementations for a particular computer language or system, the formal specification will allow someone to write one with relative ease. This means that the format can be read and understood.

If a relational database is used, the data storage is, in terms of performance characteristics, binary. However, best practice is to store data in a text-based format. This is because binary formats are fragile when attempting to share data between two unknown systems and binary formats are not easily understood by humans without a very high level of skill in computers and binary packing. Relational databases come with tools to extract the data and store it as a set of text files. These are called *dumps* of the database. For example, in MySQL, this tool is called *mysqldump*[74] and in PostgreSQL, it is called *pg_dump*. Most databases in the Humanities are very small when compared to most commercial databases and will thus be easily stored off the server and will make recovery possible in cases of catastrophic loss. Furthermore, it is advisable to make a dump from a relational database available for use by others. This is because researchers may wish to ask questions of the data, especially in the aggregate, that a simple search form on a website will not allow. Not making dumps from

---

[73] For CSV file see https://datatracker.ietf.org/doc/html/rfc4180 and for data formats in general see https://www.w3.org/TR/dwbp/#dataFormats

[74] See the MySQL manual for more information: https://dev.mysql.com/doc/refman/8.0/en/mysqldump.html. See also the PostgreSQL manual: https://www.postgresql.org/docs/13/backup-dump.html

relational databases available could stifle innovation in the future and make duplication of effort necessary which will slow the pace of research. Additionally, since the databases are generally small, the dump files can be easily shared between researchers if necessary.

Recommendation 5: Websites should be built using recognised data formats, such as XML, or in open-access relational databases, such as MySQL and PostgreSQL

Recommendation 6: Dump files should be created frequently as backup and made available to the research community to encourage innovative uses of the data.