UK-Ireland Digital Humanities Association

Multilingual DH in the UK and Ireland: Summary Report and Future Recommendations

Authors: The *Multilingual DH in the UK and Ireland* Community Interest Group ([web](web))

Pedro Nilsson-Fernàndez (Department of Digital Humanities, University College Cork)
Paul Spence (Department of Digital Humanities, King's College London)
Naomi Wells (Digital Humanities Research Hub/Institute of Languages, Cultures and Societies at the School of Advanced Study, University of London)

1. Introduction

The last few years have seen a steady growth in work acknowledging the challenges of linguistic diversity in the Digital Humanities (DH). Early research critiquing issues around global diversity and visibility in the field as a whole (Fiormonte, 2012; Galina, 2014) appeared around the same time as a significant number of new DH associations/groups were created in countries where English is not the official language.[1] There have been both efforts to problematise language diversity in DH's internal communications and wider attempts to foster greater multilingual support in its research practices. Structures in some major DH organisations have changed to accommodate the recent drive to encourage multilingual practices and policies. The Alliance of Digital Humanities Organisations (ADHO), which had a standing committee on Multilingualism and Multiculturalism as early as 2005, has recently formalised the international Multilingual DH Network initiated through the work of Dombrowski, by creating a Multilingual DH Special Interest Group in 2023 led by Till Grallert and Merve Tekgürler. Meanwhile, the creation of DARIAH's Multilingual DH Working Group in 2023, co-chaired by Alíz Horváth and Maroussia Bednarkiewicz, aims to draw on the multilingual nature of DARIAH's 22 European participant countries.

Our understanding of multilingualism in the context of DH acknowledges the fact that there are practices within the discipline that engage with and/or employ languages other than English in their research and teaching activities, while also recognising that researchers often engage with multilingual repertoires that are under-represented in the frequently culturally flattened and monolingual dynamics of digital tool development. As Susanna Nocchi noted, it is important to note that this is not simply a question of language support; limiting language choices also constrains our access to cultural systems, even within academia. This acknowledgement is key, as it understands that, even though English remains the dominant language in the field of DH, there are emergent voices that are slowly but surely diversifying the field and which are part of much broader efforts to expand digital knowledge production beyond a handful of hegemonic languages. It is often the case that attempts at performing DH through these 'other' languages encounter barriers constructed by the Anglocentric structures that dominate the field. To put this plainly: DH research and teaching in many non-anglophone parts of the world are often barely visible in the field's anglophone 'global' fora, which severely limits the kind of knowledge exchange the field is able to perform. Multilingual advocacy must then address the effects that these systemic barriers have across the different aspects of the work carried out in the Digital Humanities and digital scholarship more generally.

There have been a number of publications covering global and multilingual practices in the Digital Humanities in recent years, including broader reviews of global diversity such as *Global Debates in the Digital Humanities* (Fiormonte, Chaudhuri and Ricaurte, 2022), and more focused pieces on community translation and minimal computing (Gil and Ortega, 2016),

---

[1] The Japanese Association for Digital Humanities (JADH), the Associazione per l'Informatica Umanistica e la Cultura Digitale, the Red de Humanidades Digitales (RedHD) and Humanidades Digitales Hispánicas (HDH) were founded in 2011, followed by Digital Humanities Deutschland (DHD) in 2012 and Associação das Humanidades Digitais (AHDig) and Asociación de Humanidades Digitales en Argentina (HDA).

multilingual stakeholders and workflows (Horváth, van Lit, Wagner and Wrisley, 2024), or drives to improve the visibility of lower resourced languages (Thieberger, 2017). There has also been related research by language activist groups, such as the Whose Knowledge? Initiative's work to make the internet more multilingual (Whose Knowledge? et al., 2022).

Prior to the formation of this Community Interest Group, its convenors had been involved in various activities related to the topic. Wells and Spence have sought to bring together languages scholars engaging in digital research through the *Digital Modern Languages Seminar Series* launched in 2019, with some of its outputs crystallising in the [*Digital Modern Languages Section*](url) on the *Modern Languages Open* platform. The [*Disrupting Digital Monolingualism workshop*](url) in 2020 aimed to bring together broader discussions around digital multilingualism in four general areas: linguistic and geocultural diversity in digital knowledge infrastructures; multimodal multilingual methods and data; transcultural and translingual approaches to digital study; and artificial intelligence, machine learning and NLP in language worlds. In 2020, Pedro Nilsson-Fernández was involved in creating the [Associació d'Humanitats Digitals Catalanes (AHDCat)](url), which promotes work on the field in the context of Catalan language and culture. Members of the group have produced several research outputs on the subject (Spence & Wells, 2021; Spence & Brandao, 2021; Spence, 2021; Nilsson-Fernàndez & Dombrowski, 2022; Viola & Spence, 2023).

While in the North American context the flourishing of DH in English Departments (Kirschenbaum, 2010) favoured the dominance of anglophone approaches, the cases of the UK and Ireland are significantly different. Anglophone dominance in the field is still patent, but a wider range of approaches and disciplinary perspectives (including linguistics; cultural studies; history; modern languages; area studies) have shaped debates about multilingualism in DH. Some of the expertise we recently see being consolidated in Multilingual DH in the UK and Ireland is owed to scholars formed in the strong tradition of Modern Languages in these islands who engage with digital scholarship and often establish strong connections with scholars and institutions beyond the UK and Ireland, particularly in the European context. However, we recognise the need to cultivate a culture of collaboration between different fields of DH and, in line with the wider agenda of this association, we also believe that a multi-sector approach is needed to address the ways anglophone assumptions and biases operate across a wide range of digital practices and infrastructures.

Our objectives in initiating this CIG were both internal and external in nature. We aim to:

- Foreground multilingualism within the Association's own practices.
- Advocate for multilingualism to be supported in digital research practices and infrastructures.
- Raise the visibility of the UK and Ireland's diverse range of languages in digital research.
- Value and widen the multilingual knowledge and expertise of the UK and Ireland's DH community.
- Contribute to the Association's commitment to inclusivity and diversity, and foster politically and socially engaged understandings of multilingual DH research.

In addressing these objectives, we hope to draw on the expertise of a range of different stakeholders, including:

- Academic researchers across the (digital) humanities who conduct digital research through languages other than English.
- Digital practitioners/Research Software Engineers (RSEs)/technical leads who support and conduct multilingual research.
- Archivists, curators and librarians who work with digitised/digital materials in languages other than English.
- Academic publishers who support multilingual research.
- Digital language activists.

We are very aware that many of our objectives are shared by the several initiatives mentioned above. Thus, we aim at working closely, for example, with the ADHO and DARIAH Multilingual DH groups, in order to avoid duplicating work that is being done elsewhere. Furthermore, we do have a particular focus on multilingualism in the context of UK and Ireland's Digital Humanities practices, and hope to engage with the specific challenges in this context. One of the main concerns of the group is making sure that practitioners in the UK and Ireland are made aware of the existing support and networks available to them, further increasing their visibility and the interconnectivity of different initiatives.

With this context in mind, the group hopes to engage with practitioners working on indigenous languages from both countries, both institutionally recognised (Irish, Welsh and Scottish Gaelic) and with less institutional support (Cornish, Guernesiais, Jèrriais, Manx, Scots, Shelta, Ulster Scots, and the different regional variants of Romani), together with the languages of diasporic communities living in the islands: Polish, Romanian, Panjabi, Urdu, Portuguese, Spanish and Arabic, just to name the seven most widely spoken ones in the UK (ONS, 2021); Polish, Romanian, French, Spanish, Portuguese, Lithuanian and German in the Republic of Ireland (CSO, 2023). While the group hopes to engage with work done on as many of the above languages as possible, our scope will always be focused on approaches coming from the field of Digital Humanities. In the following sections it will become apparent that Irish, Welsh and Scottish Gaelic are the three indigenous languages most commonly featured in the digital initiatives we outline; this may be indicative of a greater availability of resources in these institutionally supported languages, but does not reflect the goals of the group inasmuch as we are committed to highlight DH initiatives in less supported languages. The same can be said about our commitment to supporting projects engaging with languages spoken by the diasporic communities in these islands; as mentioned in our initial proposal, it is key to promote Digital Humanities work that uses activism to address any form of anti-immigrant and isolationist sentiments, as well as related forms of 'linguaphobia'. This does not only reiterate the Association's dedication to inclusivity and diversity, but also encourages politically and socially engaged approaches to Multilingual DH research.[2]

---

[2] One example of the type of engagement we refer to can be observed in Saskia Huc-Hepher's web archiving project, *The London French Special Collection*. The project's impact study case report highlights the ability of such a project to engage with communities living in the diaspora and invite them to

This report draws largely on the results of an online workshop organised by the Multilingual Digital Humanities in the UK and Ireland Community Interest Group on Friday 24th November 2023. This public workshop aimed to discuss and develop recommendations for embedding multilingualism within and beyond the UK and Ireland's Digital Humanities communities. The workshop included both open participation and invited contributions from DH researchers who have taken a leading role in furthering Multilingual DH research locally and internationally. Those who attended the workshop were invited to share their own ideas and recommendations for raising the visibility of and expanding Multilingual DH by answering a series of questions around the languages they speak and use in their research, obstacles and barriers they faced in embedding multilingualism in their research communities, guidelines/models of best practice which influenced them or specific actions they felt the association should take to support and foster multilingual awareness and performance.

The invited contributors to the local panel (UK & Ireland) were Patrick Egan (Munster Technological University), Thea Pitman (University of Leeds), Martin Wynne (University of Oxford and CLARIN) and Shuang Xiao (University College Cork). The international panel consisted of Till Grallert (Universitätsbibliothek, Humboldt-Universität zu Berlin), Alíz Horváth (Eötvös Loránd University & Chair of DARIAH Multilingual DH WG), Cosima Wagner (Freie Universität Berlin) and Jonas Müller-Laackman (State and University Library Hamburg).

It is important to note that, while many of the outputs in this report were the direct result of the input of the above speakers, they were not directly involved in the production of this report at its draft stages. Any inaccuracies or misinterpretation of the points made during the workshop are the sole responsibility of the Multilingual DH CIG convenors. Feedback was also invited from other scholars not present at the workshop, and their contributions via comments are mentioned and credited in the text.

2. Creating language-inclusive DH spaces

One of the priorities of the CIG is bringing multilingual practices into the spaces in which we develop our activities, transforming teaching and research spaces alike into language-inclusive spaces. Whether conceived to gather practitioners virtually or in-person, the construction of these spaces must be a community-building exercise in itself, aiding in the consolidation of a network of experts in Multilingual DH who will ultimately ensure that these spaces help counter "language indifference" (Burdett, Burns, Duncan & Polezzi, 2018; Forsdick 2017) and facilitate wider engagement with multilingualism among scholars at all levels.

An early example of such efforts can be found in the original ADHO *Discussion Paper: Internationalization, Multi-lingual & Multi-cultural agenda* (Burr, 2006), where steps were recommended to embed more inclusive linguistic practices in the organisation of DH events. In

---

participate in digital curatorial practices, and the approach has since been extended to Russian, Chinese and Latin American communities in the UK.

the context of research events, one of the first successful examples of embedding multilingualism in a DH conference was the 2014 GO:DH whisper initiative, in which attendees could wear pins indicating the languages they could help interpret from; Élika Ortega's account of the community reactions speaks to the readiness of the DH community to engage in these practices (2014).

We are therefore committed to building on existing guidelines such as the GO:DH "Translation toolkit" but we are also aware that not every event benefits from the same multilingual composition of the attendees of DH2014, and the contexts in which we are trying to operate are of a very heterogeneous nature; in smaller or less global events it may be often the case that we find ourselves to be the only, or one of a few, multilingual scholar(s) in the room, and this needs to be accounted for when creating/building on existing guidelines.

Careful examination of the specific requirements to offer effective multilingual support for DH events is needed – this is not just about adding a list of new languages, although strategic thinking about which languages are given sustained support would also be a good approach, as well as catering for different geographical and cultural contexts on a case-by-case basis. There is not a single solution or approach that fits all contexts, and piloting new methods is something that we very much welcome; during our workshop, Till Grallert suggested the concept of code-switching during presentations as a performative way to highlight multilingualism and challenge monolingual norms. This is a practice that was also observed during the Disrupting Digital Monolingualism workshop in 2020 and that aligns with our aim at making multilingualism more visible within the DH community. Merve Tekgürler highlights the importance of using examples from languages other than English in teaching/conferencing practices as a way of decentering English from the curricula. In the UK and Irish contexts (as is the case of other English-speaking areas in the world) there is typically an unvoiced assumption that an event, if not indicated otherwise, will be delivered in English, despite the multilingual composition of our own DH communities and the international nature of many of our events.

As a CIG we strongly recommend that, even if English is the main working language of an event, this should be explicitly acknowledged (Bender, 2011). We also encourage all event organisers to consider in particular how best to support the inclusion of speakers of other languages, and reduce, whenever possible, the associated unpaid labour that typically becomes an added burden to the multilingual scholar (Nilsson-Fernàndez & Dombrowski 2022). In the second annual event of the Association, for example, we have made the bilingual CfP and interpretation from Gaeilge to English available thanks to a paid collaboration with UCC students at the Irish Department; our speaker was kind enough to share a draft of the text with us in advance, and we have used a relatively simple and free software for the live transcription. However, other event organisers who have offered similar provisions have found there has often been limited take-up of the opportunity to present in other languages and it will be important to undertake further research to understand the barriers and power dynamics that influence these decisions.

Such a relatively small step taken by event organisers can go a long way in helping scholars whose first language is not English feel that they are part of the conversation; during our workshop, Postgraduate DH student Shuang Xiao highlighted the anxiety felt by some of her peers from the Chinese DH community when presenting their work in English. Some of these barriers often go beyond the linguistic; many of the tools and platforms that are common in the anglophone world and that we take for granted (i.e. Google or Microsoft software packages) are either not commonplace, or simply not available in some geographical contexts. In this respect, we would highlight the importance of adopting an intersectional approach that recognises how language may intersect with other forms of digital exclusion. As highlighted by Grallert in our workshop, there are particular opportunities here for Multilingual DH to engage with the work associated with minimal computing in DH (Gil, 2015; Radio, 2020; Risam et al, 2022), and to further both multilingual approaches to minimal computing as well as minimal computing approaches to Multilingual DH.

3. Supporting the development and dissemination of multilingual DH tools and resources

The Digital Humanities have a long history in developing research tools, platforms and datasets across the world and in many languages, but it is widely accepted that this is subject to geolinguistic biases. Addressing this challenge requires both better awareness of the problems and concerted action from a wide range of actors: educators; students; the GLAM sector; and collaborators from the tech sector among others.

This is of course a global challenge, but, where appropriate, we will point to specific actions which can be taken in a UK-Ireland context here. We would also clarify that we are not proposing a specific multilingual infrastructure for UK and Ireland here, but rather we wish to draw attention to an already extensive global infrastructure for addressing multilingualism, in particular at European level through organisations like DARIAH and CLARIN, which researchers in the UK and Ireland can connect with and customise.

Awareness raising

There are a number of different aspects to advocating for greater awareness about multilingualism in DH resource creation and tool/platform development:

- Multilingualism, properly understood here, integrates linguistic and cultural dimensions. Multilingual DH both relies on the kind of language resource and technology creation we associate with applied linguistics and the kind of cultural criticism we associate with fields such as Modern Languages (see Section 5).
- We need to consider the wider context for tool/resource development. The practical challenges do not just relate to languages, but also scripts and directionality. The challenge includes both historic and living languages/variants. In a UK/Ireland context,

- we need to consider a range of indigenous, heritage and community languages used on these islands.
- There is a range of stakeholder perspectives to take into account here – including for example those of researchers, teachers, librarians or developers. Important recent work by Horváth, van Lit, Wagner and Wrisley (2024) has advanced our understanding of this, but some concerted work on stakeholder perspectives may be beneficial in a UK-Ireland context too. Given the long history of tool and resource creation in UK and Ireland DH, there are particular opportunities and challenges to address in that regard.
- A Multilingual DH strategy will need to include action at different levels of material scale from 'small' to 'big' data perspectives, and at different scales of computation from 'minimal computing' to AI, NLP and Large Language Models.
- Our capacity to generate multilingual resources or tools is not unlimited, and we will often need to 'be strategic' in how we address multilingual challenges, grouping together similar challenges, e.g. around language scripts, lower resourced languages, storytelling or oral communication to facilitate cross-language knowledge sharing about tool creation.
- A central challenge is how people can discover, share or disseminate information about multilingual DH since it is a challenge involving multiple fields, digital practices and professional roles.
- A multilingual turn in DH will need to consider both increasing multilingual support (more languages) for platforms in a more formal capacity-building sense, and taking more agile approaches to customise existing tools and frameworks to allow users to adapt them using a hacker-style approach.

Areas of action

In our November 2023 workshop, participants highlighted a number of different ways in which we can foster multilingual action.
- **Availability of language resources**. There are obstacles in locating digital language resources for languages other than English. Martin Wynne of CLARIN highlighted efforts to address these problems by: building a network of experts; ensuring the interoperability of digital language resources; enhancing automatic speech recognition systems and language modelling; collaboration and resource sharing. CLARIN Knowledge Centres provide contact points to centres of expertise in language resources and technologies in a number of languages and domains. As Wagner's work has highlighted, GLAM institutions are also storage locations of resources in multiple languages and we should also work to collaborate more directly with library associations and networks like RLUK and LIBER Europe to support Multilingual DH in the UK and Ireland.
- **Digital methods**. Information about digital methods can be hard to come by in languages other than English and projects such as the Open Methods Initiative are important in fostering awareness of the situated use of digital humanities methods and tools in a number of different languages.
- **Platforms and workflows**. The logic of 'localisation' often takes languages as flat and clearly bounded objects which can be mechanically translated into each other. The

Programming Historian provides an important model for how to engage with multilingual workflows, both between and within individual language communities.

- **Natural Language Processing (NLP)**. NLP and Large Language Models (LLMs) resources are heavily biased towards a small group of highly resourced languages and there is a serious barrier to using computational methods for many languages in the world (Nicholas & Bhatia, 2023). Digital humanists have contributed to wider efforts to address the lack of support for NLP-based approaches through resources such as the snapshot Multilingual DH created in 2019 and the New Languages for NLP project, which taught participants how to annotate linguistic data and train statistical language models in a number of selected historic or low resourced languages.
- **Datasets**. There are numerous Multilingual data challenges, ranging from data standards/modelling, semantic representation or visualisation to data analysis tools and workflows. A recent special collection in the Journal of Open Humanities Data (ed. Viola) aims to address "the intersection of data and workflows in Multilingual DH".

4. Supporting under-resourced languages of the UK and Ireland

One of the main concerns identified during our workshop was how under-resourced languages other than English were in the context of Digital Humanities in the UK and Ireland. Speaking to the case of *Gaeilge* in Ireland, Patrick Egan pointed to issues such as incomplete datasets and insufficient scholarly engagement with Multilingual DH to date.[3] However, there are initiatives working towards greater engagement in this area that build on existing expertise.

A notable case is the plan to create DR-LIB: a Knowledge Centre to support the use of Digital Resources for the Languages of Ireland and Britain, presented at the Association's second annual event by Martin Wynne, Mo El-Haj, Dawn Knight, Paul Rayson and Mícheál J. Ó Meachair. We welcome the much-needed initiative by the expert team behind this as an example of cross-language collaboration across Britain and Ireland that will significantly strengthen Multilingual DH infrastructures. Meachair is also currently working on the National Corpus of Irish (Corpas Náisiúnta na Gaeilge), co-founded an Ríomhacadamh, a collective that localises and translates software into Gaeilge, and compiled a corpus of 7.5-million words of educational materials (EduGA) in Gaeilge. Knight is behind the projects *ThACC – Thesawrws Ar-lein Cymraeg Cyfoes - Using Word Embeddings to Create a Thesaurus of Contemporary Welsh', 'FreeTxt: supporting bilingual free-text survey and questionnaire data analysis'*, and 'CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes (The National Corpus of Contemporary Welsh): A community driven approach to linguistic corpus construction' (2016-2020).

---

[3] Egan was referring to digital infrastructures connected to his research in ethnomusicology and in this respect it is important to also consider the importance of digital tools and platforms that are able to archive and document oral sources and traditions and the frictions playing out between oral and textual traditions (Ong & Hartley, 1982; Foley, 1996).

It is in the interest of the CIG to engage with and further visibilise these initiatives so future Multilingual DH scholars can use and build upon them. It is also of value to acknowledge the prior work done through projects such as the [Corpus of Electronic Texts (CELT)](#), previously known as *Cork University and Royal Irish Academy* (CURIA), and originally *Thesaurus Linguarum Hiberniae;* a project led by a UCC colleague now retired, Peter Flynn, and which required the creation of The Sun workstation, the first web server ever built in Ireland, and the ninth in the world, back in 1991 (O'Sullivan, 2020).

When it comes to repositories, the Digital Repository of Ireland provides insight into the availability of resources in Gaeilge, as they are the national infrastructure leading digital preservation in the Republic of Ireland. An initial look at the repository indicates that out of the nearly 70 thousand objects (69,239) there are 9,552 with the "Irish" language tag on their metadata, which corresponds to 13% of the archived data. This number goes down to 11% when you remove objects that share both the English and Irish language tag. In the case of Wales, Reyer Zwiggelaar's project *Towards the creation of a National AI-enabled Repository for Wales* (2022), a pilot to use AI to bring datasets from Welsh digital repositories together, also had the objective of bringing the AHRC born-digital data community together. The University of Glasgow hosts *The Digital Archive of Scottish* (DASG), with a comprehensive electronic corpus, vernacular materials and audio recordings.

From the point of view of our CIG, one of the main priorities is to further improve the visibility of these initiatives. As a cross-languages initiative, the CIG also has a particular role to play in promoting the sharing of approaches, resources and best practices between the UK and Ireland language-specific DH research communities, as well as actively engaging with the wider international Multilingual DH communities.

5. Multilingual DH, digital cultures and digital theory

As highlighted earlier, the emphasis on languages in Multilingual DH should not be understood as a focus on purely 'linguistic' concerns. There can be a particular danger when isolating language from the cultural and social contexts in which it is used of treating it as merely a 'tool' of communication and relatedly seeing multilingualism as merely a 'technical' problem to be solved only through the development of more effective language technologies or the presence of multiple languages on digital platforms. For as Ponce de la Vega writes in relation to digital archives, "while multilingualism is fundamental [...], disregarding the geopolitics and colonial history of cultures, languages, and systems of knowledge can lead to a reinforcement of hierarchical structures of power" (2024, 40).

It is consequently important to highlight that a key focus of the Multilingual DH CIG goes beyond an isolated focus on languages to also support and raise the visibility of digital cultural and area studies research in the UK and Ireland that focuses on non-Anglophone contexts. In the UK

context, a primary example of this work is associated with the [Digital Latin American Cultures Network](#) led by Tori Holmes, Thea Pitman and Claire Taylor. Taylor and Pitman in particular articulated the vital contribution of linguistically- and culturally-specific cultural studies approaches to DH in contesting "assumptions regarding (unstated) Anglophone models of the digital" (Pitman & Taylor 2017). In this respect, the work of Multilingual DH is also closely related to the need to value and recognise different cultural knowledge and related theoretical approaches to the digital expressed in languages other than English.

Thinking more specifically of how the UK and Ireland's DH communities can respond to these issues, we would encourage explicit reflection on how our own digital teaching and research practices may be contributing to reinforcing the hegemony of anglophone perspectives on digital culture and theory. As highlighted in our workshop, we should reflect on our DH reading lists and curriculum, and explore how in collaboration with colleagues and our students we might incorporate perspectives in other languages in our classrooms, a good example being the [Diversifying Toolkit](#) created at the University of Edinburgh. We should also work to strengthen the presence of area studies within the UK and Ireland's DH community by better communicating the mutual benefits of such collaborations to both fields, which should be connected to efforts to broaden our international collaborations and partnership beyond primarily anglophone and Global North contexts.

Lastly, understanding language as a cultural and social practice rather than a merely functional tool is vital for incorporating into Multilingual DH work more complex theories of multilingualism that are grounded in an understanding of how humans use language in context. While we use the term 'multilingual' as it is the term most widely understood, this is not to ignore the critiques of the ways it can reinforce the idea of separate bounded and often national languages that obscure the more fluid and creative ways people mix and use languages. While terms such as translanguaging and translingualism are now widely used to reflect these more fluid practices in applied and sociolinguistics, the implications of these approaches for computational linguistics and relatedly the development of language technologies remain underexplored. Bridging these perspectives and approaches, while undoubtedly challenging, must be a core focus of Multilingual DH work in order to ensure the development of digital technologies, tools and methods for different languages does not risk further reinforcing bounded and nationalist conceptualisations of language.

6. Supporting cross-sector and cross-institution Multilingual DH work

One of the key messages that came out of the discussions we held was the wide range of sectors and departments who have a stake in Multilingual DH, even if awareness among some stakeholders of the importance of multilingualism in their digital practices may be limited. It is unsurprising that in the UK and internationally, those who we might broadly describe as 'languages researchers' have taken the lead in raising awareness of the importance of

multilingualism in DH. Nevertheless, multilingualism has implications for all areas of digital practice and research, as is likely to be more openly acknowledged in officially bilingual countries such as Ireland and Wales but which also applies in predominantly anglophone countries such as England.

Participants in our workshop highlighted nevertheless that it was challenging to maintain regular cross-institutional and cross-sector engagement in Multilingual DH activities and discussions. Workshop speakers also highlighted outwardly dismissive or even hostile attitudes, and such attitudes can be particularly strong in heavily anglophone contexts where monolingualism in English is often taken for granted. In this respect, it is important for Multilingual DH to form part of wider advocacy efforts to raise language sensitivity in society more broadly.

It is nevertheless also important to acknowledge that while for languages researchers in DH multilingualism is likely to be a primary concern across their work, for many, multilingualism is likely only one of a wide range of concerns competing for their time and attention. As such, engaging such wider audiences is likely to be best achieved through specific, targeted initiatives that surface particular issues where multilingualism and digital research/practice intersect in ways that are likely to be of particular interest to specific fields or sectors. Without seeking to be exhaustive, below we highlight key areas that have emerged both from our discussions and more widely in the field as particularly productive avenues for targeted initiatives and which are likely to be the focus of future CIG activities:

- **University libraries and the GLAM sector:** Libraries were identified by Wagner and Müller-Laackman as one of the primary stakeholders in Multilingual DH in the workshop, and in fact internationally much Multilingual DH work has been led by researchers based in university libraries such as Quinn Dombrowski in the US and Cosima Wagner in Germany. For as Goodale explains 'Libraries are often positioned at the intersection of the Digital Humanities, instruction, and public outreach, and are thus in a unique position to advocate for greater multilingual representation in DH' (Goodale 2024, 103). Outside of HE, a range of GLAM institutions in the UK and Ireland hold collections that include materials in a range of languages and consequently multilingualism cannot be ignored in cataloguing and archiving decisions. Technical challenges connected to multilingual metadata and the monolingual or Latin script biases of library systems are likely to be of particular interest to the sector, and were highlighted in our workshop by Wagner as crucial work for disrupting monolingual knowledge infrastructures. This work should also be connected to wider initiatives on decolonising the library/GLAM sector through greater multilingual sensitivity.
- **Academic publishing:** As underlined by the European OPERAS project on open scholarly communication (OPERAS Multilingualism White Paper July 2021), there is a close relationship between DH's commitments to open knowledge infrastructures and work to strengthen support for Multilingual digital publishing. The Programming Historian's work in this area over a number of years is particularly notable in this respect, and while primarily a DH initiative it stands as a leading example of Multilingual digital publishing in a broader English-dominated academic publishing sector. PH's work

illustrates the multilingual affordances of open digital publishing platforms, without ignoring the practical and technical challenges that they are working to share their approaches to tackling. Building on such efforts, there is scope to engage more widely with the academic publishing sector in the UK and Ireland in approaches to supporting Multilingual digital publishing. It is also worth highlighting that AI translation tools are likely to be of increasing appeal to publishers looking to reach different linguistic markets, and Multilingual DHers are ideally placed to provide the expertise needed to inform the development of ethical and trustworthy approaches to using such tools in ways that are informed by a sophisticated understanding of machine and human translation.

- **Education and schools:** When considering the future of Multilingual DH, particularly in the UK, in our workshop Pitman highlighted that we cannot ignore the decline in language learning in UK schools and relatedly the closure of languages departments in universities across the UK.[4] While as stated, Multilingual DH is not solely the concern of languages departments and researchers, it is undeniable that the future pipeline of Multilingual DHers in the UK is closely connected to this wider languages education context. Relatedly, discussions around the continued value of languages education increasingly make reference to the impact of the digital but can often depart from superficial understandings of the technologies themselves (see for example the recent Times Higher Education headline: 'Is AI the final nail in the coffin for modern languages?'). In this respect, educational interventions that focus on combining critical multilingual and digital literacies are vital for developing young Multilingual DHers as well as contributing more broadly to the future sustainability of languages education in the UK.

- **Technologists and software engineers:** Evidently a core aim of the CIG is to ensure greater awareness of multilingualism among those who develop digital technologies, platforms and tools, with Multilingual DH having important implications for topics such as digital design, user interaction, AI, NLP and machine translation. While strengthening connections across the tech sector should be part of this work, Tekgürler highlighted the fact that establishing successful collaborations can be challenging in terms of creating a meeting point where we are able to listen and talk as equal partners with stakes in what the future of technology looks like. It would be particularly valuable here to seek the advice and input of RSEs, especially those working in DH. There is particular scope for collaboration with the Association's RSE CIG on this subject and we would propose a future event focused on connecting the interests of both groups as key to developing this area of collaboration.

---

[4] While the situation is less acute in Ireland (Bruen 2023), we are not discounting the value of similar interventions in Irish education contexts and there are a number of recent funding initiatives in Ireland attempting to strengthen the study of languages in both secondary and higher education in Ireland.

7. Multilingual DH advocacy work in and beyond the UK and Ireland

As noted already, there are a number of different projects and initiatives designed to address Multilingual DH challenges. As part of our efforts to raise awareness of and support these at local level, we will engage with international initiatives in the interest of following best practices and in order to avoid overlap/duplication of work, but with a strong focus on addressing the specific challenges in these islands. Discourse in the UK (and to a lesser extent Ireland) has traditionally held a strong bias towards assumptions about monolingualism and/or the dominance of English, and one of our main goals is to map and visibilise the actual extent and nature of multilingualism in the two countries. These practices are part of a much wider range of activities carried out within other areas of research (such as modern languages, area studies, sociolinguistics or computational linguistics) or by language communities and activists. These present a number of different models for community building and advocacy to support Multilingual DH efforts:

- **DH language communities**. Certain language communities have been historically internationally visible (such as Italian or Japanese) in the digital humanities, and others have become much more visibly prominent in recent years (such as Spanish or German). These communities often bring together researchers working in those languages, with others studying them, but they are generally highly-resourced languages.
- **Tool creation/documentation**. As seen with the NLP examples provided earlier, the DH community has often seen it as an important part of its role to document or create tools to address imbalances in digital language resources.
- **Agile/pop-up multilingual initiatives**, such as those fostered by Global Outlook DH. These have typically sought to take tactical approaches to issues in the short term in order to draw attention to multilingual challenges, or in order to crowdsource responses.
- **Multilingual DH networks**, such as the ADHO or DARIAH networks, or indeed our UK-Irish initiative. These tend to take a more strategic or programmatic approach to resolving multilingual issues.

We regard all of these efforts as vital, but we also believe that it is important to take strategic and policy perspectives on the challenge. In particular, participants felt that it is important to:

- Consider how we can best prepare the ground for good advocacy, analysing the main challenges in terms of human resources, budgets, hierarchies, cultural/linguistic assumptions, etc.
- Analyse the key expertise required to facilitate greater multilingual support.
- Offer examples of best practice, positive models, guidelines and sample multilingual workflows.
- Address interdisciplinary challenges: how to facilitate stable collaboration on Multilingual DH with a community which is highly diverse across disciplines, professional roles,

geography and languages, and which therefore has different motivations and systems of validation.

- Review knowledge infrastructures beyond DH and consider how we can disrupt the broader knowledge infrastructures which universities and other research institutions depend on – search, library/research software, text mining 'labs' or wiki platforms.
- Articulate computational research and policy areas where DH can engage with or influence wider multilingual agendas, e.g. in relation to NLP and AI.
- Use our lobbying power as an Association to identify strategies for making multilingualism a key factor in debate around the research infrastructure we need. We need to consider how we can make it what Wagner called "a fundamental condition" in the decisions taken by senior administrators and policy makers.
- Discuss how to make Multilingual DH work more visible, including translating definitions, policy briefs and action sheets.
- Identify agile multilingual approaches. Multilingual DH has often been most successful where it has taken a tactical or agile approach to its advocacy, e.g. through the DH Whisperers initiative mentioned elsewhere.
- The approaches mentioned in the previous point must also account for and acknowledge that the labour associated with facilitating multilingual practices in the field often falls on multilingual practitioners only; solutions to alleviate that additional toil will be central to our advocacy.

8. Conclusion

While as we have highlighted multilingualism is a complex topic that involves a range of stakeholders, we hope this report provides researchers and practitioners in the field with an understanding of the challenges associated with Multilingual DH research and practice, as well as the opportunities to build on and collaborate with existing initiatives in and beyond the UK and Ireland. We have also sought to highlight that rather than a mere technical issue, multilingualism is crucial to the Association's commitment to fostering inclusivity in DH and should be core to the future DH research agenda.[5] The CIG's work is also closely connected to the Association's emphasis on cross-sector collaboration, with collaborations with libraries and publishing initiatives particularly crucial for ensuring the development of more multilingual digital knowledge infrastructures.

In this respect, while acknowledging the crucial role of languages researchers in the present and future of Multilingual DH, the aim of the CIG as a DH initiative is to provide a channel for greater engagement with those for whom multilingualism may not be the core focus of their work but who share our commitment to more inclusive digital practices and infrastructures. As Horváth highlighted in our workshop there are many who feel they aren't sufficiently 'multilingual' and/or sufficiently 'DH' to be involved in multilingual DH work, and in this respect it will be crucial that the CIG's future activities explicitly invite in a range of perspectives beyond those who

---

[5] It is notable that at the 2024 Annual Event there were a wide number of presentations addressing multilingualism, which suggests this is already in progress.

explicitly identify as 'multilingual DHers'. Our focus over the next 2-3 years will consequently be on organising more targeted initiatives that focus on specific areas and issues we and others have identified as most pertinent to different sectors and/or disciplines.

Such initiatives are likely to involve international collaboration with some of the groups and initiatives mentioned in this report. As this report highlights there are a wide range of existing infrastructures and initiatives, and in a context of limited resources our workshop discussions raised the risks of duplicating or dissipating the efforts of the existing small but dedicated international community of multilingual DHers. At the same time, our aim with setting up the UK-Ireland specific CIG is to provide a channel for those in the UK and Ireland to better understand how they can connect to and contribute to such initiatives. We believe this channel is particularly crucial in the heavily anglophone and monolingual research and teaching contexts of much of the UK and to a lesser extent Ireland, where knowledge of Multilingual DH resources and initiatives can be less easily accessible.

While the CIG will seek to further develop our activities in alignment with the principles, approaches and priority areas outlined in this report, we conclude with some brief recommendations for the UK and Ireland's DH communities on how they can support or engage with Multilingual DH in their own practices and contexts:
- DH event organisers should explicitly acknowledge the working language of an event and discuss strategies for including multilingual speakers in their contexts.
- DH educators should explore with their colleagues and students how to incorporate DH resources and readings in languages other than English into their curriculums, and explicitly address in their classrooms the limitations of DH tools for working with other languages.
- DH researchers/practitioners in the UK and Ireland should further engage with and contribute to international initiatives such as CLARIN, DARIAH's Open Methods Initiative, the DARIAH Multilingual DH library and the Programming Historian to promote knowledge of and expand the resources available to support Multilingual DH research.[6]
- DH researchers/practitioners should explore possibilities to work with local language communities and the researchers who work on these languages to discuss with them digital research opportunities that support their specific needs and interests.
- DH researchers/practitioners should strengthen connections with UK- and Irish-based area studies and collaborators beyond predominantly anglophone contexts to better recognise and incorporate into our understanding of the field multilingual digital knowledges and cultures from inside and outside of the UK and Ireland.
- DH researchers and practitioners involved in the development and/or implementation of digital platforms and systems should reflect on the anglophone assumptions that

---

[6] It is important to acknowledge here that these initiatives are primarily based in Europe. This is partly connected to the CIG's location within Europe but also inevitably to the concentration of resources to sustain such initiatives in the Global North. There are nevertheless important initiatives undertaking similar work in other regions of the globe such as Masakhane, a grassroots NLP community for Africa, by Africans.

underpin these platforms/systems and explore how they can more effectively include users of and materials in languages other than English.

9. References

Bender, Emily M. 2011. "On Achieving and Evaluating Language-Independence in NLP." *Linguistic Issues in Language Technology* 6 (October). https://doi.org/10.33011/lilt.v6i.1239.

Bruen, Jennifer. 2023. "Foreign Language Learning in the Republic of Ireland: Developments, Challenges, Policy Implications". *Languages, Society & Policy.* https://www.lspjournal.com/post/foreign-language-learning-in-the-republic-of-ireland-trends-developments-challenges-and-policy

Burdett,Charles, Jennifer Burns, Derek Duncan and Loredana Polezzi. 2018. "Transnationalizing Modern Languages: Reframing Language Education for a Global Future." https://www.bristol.ac.uk/media-library/sites/policybristol/PolicyBristol-report-35-Sept18-transnationalizing-modern-languages.pdf

Burr, Elisabeth. 2006. "Discussion Paper: Internationalization, Multi-Lingual & Multi-Cultural Agenda." *Alliance of Digital Humanities Organisations*. https://adho.org/wp-content/uploads/2022/07/MLMC_DiscussionPaper_2006.pdf

Central Statistics Office, Ireland (CSO), released 2023, CSO website, Summary Results, Migration and Diversity 2022 Census. https://www.cso.ie/en/releasesandpublications/ep/p-cpsr/censusofpopulation2022-summaryresults/migrationanddiversity/

Fiormonte, Domenico, Sukanta Chaudhuri & Paola Ricaurte, eds. 2022. *Global Debates in the Digital Humanities*. Debates in the Digital Humanities Series. Minneapolis: University of Minnesota Press.

Fiormonte, Domenico. 2012. "Towards a Cultural Critique of the Digital Humanities." *Historical Social Research / Historische Sozialforschung* 37 (3 (141)): 59–76.

Foley, John Miles. 1996. "Signs, Texts, and Oral Tradition." *Journal of Folklore Research*, vol. 33, no. 1, pp. 21–29.

Forsdick, Charles. 2017. "'Translating Cultures', Translating Research," November. https://www.repository.cam.ac.uk/handle/1810/269993.

Galina Russell, Isabel. 2014. "Geographical and Linguistic Diversity in the Digital Humanities." *Literary and Linguistic Computing* 29 (3): 307–16. https://doi.org/10.1093/llc/fqu005.

Gil, Alex. 2015. "The User, the Learner and the Machines We Make." Minimal Computing
   Thought Piece. http://go-dh.github.io/mincomp/thoughts/2015/05/21/user-vs-learner/

Gil, Alex, and Élika Ortega. 2016. "Global Outlooks in Digital Humanities: Multilingual
   Practices and Minimal Computing." In. *Doing Digital Humanities: Practice, Training,
   Research*, edited by Constance Crompton, Richard J. Lane, and Raymond George
   Siemens, 58-70. Routledge.

Goodale, Ian. "Pedagogy and Praxis in Libraries. Natural Language Processing for
   Non-English Texts". In *Multilingual Digital Humanities*, edited by Lorella Viola and Paul
   Spence, 103-113. Routledge.

Horváth, Alíz, Cornelis van Lit, Cosima Wagner, and David Joseph Wrisley. 2024. 'Centring
   Multilingual Users: Thinking through UX Personas in the DH Community'. *Digital
   Studies / Le Champ Numérique* 13 (3). https://doi.org/10.16995/dscn.9608.

Kirschenbaum, Matthew G. 2010. "What Is Digital Humanities and What's It Doing in
   English Departments?" *ADE Bulletin*, no. 150.

Nicholas, Gabriel, & Aliya Bhatia. 2023. "Toward Better Automated Content Moderation in
   Low-Resource Languages". *Journal of Online Trust and Safety*, 2(1).
   https://doi.org/10.54501/jots.v2i1.150

Nilsson-Fernàndez, Pedro, and Quinn Dombrowski. 2022. "Multilingual Digital Humanities",
   in *The Bloomsbury Handbook to the Digital Humanities*, edited by James O'Sullivan,
   81-90. New York: Bloomsbury. https://hdl.handle.net/10468/14084.

Office for National Statistics (ONS), released 2021, ONS website, Statistical Bulletin,
   Language England and Wales census 2021.
   https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/bulleti
   ns/languageenglandandwales/census2021#main-languages-within-households

Ong, Walter J., and John Hartley. 2012. *Orality and Literacy: The Technologizing of the
   Word*. New York; London; Routledge.

OPERAS Multilingualism White Paper. 2021.
   https://operas-eu.org/special-interest-group-living-book/operas-multilingualism-white-p
   aper-june-2021/

O'Sullivan, James. 2020. "The Digital Humanities in Ireland," November.
   https://hdl.handle.net/10468/10746.

Ortega, Élika. 2014. "Multilingualism in DH. Notes from the DHSI2014 Unconference
   Session." *elikaortega.net*, June 17.
   https://elikaortega.net/blog/2014/multilingualism-dh/.

Pitman, Thea, and Claire Taylor. 2017. "Where's the ML in DH? And Where's the DH in ML?
   The Relationship between Modern Languages and Digital Humanities, and an
   Argument for a Critical DHML." *Digital Humanities Quarterly* 11 (1).
   https://www.digitalhumanities.org/dhq/vol/11/1/000287/000287.html

Ponce de la Vega, Lidia. 2024. "Diversifying Digital Biodiversity Knowledge: A Latin American Multilingual Perspective on the Biodiversity Heritage Library". In *Multilingual Digital Humanities*, edited by Lorella Viola and Paul Spence, 31-37. Routledge.

Radio, Erik. 2020. "Minimal Computing and Ontologies." Minimal Computing Thought Piece. https://go-dh.github.io/mincomp/thoughts/2020/07/21/minimal-ontology/

Risam, Roopika, and Alex Gil. 2022. "Introduction: The Questions of Minimal Computing." *Digital Humanities Quarterly* 16 (2). https://www.digitalhumanities.org/dhq/vol/16/2/000646/000646.html#mcwg2022.

Spence, Paul. 2021. "Disrupting Digital Monolingualism: A Report on Multilingualism in Digital Theory and Practice." London: Language Acts & Worldmaking project. https://doi.org/10.5281/zenodo.5743283.

Spence, Paul, and Renata Brandao. 2021. "Towards Language Sensitivity and Diversity in the Digital Humanities." *Digital Studies / Le Champ Numérique* 11 (1). https://doi.org/10.16995/dscn.8098.

Spence, Paul, and Naomi Wells. 2021. "Digital Modern Languages Launch Issue". *Modern Languages Open*. DOI: http://doi.org/10.3828/mlo.v0i0.412

Thieberger, Nick. 2017. "What remains to be done—Exposing invisible collections in the other 7,000 languages and why it is a DH enterprise." *Digital Scholarship in the Humanities* 32 (2): 423-434. https://doi.org/10.1093/llc/fqw006

Viola, Lorella, and Paul Spence (eds.) 2023. *Multilingual Digital Humanities*. Routledge

Whose Knowledge? et al. 2022. "The State of the Internet's Languages Report", a collaboratively researched and documented initiative, https://internetlanguages.org/

10. Key Resources/Links

Below follows a short list of some key resources related to multilingualism in DH spaces, with a particular focus on those which may be relevant to the UK-Ireland context. This list is by no means comprehensive, and we hope to expand it in future.

- Bibliographies:
    - DARIAH: The Living Biography of Multilingual DH (https://www.zotero.org/groups/4917236/the_living_bibliography_of_multilingual_dh/library)
    - Global Outlook:DH (https://www.zotero.org/groups/global_outlook_digital_humanities)
- Data, Tool Collections and Tutorials
    - Natural Language Processing for Multilingual texts (https://github.com/multilingual-dh/nlp-resources)
    - DARIAH Open Methods (https://openmethods.dariah.eu/)

- ○ CLARIN: Common Language Resources and Technology Infrastructure (https://www.clarin.eu/)
  - ○ The Programming Historian (https://programminghistorian.org)
  - ○ The Diversifying Toolkit, University of Edinburgh (https://edinburgh-uk.libguides.com/diversifyingtoolkit)
- Software localisation/translation initiatives
  - ○ An Ríomhacadamh (https://riomhacadamh.wordpress.com/)
- Digital Repositories, Archives and Corpora
  - ○ Digital Archive of Scottish Gaelic (https://dasg.ac.uk/en)
  - ○ Digital Repository of Ireland (https://dri.ie/)
  - ○ Corpas Náisiúnta na Gaeilge (National Corpus of Irish) (https://beta.corpas.ie/ga/cmg/)
  - ○ Corpws Cenedlaethol Cymraeg Cyfoes (The National Corpus of Contemporary Welsh) (https://corcencc.org/)
  - ○ Corpus of Electronic Texts (CELT) (https://celt.ucc.ie/)
  - ○ The Digital Archive of Scottish (DASG) (https://dasg.ac.uk/en)
  - ○ Thesawrws Ar-lein Cymraeg Cyfoes (ThACC) (https://github.com/CorCenCC/Thesawrws-Welsh)
- Multilingual DH Networks
  - ○ Multilingual Digital Humanities (https://multilingualdh.org/en/)
  - ○ DARIAH's Multilingual DH Working Group (https://multilingual.hypotheses.org/)
  - ○ OPERAS Multilingualism SIG: (https://operas-eu.org/special-interest-groups/multilingualism-2/)

Recommended citation:

Nilsson-Fernàndez, Pedro, Paul Spence and Naomi Wells. 2024. "Multilingual DH in the UK and Ireland: Summary Report and Future Recommendations" *UK-Ireland Digital Humanities Association*. https://doi.org/10.5281/zenodo.13683711