



Revue européenne des migrations internationales

vol. 30 - n°3 et 4 | 2014

Les traces de la dispersion / Numéro ouvert

Visualising the Emigrant Letter

Propositions pour une visualisation graphique de correspondances de migrants

Propuestas para una visualización gráfica de las correspondencias de emigrantes

Emma Moreton, Niall O'Leary and Patrick O'Sullivan



Electronic version

URL: <https://journals.openedition.org/remi/7081>

DOI: 10.4000/remi.7081

ISSN: 1777-5418

Publisher

Université de Poitiers

Printed version

Date of publication: 1 December 2014

Number of pages: 49-69

ISBN: 979-10-90426-23-8

ISSN: 0765-0752

Electronic reference

Emma Moreton, Niall O'Leary and Patrick O'Sullivan, "Visualising the Emigrant Letter", *Revue européenne des migrations internationales* [Online], vol. 30 - n°3 et 4 | 2014, Online since 01 December 2017, connection on 14 April 2022. URL: <http://journals.openedition.org/remi/7081> ; DOI: <https://doi.org/10.4000/remi.7081>

Visualising the Emigrant Letter

❖ Emma Moreton¹, Niall O'Leary²
and Patrick O'Sullivan³

Introduction

Over the past few decades, there has been a growing interest in emigrant letters. Emigrant letters have been used by social historians to investigate push/pull factors and the role of institutions and communities in the process of emigration; they have also been used to explore and understand the conditions and daily lives of emigrants and how individuals and groups adapted to the New World (see Thomas and Znaniecki, 1958; Erickson, 1972; Miller, 1985; Kamphoefner, Helbich and Sommer, 1988; and Fitzpatrick, 1994). Letters are expressive and indicative of correspondents' identities, values, preoccupations and beliefs; they are a powerful source of information and understanding about migration issues, provide a colourful picture of domestic life from an emigrant perspective, and shed light on processes of language change and variation (on this last point, see, for example, Dossena, 2007; Elspaß, 2002; and McLelland, 2007).

Much discussion of the use of "ego documents" or "documents of life" is built around the two editions of Ken Plummer's book, *Documents of Life: An introduction to the problems and literature of a humanistic method* (1983) and *Documents of Life 2: An Invitation to a Critical Humanism* (2001). Plummer has an interesting section on the use of letters and the reading of letters. However, Plummer, following on from Eugene Webb's concerns, from the 1966 book *Unobtrusive Measures*, highlights a problem, and to categorise this problem he uses the term "dross rate": "Letters are not generally focused enough to be of analytic interest – they contain far too much material that strays from the researcher's concern" (2001: 55). This would seem to be the case in some socio-historical studies. Arnold Schrier's *Ireland and the American Emigration, 1850-1900* (1958) and Kerby Miller's *Emigrants and Exiles* (1985), for instance, are less interested in the formulaic aspects of emigrant letters and more interested

1 Senior Lecturer, Coventry University, Priory Street, Coventry, CV1 5FB; emma.moreton@coventry.ac.uk

2 Digital Humanities Specialist, 48, Hyde Square, South Circular Road, Dublin 8; info@nialloleary.ie

3 Freelance Writer and Researcher, Glucksman Ireland House, New York University, One Washington Mews, New York, NY 10003, U.S.A.; patrickos@blueyonder.co.uk

in what letter writers have to say about, for example, their reasons for migration, experiences of the New World and memories of home. It is worth noting that for Webb and his colleagues “dross rate” is a social science technical term, which can be applied to any kind of data, not just letters (Webb *et al.*, 1966: 31).

Liz Stanley, however, takes issue with Plummer’s use of the phrase “dross rate”, as applied to letters: “[...]The features of letters he perceives as problems are the very things I find interesting and deserving sustained attention as analytical problematics” (Stanley, 2004: 202). For an example of how information perceived as unimportant by one scholar might be perceived as important by another scholar, David Fitzpatrick’s work on Irish emigration to Australia is especially useful here. Fitzpatrick, in the introduction to his 1994 book, *Oceans of Consolation*, explores earlier books of and on emigrant letters, the work of his precursors and peers. He looks especially at the editing practices of Blegen (Norwegian letters), Conway (Welsh letters) and Kamphoefner (German letters). Conway, “with due acknowledgement to Blegen, likewise excluded ‘dross’...” (that word again), while Kamphoefner deleted “ritualized pious reflections” and “endless lists of persons to whom the letter-writer sends his or her regards...” – although, to be fair to Kamphoefner, this was to do with space limitations of the printed book (Fitzpatrick, 1994: 20, referring to Kamphoefner, 1991: 46). Fitzpatrick’s own practice is to print his Irish emigrant letters in full, and his method is to take those lists of persons and make maps. Thus, on page 46, there is figure 10, “A Geography of Affection”, mapping the neighbours and relatives mentioned in one letter collection. Fitzpatrick’s maps are curiously moving; they provide a map of a mind, a map of memories – the landscape, as the emigrant remembers it.

There are two main issues being touched on in the discussion so far. The first relates to how letters are used in different ways across the disciplines: what is one scholar’s “dross” is another scholar’s gold. This can lead to the loss of information and/or the reduplication of work, with letter collections often being transcribed several times by different projects – projects which have their own research aims and, quite often, their own transcription and markup practices. The second issue is to do with interdisciplinarity: while different disciplines might use letter collections, the research is rarely interdisciplinary. Emigrant correspondence projects have often evolved independently of one another, and although project teams have been successful in tackling important research questions relating to, for example, social history and immigration studies they have rarely joined forces, or engaged with stakeholder groups from other disciplines.

There seems to be a good case, here, for developing a collaborative, cross-disciplinary approach to the digitisation and markup of historical emigrant letter collections. Presently, most existing digital letter collections consist of unannotated versions of original manuscripts. The digitisation process has made the letters more accessible to academics and the general public, and has also increased their searchability, at least to a certain extent. However, relatively few projects have moved beyond the digitisation stage to exploit text content and enhance usability and searchability through the use of digital technologies. Different emigrant letter collections cannot easily interconnect if they are simply digitised without markup, and some search pathways through the material will

remain unavailable if software tools are not employed to process this encoding.

In 2013 a research network was established by Emma Moreton and Hilary Nesi, Coventry University, with funding from the Arts and Humanities Research Council, to explore the digitisation and annotation of historical emigrant letter collections, with a special focus on letters of the Irish Diaspora. The project, "Digitising experiences of migration: The development of interconnected letter collections", was funded under the AHRC theme "Digital Transformations in the Arts and Humanities", which aims "to exploit the potential of digital technologies to transform research in the arts and humanities, and to ensure that arts and humanities research is at the forefront of tackling crucial issues such as intellectual property, cultural memory and identity, and communication and creativity in a digital age"⁴

The research network project started in April 2013 and finished in July 2014. Through a series of workshops, the project brought together scholars from different disciplines currently working with emigrant letters as a primary data source, to explore the digital potential of these perhaps iconic documents. In these workshops network participants examined how letters are being used across the disciplines, identifying where there are similarities and differences in transcription, digitisation and annotation practices. The aims were to discuss issues and challenges surrounding digitisation, build capacity relating to correspondence markup, and initiate the process of interconnecting resources to encourage cross-disciplinary research. Central to this was the development of a system of correspondence markup to represent the linguistic, structural, discursual, contextual and physical properties of the letters, thus offering different layers of meaning and ways-in to the texts, allowing for more sophisticated searches, and also the presentation of outputs through meaningful visualisations.

Eleven project partners were named in the original bid. The network subsequently grew to include over twenty scholars – an indication, perhaps, of an increasing interest in "histories from below", in this case from the perspective of those who directly experienced migration and its consequences. The network includes scholars from a range of disciplinary backgrounds (historians, migration studies experts, archivists, socio-linguists, corpus linguists and digital humanists), each with their own epistemological frameworks and ways of knowing. The range of backgrounds and expertise within our network offered many challenges as well as opportunities. Full details of the research network partners, together with the letter collections they are involved with, can be found on the project blog "Digitising experiences of migrations: The development of interconnected letter collections"⁵.

The following sections will briefly talk through the network process.

4 www.ahrc.ac.uk

5 www.lettersofmigration.blogspot.fr

Towards a System of Markup for Emigrant Letter Collections

The process of encoding is an intellectual activity, which involves thinking about which features of the document to represent, the relationship between those features and how those features should be named, described and categorised in a formalised way. The way a text is encoded will reveal something about what we believe to be important or salient about the original document. Encoding makes explicit our interpretation of a document and as such it is never a neutral process. The same text, then, can be encoded in many different ways, drawing out features that are most relevant to our own research interests, or projects. Encoding makes things visible – it enriches the text; it allows us to look at a text in new ways and from many disciplinary perspectives, offering different ways-in and providing different layers of interpretation. For the “Digitising experiences of migration” project, the encoding process was divided into two stages: document analysis and markup.

Stage One: Document Analysis

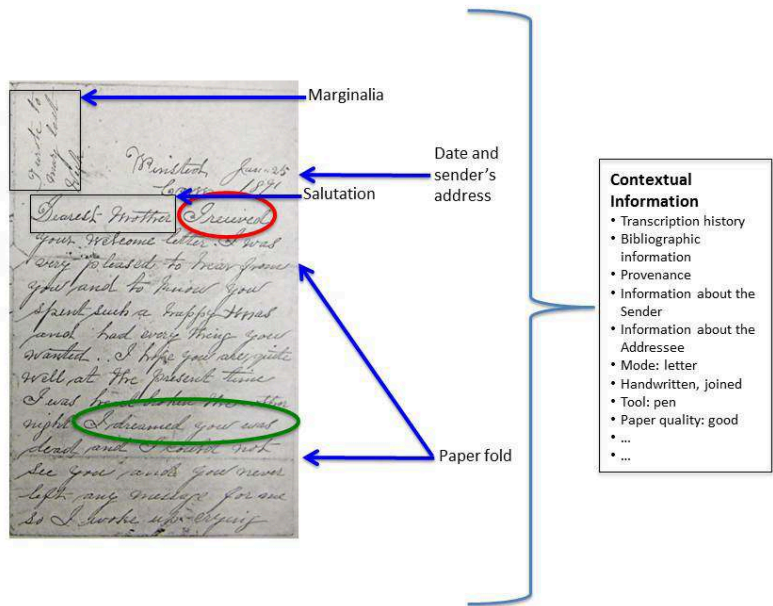
The first stage in our encoding project was to carry out a detailed analysis of the document/s to be digitised and marked-up, to identify features that are important in and across the various emigrant letter collections. This first step is necessary before encoding if the resulting markup is to be useful (Stadler, 2013). When analysing a selection of emigrant letters, the project partners were asked to consider two key questions:

- What do different researchers use correspondence collections for?
- What features of the letters are considered important across the disciplines?

In many ways, the document analysis stage of the process is the most challenging. Looking at Figure 1, for example, a letter by an Irish emigrant called Julia Lough, one can see a wealth of information that might be of interest to a broad range of scholars, with a broad range of research questions. A paper conservator or paper historian may, for instance, be interested in the “material letter” (Daybell and Hinds, 2010 and Steen, 1994), including the quality of the paper and how the paper has been folded (vertically through the centre and horizontally one third from the top/bottom of the sheet), whereas a graphologist may wish to capture information about the handwriting style, whether there are multiple authors contained within the document, whether the original letter is written in ink or in pencil and how the document is structured, perhaps distinguishing between writing that is contained in the margins and writing that is contained in the main body – information which might provide valuable clues as to the author’s circumstances and background. Additionally, whereas linguists may be more interested in capturing information about the language of the document, such as spelling variations (circled in red), unusual syntax (circled in green), omissions or repetitions, historians, on the other hand, may be more interested in any references to people, places and significant events. There will also be contextual information that is not explicitly stated within the document content, but which is useful to capture nonetheless. This information may be inferred from the document itself or it may be obtained from outside sources. Archival notes, for instance, confirm that the letter shown in Figure 1 was written by someone called Julia Lough, a female Irish emigrant from a town called

Meelick in (what was then called) Queen's County, Ireland. Additionally, previous research by immigration historians suggests that Julia had five sisters, three of whom, like Julia, emigrated to America in the 1870s and 1880s. And, finally, there is information available to us about where the Lough collection came from: a few of the letters were donated to Arnold Schrier, Professor Emeritus at the University of Cincinnati, by Canice and Eilish O'Mahony of Dundalk, Co. Louth and the remaining letters were donated to Professor Kerby Miller, University of Missouri, by Edward Dunne and Mrs Kate Tynan of Portlaoise, County Laois.

Figure 1: Document analysis
(letter by Julia Lough, Winsted, 1891 to her mother in Meelick, Ireland)



Document analysis, then, is the starting point of the encoding process, looking at how the various disciplines, with their diverse range of research questions, use emigrant letters. Through this process it was possible to identify where there were commonalities across the disciplines and where there were differences, and, finally, where reduplication was taking place. There is not enough space to go through all of the features that were examined, so this paper will focus on just three features of the letters that were deemed to be important across the disciplines, namely: person, location and date. Having examined a selection of emigrant letters, taken from different archives, the network partners agreed that – where possible – the following information should be captured in relation to person, location and date (see Table 1). There are two things to note here: 1) the lists contained within Table 1 are not exhaustive (depending on the research project other or additional information may need to be captured); and 2) it is very rare that all of this information is available for any one letter – quite often the researcher is working with a letter extract, for instance, and has very little information about the sender and/or the addressee.

Table 1: Person, location, date

Person (Sender/Addressee)	Location (Sender/Addressee)	Date
First name	Street	Day
Surname	Town	Month
Maiden name	Region/country	Year
Nickname/s	Country	
Date of birth	Geographical coordinates	
Date of death	Additional information	
Sex		
Occupation/s		
Social status		
Education		
Faith		
Relationships		
Places of residence		
Date of emigration + From/To		
Additional information		

Stage Two: Markup

Once the research network had examined the various ways in which letter collections are put to use, specifically identifying those features of letters that are important across the disciplines, the next stage was to agree on how to model that information using TEI markup, thus formalising and standardising this metainformation and allowing letter collections to potentially interconnect.

The research network used the TEI (Text Encoding Initiative) P5 Guidelines (TEI Consortium to carry out the encoding process⁶). The TEI provides a markup language and guidelines for working with digitised texts in the humanities. Specifically, the guidelines “make recommendations about suitable ways of representing those features of textual resources which need to be identified explicitly in order to facilitate processing by computer programs. In particular, they specify a set of markers (or tags) which may be inserted in the electronic representation of the text, in order to mark the text structure and other features of interest” (TEI Consortium, 2014: 23).

The TEI Special Interest Group on Correspondence, established in 2008 and led by Peter Stadler (Universität Paderborn), Marcel Illetschko (Austrian National Library, Vienna) and Sabine Seifert (Humboldt University, Berlin), “seeks to bring together scholars interested in creating digital scholarly editions of correspondence... to discuss and develop sample tagsets (including suggesting additions/modifications to the TEI Guidelines) for varying forms of correspondence” (TEI, 2014). In particular, the TEI Correspondence SIG are developing a special purpose element, <correspDesc>, which will allow features of letters such as sender and addressee to be represented in a standardised way. At the time of writing, this special purpose element is embedded under <profileDesc> within the TEI header.

The purpose of the <profileDesc> section of the TEI header is to provide a text profile “containing classificatory and contextual information about the text, such as its subject matter, the situation in which it was produced, the individuals

⁶ Document freely available online at: <http://www.tei-c.org/Guidelines/P5/index.xml>

described by or participating in producing it, and so forth” (TEI Consortium, 2014: 17). It should be noted that the markup being proposed in this paper is what was discussed and agreed on during the workshops; for the most up to date information about using TEI to model correspondence, readers should visit the Correspondence SIG wiki⁷ and for example xml files, showing applications of the correspDesc proposal.⁸

Figure 2: Example markup (person, location and date)

```
<profileDesc>
  <ct:correspDesc>
    <ct:sender>
      <persName key="LOUGHPers_0001"/>
    </ct:sender>
    <ct:addressee>
      <persName key="LOUGHPers_0002"/>
    </ct:addressee>
    <ct:placeSender>
      <placeName key="LOUGHPlace_0001"/>
    </ct:placeSender>
    <ct:dateSender>
      <date when="1891-10-25"/>
    </ct:dateSender>
  </ct:correspDesc>
</profileDesc>
```

The markup shown in Figure 2, above, captures information about the sender <ct:sender>, the addressee <ct:addressee>, and the date of the letter <ct:dateSender>. Embedded within <ct:sender> is the element <persName> that contains a unique identifier, or “key”, which links to a separate personography file (in this case LOUGHPers_0001). This individual personography file contains more detailed information about the sender (information such as that described in Table 1). Similarly, embedded within <ct:placeSender> is the element <placeName> that contains a unique identifier, or “key”, which links to a separate placeography file (in this case LOUGHPlace_0001). This individual placeography file contains more detailed information about the sender’s location (information such as that described in Table 1). Information about the addressee <ct:addressee> is organised in the same way.

The personography and placeography files are effectively the same as authority files – a term used by archivists and librarians to describe bibliographic “master” files. There is one personography file (or authority file) for each person and there is one placeography file (or authority file) for each place. These authority files for person and place are each given a unique identifier, which is then referred to in the markup. In other words, LOUGHPers_0001, for instance, points to a separate file which contains information about this particular participant (Julia Lough) such as date of birth, first name, surname, maiden name, nicknames (including all spelling variations), sex, occupations, date of emigration, date of death, etc. Similarly, LOUGHPlace_0001 points to a separate file which contains information about this particular place (Winsted) including, for

⁷ <http://wiki.tei-c.org/index.php/SIG:Correspondence>

⁸ <https://github.com/TEI-Correspondence-SIG/respDesc>

instance, its geographical coordinates. Having separate personography and placeography files (or authority files) for each person and place makes it much easier to manage changes to related metainformation at a later date. It is easier to change one master document – i.e. the personography or placeography file – than it is to change hundreds of documents.

In addition to the personography and placeography information `<ct:dateSender>` contains details of when the letter was dated. There are different ways to capture this information within the header. If the letter contains a date, then the day, month and year can be represented in the markup (as in Figure 2). However, quite often an exact date is missing and it is up to the researcher to make an educated guess as to when the letter was written. In such instances, the `notBefore` and `notAfter` attributes can be used to place the letter within an approximate timeframe:

```
<date notBefore="1800" notAfter="1899"/>
```

It is not possible, in the space of this paper, to discuss how the personography and placeography information can be modelled using TEI, however further information about modelling the emigrant letter, together with links to example xml files, can be found on the project blog.

Interconnecting Letter Collections

Having agreed on how best to model information relating to person, location and date, within the TEI header, it was then possible to start the process of interconnecting some of the emigrant letter collections the project partners are involved with. We focused on two collections in particular: the Irish Emigration Database (IED) – a collection of over 4,000 letters by Irish emigrants, held at the Mellon Centre for Migration Studies at the Ulster American Folk Park Museum in Omagh, Northern Ireland;⁹ and letters from the Digitizing Immigrant Letters (DIL) project at the Immigration History Research Centre at the University of Minnesota – a collection of eighty-five letters by migrants and their families in Europe and North America.¹⁰ The first step was to standardise the metainformation that is available for these two collections. An example of the raw metainformation can be seen in Figure 3, below. Project participant Luis Anke, Universitat Pompeu Fabra, then standardised this metainformation, converting it into CSV format (see Figure 4).

The standardised metainformation was then passed to Peter Stadler, Universität Paderborn, to be made TEI compliant. This involved two stages: 1) converting the CSV file to TEI via OxGarage;¹¹ 2) transforming generic TEI to correspDesc compatible TEI via XSLT.

9 <http://www.dippam.ac.uk/ied>

10 <http://www.ihrumn.edu/research/dil/aboutDIL.htm>

11 <http://www.tei-c.org/oxgarage/>

Figure 3: Example of the raw metainformation from the letter collection held at the IHRC, University of Minnesota

id	url	date	sender	place_of_sender	addressee	place_of_addressee
1	+id					
2	+472436	http://umedia.lib.umn.edu/node/472436	April 30, 1924	Grebenstchikoff, George	New York City, New York	Nicholas Roerich, Kullu, India
3	+710673	http://umedia.lib.umn.edu/node/710673	June 6, 1950	Budzka, Eduard	Wentorf, Germany	Klara and Vaclau Panucevich
4	+88790	http://umedia.lib.umn.edu/node/88790	September 10, 1957	Paikens, Helena	Minneapolis, Minnesota	her mother-in-law Anna Paikens
5	+88811	http://umedia.lib.umn.edu/node/88811	January 19, 1929	Nepriysky-Hranovsky, Serhii	Berezhtsi, Ukraine	his brother Alexander Granovsky
6	+88833	http://umedia.lib.umn.edu/node/88833	January 12, 1912	Aalto, Bert	Big Falls, Minnesota	his friend Hilma Aerila
7	+88831	http://umedia.lib.umn.edu/node/88831	August 26, 1911	Aalto, Bert	Big Falls, Minnesota	his friend Hilma Aerila
8	+472442	http://umedia.lib.umn.edu/node/472442	May 16, 1936	Roerich, Nicholas	Kullu, India	George and Tatiana Grebenstchikoff
9	+88793	http://umedia.lib.umn.edu/node/88793	December 11, 1898	Fazio, Lucia	Hoboken, New Jersey	Alessandro
10	+553855	http://umedia.lib.umn.edu/node/553855	July 5, 1927	Kovac, T. Paul	Hawleyville, Connecticut	his mother
11	+472455	http://umedia.lib.umn.edu/node/472455	November 8, 1949	Douline, Nikolay	Southbury, Connecticut	George
12	+710636	http://umedia.lib.umn.edu/node/472455	August 14, 1914	Archur Sch.	probably Buenos Aires	Emilie Mehle
13	+552682	http://umedia.lib.umn.edu/node/552682	1948-09-30	Petrus, Antonietta	Montreal, Canada	Loris Palmi

Figure 4: Standardised metainformation

id	url	date	sender	place_of_sender	addressee	place_of_addressee
472436	http://umedia.lib.umn.edu/node/472436	April 30, 1924	Grebenstchikoff George	New York City, New York	Nicholas Roerich, Kullu	Kullu, India
710673	http://umedia.lib.umn.edu/node/710673	June 6, 1950	Budzka Eduard	Wentorf, Germany	Klara and Vaclau Panucevich	Chicago, Illinois
88790	http://umedia.lib.umn.edu/node/88790	September 10, 1957	Paikens Helena	Minneapolis, Minnesota	her mother-in-law Anna Paikens	Lencini, Latvia
88811	http://umedia.lib.umn.edu/node/88811	January 19, 1929	Nepriysky-Hranovsky Serhii	Berezhtsi, Ukraine	his brother Alexander Granovsky	Madison, Wisconsin
88833	http://umedia.lib.umn.edu/node/88833	January 12, 1912	Aalto Bert	Big Falls, Minnesota	his friend Hilma Aerila	Laitila, Finland
88831	http://umedia.lib.umn.edu/node/88831	August 26, 1911	Aalto Bert	Big Falls, Minnesota	his friend Hilma Aerila, Laitila	Laitila, Finland
472442	http://umedia.lib.umn.edu/node/472442	May 16, 1936	Roerich Nicholas	Kullu, India	George and Tatiana Grebenstchikoff	Southbury, Connecticut

Figure 5: TEI compliant metainformation

40	+	<ct:correspdesc xalns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspdesc" n="710673">
41	+	corresp="http://umedia.lib.umn.edu/node/710673">
42	+	<ct:sender>Budzka, Eduard</ct:sender>
43	+	<ct:address>Klara and Vaclau Panucevich</ct:address>
44	+	<ct:placeSender>Wentorf, Germany</ct:placeSender>
45	+	<ct:placeAddressee>Chicago, Illinois</ct:placeAddressee>
46	+	<ct:datesender when="1950-06-06"/>
47	+	</ct:correspdesc>
48	+	<ct:correspdesc xalns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspdesc" n="88790">
49	+	corresp="http://umedia.lib.umn.edu/node/88790">
50	+	<ct:sender>Paikens, Helena</ct:sender>
51	+	<ct:address>her mother-in-law, Anna Paikens</ct:address>
52	+	<ct:placeSender>Minneapolis, Minnesota</ct:placeSender>
53	+	<ct:placeAddressee>Lencini, Latvia</ct:placeAddressee>
54	+	<ct:datesender when="1957-09-10"/>
55	+	</ct:correspdesc>
56	+	</ct:correspdesc>
57	+	</ct:correspdesc>

Through standardising this metainformation and making it TEI compliant, it was then possible for Niall O'Leary (freelance IT consultant) to create a range of visualisations, exploring aspects of migration such as the movement of migrants over time, or letter writing networks. Examples and full details of the visualisations that were created for this project can be found on Niall O'Leary's website.¹² The visualisations use a variety of open-source libraries, including Leaflet, Exhibit, JQuery, D3, Dimple and Peplemovin,¹³ but have been chosen and tailored specifically for spatial, temporal and personal attributes.

Even with minimal metadata it is possible to create meaningful visualisations which allow the user to notice patterns within the data (i.e. what is there) and, just as important, gaps (i.e. what is not there). This paper focuses on metainformation relating to participants (the sender and the addressee – their names and locations) and the date of the letter. This metainformation can be stored in XML documents, spreadsheets or databases, and (provided it is well structured, with established fields relating to sender, addressee, location and date) it is relatively easy, from a technical perspective, to access and retrieve that information. Additionally, the metainformation relating to lots of letters can be stored and retrieved at once. The key question, then, is how to “read” this metainformation.

Obviously it is possible to read the metainformation relating to each letter one at a time (see Figure 6). One might even be able to identify trends through close reading of the data in this format. Ultimately, though, very large datasets may prove hard to grasp in their entirety, and spotting trends can be virtually impossible. Even a relatively small dataset of, for example, 5,000 letters, can be difficult to analyse in this way.

This is where data visualisation comes in. By processing the metainformation into a graphical form it is possible to navigate the content more easily, spot trends, concentrate on particular stories within the whole, and bring together content from several different sources. The map shown in Figure 7 (developed using the open source map library, Leaflet), for example, gives a general overview of the distribution of correspondence from both the IED and DIL collections. The blue dots represent destinations while the red dots represent the origins of the letters. Very quickly it is possible to see a cluster of red dots around Northern Ireland – something that one might expect given the nature of the data being used (most of the metainformation here comes from the IED). This is an important point when looking at any visualisation: the visualisation can only reflect the dataset being analysed, including its biases. In other words, the picture is only as holistic as the collection is inclusive. In Figure 7, for instance, there is little by way of correspondence with Asia. This is not necessarily because there were no migrants from Asia to the USA; rather, there are no letters from Asia to the USA within our particular dataset.

12 <http://development.nialloleary.ie/correspondence/correspondence.php>

13 Leaflet (<http://leafletjs.com/>), Exhibit (<http://www.simile-widgets.org/exhibit3/>), JQuery (<http://jquery.com/>), D3 (<http://d3js.org/>), Dimple (<http://dimplejs.org/>) and Peplemovin (<http://peplemov.in/>).

Figure 6: TEI compliant metadata from the DIL project, University of Minnesota

```
<ct:correspDesc xmlns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspDesc"
  n="88829"
  corresp="http://umedia.lib.umn.edu/node/88829">
  <ct:sender>Aalto, August</ct:sender>
  <ct:addressee>his friend Hilma Aerila</ct:addressee>
  <ct:placeSender>Humboldt, California</ct:placeSender>
  <ct:placeAddressee>Laitila, Finland</ct:placeAddressee>
  <ct:dateSender when="1906-04-27"/>
</ct:correspDesc>
<ct:correspDesc xmlns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspDesc"
  n="552683"
  corresp="http://umedia.lib.umn.edu/node/552683">
  <ct:sender>Petrís, Antonietta</ct:sender>
  <ct:addressee>Loris Palma</ct:addressee>
  <ct:placeSender>Montreal, Canada</ct:placeSender>
  <ct:placeAddressee>Venice, Italy</ct:placeAddressee>
  <ct:dateSender when="1948-10-05"/>
</ct:correspDesc>
<ct:correspDesc xmlns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspDesc"
  n="88812"
  corresp="http://umedia.lib.umn.edu/node/88812">
  <ct:sender>Nepriysky-Hranovsky, Serhii</ct:sender>
  <ct:addressee>his brother Alexander Granovsky</ct:addressee>
  <ct:placeSender>Berezhtsi, Ukraine</ct:placeSender>
  <ct:placeAddressee>Saint Paul, Minnesota</ct:placeAddressee>
  <ct:dateSender when="1931-03-01"/>
</ct:correspDesc>
<ct:correspDesc xmlns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspDesc"
  n="552680"
  corresp="http://umedia.lib.umn.edu/node/552680">
  <ct:sender>Palma, Loris</ct:sender>
  <ct:addressee>Antonietta Petrís</ct:addressee>
  <ct:placeSender>Venice, Italy</ct:placeSender>
  <ct:placeAddressee>Montreal, Canada</ct:placeAddressee>
  <ct:dateSender when="1948-09-23"/>
</ct:correspDesc>
<ct:correspDesc xmlns:ct="http://wiki.tei-c.org/index.php/SIG:Correspondence/task-force-correspDesc"
  n="710637"
  corresp="http://umedia.lib.umn.edu/node/710637">
  <ct:sender>Malvine R.</ct:sender>
  <ct:addressee>Emilie Wehle</ct:addressee>
  <ct:placeSender>Champaign, Illinois</ct:placeSender>
  <ct:placeAddressee>Vienna, Austria</ct:placeAddressee>
  <ct:dateSender when="1946-11-28"/>
</ct:correspDesc>
```

Figure 7 is informative, but it is difficult to appreciate the level of correspondence coming from individual locations. A simple pie chart, however, can make this clear. A quick glance at Figure 8 and Figure 9 shows that a majority of letters came from Ontario (6.8%) and New York (6.7%) and most letters were sent to Belfast (5.8%) and Ontario (6.1%). Note that the 12% and 30.7% segments represent “unknown” (i.e. no sender or addressee location information was available for these letters).¹⁴

To look more closely at the participants involved in the letter writing process, it is necessary to return to the mapping interface, but this time adding a level of interaction for the user that allows them to identify particular places, people and times. It is difficult to show this with a static image, but Figure 10 is an interactive map and by clicking on one of the bubbles, the user can retrieve additional information about the participants involved in the correspondence.

It is also possible to display the interaction between participants in another way, focusing more on the individuals that are involved in the correspondence (see Figure 11).

¹⁴ One cannot expect to have all metadata for all letters, but for what metadata there is it is important that data be standardised. “New York City, New York” is distinct from “New York, New York” or even just “New York”; each entry will be dealt with as a separate location. The placeography files previously mentioned might help resolve this issue.

Figure 7: Visualisation: map showing the distribution of correspondence (IED and DIL collections)

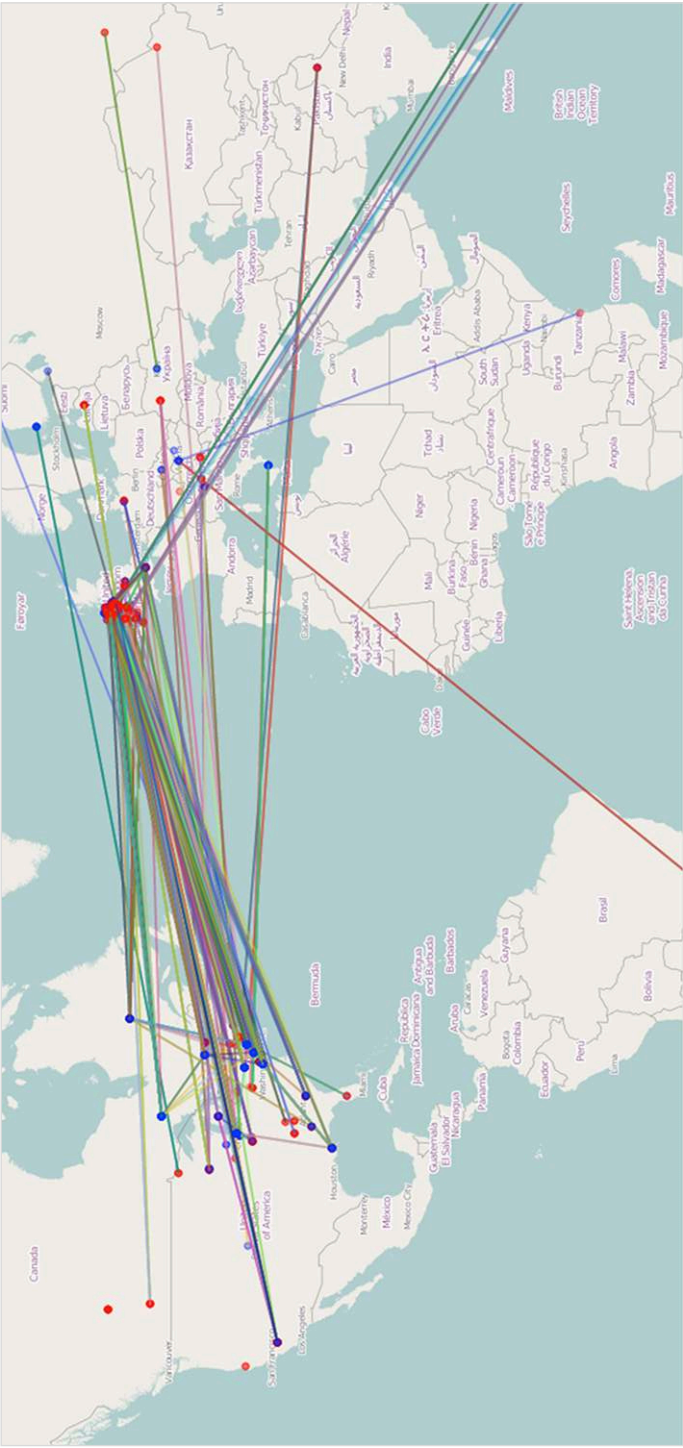


Figure 8: Visualisation: locations letters are sent from

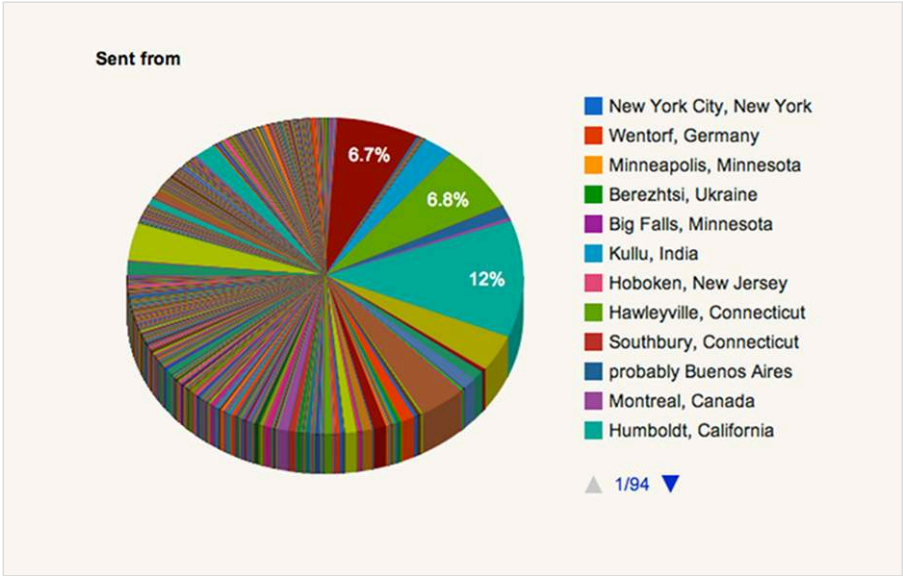


Figure 9: Visualisation: locations letters are sent to

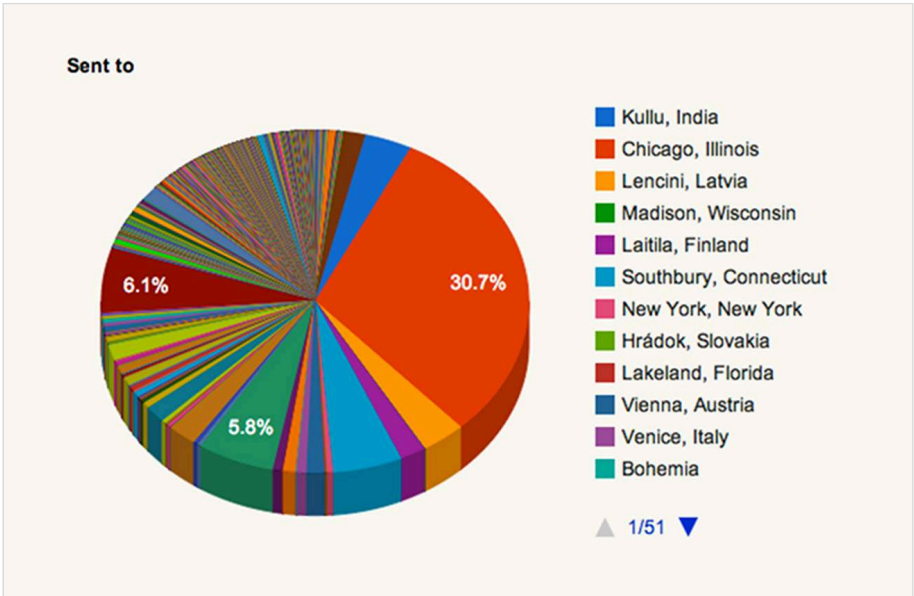


Figure 10: Interactive visualisation: participants and locations

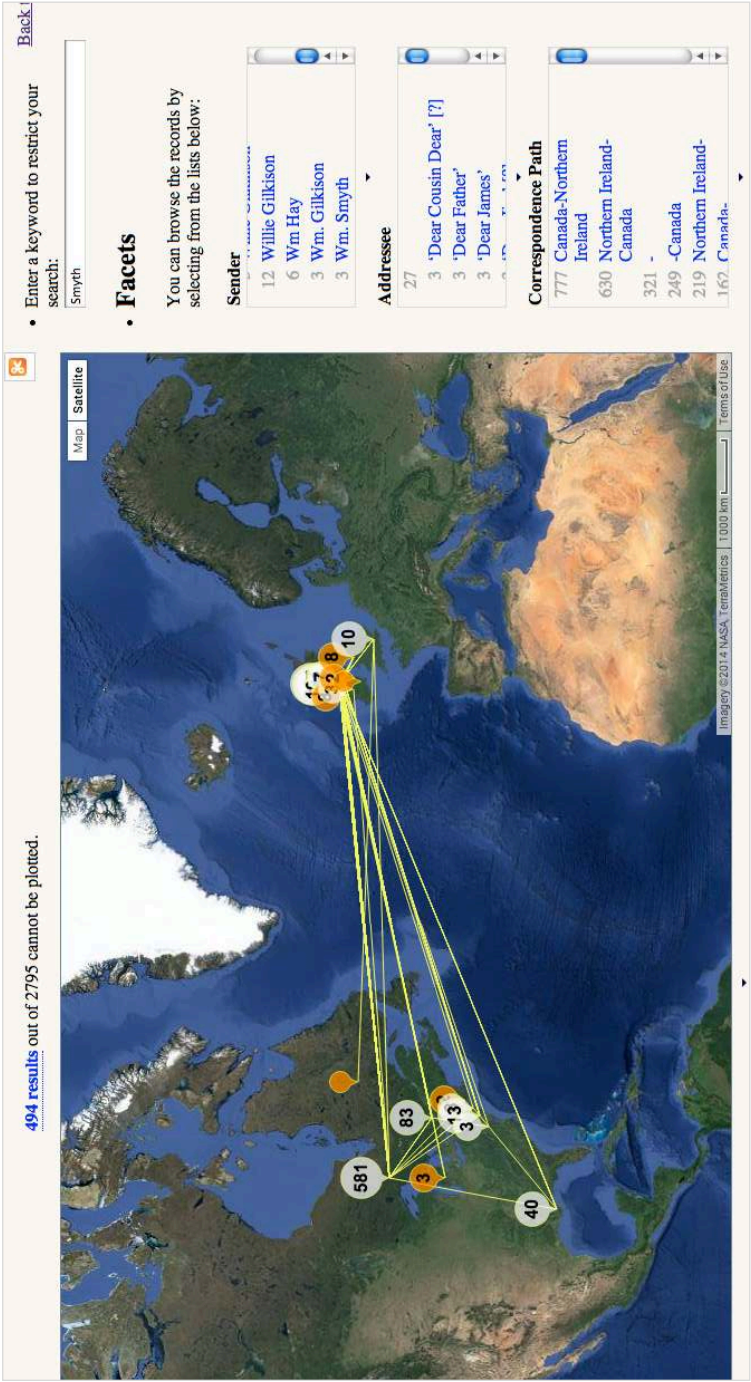


Figure 11: Interactive visualisation: participants

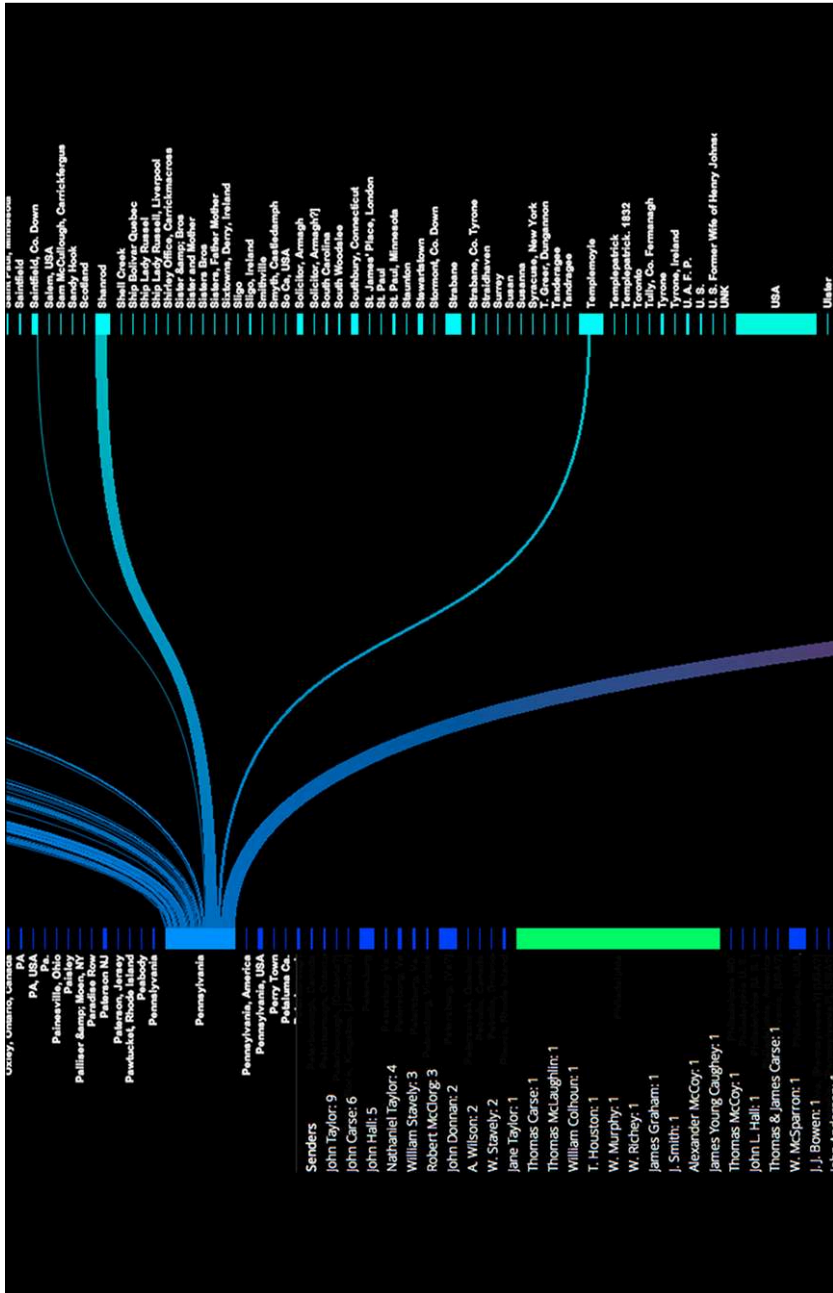


Figure 12: Interactive visualisation: letter writing networks

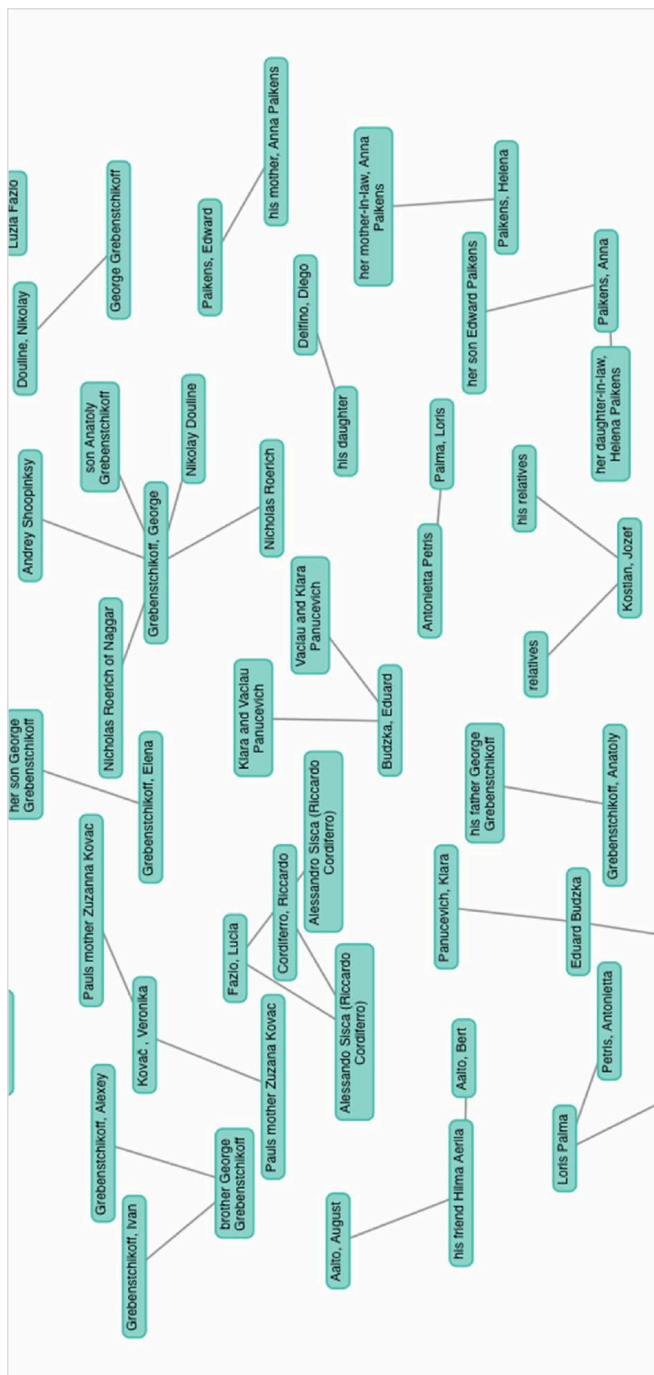


Figure 13: Visualisation: participants, locations, time

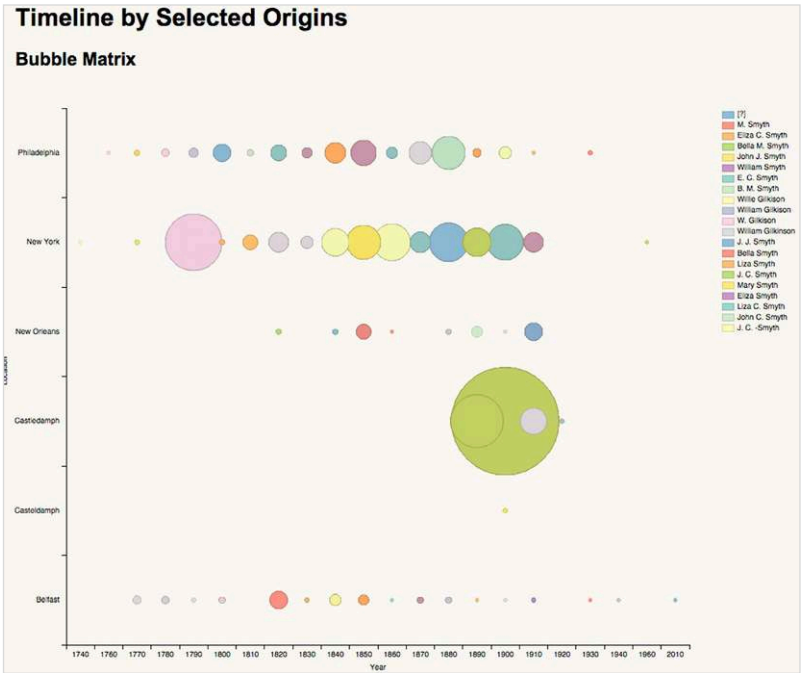
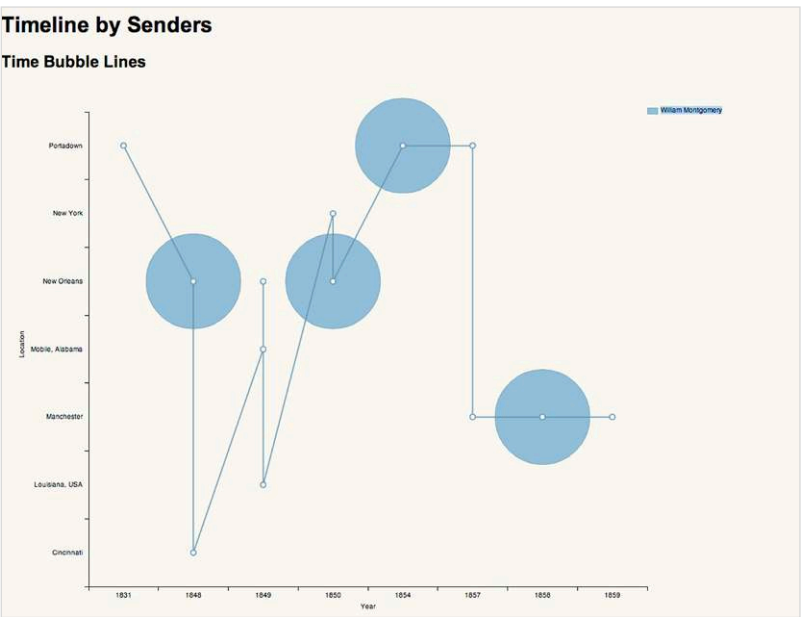


Figure 14: Visualisation: the example of William Montgomery



Narrowing down further on the participants, Figure 12 shows letter writing networks. By clicking on an individual's name, it is possible to view all persons the individual wrote to, or received letters from.

The visualisations so far have concentrated on people and places, but there is another aspect to the letters to consider – time. Again, it is possible to take an overview of each collection. This gives us an appreciation of when the main bulk of the letters were written for each dataset. However, by bringing in a faceted means of filtering the data, it is possible to interactively narrow down the analysis. This can be done in a number of different ways, by line-charts, area-charts, step-charts, bubble-matrices and bubble-lines. In Figure 13, a bubble chart is used to track a person's movements over time. The vertical axis represents location and the horizontal axis represents time. Each bubble represents a person. This is perhaps detailed more clearly in Figure 14, which shows the movements of William Montgomery (an Irish emigrant from Portadown, Ireland).

William Montgomery appears to have travelled from Portadown to New Orleans sometime between 1831 and 1848. He stayed in America until at least 1850 before returning to Portadown sometime between 1854 and 1857. He then moves to Manchester, after which time there are no more letters.

Such analysis will not provide all the answers, but it gives a firm basis for further research. Data visualisation allows the user to discover new insights into some very old data, hopefully giving it a new resonance for today. It can also be very cost-effective; all these visualisations shown in this paper use open-source (free) technologies. The scalability of such tools make it possible for multiple projects, with similar content, to be brought together.

Conclusion

What this encoding process highlighted, in particular, is the potential for working with header information (in this case information embedded within <correspDesc>) relating to person (sender and addressee), location and date, without necessarily having access to the letter itself. One of the biggest challenges that came out of the network process related to accessibility of letter collections and issues to do with intellectual property. In other words, it is often difficult to get access to collections and even more difficult to make collections freely available online – especially when working across disciplines and across cultures. By focusing on meta-information about the letter (rather than the letter itself) there were fewer barriers to overcome with regards interconnecting resources.

Once the information is digitised, computer programmes can search in and across collections to notice patterns. It becomes possible to step near and step back. Step Near, or step closer, in that we can explore our research material in more detail and explore that detail in new, different ways. We can, to use film metaphors, zoom in, pan round. The detail becomes available to a multitude of disciplines. And Step Back, in that we can observe our material from a distance, observe patterns that might not otherwise be easily visible, though they might have been postulated. Digital technologies can order, and make visual, patterns within our research material – so that, at every stage, there is the ability to Step Near or Step Back. Possible patterns can be explored in detail.

This is just the first step in terms of developing fully interoperable resources, but hopefully – with further funding – more letter collections will be able to inter-connect in this way. Increasingly funding bodies do look for “interdisciplinarity”, but it is not always clear quite what it is that is being looked for, and there are few formal academic careers in interdisciplinary studies (Griffin *et al.*, 2006). Nonetheless, the role of the interdisciplinary interlocutor is worth developing and there are some very simple and obvious rewards to interdisciplinary dialogue: just from the first workshop, for instance, it became apparent that different scholars were working on the same letter collection. For some disciplines a letter is simply a piece of text, trapped on paper and yearning to breathe free. Other disciplines are interested in the actual piece of paper, and argue – quite reasonably – that the physical object is of interest and may hold information that will, in fact, clarify the text and its history. Through examining the emigrant letter within a digital humanities framework it was possible for different disciplines to come together to discuss some of the problems and opportunities of working with correspondence collections, and to discuss best practice within and across disciplines – alliances were formed and the groundwork for future collaborative projects is starting to take place.

References

- Arts and Humanities Research Council (AHRC)** (2012) *Digital Transformations in the Arts and Humanities*, [online] last checked on 01/09/2014. URL: <http://www.ahrc.ac.uk/Funding-Opportunities/Research-funding/Themes/Digital-Transformations/Pages/Digital-Transformations.aspx>
- Daybell James and Hinds Peter** (2010) *Material Readings of Early Modern Culture: Texts and Social Practices, 1580-1730*, Basingstoke, Palgrave Macmillan, 277 p.
- Dossena Marina** (2007) “As This Leaves Me at Present” – Formulaic Usage, Politeness and Social Proximity in Nineteenth-Century Scottish Emigrants’ Letters, in Stephan Elspaß, Nils Langer and Wim Vandebussche Eds., *Germanic Histories from Below (1700–2000)*, Berlin, De Gruyter, pp. 1-30.
- Elspaß Stephan** (2002) Standard German in the 19th-century? (Counter-) Evidence from the Private Correspondence of “Ordinary People”, in Andrew R. Linn and Nicola McLelland Eds., *Standardization: Studies from the Germanic Languages*, Amsterdam, Benjamins, pp. 43-65.
- Erickson Charlotte** (1972) *Invisible Immigrants: The Adaptation of English and Scottish Immigrants in Nineteenth-Century America*, London, Weidenfeld and Nicolson, The London School of Economics and Political Science, 531 p.
- Fitzpatrick David** (1994) *Oceans of Consolation: Personal Accounts of Irish Migration to Australia*, Cork, Cork University Press, 649 p.
- Griffin Gabriele, Medhurst Pam and Green Trish** (2006) *Interdisciplinarity in Interdisciplinary Research Programmes in the UK*, Hull, University of Hull, 80 p.
- Kamphoefner Walter D., Helbich Wolfgang and Sommer Ulrike** (1988) *News from the Land of Freedom: German Immigrants Write Home*, Ithaca and London, Cornell University Press, 645 p.

McLelland Nicola (2007) "Doch mein Mann möchte doch mal wissen..." A Discourse Analysis of 19th-century Emigrant Men and Women's Private Correspondence, in Stephan Elspaß, Nils Langer, Joachim Scharloth and Wim Vandenbussche Eds., *Germanic Language Histories from Below (1700-2000)*, Berlin, De Gruyter, pp. 45-68.

Miller Kerby A. (1985) *Emigrants and Exiles: Ireland and the Irish Exodus to North America, 1815-1920*, Oxford, Oxford University Press, 684 p.

Plummer Kenneth (1983) *Documents of Life: An introduction to the problems and literature of a humanistic method*, London, Allen & Unwin, 175 p.

Plummer Kenneth (2001) *Documents of Life 2: An Invitation to a Critical Humanism*, London, Sage Publications, 306 p.

Schrier Arnold (1958) *Ireland and the American Emigration, 1850-1900*, Minneapolis, University of Minnesota Press, 210 p.

Stadler Peter (2013) *Introduction to TEI. Presented at the Utrecht Workshop "Digitising experiences of migration"*, 29-30 May, Utrecht University, [online]. URL: <http://prezi.com/d2vahzuektx/introduction-to-tei/>

Stanley Liz (2004) The epistolarium: On theorizing letters and correspondences, *Auto/Biography*, 12, pp. 201-235.

Steen Sara J. (1994) Manuscript Matters: Reading the Letters of Lady Arbella Stuart, *South Central Review*, 11, pp. 24-38.

TEI Consortium (Ed.) (2014) *The TEI Header*, [online] last checked on 01/09/2014. URL: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>

TEI Consortium (Ed.) (2008) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Version 1.1.0., 4 July.

TEI (2014) *SIG: Correspondence*, [online] last checked on 01/09/2014. URL: <http://wiki.tei-c.org/index.php/SIG:Correspondence#Introduction>

Thomas William I. and Florian Znaniecki (1958) *The Polish Peasant in Europe and America*, New York, Dover Publications [original edition, 1918-1920], 2250 p.

Webb Eugene J., Campbell Donald T., Schwartz Richard D. and Sechrest Lee (1966) *Unobtrusive methods: Nonreactive research in the social sciences*, London, Sage Publications, 240 p.

Emma Moreton, Niall O'Leary and Patrick O'Sullivan

✦ **Visualising the Emigrant Letter**

Emigrant letters are a rich resource for teaching and learning, transcending disciplinary and methodological boundaries. They are expressive and indicative of correspondents' identities, values, preoccupations and beliefs, providing a powerful source of information about migration issues and shedding light on processes of language change and variation. Although many emigrant letter collections have now been digitised, not all are properly archived; some are reduplicated and others are in danger of being lost. The documentation and preservation of such letters is, therefore, a particularly pressing need. In 2013, an AHRC research network was established to look at ways of improving interconnectivity between digital collections of migrant correspondence. This paper reports on work carried out so far, focusing on how emigrant letter projects might move beyond the digitisation stage to exploit text content and enhance usability and searchability through the use of visualisation tools.

✦ **Propositions pour une visualisation graphique de correspondances de migrants**

Les lettres de migrants constituent un matériau riche pour l'enseignement et l'apprentissage en ce qu'elles permettent et exigent le dépassement des frontières disciplinaires et méthodologiques. Elles nous fournissent des éléments d'information sur l'identité des correspondants, sur leurs valeurs, leurs préoccupations et leurs croyances, nous renseignent sur les problématiques associées à la migration et nous éclairent sur les variations et changements que peut connaître la langue. Bien que de nombreuses collections de lettres de migrants aient été numérisées, toutes ne sont pas à ce jour correctement archivées ; certaines sont simplement redupliquées et d'autres courent le risque d'être perdues. La conservation et l'étude documentaire de ces lettres constituent véritablement deux urgences. En 2013, un réseau de recherche AHRC a été mis en place afin d'évaluer les actions qu'il est nécessaire d'apporter pour améliorer l'interconnexion entre les banques de données numériques consacrées aux correspondances des migrants. Cet article dresse un état des lieux des travaux réalisés et se concentrera à la fois sur les prolongements de ces projets vers une exploitation approfondie du contenu des textes et sur les bénéfices apportés par le développement d'outils avancés de visualisation graphique.

✦ **Propuestas para una visualización gráfica de las correspondencias de emigrantes**

Las cartas de emigrantes constituyen un material rico para la enseñanza y el aprendizaje en tanto que permiten y exigen la superación de fronteras tanto disciplinarias como metodológicas. Nos abastecen de elementos de información sobre la identidad de los correspondientes, sobre sus valores, sus preocupaciones y sus creencias, nos informan sobre las problemáticas asociadas con la migración y nos alumbran sobre las variaciones y los cambios que pueden darse en la lengua. Aunque las numerosas colecciones de cartas de emigrantes han sido digitalizadas, no todas están en la actualidad correctamente archivadas; algunas están simplemente reproducidas y otras corren el riesgo de perderse. La conservación y el estudio documental de estas cartas constituyen verdaderamente dos urgencias. En 2013 se puso en marcha una red de investigación AHRC con el fin de evaluar las acciones que es necesario aportar para mejorar la interconexión entre los bancos de datos numéricos dedicados a las correspondencias de los emigrantes. Este artículo presenta un estado de la cuestión sobre los trabajos realizados y se concentrará al mismo tiempo en los prolongamientos de estos proyectos hacia una explotación profunda del contenido de los textos y también en los beneficios aportados por el desarrollo de herramientas avanzadas de visualización gráfica.