

# Databases for Quantitative History

Luke Kirwan

University College Cork, Ireland

luke.kirwan@gmail.com

## Abstract

This paper will propose that, rather than sitting on silos of data, historians that utilise quantitative methods should endeavour to make their data accessible through databases, and treat this as a new form off bibliographic entry. Of course in many instances historical data does not lend itself easily to the creation of such data sets. With this in mind some of the issues regarding normalising raw historical data will be looked at with reference to current work on nineteenth century Irish trade. These issues encompass (but are not limited to) measurement systems, geographic locations, and potential problems that may arise in attempting to unify disaggregated sources. It will discuss the need for a concerted effort by historians to define what is required from digital resources for them to be considered accurate, and to what extent the normalisation requirements for database systems may conflict with the desire for accuracy. Many of the issues that the historian may encounter engaging with databases will be common to all historians, and there would be merit in having defined standards for referencing items, such as people, places, locations, and measurements.

## 1 Introduction

In a 2007 article Steve Bailey noted that regarding the archival profession, “Where we do let ourselves down is in our failure to reassess our methodologies and to reapply them to a new and rapidly changing world”(Bailey 2007, p. 123). This applies not only to archives, but also to historians. Our methodologies and standards for assessment are predominantly derived from the nineteenth and Twentieth centuries. The digital age has profoundly impacted every facet of human existence and it could revolutionise the historical method in terms of how we research, analyse and disseminate our subjects. As such we must reassess our methods and tools and develop new methodologies and guidelines for their implementation.

This paper will examine two key components in using databases to further historical research, appropriate standards and methodological constraints and data accessibility. For databases to truly drive forward historical research, digital humanists have an obligation to demonstrate to the wider historical community how they can benefit from sharing the raw data they have collated over years of research through the use of databases. For this to happen we need an accepted set of historical standards and methodological approaches in database creation. A well-structured, well-designed database can benefit the four major components of scholarship identified by Ernest Boyers: discovery, integration, application, and teaching (Pearce, Weller, Scanlon, Kinsley 2012, pp. 35–36).

## 2 Use of databases in History

In terms of Irish historical research, databases remain underdeveloped. The History Data Service maintains a database of Irish Historical Statistics that covers a broad swath of nineteenth and twentieth century social and economic life in Ireland (Clarkson, L.A., et al 1997). However, as this database is no longer live on the internet it must be downloaded from the UK Data Archive and painstakingly reconstituted from separate sections, no small feat as the documentation on the original structure is rather sparse. Two further databases, the Historical National Accounting Group's (HNAG) *Database of Irish Historical Statistics* and *Duanaire* contain further quantitative data for Irish economic history (Begley, Geary, O'Rourke [no date]; *Duanaire: A Treasury of Data for Irish Economic History* [no date]). They are invaluable sources for Irish historians, but provide a macro view of Irish society; no database produced so far has mined the available material for regional Irish statistical data. Furthermore, they are currently works in progress. *Duanaire* is producing material on the state of eighteenth century Irish trade and the HNAG data is quite strong for labour statistics for twentieth century Ireland. The database that I am currently working on will be a micro view of regional trade in Cork city during the early nineteenth century, complementing the earlier work produced by the aforementioned projects.

Databases are not the key to solving all riddles in the humanities. As noted by the *Historical Data Service* "A database is not a historical source" (History Data Service [no date]). A database may comprise a multitude of elements from historical sources, but a researcher must still, at some point, enter an archive and engage with primary source material. What it can do is grant access to representations of disparate sources to a wider community, and that is the true value for databases in historical research. From here it can be integrated with other database representations of historical sources and become a very powerful research tool. The methodologies we currently use to evaluate and reference records were designed for analogue media. Professionals must develop new methodologies for evaluating the representation of historical scholarship in the digital ether.

For example, when referencing and creating visualisations from my own data I try to keep the figures as close to the original figure, without rounding up or down. As this often entails converting from imperial to metric I have already intervened and imposed an 'interpretation' on the primary source data. Some researchers choose to round the figures up or down, which means, when it comes to comparing the conclusions of different historiographical interpretations, there is already an inherent divergence, even if a minute one, between the data. If the raw data is provided one can remove the historian's intervention and probe the data and the analytical method in more depth and with greater perspective. It also provides the opportunity to impose one's own methodology on another's interpretation to glean new insights. The data can be further tested to see if it holds up in the face of rigorous examination. After all, this is the cornerstone of the scientific method, and one to which all researchers, no matter the discipline, should be accountable. Of course there are aspects of humanities research that are subjective, especially in the interpretations and perceptions of events, but any quantitative research should be replicable to ensure the foundations on which subjective interpretations are built are correct.

Not only should we be willing and active in providing our raw data to others, we need to encourage others in the academy to join us. The hoarding of raw data, produced from primary sources, is akin to an archive publishing a catalogue of inaccessible material. It helps no one and frustrates many. Historians and humanists generally do not often work collaboratively. It is not part of a historians training and not a skill that is developed throughout many historians education. Even an edited book sees a broad selection of expert historians writing their sections independently of one another before the final results are compiled. This is not collaboration. As collaborative work is not typically part of the historians training, so how could we expect seasoned academics to share data they laboriously compiled in isolation over many years? There are other reasons why data may not be shared, there might be further work that the creator wants to produce from the data or possibly there are doubts as to whether they would

receive appropriate credit for the work involved in compiling the data. The former is an understandable reticence that very few would argue with. The fear of not being appropriately credited has certain merits, but this fear is a more general academic concern. It is plagiarism to not appropriately attribute work to its creator. The use of facilities such as the History Data Service at the University of Essex could potentially be a way to balance this fear, as access to the data involves agreeing to certain fair use policies (*UK Data Service-Terms and conditions of access* [no date]). The threat of plagiarism is one that always will exist in academia, but it is something that the majority of academics do not engage in, and the same would be true for providing digital data. If it is clearly known that this work has to be acknowledged it will be by the majority.

Two interventions may help: The first is being dealt with through the digital humanities community; demonstrating the power and benefit of large-scale research collaboration. I'm thinking in particular of Stanford's Republic of Letters project, where they accept the criticisms from other academics and then absorb them into renewing and rejuvenating the project (Stanford [no date]). Sharing Ancient Wisdom is another, where multiple projects across the classics community are aggregated to create a more useful resource and one with more depth and breadth than any of the individual projects could ever have achieved (HERA [no date]). We need to actively demonstrate how our digital work can be integrated into previous scholarship, and how the potential for incorporating data from historians that are reticent regarding the 'digital' can be of benefit to them.

The second intervention is ensuring that all of this data is compiled and structured in a manner that respects and engages with good historical practices. We have various accepted historical methodologies, but where are our standards for the application of digital technology? In the past number of years the emergence of new technologies has allowed researchers to explore traditional historical questions through the prism of vast datasets and visualisations, helping to increase our understanding of the past. The increase in accessibility and the sharing of such tools and data has also led to the emergence of new questions and concepts about the past. Digital history is both the application of technology to traditional historical inquiries, as well as the creation of entirely new perspectives of the past. The application of digital humanities to the study of history is considered a new methodological approach, but it remains rather ad-hoc. How should a database be structured? How should 'nouns' be encoded? How do we tie ancient, sometimes no longer extant regions, into a modern GIS system? If our data is to be useful and collaborative we need standardised ground rules that are agreed upon, even in a quite contentious discipline, and through which we can share and amalgamate our data. This is not a new issue: In a 1988 article Manfred Thaller, discussing the objectives of the Max Planck Institute, stated that it aimed to create different projects that agree on a basic design model (Thaller 1988, p. 135). This is quite an interesting article regarding the potential for digital history, as some of what he predicted now exists. The historical workstation he proposed is essentially Wikipedia. Why though, is there not yet a model, a methodology, for digital history? Is this because Wikipedia originated outside the academy, is open source, and in some ways is a "denigration of expertise" (Rosenzweig 2006, p. 122), whereas a model for 'digital history' would require academic historians to take control of its direction in a collaborative manner?

### 3 Standards/Methodology issues

As the core of historical methodologies derives from the nineteenth century, perhaps for new inspiration we should look towards other areas of emerging historical research. Oral history is, in many respects, quite a different type of qualitative research from more 'traditional' historical studies. By its very nature the primary source material has inherent flaws and intrinsic biases, much like any other source materials. Due to the fundamental differences in source compilation, oral history associations provide numerous guides to best practice and interview techniques (Oral History Association [no date]; Oral History Society [no date]). This helps mitigate against too much overt interference in the collection process by the historian. The Oral History Association also collects essays on their

website about oral history in the digital era and maintains a wiki page that allows practitioners to contribute and learn from one another (Oral History Association [no date]). This is one area in which oral historians appear to be leading from the front, collaboration and dissemination within their professional network. If “digital history” is to become more closely aligned with the “traditional” historians this is where we must focus. The skill sets and tools used should be made as transparent as possible in order to foster collaboration and engagement and to demonstrate the benefits that derive from moving data freely. Christine Borgman acknowledges this, stating that “Until analytical tools and services are more sophisticated, robust, transparent, and easy to use for the motivated humanities researcher, it will be difficult to attract a broad base of interest within the humanities community”(Borgman 2009, p. paragraph 5).

So what types of standards are required, at a basic level, to construct a quality historical database?

1. Guidelines for what constitutes appropriate documentation that tracks all stages of the process and error normalisation or omissions.
2. Standards for codification of ‘nouns’: for example, people or places that can be cross-referenced to other systems.
3. Normalisation: to what level should/can data be normalised and how does this impact upon the veracity of the database as compared to the source?
4. How and where the conversion of terminology, individuals, locations, and measurements to modern representations should take place or is appropriate.
5. Referencing to original sources and the interpolation of multiple sources.
6. Which keys should be used to link disparate sources together? If we can agree on the second point that would contribute to this point.

One potential solution to these issues would be to try and adapt modern codification conventions to historical data. My own work involves collating data from the CUST 15 series of imports and exports in the British National Archives. This collection covers details to item level the type, quantity, and value of goods arriving and departing from Irish ports from 1698-1829. There are a number of other projects working from the same, or related, source material. Two that are very closely aligned to my own work are the Duanaire project (*Duanaire: A Treasury of Data for Irish Economic History* [no date]) which aims to collect and disseminate Eighteenth century Irish economic data, and the Trading Consequences project that explores commodity trading in the nineteenth century British Empire (*Trading Consequences* [no date]). These are two very different projects but they both rely on commodity listings from the CUST series, CUST 15 and 5 respectively. Since commodity trading is central to both, there is a requirement for a primary goods table, with all the necessary keys and data that entails. Although temporally these projects deal with different worlds, the broad swath of commodities being traded remains broadly similar. This is based on a comparison of the data being used in Duanaire for the eighteenth century with the data for trading consequences taken from 1862, and my data which covers the first two decades of the nineteenth century, in the middle of both. There are of course variations, but the major commodities are relatively static. If the purpose of creating such resources is to allow them to be collated into a larger resource, which I believe it should be, then projects like these are ideal for such integration. They are relying on a similar source, have broadly similar objectives, and they utilise the same list of ‘nouns’ for database primary keys.

Unfortunately they take similar, but different approaches in how they code the primary keys. An example would be that in the Trading Consequences commodities list there is a single entry for Salt, with the code ‘salt’. In my own commodities list there are five separate entries, varying from Salt Barrel (‘saltbarr’), to Salt Foreign Bushel (‘saltforbus’). These differences are predominantly due to the different manners through which each project collected commodity data, but it serves to highlight how codification for a relatively basic commodity can have a

wide variance due to personal preferences and decisions. Personally I like to include measurement data with the Primary Key code as it helps me track changing measurements for produce. This is more important for my project than the Trading Consequences project. However, Trading Consequences uses a lot of R and frequency analysis, so the commodity list is a base lexicon derived from DBpedia that includes preferred naming conventions and alternative names for a specified product. Being able to integrate this base lexicon into my own data, or indeed any project concerned with nineteenth century trade, would be invaluable. Unfortunately this would involve either using the Trading Consequences keys from the outset, or fundamentally altering existing databases, with all the problems that entails. What is required is a common reference point, from which any commodity based database can be linked through.

One solution to this, that I think has a very strong merit, is to use the Standard International Trade Classification (SITC) (United Nations Statistics Division - Classifications Registry [no date] I must thank Damian Malone from the Irish Central Statistics Office for bringing my attention to the benefits of using SITC for this purpose).(United Nations Statistics Division - Classifications Registry [no date] I must thank Damian Malone from the Irish Central Statistics Office for bringing my attention to the benefits of using SITC for this purpose) There are numerous benefits to using this system. SITC relies on a hierarchical codification system. Nineteenth century commodities do not necessarily always have a contemporary counterpart. By using SITC though you can classify items to a level that at least represents the *purpose* of the product. To return to the different codification for salt, the SITC reference would be 598.99, which covers salt for curing. This would mean that the five different primary keys I use for different types of preservation salt would all be quickly related to the same single entry from Trading Consequences. Of course this is an example of a product that is still in modern usage and has an appropriate classification. The flexibility of this system comes in when there are products that don't have an easily identifiable equivalent. Madeira Wine is an import that has some significance in Irish trade. Unfortunately the nineteenth century ledgers do not record the type of fermentation used or grapes used. If it was a wine of fresh grapes SITC would classify it as 112.17, whereas if the grape was in fermentation it would be 112.11. As the type cannot be identified that specifically this system allows for the use of the code for one level up, 112.1, wine of grapes. This ensures that the item has a relevant code, and one that covers all wines. Not only does this allow for better integration of datasets, it also allows for a far more refined querying system, so the user can search for wines (112.1) or for all alcoholic drinks (112). Furthermore, SITC maintains a one to one relationship with Eurostats Combined Nomenclature (CN) classification system, so it is possible to move readily between the two systems with a conversion table. In a similar vein the RICardo World Trade History project uses the Correlates of War datasets to normalise locations and correlate historical locations with their contemporary equivalents, as well as to track their position in the world at the time, whether they were an empire or a dependency (RICardo [no date], *Correlates of War* [no date]). This provides for a partial solution to the issues of regional classifications.

## 4 Accessibility

Certain accessibility and interoperability issues arise when it comes to converting historical measurement systems into something usable in a database. This problem swells when one considers regional changes, linguistic changes, changing naming conventions and so on. One prime example of this would be the wide variety of measurement systems being used in Britain during the eighteenth and nineteenth centuries, both before and after standardisation in 1825 (For an excellent example of this see Velkar 2008). Kathryn Tomasek has suggested that the harmonised system(World Customs Organisation [no date]) could be used for commodities, but identifies a flaw in that it does not take into account historical commodities (Tomasek 2013). We cannot overtly rely on solutions coming from outside the field of history. Digital humanists, and more specifically digital historians, must share their own data, and solutions they have developed for handling historical data, without restriction. Without doing this we

run the risk of constantly re-inventing the wheel and recreating each other's work over and over again. A defined set of standards enhances interoperability and the potential to build upon previously compiled datasets, increasing both what can be achieved from limited resources as well as the scope of the dataset. We need to agree on acceptable practices in standardisation and normalisation. The 'American Council of Learned Societies' report on online collaboration brings this to the fore when they note that "lone scholars... are working in relative isolation, building their own content and tools, struggling with their own intellectual property issues, and creating their own archiving solutions", later adding that this results in "unnecessary redundancy of effort" (American Council of Learned Societies 2006, p. 21;36).

A couple issues that should be addressed regarding data standardisation and normalisation are:

1. The levels of intervention that is accepted good practice.
2. The importance of remaining sympathetic to the original source and media, as much as possible or practicable.
3. Appropriate documentation, potentially including the raw, un-normalised data.
4. Appropriate conversion methodologies.

"Openness matters for the digital humanities for reasons of interoperability, discovery, usability and reusability ...in addition to requirements for new skills, more open practices may lead to increased possibilities for interdisciplinary working"(Scanlon 2012, pp. 180–181). Most people's engagement with databases is mediated through a user interface, so they rarely see the nuts and bolts of how they work. Open resources, and openness in their construction methods could also encourage 'traditional' historians to engage more readily with those involved in digital humanities research. Historians hold a vast amount of data in personal 'silos'. For some, the perspective holds that to do their own research, each individual must collect this data for themselves. This is true, to a limited extent, but it overlooks the benefits of sharing research to further our historical understanding. Sharing our data provides future researchers with the time, and the space, to develop new source materials and new perspectives.

In a comment on a section of *Fielding History* that discusses database design William Caraher notes that "Transparency in database design is as important to understanding databases as a historical publication as transparency is in traditional historical work...In other words, can you drill down through your database to your source in an efficient way?" (Dougherty 2011, p. paragraph 15). Such transparency in process is required to 'sell' our digital work. This mainly requires two core items, documentation and referencing. Unfortunately digital humanists appear so engaged with the humanities side of their work that it appears documenting the digital aspect is an afterthought. This is more apparent with some of the older databases that are available, but we also need to consider how we justify our own data normalisation and processes. One of the best ways to do this for both ourselves and for future researchers is to write up the process in a manner aimed towards non-specialists, simple, clear, and concise. Then we can see if our justifications are adequate.

## 5 Referencing

For databases and the data contained within to be useful beyond their original purpose we need to ensure that our data can be tested against the original source material. This requires an appropriate referencing system. Digital historians do not 'create' new information; we create a new curation process and interpretation, which requires an appropriate referencing system. The organisation and classification of data is not traditionally considered part of the historical process, but the digital humanities are forcing us into areas previously the domain of archivists and records managers. A crucial difference is that, according to traditional archival scholarship and practice materials from

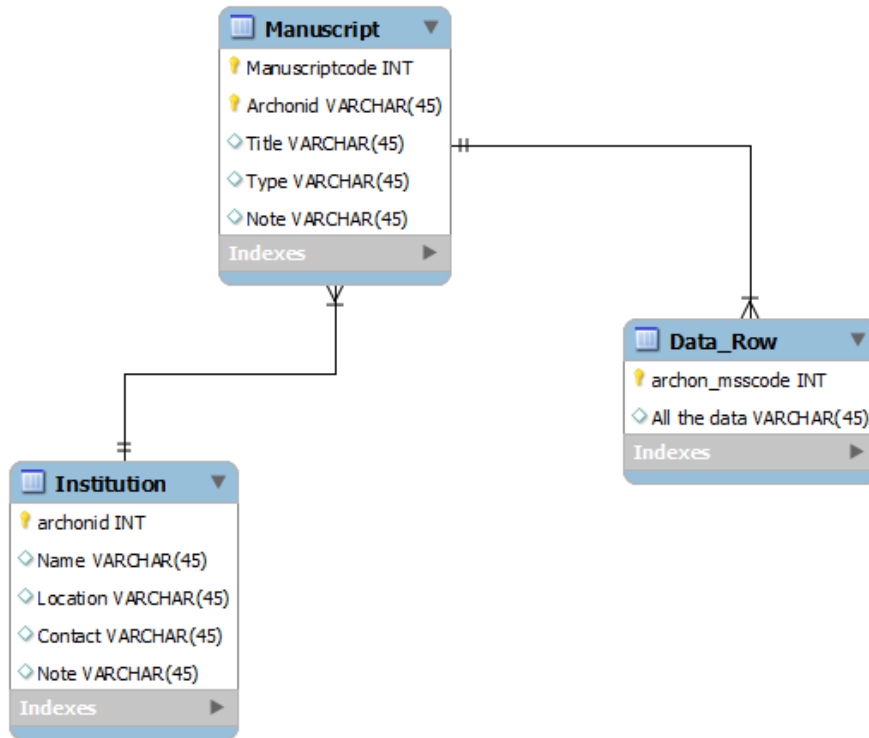
different organisations or collecting bodies should not be organised together, as this would breach their archival provenance, and affect their 'recordness'. In this new curation method we must ensure that we can extricate the intermingled data in a manner that enables easily tracing it back to its source; we need an agreed upon referencing system.

This issue is not unique to the field of digital humanities/digital history. Toby Green of the Organization for Economic Cooperation and Development has written extensively regarding the problem of referencing of their data, not just by academics and researchers, but by libraries as well (Green 2009, 2012). His solution is to use DOI identifiers for the data. This may work adequately, or even quite well, for data held by the OECD, which is relatively consistent and already in a digital format, but for historical data it does not suit. Of course his main problem is referencing of the datasets themselves, rather than the raw data they contain, so his call for a bibliographic dataset will not solve all of the referencing issues for historical datasets, but perhaps some of them. Similar issues occur in the sciences, where it can be difficult to get the data without rerunning experiments. A six-point citation system has been proposed, including DOI and Universal Numeric Fingerprint (UNF), to reference data (Altman, King 2007, pp. 1–4). I propose an alternative that is specific to databases and far simpler in its implementation.

The problem with using these systems for historical data lies in the manner in which historical data has been compiled. Archival materials, especially those still in paper or card catalogues, do not have persistent digital identifiers, for self-evident reasons. It is entirely in the realm of possibility for individual researchers to assign persistent identifiers to this material, but other researchers may not consistently adopt these identifiers and almost assuredly the archival institutions would not do so.

One possible solution, not permanent but flexible enough to deal with the variety of archival materials a researcher may consult, is to create reference tables. Though not innovative, it could be used effectively in a number of ways. The reference table will contain all the pertinent data, institution, collection, dates, covering notes—all the pre-requisite material for a solid annotated archival reference, and a primary key for the relation. What I propose is that, rather than the more complicated system of triplet references that one gets with RDF, a two-part primary key is used.

Archives, as part of the context statement for their collections, use an institutional repository code. This code identifies each individual archive, through a system like 'archon' (The National Archives [no date]). This is a persistent identifier for the repository and part of the ISAD (G) guidelines (International Council on Archives 2000). I combine this with the manuscript or collection reference number to get a code like 624/371, which references the National Library of Ireland, *Exports and Imports ledger for 1813-1814*.



**Figure 1 Sample version for permanent archival DOI**

This system has several advantages. The concept behind the code is very basic and therefore easy to grasp, involving no detailed explanations for people of any level of computer literacy. Even in the unlikely event that an archive changed manuscript codes, old and new could be easily cross-referenced. Furthermore, in the event that the manuscript reference could not be located the repository reference will be the same, so the holding institution is always known. To support this data the primary key references a table that contains all the traditional elements needed for an archival reference, including details of any changes or alterations to the data. As repository data is reasonably persistent the entries in this table only need to be created once. The code is simple enough to be manipulated to fit into any other systems as required or repurposed. If any issues arise regarding reproduction rights the SQL can be adapted to remove open access on the basis of either the repository (identified in the first three digits) or the item itself (the final digits).

When creating a large, single-source dataset the incentive for such a referencing system may not be evident. But when a researcher is trying to create a dataset of disaggregated data, and one that can be adjusted and fitted into similar datasets, a simple, easy to use system is crucial to provide for some level of historical, archival, and source accountability. Without such a system the data runs the risk of becoming unmanageable, unusable, and unaccountable. It also can be retroactively applied to other datasets with relative ease.

## 6 Conclusion

“Research practices and associated needs have evolved in sometimes subtle but significant ways, requiring parallel adjustments for those supporting history research”(Rutner, Schonfeld 2012, p. 15). By sharing our data in an



open and accessible manner we can expand our networks and resources. Access to materials previously unthought-of, or unknown, could immensely benefit researchers outside the original collation process. The burden for ensuring that this process is accountable and historically viable lies with the digital humanities community involved in historical studies. One of the first steps should be to assume responsibility for the creation of appropriate methodological processes. If this is done there is the potential to encourage our less technically-inclined colleagues to embrace these new systems and allow the importation of their data. We need accepted standardisation guidelines, and more importantly, we need to demonstrate the value of digital history through its practice.

## References

- ARCHIVES, International Council on, 2000, *ISAD(G): General International Standard Archival Description* [online]. Ottawa : International Council on Archives. [Accessed 30 November 2013]. Available from: [http://www.icacds.org.uk/eng/ISAD\(G\).pdf](http://www.icacds.org.uk/eng/ISAD(G).pdf)
- ASSOCIATION, Oral History, [no date], *OHDA Essays* [online]. [Accessed 26 November 2013 a]. Available from: <http://www.oralhistory.org/ohda-essays/>
- ASSOCIATION, Oral History, [no date], *Principles and Best Practices* [online]. [Accessed 26 November 2013 b]. Available from: <http://www.oralhistory.org/about/principles-and-practices/>
- BAILEY, Steve, 2007, Taking the Road Less Travelled By: The Future of the Archive and Records Management Profession in the Digital Age. *Journal of the Society of Archivists* [online]. 2007. Vol. 28, no. 2, p. 117–124. [Accessed 11 November 2013]. DOI 10.1080/00379810701607777. Available from: <http://www.tandfonline.com/doi/abs/10.1080/00379810701607777>
- BEGLEY, Jason, GEARY, Frank and O'ROURKE, Kevin, [no date], *HNAG Database of Irish Historical Statistics* [online]. Available from: [http://www.tcd.ie/iis/HNAG/HNAG\\_database.htm](http://www.tcd.ie/iis/HNAG/HNAG_database.htm)
- BORGMAN, Christine L., 2009, The Digital Future is Now: A Call to Action for the Humanities. [online]. 2009. Vol. 3, no. 4. [Accessed 10 November 2013]. Available from: <http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>
- BYRNE, Kate, 2008, Relational Database to RDF Translation in the Cultural Heritage Domain. *Internet, may* [online]. 2008. [Accessed 26 November 2013]. Available from: <http://homepages.inf.ed.ac.uk/kbyrne3/docs/rdb2rdfForCH.pdf>
- CLARKSON, L.A. and ET AL, 1997, *Database of Irish Historical Statistics*. Colchester, Essex : UK Data Archive.
- COLEMAN, Nicole, [no date], *Mapping the Republic of Letters* [online]. [Accessed 2 December 2013]. Available from: <http://republicofletters.stanford.edu/>
- Correlates of War*, [no date]. [online], [Accessed 5 October 2014]. Available from: <http://www.correlatesofwar.org/>
- DOUGHERTY, Jack, 2011, *Fielding History (Bauer) Fall 2011* [online]. [Accessed 10 November 2013]. Available from: <http://writinghistory.trincoll.edu/data/fielding-history-bauer/>
- Duanaire: A Treasury of Data for Irish Economic History*, [no date]. [online], [Accessed 5 October 2014]. Available from: <http://duanaire.ie/>

- GREEN, Toby, 2009, We need publishing standards for datasets and data tables. *Learned Publishing*. September 2009. Vol. 22, no. 4, p. 325–327. DOI 10.1087/20090411.
- GREEN, Toby, 2012, *Publishing Data Alongside Analysis, Books and Journals* [online]. OECD Publishing White Paper. [Accessed 14 November 2013]. Available from: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1072&context=charleston>
- HERA, [no date], *Sharing Ancient Wisdoms* [online]. [Accessed 2 December 2013]. Available from: <http://www.ancientwisdoms.ac.uk/>
- NETWORK, EHPS, [no date], *European Historical Population Samples Network* [online]. [Accessed 26 November 2013]. Available from: <http://www.ehps-net.eu/content/about>
- ORGANISATION, World Customs, [no date], *Harmonized System Database Online* [online]. [Accessed 3 December 2013]. Available from: <http://www.wcoomd.org/en/topics/nomenclature/instrument-and-tools/hs-online.aspx>
- PEARCE, Nick, WELLER, Martin, SCANLON, Eileen and KINSLEY, Sam, 2012, Digital Scholarship Considered: How New Technologies Could Transform Academic Work. *in education* [online]. December 2012. Vol. 16, no. 1. [Accessed 10 November 2013]. Available from: <http://www.ineducation.ca/index.php/ineducation/article/view/44>
- RICardo*, [no date]. [online]. [Accessed 5 October 2014]. Available from: [http://graduateinstitute.ch/home/research/projets/historical\\_imagination/research-leads/ricardo.html](http://graduateinstitute.ch/home/research/projets/historical_imagination/research-leads/ricardo.html)
- ROSENZWEIG, Roy, 2006, Can History Be Open Source? Wikipedia and the Future of the Past. *The Journal of American History* [online]. June 2006. Vol. 93, no. 1, p. 117–146. [Accessed 14 November 2013]. DOI 10.2307/4486062. Available from: <http://www.jstor.org/stable/4486062>
- RUTNER, Jennifer and SCHONFELD, Roger, 2012, *Supporting the Changing Research Practices of Historians* [online]. [Accessed 5 November 2013]. Available from: [www.sr.ithaka.org/download/file/fid/693](http://www.sr.ithaka.org/download/file/fid/693)
- SERVICE, History Data, [no date], From Source to Database. In : *Digitising History* [online]. Available from: [http://hds.essex.ac.uk/g2gp/digitising\\_history/sect33.asp](http://hds.essex.ac.uk/g2gp/digitising_history/sect33.asp)
- SOCIETIES, American Council of Learned, 2006, Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences [online]. American Council of Learned Societies. Available from: [http://www.acls.org/uploadedFiles/Publications/Programs/Our\\_Cultural\\_Commonwealth.pdf](http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf)
- SOCIETY, Oral History, [no date], *Oral History - Practical Advice* [online]. [Accessed 26 November 2013]. Available from: <http://www.ohs.org.uk/practical-advice.php#plan>
- THALLER, Manfred, 1988, Data Bases v. Critical Editions. *Historical Social Research / Historische Sozialforschung* [online]. January 1988. Vol. 13, no. 3 (47), p. 129–139. [Accessed 13 November 2013]. Available from: <http://www.jstor.org/stable/20754321>
- TOMASEK, Kathryn, 2013, *Encoding Financial Records* [online]. [Accessed 17 November 2013]. Available from: <http://journalofdigitalhumanities.org/2-2/encoding-financial-records-by-kathryn-tomasek/>

*Trading Consequences*, [no date]. [online], [Accessed 5 October 2014]. Available from:  
<http://tradingconsequences.blogs.edina.ac.uk/>

*UK Data Service-Terms and conditions of access*, [no date]. [online], [Accessed 2 March 2014]. Available from:  
<http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx>

W, Carus A. and OGILVIE, S., 2005, 0512: *Turning Qualitative into Quantitative Evidence: A Well-Used Method Made Explicit* [online]. Cambridge Working Papers in Economics. Faculty of Economics, University of Cambridge. [Accessed 19 November 2013]. Available from: <http://ideas.repec.org/p/cam/camdae/0512.html>

## Biography of the author

**Luke Kirwan** is a Digital Arts and Humanities PhD candidate in University College Cork. He is examining the impact international trade had on the development of Cork city during the early nineteenth century. From this research he will be creating an open access database of trade figures for the region.