

e-research resources, tools and methods for historic place name analysis

JISC – funded Workshop Report Institute for Name Studies, University of Nottingham

3 September 2010

Executive Summary

A substantial amount of the 'critical mass' of digital data available to scholarship contains place-names, and it is now recognised that spatial and temporal data points, including place-names, are an vital part of the e-research infrastructure that supports the uses, re-use and advanced analysis of data using ICT tools and methods. Place-names can also be linked semantically to contribute to the web of data, and to enrich existing content through linking existing data, and identifying new collections for digitization to strategically enhance existing digital collections. However, existing e-projects rely on modern gazetteers limiting them to the modern and the near-contemporary. This workshop explored how to further integrate the wealth of historical place-name scholarship, and the resulting digital resources generated within UK academia, so enabling integration of local knowledge over much longer periods.

- **Presentations** came from (a) projects led by historical researchers, generally constructing new gazetteers via archival research, and (b) e-research and e-infrastructure projects, generally working with existing digital gazetteers. The meeting explored alternative approaches but did not identify an overall consensus.
- **Existing digital gazetteers** derive mainly from topographic mapping. Size is expressed solely as numbers of entries, quality as positional accuracy, and the extent and detail of “feature typing”: the names are labels for features on the map, and feature type information records the kind of feature: a lake or a building (or, more specifically, a church or a power station). The accuracy of the names themselves is not addressed. The lack of variant names, and especially the earlier forms of names, makes the identification of geographical names within historical texts unreliable.
- **Historical gazetteers constructed by place-name researchers** contain many variant names per entry, so their size is not accurately measured by numbers of entries. Quality comes from detailed attribution of names, i.e. the identification of the historical sources in which they appear, so dates appear systematically being the dates at which those documents were created. Feature typing is not a major concern, with the exception of the GBH GIS’s administrative unit gazetteer which records the legal status of parishes, districts etc.

- **Data models** employed for both types of gazetteer are usually simple but for different reasons: large digital gazetteers often contain little more than a name, a coordinate and a feature type per entry, so a simple “flat” structure is inevitable; digital gazetteers constructed by place-name researchers hold richer information, but tend to be modelled on earlier paper publications, with extensive free text. The GBH GIS is again an exception which enables easier quantification: it identifies only c. 18,000 UK “places”, but these are linked to c. 150,000 attributed names drawn mainly from statistical reports, 19th century gazetteers and earlier travel writers.
- **Coverage:** Gazetteers from the Ordnance Survey and GeoNames provide national coverage. All major new historical place-name projects are constructing digital resources, but earlier publications of the Survey of English Place Names need to be computerised (the CHALICE project makes a start on this, but copyright issues are un-resolved). Only Somerset has yet to be researched, but others are still in progress. More work is needed in Scotland, Wales and Ireland. Even with complete UK coverage, different file structures would need to be integrated.
- **Access:** Historical place-name databases are increasingly accessible on-line, but only as human-searchable web sites. JISC-funded e-resources based on existing digital gazetteers are accessible via web services and increasingly as linked data, and are consequently machine-accessible. Much of the informal discussion at the meeting focused on how to make genuinely authoritative resources more systematically accessible: AHRC funding has supported scholarly research but not standardisation or sophisticated access mechanism; JISC has funded such mechanisms but has tended to rely on off-the-shelf data from non-academic sources.

e-research resources, tools and methods for historic place name analysis

Report

Paul S Ell, Lorna Hughes, Humphrey R Southall

Hosted by Jayne Carroll, EPNS; Paul S. Ell, Centre for Data Digitisation and Analysis, Queen's University Belfast; and Lorna Hughes, Centre for e-Research, King's College, London

The Arts and Humanities now have access to a vast array of electronic research resources funded by, for example, the Arts and Humanities Research Council's (AHRC) Resource Enhancement Scheme and numerous Joint Information Systems Committee (JISC) grant calls. Whilst the AHRC Resource Enhancement Programme has now come to an end still around 50 per cent of AHRC-funded research results in some form of e-resource. Whilst many resources exist, they are largely underused and we have yet to see a step-change in scholarship promised by the digitisation of key research assets. One explanation for this is that almost all of the e-resources created form isolated data silos with, typically, individual bespoke websites for each resource with its own interface and idiosyncrasies. The result is that resource discovery is difficult – there is no master list of all e-resources available – and scholars are required to interact with a unique interface for each resource to extract information.

Grid Computing, particularly the Data Grid, and e-Science, offer the potential to make information discovery, harvesting and analysis far easier than is currently the case. AHRC have suggested

'Digital resources in the A&H have grown at an astonishing rate in the last ten or twenty years; out of over £100m spent by the AHRC since 1999 on research project awards, half has been given to projects with some kind of digital output. The problem is that researchers do not yet have the technology to make the fullest use of these resources, because they are generally not connected together. e-Science provides a set of solutions for this problem, and for the related development of facilities for research collaboration using the Internet.

e-Science thus stands for a specific set of advanced technologies for Internet resource-sharing and collaboration: so-called grid technologies, and technologies integrated with them, for instance for authentication, data-mining and visualisation. This has allowed more powerful and innovative research designs in many areas of scientific research, and is capable of transforming the A&H as well.ⁱ

One of the vital elements in linking disparate resources in the Arts and Humanities, and to a large extent the Social Sciences, is geographical location. All Arts and Humanities data has some spatial presence whether expressed precisely as an administrative unit for which boundaries can be plotted,

or less exactly through a place-name. It is, therefore, possible, to relate data by location. However, there are many challenges, the most significant being the lack of comprehensive electronic gazetteers which have sufficiently detailed spatial granularity and take account of changes in place-name spelling over time. If such gazetteers were available, and were structured in such a way as to facilitate links between resources, the reuse of existing electronic data would be enhanced and cross-disciplinary work promoted.

The focus of this Workshop was to examine the development of historical place-name gazetteers to interlink disparate electronic resources using sources such as the English Place-Name Society volumes together with developing similar resources for areas outside England within the British Isles. The workshop aimed to bring together key stakeholders – holders of relevant gazetteer information (in other words place-name scholars); Digital Humanists who would benefit from better integration of e-resources; e-resource aggregators – linking materials drawn for different sources by location, and e-resource providers including digitisation units, data archives and data disseminators.

The workshop objectives focussed around the following themes:

- Identifying the key stakeholders, and their requirements for digital gazetteers
- Reviewing existing digital gazetteer projects and resources, and assessing their suitability for historical use
- Reviewing current UK research into historical place-names, and assessing how the results can best be integrated into developing e-infrastructure
- Assessing alternative data structures for holding digital gazetteer data
- Developing a collaborative framework for comprehensive funding to develop electronic gazetteers

Attendees

Jean Anderson – Humanities Advanced Technology and Information Institute, University of Glasgow

Jayne Carroll – School of English Studies and Institute for Name Studies, University of Nottingham

Paul Carvill – School of English Studies and Institute for Name Studies, University of Nottingham

Richard Deswarte – History Data Service, University of Essex

Stuart Dunn – Centre for e-Research, King's College London

Alastair Dunning – e-Content Programme, JISC

Claire Grover – School of Informatics, University of Edinburgh

Lorna Hughes – Centre for e-Research, King's College London

Leif Isaksen – School of Electronics and Computer Science, University of Southampton

Jeremy Morely – Centre for Geospatial Science, University of Nottingham

Kay Muhr – Queen's University, Belfast

David Parsons – Centre for Advanced Celtic Research, University of Wales

Humphrey Southall – Department of Geography, University of Portsmouth

Jonathan Tedds – Department of Physics and Astronomy, University of Leicester

Jo Walsh – EDINA, University of Edinburgh

Apologies

Paul S Ell – Centre for Data Digitisation and Analysis, Queen’s University, Belfast

Ian N Gregory – Department of History, University of Lancaster

Keith Lilley – School of Geography, Archaeology and Palaeoecology, Queen’s University, Belfast

Programme

10:00 Coffee and arrival

10:30 Lorna Hughes, Jane Carroll: Welcome and introduction

10:40 Lorna Hughes: e-Research for historic place names: strategic developments

10:50 The state of the art in historic place name research

1. Jayne Carroll and Paul Carvill – An overview of EPNS and the Key
2. Kay Muhr - Place-Names in Northern Ireland, current state-of-the Art
3. Jean Anderson - Place name research and resources in Scotland
4. David Parsons - Welsh Place names initiatives

12:00 Humphrey Southall - Data Models for Historical Gazetteers: Administrative units, locations, ‘places’ and place-names

12:25 Discussion

1:00 Lunch

1:45 e-Research and historic place-name research: recent initiatives

Alastair Dunning – JISC and e-infrastructure for e-resource Integration

Jo Walsh and Clair Grover – the JISC-funded CHALICE Project: linked data

Leif Isaksen, Southampton – HESTIA: Place name extraction from text

2:45 Concluding remarks: Alastair Dunning, and discussion

3:30 Tea and departure

Introduction

Lorna Hughes (Kings College, London)

The introduction set the aims of the workshop in the wider context of e-research, including the use of ICT tools, methods and infrastructure. In particular, there was an overview of the use of digital resources through the implementation of existing ICT tools and methods through projects based at the Centre for e-Research, King's College, London. CeRch has developed a research programme around the creation and use of digital resources. A key interest is complex data, especially from the arts and humanities. CeRch is also investigating ways of managing digital content through data and metadata systems. It is involved in several national and international initiatives, including DARIAH and arts-humanities.net. Hughes highlighted several projects based around digital collections where there is potential to scope current practice and developments in e-research for place name research, including the 'Stormont Papers', developed in collaboration with the Centre for Data Digitisation and Analysis at Queen's Belfast, which has already investigated the use of GeoNames through the JISC funded 'Embedding GeoCrossWalk' project, which is extracting place-name data from existing digital content by embedding geo-referencing; and the SAILS (Sailors Archives and Logbooks of Ships) project which is structuring and linking archival records including ship's logs.

The CeRch projects highlight a need for building sustainable digital collections, with an increased emphasis on using, enhancing and re-using existing digital resources, especially linking content across collections. New digitisation initiatives should focus on projects with widest interdisciplinary interest, accessing resources from different disciplinary areas, emphasis on how digital content can be modelled and represented for collaboration and interoperability between existing resources. These projects highlight the fundamental step change for accessing different sources, especially through deep linking to key source materials.

Existing historical gazetteer work

The first part of Workshop was devoted to assessing the current availability of historical gazetteers for the constituent countries of the United Kingdom whether in analogue or digital form. From discussions over a number of years it was clear that developments in England were far ahead of any other work in the United Kingdom.

The Institute for English Name Studies and the English Place-Names Survey (EPNS) Jayne Carroll (University of Nottingham)

Jayne reported on the work of the Institute for Name Studies at the University of Nottingham. Nottingham has long been a focus for place-work, recognised informally in the 1960s and through the establishment of the Centre for English Name Studies. In 1992 the Institute was formed which hosts the English Place-Names Survey. However, EPNS represents an element of the work of the English Place-Names Society which is an independent learned society.

The Institute has a significant research profile. Its main interests are in tracing place-names back to their earliest form, and to the languages in which they were coined. Most were first used in the Anglo-Saxon period; but others are in Old Norse etc. Historical documents are very important in terms of sourcing place-names. Projects include the Survey of English Place-Names (EPNS) and the Vocabulary of English Place-Names, which is essentially a dictionary of place-name elements. Funding from the Leverhulme Trust has also been made available for, for example, Landscapes of Governance, which focuses on the vocabulary and archaeology of Anglo Saxon assembly sites, and

British Academy grants have also been key to sustaining EPNS work. The Institute is active in the teaching activities at Nottingham in place-names and person-names and also provides PhD supervision in these areas.

The English Place-Names Survey **Paul Cavill (University of Nottingham)**

Paul Cavill specifically described the work of the English Place-Names Survey. The Survey was established in 1923 and seeks to provide coverage, by county, of the whole of England with several volumes devoted to larger counties. Currently there are 85 volumes in print with more in preparation. Every county is either complete or is being worked on with the exception of Somerset.

Within the overall county structure the volumes are broken down into sub-county units including Wapentakes, Hundreds, Rapes, and Lathes. The common reporting unit is the parish with names below this broken down into sub-parish units where appropriate (townships, chapelries, hamlets etc). Field names are also recorded for later volumes. There are different levels of geo-referencing for different divisions, and there are no explicit references to OS maps.

As might be expected for such a long-running project the volumes vary over time. For Bedfordshire, for example, which is an early volume, all the references to place-names are taken from major national administrative resources such as the Domesday Book. Organisation within each parish is by date and significant differences in the spellings and their interpretation are noted. Later volumes, such as Shropshire, provide more information such as the formation of the parish, and information drawn from cartographic sources. There is far more reliance on local resources, i.e. publication of the Shropshire archaeological society, which enhance the entry. Contemporary spellings often appear first in local sources. A distinction is made between manuscript sources and printed sources and also a distinction between personal names and topographic names. Field names are included.

Place-Name work in Wales **David Parsons (University of Wales Aberystwyth)**

David previously directed the Institute for Name Studies before taking on a new post at the University of Wales Aberystwyth, examining the development of Welsh place-names. In introduction he commented on how the data rich resources of EPNS might be developed, following long discussions with the Centre for Data Digitisation and Analysis at Queen's University Belfast, and that the Institute for Name Studies was key to any work on historical English place-names. He emphasised that the University of Nottingham could not simply agree to digitise the volumes, as they are owned by the English Place-Name Society, a learned society with a Council of academics whose formal agreement would need to be sought. While they have not yet seen the value of digitisation, they did agree to an earlier Leverhulme-funded project aimed at developing a complete vocabulary of English place-names. During this project the complexity of digitising all place-names within EPNS was appreciated. The work made available in digital form included the first place-name variant recorded but not later variations, something that will be vital in the construction of a comprehensive online gazetteer. Even so a complex database was created which has been made available as an online clickable map of English parishes (The Key to English Place Names). This

initiative also demonstrated the opportunities for any project that digitised the EPNS place-name data to link the content with related, analytical information.

David's current work relates to Wales where he is developing the Welsh place-name studies programme at the University of Wales Celtic Studies Research Centre, the Canolfan. There is nothing like the EPNS in Wales. Existing place-name research is mainly focused on local efforts, resulting in uneven publications over the years. Some are very detailed/scholarly, lots more independent amateur work. Coverage is very patchy, and the dual system of administration makes place-name work especially complex for Wales. The leading academic in the field was Professor Melville Richards of the University of Bangor who started systematic research into Welsh place-names starting with the administrative units of Wales. This in fact represents an imposed administrative system by England over a native Welsh one, which complicates matters. Melville Richards developed a list of spellings, but these were not connected to a gazetteer and there is no way of searching it in a structured way. David intends to build on the work begun by Melville Richards on a dictionary of Welsh place-names, developing a database from this linking the material to a Welsh administrative gazetteer. There is a need to take it forward at a 'British' level – negotiate how English Place Name researchers can link up with Welsh equivalents to provide seamless data interrogations.

[It emerged during the meeting that the GB Historical GIS project had already constructed a database from Richards' book, and substantially extended it. This Welsh administrative gazetteer has now been copied to David.]

Northern Ireland Place-Name Research **Kay Muhr (Queen's University Belfast)**

Kay Muhr led work on Northern Ireland place-names which started in 1985. There are Ireland-specific problems of community and language relating to NI place-names. In the late 1970s the traditional townlands, which cover the whole of Ireland as around 60,000 small spatial units, 16,000 of which are in Northern Ireland, were demoted from the government's administrative system and replaced with road names. Whilst not very visible to the outsider most natives would identify themselves as being from a specific townland, and there was concern at their apparent abandonment. This led to funding from the N. Ireland government for research into the origins of place-names in Northern Ireland. There are some particular issues relating to Ulster place-names associated with the historical development of the area. Gaelic place names were recorded by an English-language administration, which changed throughout the complex historical development of the region. These developments are reflected in the way that place-names are recorded in archaeological, religious and literary sources.

So far more than 30,000 place-names have been gathered for Northern Ireland with minor names being taken from 1:50,000 and 6 inch to the miles Ordnance Survey maps. These names have been placed within a database with a clear structure, to a degree following the analogue format of the EPNS volumes. This structure is place-name > historical forms > bibliography > elements.

Of the place-names gathered eighty per cent are in Gaelic, but they are recorded in English language systems and English language spellings. The names reflect the natural history, landscape archaeology, cultural history, customs and traditions, church history and literature of Ireland.

More recently Ordnance Survey have become more interested in place-names from a cultural perspective and have a project to research the nature and meaning of place-names on their maps. This is part of promoting the townland back to its former place. In Ireland, Kay emphasised that typically work has been carried out by researchers with an interest in the origin and meaning of a place-name and that there has been less concern with variation of place-name spelling over time. There were also difficulties in developing a historical gazetteer as spatial entities could, and did, have the same name. It was problematic, in addition, to work out what the entities specifically were. Crucially, the way that the project has developed has seen digitisation as a process, which has meant that the data must be updated over time. Kay stated it was important that the database structure did not dominate the data gathering – it has to be reflexive and responsive. The data should be kept as ‘raw’, or open, as possible. Comments fields should be used as little as possible, but they are necessary, as this is where the researcher keeps their growing knowledge.

Kay recommended the following websites to Workshop participants: www.ulsterplacenames.org, www.ulsterplacenamesni.org and www.logainm.ie.

Scottish Place-Name Research

Jean Anderson (University of Glasgow)

As in Northern Ireland, much of the imperative behind place-name research has been directed at understanding the origin of place-names and interpreting their meanings. In particular, Jean referred the group to the work of Thomas Clanchy and Carol Duff at Glasgow University.

Research at Glasgow into place-names has centred around interest in place-names and their boundaries with other spatial units and the semantics of these boundaries. As PI of the JISC-funded ENROLLER project, Jean has gathered large datasets at Glasgow University that include place-names and have been useful to scholars with interests in this area, and Jean has an interest in bringing datasets together to enable searching across collections, possibly using ENROLLER as a platform. The Scottish Place name Society has a Scottish place-name database and researchers at Glasgow have an interest on building on it to encompass more historical variations in names in collaboration with users and potential users.

Data Models for Historical Gazetteers: Administrative units, locations, "places" and place-names

Humphrey Southall (University of Portsmouth)

Humphrey began with some general observations:

- Mainstream GIS software, e.g. ArcGIS, is unsuited to the representation of historical sources because it is built on a data model in which location is the most certain attribute of entities, so entities are defined by location, names existing only as "labels".
- Historical GIS is distinguished not by a time dimension (even utility companies must record change), but by the uncertainty of our knowledge; so we must document not just what we

know but how we know it. Historical gazetteers should systematically include dates, but these dates will appear as parts of references to specific historical documents; and they must be able to hold information about places with unknown locations (e.g. Avalon).

- Most existing digital gazetteers are based on too simple a data model. Gazetteers from mapping agencies (OS, NGIA, etc.) have a simple model because the content is too limited: an ID, just one name, a coordinate and sometimes a feature type; no attribution data, because the agencies have not maintained a systematic record of when, where and from whom names were obtained. Academic place-name databases are far richer, but tend to similarly flat structures closely based on paper publications, storing too much as unstructured text.
- Size is not everything: gazetteers with many variant names per "entry", and systematic attribution data, may be more useful to more people than those with the most entries.

He then described the current structure of the GBHGIS, and how it addresses these issues. It is now neither British nor a GIS: the Postgres-based system was developed with EU funding and is European in scope, also providing the data model for a global system being planned jointly with Harvard and other US universities; and the model is fundamentally an ontology with optional GIS features. It includes two linked listings, of 79,176 administrative units (AUs) and 17,786 "places"; but they share a single separate listing of 150,007 geographical names. Explicit relationships between AUs are mandatory and total 250,032; locations are optional, but can include dated sequences of polygons. Every unit, name, relationship and polygon is required to be linked to an *authority*, defined in a central table based on Dublin Core; and the system allows for both an *immediate* authority, such as Frederick Youngs' *Guide to the Local Administrative Units of England* (Royal Historical Society, 1979 and 1991) and, if known, an *ultimate authority*, meaning a treaty or statutory instrument. "Places" have a one-to-many relationship with AUs, but some places have no linked AUs. Places inherit all the names of linked AUs plus additional names e.g. from travel writers. Web pages for "places" include place-name pages, each name being hyperlinked to the original source.

Discussion

The morning was completed by a general discussion reflecting the vision of developing a spatio-temporal gazetteer and the current work taking place in the UK and Republic of Ireland.

It was noted that:

- EDINA hold contemporary and "near contemporary" gazetteer data.
- The GBHGIS is primarily historical, but includes current local authorities and NUTS 1-3.
- Software to automate harvesting place-names from historical maps would be extremely useful, but this is technically extremely difficult because of background clutter and great variation in orientation, alignment and fonts. A project linking Harvard's Center for Geographic Analysis with András Kornai of the Budapest Institute of Technology hopes to investigate this. Paul Ell noted that he was part of the ISTHMUS FP7 EU grant bid to draw disparate spatially-referenced data for urban areas in the EU together and that he has a Workpackage directed at harvesting names from historical maps. EPNS of course include historical maps in their range of sources for place-name spelling.

- One way forward could be to digitise place names from historic maps using crowd-sourcing. This would give a good gazetteer from historical sources. There would be a need to cross-coordinate the maps – different data/projections systems etc, but it should be fairly easy to relate them back to WGS84. This reflected Ell's FP7 WP.
- Leif Isakson noted that the Pleiades project is working with the Barrington Atlas using URIs.
- It was very important that any system of geo-ontologies should publish stable and persistent URIs. Thus any system would need a good sustainability model.
- Resource developers dealing with individual collections have, for historical reasons, despaired at the non-standard mappings from different sources. There is a need to record how different units, entities etc are linked together.
- It was noted that there was a significant difference between the quality of published gazetteers which are rich and academically rigorous versus open gazetteers such as GeoNames. There is a need for quality assured/rigorous ones such as EPNS.
- There is a significant gulf between the place-name requirements of academic researchers and off-the-shelf GIS 'solutions'.
- The computerisation of traditional scholarly reference works is highly desirable but may be hard to automate. Youngs' *Local Administrative Units* was converted into the GBHGIS ontology manually, by closely supervised clerical assistants, as its structure was both very complex and never quite systematic enough for OCR followed by machine parsing without an unacceptable loss of knowledge.
- It was asked how big is the gap between what place-names scholars need and what existing digital gazetteers provide. One issue is how far a gazetteer should be seen as a fixed body of information versus something to be steadily extended. The latter has obvious merits, but may inhibit use as an authority. Should some elements, such as coordinates, be seen as fixed? Kay Muhr noted that there are names that have moved in Northern Ireland. Further, spatial footprints of cities in particular change over time.
- One issue is how to represent "places" geographically, by points or by polygons. AUs can be given firm polygons, but place are vaguer. Higher-level places, like "East Midlands" are particularly problematic.
- There is a need to decide how uncertainty is represented, although this can be obvious if there is a big enough collection of AUs.
- If there is a hierarchal structure, you can apply metrologies of places within places: e.g. 'the communities of the Black Sea' – you know what this means, but it is an aggregation of different communities. In contemporary mapping you have the same issues of a tree hierarchy not working because of ambiguities.

e-Research and historic place-name research: recent initiatives

JISC and e-infrastructure for e-resource Integration

Alastair Dunning (JISC)

Alastair related these discussions to JISC activities, especially through the e-Content programme. JISC was interested in how place-name work could be exploited to provide broader services. Most JISC digitisation project contained spatially referenced data but this information was not used to its full potential. Linking information by place could aid resource discovery and research. Examples of

key projects included the British Library Archival sound recordings, GeoDigRef and the Unlock service. A key question is: how can information about historic place-names be put together with data from other sources? Geo-crosswalk, for example, is unsuitable for historic data: it is a data architecture project, based on GIS, and is not using ontologies.

It would be good if Unlock expanded to cope with historical information – there is lots of online information concerned with historic places. How can Unlock text access this? There are many challenges such as licensing, large amounts of content and different research questions. However, place-name research could help resolve some issues. More importantly, how can historic gazetteers ‘unlock’ digital content created without geo-tagging? And who should fund these gazetteers: they are the core ‘building-block’ of much research.

Alastair argued that EDNA and UKDA have implied in the past that modern place-names can be used for interrogating historical sources, but they cannot.

The JISC-funded CHALICE Project: linked data Jo Walsh and Claire Grover (University of Edinburgh)

Jo.

- CHALICE came out of the Embedding Geo-X-Walk workshop in January. There is not good historical coverage in GeoNames etc, which was problematic. CHALICE wanted to use text mining to produce a gazetteer that would improve historic text mining, in a virtuous circle; and to develop open source tools to do data mining from EPNS.
- 4 Partners: CDDA – providing and digitising the material through its partnership with the Institute for Name Studies; LTG – doing the text mining; EDINA as project managers; CeRch developing use cases for impact and value.
- Linked data: based on URIs, and will link to GeoNames on the ‘things as a point’ principle, using URIs as names for things.
- When someone looks up a URI, they can provide more information. Need to be able to exchange, annotate etc.
- E.g. Sameas.org – find out how many things on the Internet ‘are’ Nottingham.
- Not sure yet where it would be most appropriate for the namespaces to be hosted.

Claire:

- Basic proposal of chalice was to take sample volumes from EPNS volumes and produce something that resembles an historical gazetteer.
- Uses a subset of the Edinburgh geoparser in Unlock text. Splits into two components: the geotagger, which finds all the place names in the document, and the georesolver, which compares them against a gazetteer (GeoNames or the Unlock gazetteer).
- Needs rules in the background – if you give it a coherent document talking about an area, it can use them to structure the resolution in a meaningful way. Two gazetteers – GeoNames and OS.
- Could configure it to look for different things, such as feature types, and the user can specify which locality to search in.
- Unlock deals with HTTP requests and returns a piece of XML.
- Some EPNS volumes are in digital form – want to convert the content into a structured thing that could be a gazetteer, or be converted into a gazetteer.

- Were given one volume, for Shropshire, that was born digital; one further volume (Cheshire) is being digitised by Belfast.
- Has not looked at field names at all, just on the name of the township; the headline that gives more detail about the names, and on the historic spelling of the place name.
- Some lower-level (e.g. field) names have been transcribed from the OS map.
- The XML has a lot of info about the structure of the volume – font sizes, styles etc in the original – very useful and important. Volumes/layouts/conventions: for each volume, a small amount of tweaking is needed to get the parser to work. Typefaces etc have changed very much over the years of the survey.
- Some EPNS entries have OS grid references, but some don't.
- Also want to do work on the different sources from abbreviations, but the abbreviation lists might not be complete/correct.

In discussion, concern was expressed that, despite the evident sophistication of the software, the apparent reliance of the CHALICE project on existing digital gazetteers, notably GeoNames, to identify the geographical names within the text of the EPNS volumes, meant that the end result would only be as historically authoritative as GeoNames rather than as the EPNS.

Hestia

Leif Isaksen (University of Southampton)

The Hestia project was funded by the AHRC and used Perseus project texts of Herodotus, which are original language and translated, and XML encoded, to extract geospatial information. Herodotus has several thousand references to spatial locations. There are problems of disambiguation – e.g. islands are often also settlements *and* spatial entities. The project worked with the Open Context project at Berkeley, which extracts references in ancient texts to sites – e.g. if you have an excavation report about Carthage, you can mine references to Carthage in the literature.

GAP project – looking at Classical texts in the Google books mass digitisation corpus, funded through the Google award programme.

-Identifying toponyms in classical texts

-Using multilingual gazetteers

-Disambiguation

-Saving metadata

-Text discovery

The project is deriving URIs, rather than co-ordinates, and demonstrates an approach more suited to narratives than database records of thesauri.

Narrative map texts widget

- Adopting a Pleiades type approach – e.g. London and Londinium refer to the same concept. Will derive URIs rather than coordinates.
- Asked how far can you capture existing variants in the existing gazetteer? Hestia is not looking to find stuff outside gazetteers.
- Asked if the project is working from translations. In large part it is but Perseus has stuff in Greek as well (but the Greek might have variant spellings).

Discussion

An open discussion reflecting afternoon presentations raised several issues.

- How can we develop gazetteers suitable for wider use?
- Getty Thesauri – cf. LI's talk – is an example of a stable gazetteer. The problem with the TGN is that they did not pay anyone for content. You cannot properly reference the TGN and it is expensive. There are known problems with GeoNames, but it is used because it is freely available, but most importantly, it uses a stable URI. It is a waste of time if there is not a stable URI, or a stable way to access data.
- This is why we have to get gazetteer data out as linked data.
- It was noted that while the concept is constant, URLs can become inappropriate – e.g. the "Vision of **Britain**" website has data from Estonia.
- OS research has looked at issues such as namespace hosting - Chris Jones in Cardiff. It is not a purely British issue- need to address this.
- Different people produce different things: are these different resources, or can they be brought together as a single resource? Theoretically they can, but much of all the EPNS's material is on paper. There is nothing in the structure that would forbid it, but it is not digital.
- The public do not always want things as complex as linguistic analysis... plan in NI is to put in digital clips of pronunciation and photos of the place/boundary stream etc to make it easier for public consumption.

Halogen

Jonathan Tedds (University of Leicester)

Jonathan presented the Halogen project, building a research infrastructure funded by JISC Managing Research Data Programme. The project is building links between genetics and surnames through correlating farming records with genetics, to build relationships between surnames and locations (see <http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmp/halogen.aspx>). There is a strong correlation between surnames and geographical location if you remove the 20 most common surnames. Issues reported included the use of data points in different places and variable resolution of geographic data. Possible solutions include using the 'Key to English Place-names' and ontologies for this material. Tedds noted that there was a good deal of interest from the scientific community in this work.

Concluding Discussion

Alastair Dunning - summing up and leading concluding discussion

- The Unlock Service has great possibilities if it could be used with different gazetteers. Could someone upload a gazetteer as well as a text?
- There is a need to get a registry of gazetteers, as well as texts, etc, to which they could be applied. If you are an academic who wishes to use a gazetteer, this would be a good resource.
- People with no interest in geography could use such a thing to mine/manage their information if it has geographic metadata.

- Need gazetteers of different levels – different users (e.g. literature) have different information needs (e.g. from place-name scholars, or GIS practitioners).
- Training (for both researchers and graduate students) in representing content spatially, which applies to various other areas. Noted that much money has already been put into GIS training, to little effect. Also noted that widely available geo web services make it easier to use these things.
- Content creation – will never be 100% accurate to create gazetteers automatically using NLP, but there is a model for doing where you get the system to do its best guess, then have an interface to allow the expert to curate the information. Need expert input to underpin ‘assisted gazetteer curation’. Noted that in some cases you *do not* need 100% accuracy. E.g. for resource discovery you don’t need such a high level of accuracy. However you sometimes have general readerships AND specialist ones. There is an impetus to e.g. write books for the RAE Noted that have been projects specifically to engage the public.
- There needs to be funding for interfaces for gazetteer development, possibly with support from EPNS, Nottingham, or similar organisations providing quality control
- The needs/issues encountered by geoparser for the format could inform future development of layout of EPNS volumes. There is a need for broader, resource discovery tools, and there should also be future research tools with deeper or narrower use.
- What would something like unlock require from the next generation of EPNS?
- Crowdsourcing is not a source of wisdom, but of collation labour, and knitting together resources.
- Money from Google ads on these resources (can lead to legal problems for HEIs).
- Groups are quite proprietary about their gazetteers. E.g. the RCAM does not know if it can give its gazetteer away, because it comes from multiple sources. It would be good to make Unlock open source, but this might mean not being able to supporting the software.
- Could be useful to have a mechanism where users can highlight text they’re not interest in.

Developing a roadmap

There was consensus that spatio-temporal place-name gazetteers deployed to link disparate e-resources by location were essential e-infrastructure to facilitate the integration and interrogation of existing e-content. Without this infrastructure resource discovery was likely to be haphazard, data silos were likely to be an on going issue, and e-content was unlikely to be sufficiently embedded in teaching and research to result in a step-change in research methodologies and scholarly findings. Whilst the workshop neither sought, or arrived at, conclusions on how this e-infrastructure should be created in practical terms, the authors of this report have drawn on the workshop discussions to summarise the current position and recommend steps to developing a roadmap.

- Current contemporary place-name gazetteers are not fit for purpose in linking historical resources. They lack chronological depth and do not attempt comprehensively to record variant place-names. Spatio-temporal gazetteers should link to contemporary gazetteers but cannot be seen as merely an extension of existing resources.

- England is unique in the British Isles in that through the work of the English Place-Names Survey a very comprehensive analogue historical gazetteer exists. Interim findings from the CHALICE Project indicate that EPNS content can be digitised using optical character recognition software and the content restructured to form an electronic gazetteer. The development of a gazetteer will be a significant and costly undertaking but, reflecting close to 80-years of detailed archival work by EPNS, such a project represents very good value for money.
- For Ireland, Scotland and Wales nothing approaching the comprehensiveness of EPNS exists although place-name work is taking place. Here it would be desirable to access current work and augment this with readily available place-name lists. For Ireland a key source is likely to be the Census which from 1861 publishes a hierarchical gazetteer for townlands. With 60,000 townlands listed this in itself provides significant content. For Wales and Scotland, Vision of Britain provides a growing number of place-names. Separate, but interlinked, projects for each country would most likely attract funding. The extensiveness of a gazetteer will depend on the ingestion of existing digital content (which for Ireland is extensive), the work of place-name scholars to interlink variant names over time, and crowd sourcing to further populate and verify the content. A ready crowd sourcing audience is most clearly in place for Ireland reflecting the interest in Irish Studies and the relevance to genealogists reflecting the Irish diaspora.

It is recommended that a working group of stakeholders present formal applications to funders adopting the approaches outlines above. Such a group should:

- Act as advocates for investment in historical gazetteers as key infrastructure for e-scholarship with funders
- Scope existing analogue and digital place-name sources, particularly for Ireland, Scotland and Wales
- Seek input from the wider academic community about how locationally-integrated e-resources can better be imbedded in research and teaching practice
- Identify key areas for immediate strategic investment

ⁱ See <http://www.ahrcict.rdg.ac.uk/activities/e-science/background.htm>.