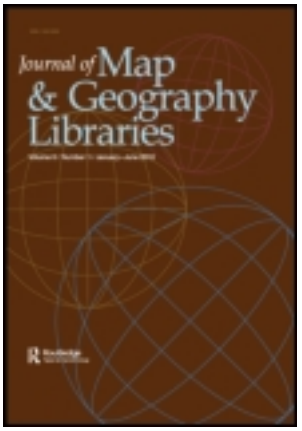
This article was downloaded by: [University of Edinburgh]
On: 03 October 2013, At: 07:33
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wmgl20

AddressingHistory—Crowdsourcing a Nation's Past

Stuart Macdonald [a] & Nicola Osborne [a]

[a] University of Edinburgh , Edinburgh , Scotland

Published online: 17 Apr 2013.

To cite this article: Stuart Macdonald & Nicola Osborne (2013) AddressingHistory—Crowdsourcing a Nation's Past, Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives, 9:1-2, 194-214, DOI: 10.1080/15420353.2013.768191

To link to this article: http://dx.doi.org/10.1080/15420353.2013.768191

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Journal of Map And Geography Libraries, 9:194–214, 2013
Published with license by Taylor & Francis
ISSN: 1542-0353 print / 1542-0361 online
DOI: 10.1080/15420353.2013.768191

Routledge
Taylor & Francis Group

AddressingHistory—Crowdsourcing a Nation's Past

STUART MACDONALD and NICOLA OSBORNE
University of Edinburgh, Edinburgh, Scotland

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

This paper charts the development and delivery of a Web 2.0 informed community engagement tool and application programming interface (API) developed at EDINA in partnership with the National Library of Scotland, as part of the JISC Digitisation and e-Content Programme. The AddressingHistory Web tool enables members of the community, both within and beyond academia (particularly local history groups and genealogists), to enhance and combine data from digitized historical Scottish Post Office Directories with contemporaneous large-scale historical maps. The paper discusses the background of post office directories and the corresponding georeferenced old maps for Scotland, the technical platforms deployed including sustainable software components, and Web applications and services. It also examines issues relating to data parsing, user generated content (UGC) created by the community including georeferencing, and the use of social media amplification for community engagement and future directions. To conclude, the paper argues that to be successful online, crowdsourcing tools such as the one developed for this project require a critical mass of content to fully engage the user community and that such success will ultimately be measured by continual and extended use within the wider community.

KEYWORDS crowdsourcing, genealogy, Scotland, open source, historic mapping, georeferencing, text parsing, community engagement

© 2013 Crown copyright

Acknowledgments: Chris Fleet, George Hamilton, Ian Fieldhouse James Reid, Lasma Sietinsone, and Fiona Hemsley-Flint

Address correspondence to Stuart Macdonald, EDINA & Data Library, University of Edinburgh, Causewayside House, 160 Causewayside, Edinburgh EH9 1PR, Scotland. E-mail: stuart.macdonald@ed.ac.uk

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

INTRODUCTION

The AddressingHistory project was funded as part of the Developing Community Content strand of the JISC Digitisation and e-Content Programme[1] and led by EDINA National Data Centre[2] at the University of Edinburgh, exploiting substantial in-house experience with mapping, geoparsing and georeferencing data.

The aim of the project was to create an online engagement tool built, using open standards, to enable members of the community, both within and beyond academia (particularly local history groups and genealogists), to combine and enhance data from digitized historical Scottish post office directories with contemporaneous large-scale historical maps. Both digitized resources (maps and directories) were provided, without restriction, by project partner National Library of Scotland (NLS) which also shared extensive digital mapping and geocoding expertise.

AddressingHistory was informed by a number of initiatives exploring the possibilities of crowdsourcing data or volunteering information, in particular, projects that have engaged wider audiences in detecting errors in OCR and the editing or adding of locations.

The concept of crowdsourcing, the breaking up of large data tasks into smaller distributable subtasks, is certainly not new. The compilation of the *Oxford English Dictionary* in the late nineteenth century (as chronicled in Winchester 1998), the Mass Observation Archive at the University of Sussex[3] (as discussed in, e.g., Sheridan 1994), and Sir Dudley Stamp's 1930s *Land Utilisation Survey of Britain* (Stamp 1931) offer famous examples of the power of combining data collected or validated by keen amateur volunteers. The Internet and the emergence of Web 2.0 (O'Reilly 2005) technologies have enabled crowdsourcing projects to take place on a much more rapid and global scale. Academic applications of crowdsourcing were led by citizen science projects, such as FoldIt[4] and GalaxyZoo,[5] which enable participants to directly contribute user-generated content or edit controlled and constrained types of data, communicate with others in the community, and develop their understanding of the topic as a result of their participation.

The Australian Newspapers Digitisation Programme, which began in 2007 and now forms part of Trove,[6] was one of the most successful early examples of crowdsourcing to detect and correct OCR errors perhaps because of the attempt to preempt and mitigate for risks associated with public contribution of data (as discussed in Holley 2009). Transcribe Bentham, a project to transcribe the digitized Jeremy Bentham archive, has further explored the technological, community participation, and validation challenges associated with OCR crowdsourcing projects (Terras 2010).

Motivations for participating in crowdsourcing projects can vary wildly from very personal and immediate needs to access and record data (Ziemke 2012) to enthusiasm for a hobby or interest (Chris Batt Consulting 2009),

to long-term interests in being part of a community (Zeitlyn 2003). Raddick et al. (2010) found that most GalaxyZoo contributors did not have a single motivation with an interest in the subject, astronomy, being a key driver but also sitting alongside an interest in contributing to original research or the fun of the task at hand. Causer and Wallace (2012), writing about Transcribe Bentham, note the distinction between a "community" and a "crowd" project and define their own participants as including both a tight-knit community of regular contributors and a larger crowd of occasional, or one-off, contributors. Echoing the findings of RunCoco initiative's guidance (RunCoCo 2011), Causer and Wallace also found that a combination of traditional media, Web 2.0 tools, and in-person communications were crucial for reaching out to different types of audiences and potential contributors.

Volunteered geographic information has risen to prominence through the emergence of a range of interactive Web tools and services harnessing the Geoweb[7] to create and disseminate geographic data provided voluntarily by individuals (Goodchild et al., 2007). FieldScope,[8] a Web-based platform to support the National Geographic Community Geography Initiative is a good example of a collaboration tool designed to support and engage citizen scientists in geographical investigations of real-world issues. The publication *Crowdsourcing Geographic Knowledge* (Sui et al., 2013) further explores a range of public participation initiatives, emerging technologies, and new challenges in a range of spatial settings, some of which find resonance with the conceptual development of the AddressingHistory Web tool.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

SCOTTISH POST OFFICE DIRECTORIES

Historical post office directories, precursors to modern-day yellow pages, offer a fine-grained spatial and temporal view of important social, economic, and demographic circumstances. For Scotland there are currently over 700 such directories spanning the period 1783–1912 (see Figure 1), which have been scanned, undergone optical character recognition (OCR) processing, and been published by the NLS in conjunction with the nonprofit Internet Archive in an ongoing initiative.

The directories emerged during the late seventeenth century to meet the demand for accurate information about trade and industry due to the expansion of commerce during this period. They were published more frequently than the census and generally had information about local facilities, institutions, and associations; listings for private residents, traders, trades, and professions; sometimes details of notable people and advertisements. They are also one of relatively few historical records of the lives of women, particularly widowed women.

The ways in which publishers collected data for the directories varied considerably. Some obtained information by personal canvassing and

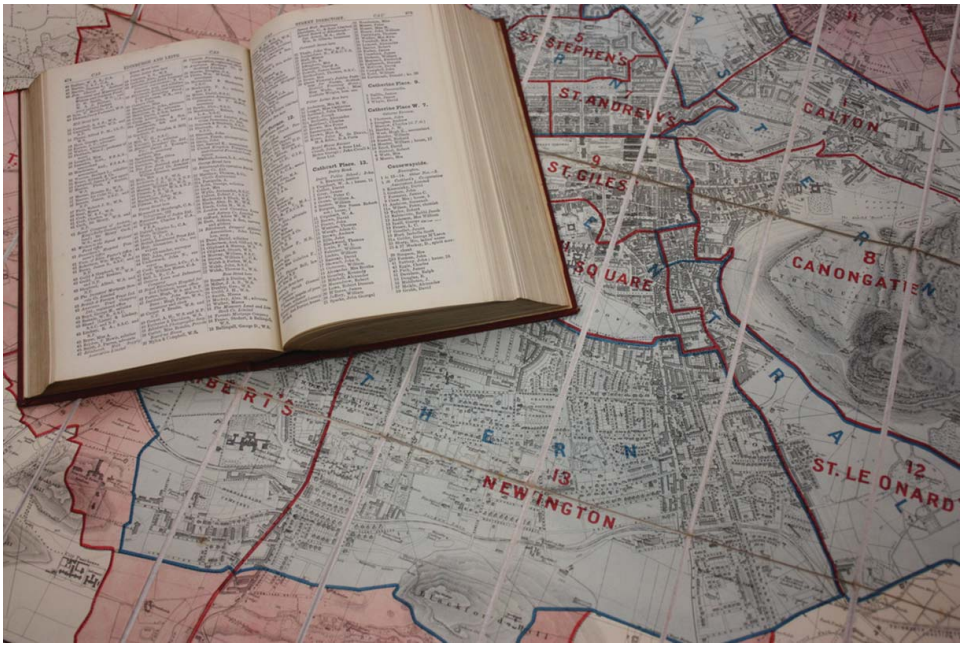


FIGURE 1 An Edinburgh post office directory and map from the National Library of Scotland's collection.

combined the results with existing trade listings. Other publishers invited citizens to send in their names together with a small payment if they wanted to be included in the directory.

By the early nineteenth century, methods of compilation were better organized. In part, this reflected the growing links between directories and the post office. Many postal officials turned their hand to directory publishing as a means of both aiding their work and augmenting their income. Information was collected by letter carriers, who circulated forms during their postal rounds and delivered the finished directory on commission.

Each annual Scottish post office directory includes an alphabetical list, known as the general directory of a town's or county's inhabitants, making it easy to find out where people lived and, in some cases, how they were employed at that time.

In addition to the general directory, most of Scotland's post office directories include a street directory and a trades directory, an alphabetical listing of people by their address and professions. A range of other directories can also be found including those listing banks and banking companies, clergymen and places of worship, educational institutions and teachers by their subject, jurisdicial institutions and practitioners, medical and surgical institutions and practitioners, chief officers of state, peers of Scotland, and members of Parliament. Additionally, one of the most appealing parts of the

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

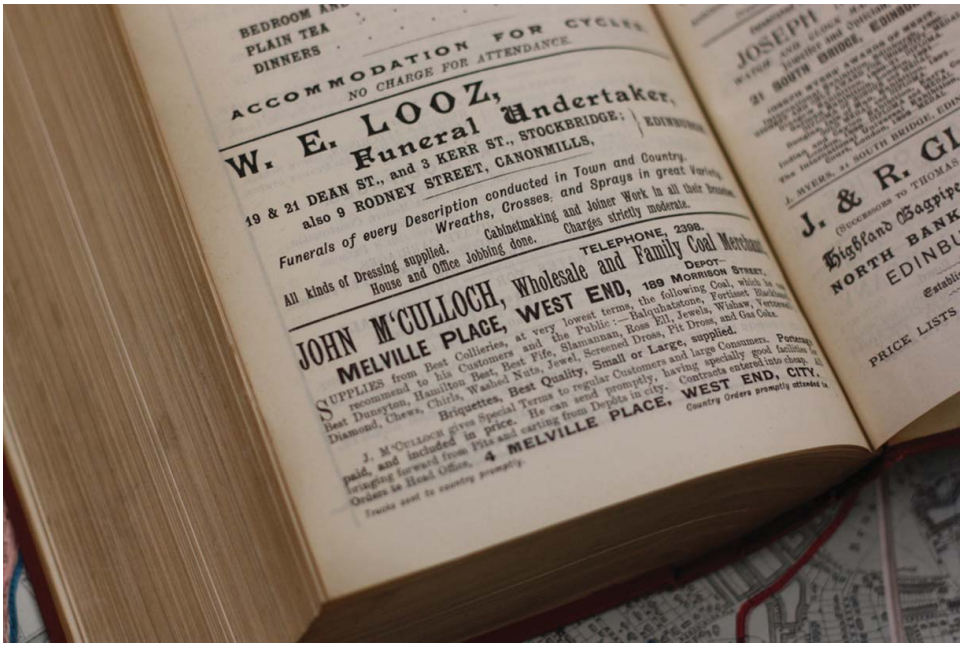


FIGURE 2 "W.E. Looz, Funeral Undertaker" ad from an Edinburgh post office directory in the National Library of Scotland's collection.

post office directories is the advertisements, which give a valuable insight into society and lifestyle for a particular era (see Figure 2).

The directories offer a wealth of detailed information regarding residential names, occupations, and addresses and are therefore a fitting resource for genealogical study. Post office directories are also recognized as being of great importance in establishing and understanding commerce and trading patterns within Scotland during the previous centuries.

Phase 1 (April–September 2010) of the project focused on three volumes (1784–85, 1865, 1905–06) of the Edinburgh digitized post office directories (held in XML format in a database structure) and mapping from the same periods. The interface and back-end storage solutions developed were scalable and as far as is practicable, designed to be self-standing so that multiple independent instances of the tool and application programming interface (API) could be supported and customized for different audiences and purposes.

One significant deficiency of the digitized directories, which the AddressingHistory online tool aimed to redress by crowdsourcing, is that addresses are not georeferenced. Georeferencing has the potential to enhance the collection by enabling spatial search and discovery. For instance, addition of the georeferencing factor permits a map-based metaphor to be used in the exploration and visualization of the resource, for example, the historical distribution of shipwrights in Edinburgh can be plotted on a base map or the map itself can be used to explore the spatial distribution of selected

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

phenomena (and their variation over time). Georeferencing the directories also opens up the potential for creating personalized maps illustrating family histories, maps tracking changes in local communities, and maps linking to other digitized materials such as census records and georeferenced images.

DIGITIZED HISTORICAL MAPPING

The second set of resources that the project exploits are digitized maps. The National Library of Scotland (NLS) Map Library is one of the ten largest in the world, with unrivalled collections: When it was known as the library of the Faculty of Advocates (from 1689), maps of Edinburgh were actively collected; as a copyright library (following the 1710 Copyright Act), the collections became particularly strong in the printed mapping of Scotland. Since 1998, the NLS Map Library has scanned over twenty thousand historical maps of Scotland,[9] including over five hundred of Edinburgh and its environs available for viewing as high-resolution color images. It is the preexistence of large-scale georeferenced maps, against which the historical post office directories are assessed and contextualized, that allows georeferencing down to individual house addresses to be accomplished. The availability of large-scale historical maps of the same era allows users of the online tool to attach a georeference to a particular post office directory address by simply clicking on the map.

The user interface to the tool and associated API was designed to be intuitive and easy to use to encourage researchers, local historical societies, genealogists, and members of the wider community to discover, explore, and contribute to rich records of social history and to enable them to create their own related maps and data sets for both academic and personal research. These were also developed to be sympathetic to tools developed by related projects, including Visualising Urban Geographies,[10] an online resource developing new insights into the spatial character and historical development of Edinburgh.

A combination of automated georeferencing (using the Google geocoding API) and crowdsourcing of the post office directory records has the potential to lead to a fully geocoded version of the digitized directories, meaning that a key element in determining the success of the project has been public participation and collaboration. This was encouraged and supported through extensive community outreach, particularly via ongoing communication with a Project Advisory Board, Edinburgh Beltane,[i] the College

[i] Edinburgh Beltane was established as a National Co-ordinating Centre for Public Engagement (NCCPE) Beacon for Public Engagement. More information can be found at http://www.publicengagement.ac.uk/about/beacons/edinburgh-beltane. The NCCPE Beacons Programme has now come to the end and Edinburgh Beltane has since rebranded to Beltane Public Engagement Network and may be found at http://www.beltanenetwork.org/

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

of Humanities, and Social Sciences Knowledge Transfer Office, as well as online through social media channels. These activities are described in more detail in the Social Media and Community Engagement sections later in this article.

A second phase of internal funding sought to develop functionality complementary to the original work while focusing on both sustainability and interoperability to resonate with JISC's vision to build durable deliverables.[11] Work included streamlining data preprocessing and loading processes with a view to providing cleaner output, spatial searching using bounding boxes, a re-evaluation and enhancement of the post office directory, and geocoding parsing processes. To streamline and improve the searchability of content, computer-generated metadata were incorporated into post office directory entries such as categorizing places and professions (through the use of standard industrial classification [SIC] codes[12]). Support for entries with multiple addresses (i.e., entries in which individuals may have a domestic address and one or more business addresses) was also improved for searching and editing purposes.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

PARSING

To assist the georeferencing process, address entries from each of the directories were parsed using Google's geocoding API[13] to assign a georeference to each address encountered in the XML database. There were issues with the legibility of the OCR-processed text (especially for the directory for the earlier period) including addresses, in addition to period addresses no longer being in existence or having suffered name changes. Thus, within the interface, a ranking mechanism makes explicit the relative accuracy of this mass-geocoded content.

In this phase of work, post office directory configuration files, which are used to configure the parser settings, were taken outside the parser itself. This enables the parser to be adapted to different formats and conventions in different directories. Most importantly, the configuration files are now editable, meaning that rules and conventions required to successfully parse the directory content can be augmented and refined by expert users. For instance, if a user observes repeated OCR errors resulting from the structure of a post office directory, such as anomalous line returns, misspellings, textual abbreviations, and the like, they are empowered to make the appropriate edits to address those issues.

The tool and API have also been developed alongside improvements to the technical platform. The improvements facilitate post office directory parsing on request (i.e., for an area of the country or era not currently covered by the tool and API) with the understanding that the requesting body or individual has sufficient technical expertise to initiate and manage the

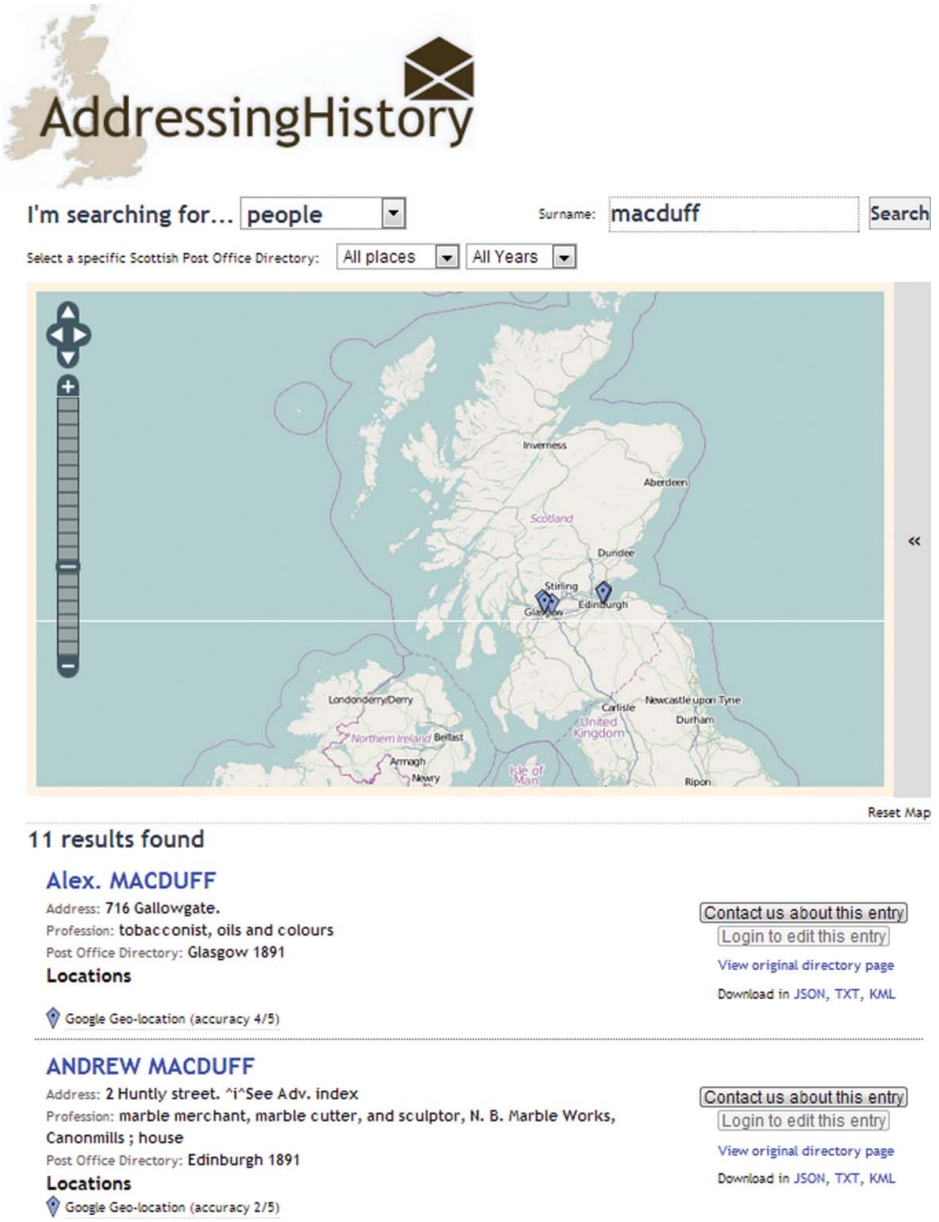


FIGURE 3 Screenshot of AddressingHistory search results for a logged-in user showing the micro level "Edit this entry" and "Contact us about this entry" crowdsourcing buttons.

parsing process (monitored by EDINA). Thus, in effect three crowdsourcing opportunities are now supported: edits at the level of the record (micro level); at the level of the configuration files (meso level); and at the level of the individual post office directory (macro level) (see Figure 3).

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

A key element to any crowdsourcing initiative is a critical mass of content sufficient to engender continued community interest and subsequent engagement. With this in mind, additional content for Edinburgh, as well as Glasgow and Aberdeen, for 1881 and 1891 to coincide with census years was added to the tool and API. The addition of Glasgow introduces Scotland's biggest population center, and the addition of Aberdeen expands AddressingHistory beyond Scotland's "Central Belt."

At the time of writing (October 2012), it is too early to say whether extending both spatial and temporal extents have had a positive impact on user participation (e.g., for images, etc.).

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

TECHNOLOGIES

The AddressingHistory tool and API were built with resilience and sustainability in mind, opting for open-source software where possible to allow greater flexibility and a feature-rich application.

AddressingHistory is built as a typical three-tier Web application. For the user-facing client presentation component, standards such as XHTML, CSS, and other relevant W3C Web standards (e.g., images) were employed as recommended by JISC. The Web interface is supported by mainstream browsers and OGC-compliant applications, including the Web Map Service interface standard.

OpenLayers, an open-source JavaScript library for displaying map data in Web browsers (with no server-side dependencies), was used for the mapping components.

The API, allowing access to the raw data in multiple output formats, is accessible via a RESTful Web service. The terms of the API and the variety of formats available ensures that the content is highly accessible and may be reused in standard research tools including GIS, combined with data from other projects, included in mash-ups or reimagined via novel interfaces.

Development

The project followed best practice for technical development, making extensive use of a number of common and well-established resources including Java SDK, Spring MVC framework, and the jQuery Javascript libraries. Unit testing was performed via the jUnit libraries.

Development initially began by scoping the application's requirements, designing a database structure to store the information contained in the post office directories in conjunction with preprocessing and data-loading software. The structural interpretation and translation of the varied content from the directories, through the development of the post office directory

parser, proved to be a time-consuming exercise. The directory data were processed and additional metadata such as the locations of addresses were added to the database.

The API, following JISC recommendations for API Good Practice,[14] was designed to allow access to the raw data using a number of HTTP GET queries, including a parameter that allows Web developers to specify the format (JSON, KML, or TXT) they want the result returned in.

The client application was built on the API, featuring Web-based mapping. To the OpenLayers mapping, we added a collection of historical maps from NLS, contemporary to post office directories of interest. A user registration system and facilities to edit the stored data and suggest specific changes were added toward the end of the development, together with various enhancements—including a view to the original scanned directory pages, a feature that provides crucial context and verification for academic researchers.

All components of the Web-accessible service and API are hosted via a Solaris 10 virtual container, together with an established PostgreSQL database, hosted at EDINA. Throughout the project, the source code, tests, and configuration files were stored in a GIT source-control repository. Software builds and releases were automated via the Apache Maven software-project management tool. Documentation is stored in a shared repository.

The post office directory parser code is available on Github under a general public license[14] and is also hosted on the Official Python Programming Web site.[15]

Augmented Reality

Using the BuildAR CMS tool, an AddressingHistory layer has been developed and published for use with the Layar Application[16] for a range of mobile platforms including iPhone and Android. Raw ASCII points of interest (POIs) and associated metadata are uploaded as a set of Google Map coordinates. POIs (e.g., each profession or SIC code) are then associated with an image or symbol representative of said POI. The AddressingHistory layer works with the Layar application to compare information about your current location (from your phone) and the georeferenced entries in AddressingHistory to work out which historical residents and businesses used to be located near where you are standing at that moment. This means that you can point your phone's camera in a direction and see clickable markers appear for each historical address superimposed on a live view of the street. Currently, the Layar application is configured for content from Edinburgh only (see Figure 4).

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

FIGURE 4 Screen capture of the AddressingHistory augmented reality layer.

User Generated Content

The AddressingHistory project raised a number of issues regarding user generated content (UGC) created by the community such as mediation, validation, and cross-checking. At present, the AddressingHistory team retains the option to check UGC on a periodic basis. As part of a sustainability plan, it is envisaged that a more routine process for validation can be established using a combination of automatic validity checks wherein sufficient user

contributions exist for multiple edits to be compared. Opportunities for an engaged user group comprising active members of the user community to voluntarily conduct validation and cross-checking of UGC through a devolved mediation process—as used in the Wikipedia model—may also be an appropriate approach, given the large number of motivated, expert amateurs within the local and family history community.

Users must register to edit entries in AddressingHistory, and any edit is therefore easily traced back to the user. Data changes are logged, meaning that inappropriate behavior (e.g., spam) or inaccurate UGC can be identified and the user can be contacted to justify that behavior or explain particularly unusual editing patterns. A perpetrator can theoretically be prevented from editing further, but at the time of writing no problems with malicious edits have been experienced.

Although UGC is displayed on the Web interface, AddressingHistory has been designed so that the original database and the database containing the UGC are maintained as separate instances, which ensures that any inaccurate or inappropriate user-generated content can easily be removed without compromising the authority of the core database (see Figure 5).

FIGURE 5 The AddressingHistory "community" badge designed for use by bloggers, and those with local history Web sites.

Social Media

From the outset of the project, there has been an understanding of the importance of engaging with a broad community of users and doing so rapidly. At that time, a steering committee informed development of a mechanism whereby the "crowd" could contribute to the creation of a fully geocoded version of the digitized directories. There was also a need to raise awareness of the project while the AddressingHistory tool was still under development to ensure that potential users and advocates were actively involved in the project before the tool was launched.

During the initial scoping process, the project team began to work with Edinburgh Beltane and the University of Edinburgh College of Humanities and Social Sciences Knowledge Transfer Office, and Communication and Marketing. Each organization provided advice on best practice and support in identifying key stakeholders for the project and in connecting to a wide

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

range of interested organizations and stakeholders. Working in partnership in this way enabled the project team to quickly establish effective approaches to engage several identified key audiences: researchers and academics concerned with the quality and coverage of the data, local historians with a specific interest in Edinburgh history, genealogists investigating specific names within the data, and under- and postgraduate students undertaking focused projects related to the data.

Social media channels were deployed to engage the public, to develop links within the local and family history communities, and to act as a vehicle to expose the tool and API to a wider audience. The following section describes both method and mechanism used to engender public collaboration and community engagement.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

Building and Developing Community Connections

At the outset of the project, an information page was created on the EDINA Web site[17] and later updated to connect to additional AddressingHistory presences. A locally hosted WordPress blog[18] was deployed as the project's Web site and hub for communicating and engaging with our target audiences.

One of the central challenges was the need to establish a motivated crowd to contribute content to AddressingHistory postlaunch. The blog was therefore used to share both core project progress and content with broader appeal such as guest blog posts and features on upcoming events or reports from those attended.

The blog had to remain relevant for funders and academic and organizational audiences. However, the establishment of a Facebook page[19] provided a more informal and accessible space for audiences, particularly those from local and family history backgrounds, to engage, share, and recommend blog posts, updates, and related materials. These presences were complemented with a Twitter profile (@addresshistory) which proved to be unexpectedly effective for building the network around the project, virally sharing updates and reaching out to key figures from target audiences such as notable academics, widely read bloggers, and heritage organizations.

Community Engagement

From the outset, the project team encouraged blogging and discussion of the project, and the project officer proactively sought out potential contacts, followers, and bloggers, responding to comments on the project to ensure that mentions were complemented with links to the Web site and that questions were answered. Rewarding interest and participation encouraged ongoing support and engagement with the project. The intent was to create an atmosphere of openness and participation, recognizing contributions

FIGURE 6 Still from the "Digitising the Post Office Directories for AddressingHistory" video hosted on YouTube.[26] The still shows Zaria from the Internet Archive, scanning post office directories while talking about the digitization.

from stakeholders, thus reflecting that the tool being created would be editable, open, and accessible for anyone wishing to explore or participate in improving it (see Figure 6).

In the run-up to launch, a number of key resources were made available to encourage both bloggers and mainstream press to engage with the online tool and the coverage of its launch. High-resolution images[20] of the physical directories and maps were captured and made available under Creative Commons license with an encouragement to bloggers and journalists to use these images. Two videos[21] were also created and shared via YouTube—again with an explicit encouragement to share—in readiness for launch, with one explaining the process of digitizing the resources included in AddressingHistory and the other placing these resources and their relevance into context. Finally, several online badges[22] were created to provide a visual way for users, key bloggers, and organizational Web sites to connect to and show support for AddressingHistory (see Figure 7).

AddressingHistory benefited greatly from the support of existing genealogy and local history bloggers and online communities, receiving regular

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

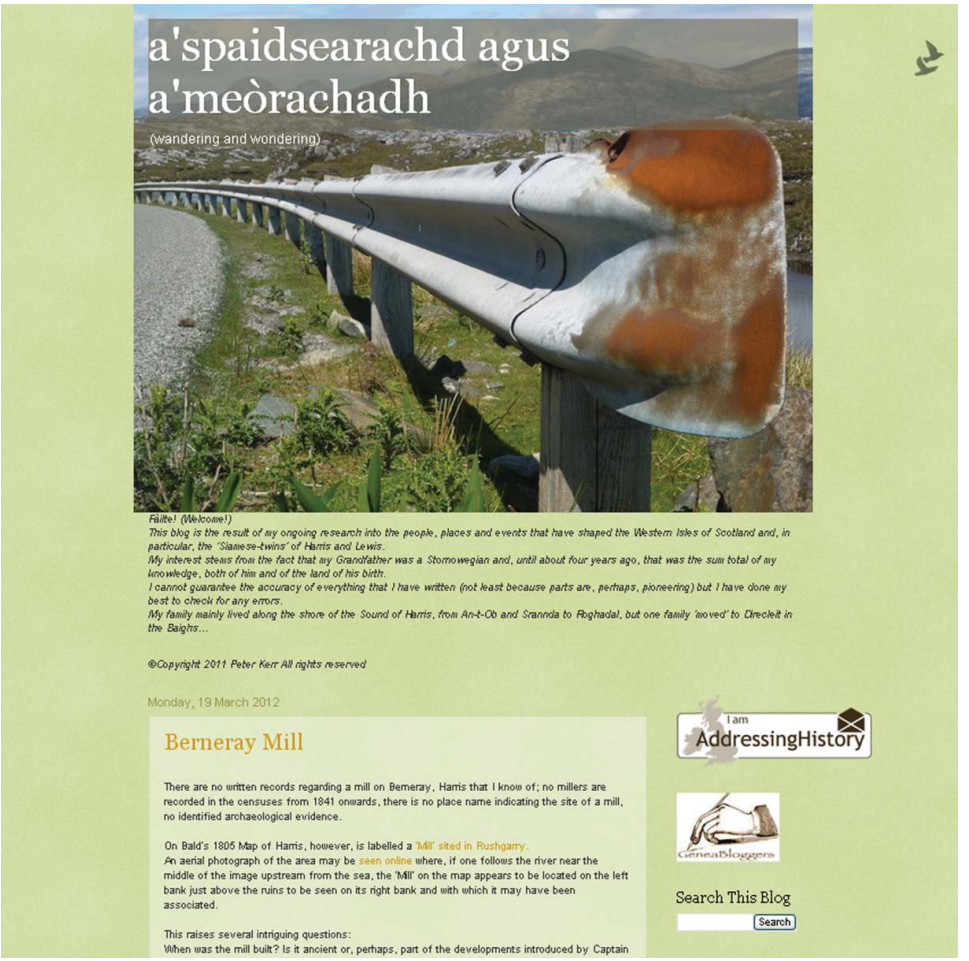


FIGURE 7 An AddressingHistory badge in situ on Peter Kerr's "a'spaidsearachd agus a'meòrachadh (wandering and wondering)" blog.[25]

mentions and links from a wide variety of sites and discussion boards.[ii] The social media and Web presences helped reach out to many interested parties,[iii] including advocates who wrote about the project online and in several high-profile local and family history print publications. Indeed, online channels never sat alone; they have always been complemented by

[ii] For example, Clan MacLea/Livingstone (http://clanlivingstone.info/forum/viewtopic.php?f=5&t=1084&p=9800&hilit=addressinghistor y#p9800), Rootschat (http://www.rootschat.com/forum/index. php?topic=496764)

[iii] Indeed, many of the social media monitoring techniques tried on AddressingHistory are now successfully being used to better monitor social media mentions of other EDINA projects and services.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

outreach activities such as events, presentations, and print communication, which were also instrumental in exposing the project to a wider audience.

Ongoing Activity

As a longer term strategy, we intend to maintain, where practicable, blog activity and Facebook and Twitter presences. The project team continues to respond to comments and reward contributions to the project whenever possible. A mailing list has been set up to ensure that we can remain in contact with those interested in AddressingHistory developments, and a Google group has been established that is aimed at users interested in using the AddressingHistory API for their own Web sites, projects, or mash-ups.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

NEXT STEPS

Features and Functionality

Initial investigative work has begun with the University of Leicester, which hosts Historical Directories,[23] a digital library of local and trade directories for England and Wales from 1750 to 1919, to pilot sample data for an English directory contemporaneous with those Scottish directories in service. Further discussion will scope both suitability and scalability of AddressingHistory as the mechanism to deliver the corpus of content for Great Britain.

In addition to upgrading OpenLayers and broadening the geographic scope of the Augmented Reality app, interesting further work may also involve investigating how best to expand and develop the crowdsourcing aspects of AddressingHistory. For example, there are opportunities to draw upon game mechanics, rewarding users immediately for their edits and additions to the data and challenging them to contribute to a broader array of edits for fun and competition rather than focusing only on the small number of addresses of immediate research interest.

Another avenue of development under consideration is the inclusion of facilities to upload and attach georeferenced content such as images, census records, videos, and sound files to AddressingHistory entries in the database. In this way, the directories would be extended to include photographs of people, buildings, and landmarks, thus enriching the resource, broadening both utility and appeal, and better enabling contributors to share and showcase their contributions to AddressingHistory in the context of their personal research.

Context versus Content

Robert Morris, emeritus professor of social and economic history at the University of Edinburgh, who provided the introductory presentation at the

AddressingHistory launch,[24] had reservations regarding context versus content. He said,

> A major feature of this project is the offer of maps, and maps which enable the user to explore and present historical information spatially. The outcome is visually attractive and exciting. There is a danger that the fun of producing the map acts as a barrier to thinking about what is happening.

He indicated that, where applicable, explanatory notes providing information about background, construct, and content of the original directory listings should be made explicit. In addition, underlying assumptions and rules about both the structure of the processed data and translation into a consumable and interactive format should be made clear.

The project team continues to be committed to making AddressingHistory useful, relevant, and interesting both to those engaged in academic research and those engaging in personal research or discovery. Prof. Morris's concerns are partly addressed through the inclusion within the interface of a range of Help documents (including a Post Office Directory Guide and People, Place & Profession Search Guides), an API Guide, and Frequently Asked Questions, although there is scope for further work to explain and contextualize the AddressingHistory data.

Sustainability

In accordance with the project plan, the AddressingHistory project partners were committed to supporting the resource for a minimum of one year while it gathered community traction. During this time, consideration was made to the processes necessary for ongoing dissemination, community take-up of the deliverables, and their adoption by the community. AddressingHistory aims to achieve this through those social media channels established as part of the project and an ongoing relationship with Edinburgh Beltane[iv] and, in turn, to appropriate organizations engaged in local and family history projects.

Given the broad applicability of the resource, it is envisaged that a range of communities may be interested in the longer-term curation and continuance of the project tools, for example, the OpenStreetMap[25] community has an active user base interested in both contemporary and historical addresses. It is also anticipated that the active involvement of engaged users throughout

[iv] The partnership cultivated between AddressingHistory and Edinburgh Research & Innovation and Edinburgh Beltane has initiated ongoing communications between EDINA and both organizations with a view to enhancing community engagement from broader service level perspectives.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

the project and beyond will provide direction on longer-term sustainability issues.

The project team is scoping a number of sustainable business models based on levels of demand provided they remain consistent with the underlying open philosophy, namely: revenue generation through an online donations facility, subscription model (e.g., per annum, per month, per use), a "freemium" model (e.g., free API download of a certain number of records with payment being required for further downloads), and academic advertising.

FIGURE 8 Screenshot of the AddressingHistory online tool.[29]

CONCLUSION

AddressingHistory was an ambitious project that combined a range of technologies from data processing and database design to Web 2.0 and Web-mapping services (see Figure 8). The initial processing of data extracted from the historical directories through OCR presented a unique challenge in terms of data errors, lack of structure, and data inconsistencies both within and across individual post office directories. Much was achieved within the relatively short project phases in terms of public engagement and amplification

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

through social media channels and the delivery of robust and scalable Web site and APIs capable of empowering the crowd with the facility to search and edit georeferenced content from the Scottish post office directories and digitized historic maps from the same era. However, gauging the success of the project goes beyond the delivery of engaging and innovative online tools. It will ultimately be measured by continual and extended use within the wider community.[v]

NOTES

1. JISC. "JISC Developing community content," http://www.jisc.ac.uk/whatwedo/programmes/digitization/communitycontent.aspx (accessed 8 Jan 2013).

2. EDINA. "EDINA Web site." http://www.edina.ac.uk/ (accessed 8 Jan 2013).

3. University of Sussex. 2012. "Mass Observation." http://www.massobs.org.uk/index.htm (accessed 8 Jan 2013).

4. University of Washington. 2012. "foldit." http://fold.it/portal/ (accessed 8 Jan 2013).

5. Zooniverse. 2012. "GalaxyZoo." http://www.galaxyzoo.org/ (accessed 8 Jan 2013).

6. National Library of Australia. "Trove". http://trove.nla.gov.au/newspaper/ (accessed 8 Jan 2013).

7. Wikipedia. http://en.wikipedia.org/wiki/GeoWeb (accessed 8 Jan 2013).

8. National Geographic. Fieldscope: Online Mapping for Citizen Science Investigation, http://education.nationalgeographic.com/education/program/fieldscope/ (accessed 8 Jan 2013).

9. National Library of Scotland. "National Library of Scotland Georeferenced Maps". http://geo.nls.uk/maps (accessed 8 Jan 2013).

10. National Library of Scotland. "Visualising Urban Geographies." http://geo.nls.uk/urbhist/ (accessed 8 Jan 2013).

11. JISC. "JISC Sustaining and embedding innovations Good Practice Guide." https://sustainembed.pbworks.com/w/page/31632855/Welcome (accessed 8 Jan 2013).

12. Office for National Statistics. "UK Standard Industrial Classification 2007 (UK SIC 2007)." http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/standard-industrial-classification/index.html (accessed 8 Jan 2013).

13. Google. "The Google Geocoding API." https://developers.google.com/maps/documentation/geocoding/ (accessed 8 Jan 2013).

14. Guy, M. 2009. "APi Good Practice." http://repository.jisc.ac.uk/344/ (accessed 8 Jan 2013).

15. EDINA. "Post Office Directory Parser GitHub repository." https://github.com/gmh04/podparser (accessed 8 Jan 2013).

16. EDINA. "Post Office Directory Parser python package." http://pypi.python.org/pypi/podparser/ (accessed 8 Jan 2013).

17. Layar. "Layar". http://www.layar.com/ (accessed 8 Jan 2013).

18. EDINA. "AddressingHistory project." http://edina.ac.uk/projects/addressinghistory_summary.html (accessed 8 Jan 2013).

19. EDINA. "AddressingHistory." http://addressinghistory.blogs.edina.ac.uk/ (accessed 8 Jan 2013).

20. EDINA. "AddressingHistory Facebook page." http://www.facebook.com/AddressingHistory/ (accessed 8 Jan 2013).

21. EDINA. "Digitising the Post Office Directories for AddressingHistory." http://www.youtube.com/watch?v=Jb3HwIUge88&feature=plcp (accessed 8 Jan 2013).

22. EDINA. "AddressingHistory's photostream." http://www.flickr.com/photos/addressinghistory (accessed 8 Jan 2013).

23. EDINA. "AddressingHistory YouTube channel." http://www.youtube.com/user/AddressingHistory (accessed 8 Jan 2013).

[v] Note: A free-to-access index for the Glasgow Post Office directories from 1783 to 1911 is now available—http://bizdirs.from-mt.com/ glasgow/

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

24. EDINA. "AddressingHistory Badges." http://addressinghistory.blogs.edina.ac.uk/how-do-i-get-involved/badges/ (accessed 8 Jan 2013).

25. Kerr, P. "a'spaidsearachd agus a'meòrachadh (wandering and wondering)." http://direcleit.blogspot.co.uk/ (accessed 8 Jan 2013).

26. University of Leicester. "Historical Directories." http://www.historicaldirectories.org/hd/ (accessed 8 Jan 2013).

27. EDINA. "AddressingHistory Launch Live Blog—NLS Maps, Tobar an Dualchais, Post Office Directory Scanning Project, Close." http://addressinghistory.blogs.edina.ac.uk/2010/11/17/ (accessed 8 Jan 2013).

28. OpenStreetMap Contributors. "OpenStreetMap." http://www.openstreetmap.org/(accessed 8 Jan 2013) http://www.openstreetmap.org/

29. EDINA. "AddressingHistory/" http://addressinghistory.edina.ac.uk/ (accessed 8 Jan 2013).

Downloaded by [University of Edinburgh] at 07:33 03 October 2013

BIBLIOGRAPHY

Causer, T., and V. Wallace. 2012. Building a volunteer community: Results and findings from. *Transcribe Bentham, Digital Humanities Quarterly 6*(2). http://www.digitalhumanities.org/dhg/vol/6/2/000125/000125.html

Chris Batt Consulting. 2009. Digitisation, curation & two-way engagement. JISC Report. http://www.jisc.ac.uk/publications/reports/2009/digicurationfinalreport.aspx

Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal 69*(4): 211–221. doi:10.1007/s10708-007-9111-y

Holley, R. 2009. Many hands make light work: Public collaborative OCR text correction in Australian Historic Newpapers. National Library of Australia Staff Papers, 2009. http://www-prod.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/1406

O'Reilly, T. 2005. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O'Reilly, 30 September 2005. http://oreilly.com/Web2/archive/what-is-Web-20.html

Raddick, M. J., et al 2010. Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review* 9(1). http://aer.aas.org/resource/1/aerscz/v9/i1/p010103_s1

RunCoCo. 2011. How to run a community collection online: RunCoCo—Final Report— v1.0, March 2011. http://projects.oucs.ox.ac.uk/runcoco/resources/RunCoCo_Report.pdf

Sheridan, D. 1994. Using the mass-observation archive as a source for women's studies, *Women's History Review 3*(1): 101–113.

Stamp, D. L. 1931. The land utilisation survey of Britain. *Scottish Geographical Magazine 47*(3): 144–150.

Sui, D., et al (Eds.). 2013. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer. http://link.springer.com/book/10.1007/978-94-007-4587-2/page/1

Terras, M. 2010. *Crowdsourcing Cultural Heritage: UCL's Transcribe Bentham Project. Seeing Is Believing: New Technologies for Cultural Heritage*. London: International Society for Knowledge Organization. University College London, 9 June 2010.

Winchester, S. 1998. *The Professor and the Madman: A Tale of Murder, Madness, and the Making of the Oxford English Dictionary*. New York: HarperCollins.

Zeitlyn, D. 2003. Gift economies in the development of open source software: Anthropological reflections. *Research Policy 32*: 1287–1291. doi:10.1016/S0048-7333(03)00053-2

Ziemke, J. 2012. Crisis Mapping: The construction of a new interdisciplinary field? *Journal of Map and Geography Libraries 8*(2), 101–117.

Downloaded by [University of Edinburgh] at 07:33 03 October 2013