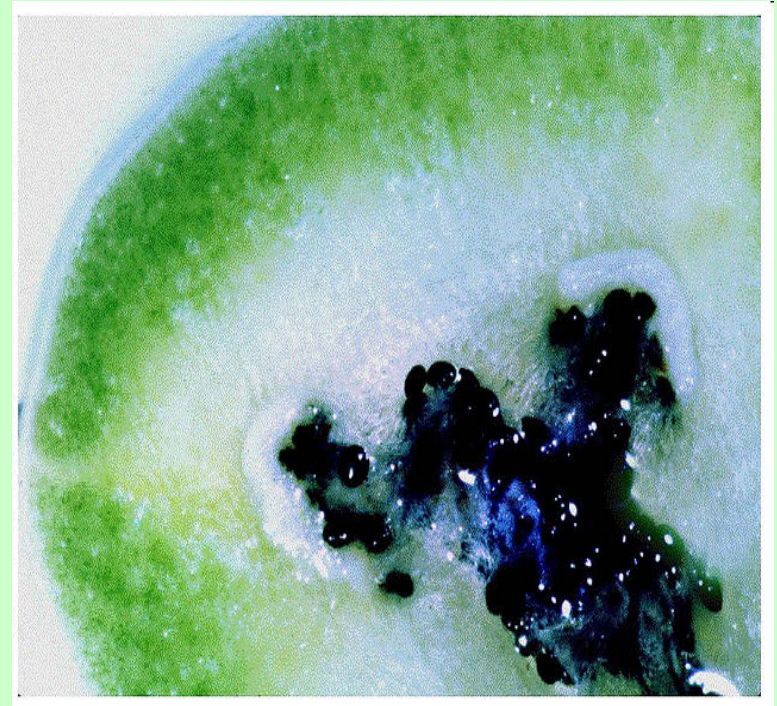




Crowdsourcing the Past with AddressingHistory

Stuart Macdonald
Project Manager
EDINA & Data Library
University of Edinburgh

stuart.macdonald@ed.ac.uk





Phase 1

**JISC-funded Community Content
project**

**6 months (April 2010 – September
2010)**

**Partner with National Library of
Scotland**

Advisory Board



To create an online crowdsourcing tool which will combine data from digitised historical Scottish Post Office Directories (PODs) with contemporaneous historical maps

Similar to Australian Historic Newspapers project provided by National Library of Australia where members of the public correct and improve OCR'd text of old newspapers
- *http://www.nla.gov.au/ndp/project_details/*



PODs offer a fine-grained spatial and temporal view on social, economic and demographic circumstances

They also provide residential names, occupations, and addresses.

Each contain 3 directories: general, street, and trades



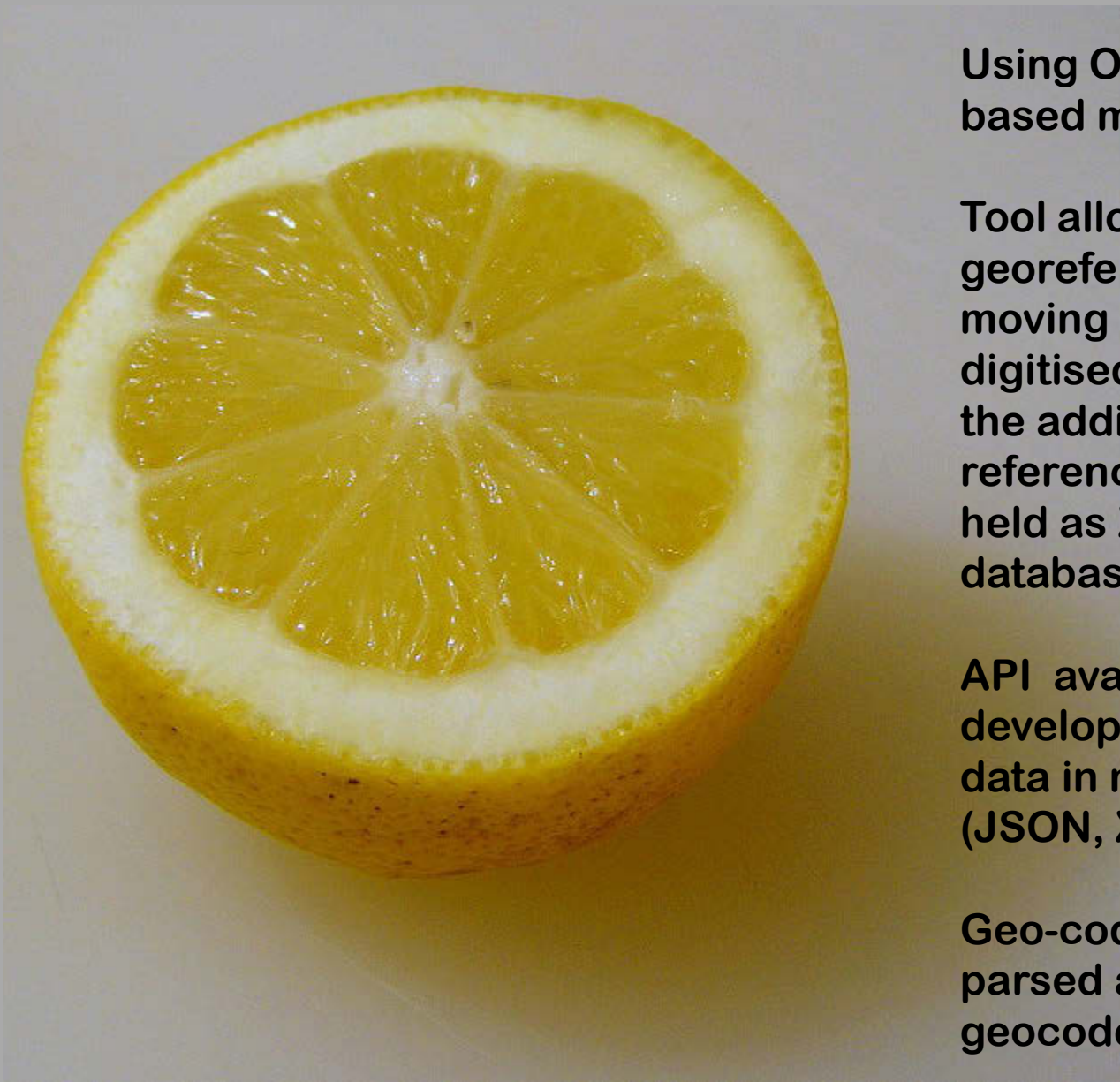
**Phase 1 focussed on 3 vols. of
Edinburgh PODs: 1784-5; 1865;
1905-6**

**Historic Scottish maps geo-
referenced by NLS**

**PODs digitised by NLS in
conjunction with the Internet
Archive**

**694 PODs (1773 to 1911) covering
28 of Scotland's towns and
counties now online**

Public domain (CC BY-NC-SA 2.5)

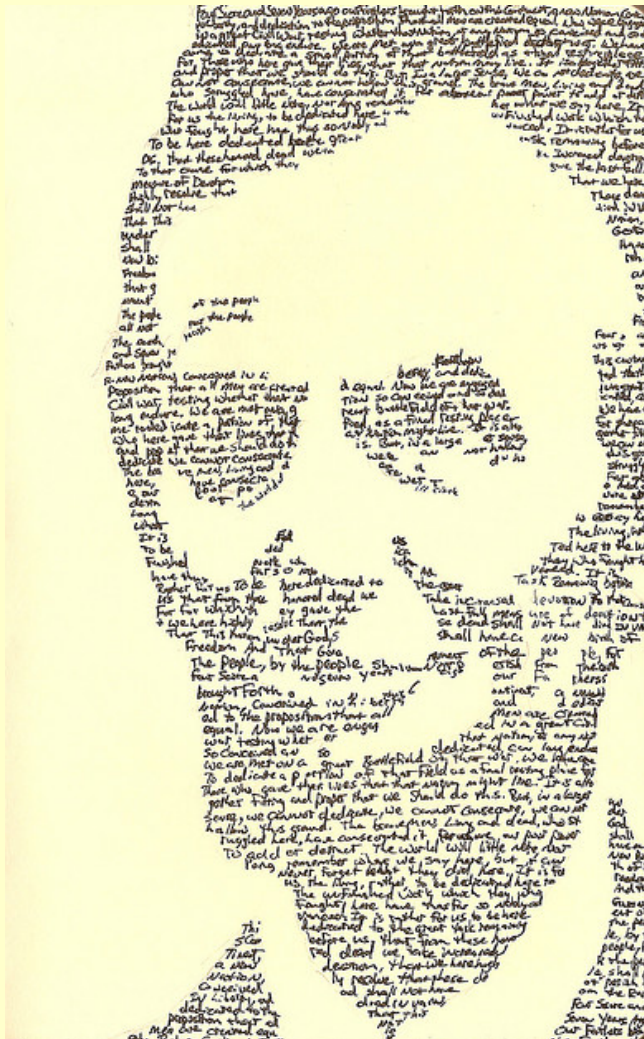


Using Open Layers as web-based mapping client

Tool allows 'the crowd' to georeference a POD entry by moving a 'map pin' on a digitised map thus facilitating the addition of an grid reference to the OCR'd POD held as XML in PostGreSQL database

API available allowing web developers access to the raw data in multiple output formats (JSON, XML, CSV)

Geo-coding of POD addresses parsed against Google geocoder



Interface had to be easy-to-use for a range of users

Robust and scalable to accommodate c.700 digitised Scottish PODs

Mechanism to check user-generated content such as geo-references, name or address edits/annotations

View original scanned directory page

Amplification of tool and API via Social Media Channels – Facebook, Twitter, Blog, Flickr, YouTube

Phase 2 sought to develop functionality to resonate with JISC's vision to build sustainable and durable deliverables and to compliment phase 1 by broadening both geographic and temporal coverage

Feb. – Sept. 2011 (EDINA Sustainability Funding)

New content (Aberdeen, Glasgow, Edinburgh for 1881 & 1891)

Re-evaluate (and enhance) parsing tool performance



Phase 2

Other additional features include:

Spatial searching (bounding box)

Associate map pin with search results

Search across multiple address

Aid searching by applying Standard Industrial Classification (SIC) codes to Professions

Augmented Reality - an AH layer has been created and published for use with the 'Layar' Application for either iPhone or Android

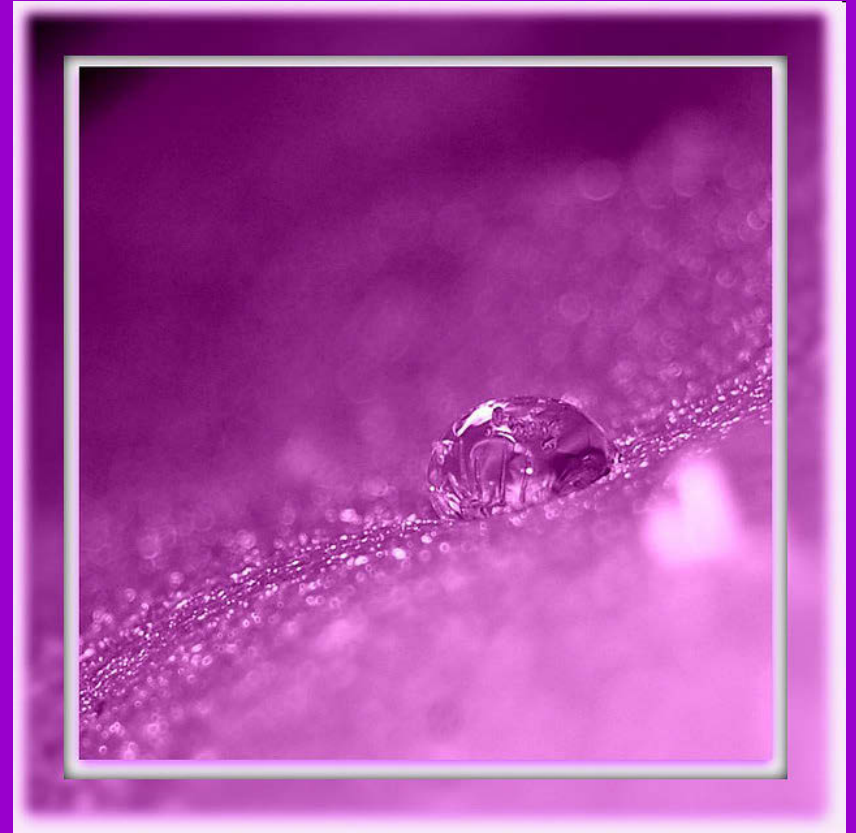
Augmented Reality Application

Using the BuildAR CMS tool an AddressingHistory layer has been created and published for use with the 'Layar' Application for a range of mobile platforms including iPhone or Android

Raw ASCII Points of Interest (POIs) and associated metadata are uploaded as a set of Google Map co-ordinates

POIs (e.g. each profession or SIC Code) have an image associated with it

The AddressingHistory layer works with the Layar App to compare information about your current location (from your phone) and the geo-referenced entries in AddressingHistory to work out which historical residents and businesses used to be located near where you are standing at that moment



Crowdsourcing on 3 levels

1. Individual record level –
georeference, address, name, occupation
2. Configuration file level -
edit and augment OCR errors / inconsistencies
to run in conjunction with parsing process for
future PODs
3. POD level -
User can request POD of interest and can be
potentially be given access to parser

(2 & 3 require modest technical
understanding and are 'policed' by
EDINA)

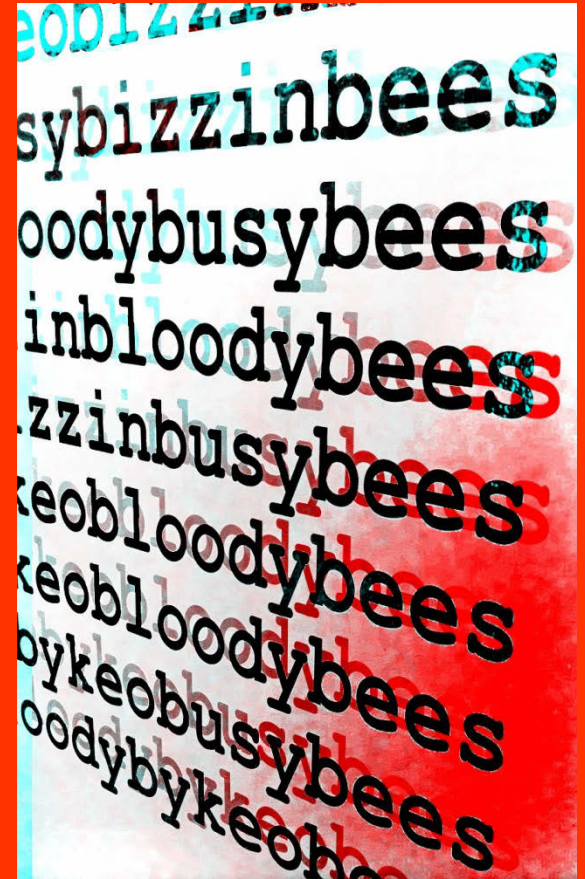
Lessons Learned

Critical mass – does geographic & temporal coverage attract and engage the crowd?

Separate out parsing from interface and back end storage - to allow any refinements to be implemented without impacting on tool and API

Externalise ‘configuration’ files – editable XML-based files that identify repeated OCR and content inconsistencies – these are run in conjunction with the POD parser to refine the parsed content hence improved searching

Parsing and refining process is almost unending - Identify what is realistically achievable with available resources and time constraints
- i.e. perform proper requirements analysis



Sustainability

Given the broad applicability of the resource a range of communities may be interested in the longer term curation of the project tools e.g. the Open Street Map community, NLS

Evaluation of possible business models for sustainability:

revenue generation via online donations

subscription model (e.g. per annum, per month, per use)

‘freemium model’ (e.g. free API download of a certain number of records with payment for further downloads)

academic advertising.



Second last slide...

Gauging the success of the project goes beyond the delivery of engaging and innovative online tools. It will be ultimately be measured by continual and extended use within the wider community.



Website:

<http://addressinghistory.edina.ac.uk/>

THANKING YOU!

Credits:

Image by aroid - <http://www.flickr.com/photos/selago/34843234/> - CC BY 2.0

Image by konqui - <http://www.flickr.com/photos/konqui/2301314089/> - CC BY-NC 2.0

Image by mosilager - <http://www.flickr.com/photos/mosilager/2260598271/> - CC BY-NC-SA 2.0

Image by racoles - <http://www.flickr.com/photos/racoles/5719938981/> - CC BY-NC 2.0

Image by James Bowe - <http://www.flickr.com/photos/jamesbowe/3351247547/> - (CC BY 2.0)

Image by yelnoc - <http://www.flickr.com/photos/yelnoc/361303918/> - CC BY-NC-SA 2.0

Image by epSos.de - <http://www.flickr.com/photos/epsos/3384297473/> - CC BY 2.0

Image by bek30 - <http://www.flickr.com/photos/bek30/6107854810/> - CC BY-NC 2.0

Image by karen horton - <http://www.flickr.com/photos/karenhorton/3261277303/> - CC BY-NC 2.0

Image by lofaesofa - <http://www.flickr.com/photos/lofaesofa/227019975/> - CC BY 2.0

Image by Psycho Delia - <http://www.flickr.com/photos/24557420@N05/5588473657/> - CC BY-NC 2.0

Image by wdj(0) - <http://www.flickr.com/photos/davidjoyner/534893725/> - CC BY-SA 2.0

Image by Symic - <http://www.flickr.com/photos/symic/2870349309/> - CC BY-SA 2.0

Image by ~milj - <http://www.flickr.com/photos/21989292@N07/4938052014/> - CC BY-NC-SA 2.0

Acknowledgements:

JISC - <http://www.jisc.ac.uk/>

NLS Geo-referenced maps and applications - <http://cpl.edina.ac.uk/>

Visualising Urban Geographies (VUG) project - <http://geonls.ukurbhist.org/>

Edinburgh City Libraries - <http://www.edinburgh.gov.uk/libraries/>