

# Creating a Definitive Place Name Gazetteer for Scotland

Bruce M. Gittings

School of GeoSciences, University of Edinburgh, Edinburgh, UK  
[bruce@ed.ac.uk](mailto:bruce@ed.ac.uk)

Place-names represent a fundamental geographical identifier, which also have considerable cultural, historical and linguistic importance. Scotland had great tradition of publishing descriptive (long-form) gazetteers in the 19th century. This tradition has continued into the present day, with the creation of the Gazetteer for Scotland from 1995 (Gittings, 2009).

The standardisation of place-names becomes crucial when comparing and correlating documents such as news articles or historical texts, and in GIS. In the UK there is no tradition of names standardisation and this has given rise to a range of different and potentially incompatible place-name gazetteers. In a world of electronic information, place-name gazetteers provide a vital method of providing a spatial footprint for datasets (Hill, 2006). They are also necessary to provide linkages between datasets and databases, exemplified by the CYGnus system, which provided the core technology for *Scotland's Places* (Gittings & Carter, 2009). This paper describes work funded by the Scottish Government to create a Definitive Place-Name Gazetteer for Scotland, which will assist in solving these very practical problems, as well as meeting the INSPIRE requirements for a place-name layer.

Despite the United Nations having resolved that member countries should establish national place-name authorities capable of standardising place-names in the form of nationally-approved gazetteers in 1967, the UK never implemented this resolution. Instead the naming of settlements remains the responsibility of individual local authorities, while the names of other geographical features are maintained by the Ordnance Survey, using what is arguably a somewhat *ad hoc* decision process.

In 2009, the author organised a workshop in conjunction with the Royal Commission on the Ancient and Historic Monuments of Scotland (RCAHMS) to examine the range of gazetteers in use in Scotland, together with a broad set of requirements. This identified a number of organisations which hold or maintain at least 15 different gazetteers that include geographical names for Scotland. The work of Barber (2010) showed that while most of these covered the same places there were important differences. The two most significant gazetteers were the Gazetteer for Scotland and the Ordnance Survey 1:50000 (OS 1:50K) product. The Gazetteer for Scotland is a descriptive gazetteer, with a modest number (22,000) of rich entries, including a textual description and rich feature-typing. The OS 1:50K gazetteer has long been Ordnance Survey's only place-name gazetteer, used as part of numerous applications.

There has been the assumption that the OS 1:50K gazetteer would be used to meet the UK's INSPIRE obligations by 2013. This suggestion became especially appealing after the OS 1:50K gazetteer became part of the free-to-use OS OpenData, meaning its use by

any and all could be encouraged as a 'standard' reference. The fact that the OS 1:50K contained by far the largest number of names (86095 for Scotland) seemed to confirm this view.

It can be argued that this process is more complex in Scotland than in other parts of the UK because Scotland exists within a rich cultural and linguistic landscape, with three languages in current use (English, Scots and Gaelic), and names drawn from others such as Norse, French, German and Anglo-Saxon. In the context of the Gaelic language, we would argue that the problems have been exacerbated by a move towards a standardised version of the language, brought about by the Scottish Qualifications Authority, which provides the school curriculum. Such a standardised version of the language loses the local peculiarities which contribute so much to place-names.

The first step in the process of providing a comprehensive gazetteer was to define a place-name. Even this is not easy. The INSPIRE definition is: "names of areas, regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest." We would certainly suggest the addition of "professional interest". Yet, if we include buildings such as Edinburgh Castle, which serve as landmarks, should we also include monuments which serve a similar purpose? If so, where do we stop? Equally if named features include bridges should they include roads? In which case, we begin to overlap with the purpose of an address gazetteer. There is also the question of historical information: of previous administrative divisions (eg. counties and parishes), or where either a feature no longer exists (eg. Hamilton Palace, demolished in 1921) or where an existing feature has a name which is no longer used (eg. Pomona for the Orkney Mainland). We have concluded that any named feature should potentially be included, but accepted the implication of this decision (ie. the list will always be incomplete). We have also identified the need for *three* fundamental gazetteers; place-names, addresses and historical names, which must be intimately linked.

In terms of meeting INSPIRE requirements, detailed examination has revealed that the OS 1:50K represents an index of the names on the map, not a gazetteer in either the traditional or modern sense. This has a series of implications for the utility of that product:

1. Spatial referencing is imprecise: names (not necessarily the features these name represent) are only located only to the centre of a 1 km grid-square
2. A lack of alternative names, including Gaelic names. There are only 25 alternative names within the 86,095 names recorded in Scotland
3. Poor feature-typing: there are only ten classes; two of these are 'other' and 'all other features', which together account for 56,060 of the 86,095 features; another of the ten feature-types is "Roman Antiquity" which has only 13 occurrences in Scotland

OS Feature Code	Number
All the others	50493
Water Features	15850
Hills	10361
Others	5573
Farms	1287
Antiquities	1188
Forests	1120
Towns	212
Roman Antiquities	13
Cities	7

4. Un-named features eg. "monument" are included
5. Uneven coverage of the country; few names are recorded within towns and cities, because there is no space for these on the Landranger map
6. There are notable errors; for example, seven cities are recorded in Scotland, when at the time there were only six
7. Occasional spelling / typographical errors
8. Repeated names eg. river names which appear more than once on the map
9. There is a one-to-one relationship between features and feature-types, for example cannot be a farm and an antiquity
10. A one-to-one relationship between a name and an administrative geography. For example, hills located on local authority (Council Area) boundaries are randomly allocated to one or other local authority, rather than being linked to both
11. Many (significant) named places not recorded in OS gazetteer
12. No historical geographical units
13. Historical / archaic toponyms not retained
14. More sophisticated relationships certainly absent (eg. 'part of')
15. The format of the OS gazetteer is far from the INSPIRE requirements (INSPIRE Data Specification on Geographical Names)

This work has addressed these issues by (1) automated methods (2) manual editing and (3) integrating the GfS and OS 1:50 to create a more 'definitive' name layer is being produced in conjunction with the Scottish Government.

## 1. Automated Methods

Automated methods were used in feature classification and duplicate detection. A glossary was used to pre-classify as many features as possible. For example, toponyms

containing word such as “river”, “abhainn” or “burn” are most likely rivers and toponyms containing the word “loch” are likely to be water bodies. There are however many exceptions, for example “Abhainn Loch Fhuaran” is a river and “Lochnagar” is both a well-known mountain, but also an adjacent corrie-lochan. Thus this classification became an iterative process, using a carefully tested succession of SQL UPDATE statements. A perl program with embedded SQL statements was used to identify similar names which were spatially ‘near’, both examining the OS gazetteer alone and making comparisons with the GfS. The “Lochnagar” example also illustrates one of the problems with this approach. Thus automated methods were used as a means of speeding manual checking rather than providing a definitive classification.

## **2. Manual Editing**

Manual editing was required both to confirm or provide a feature classification, but also to improve the spatial referencing. Considerable effort was invested in an editing tool designed to minimise the time taken to classify individual features. With more than 100,000 features to be checked (many of the GfS toponyms were checked as well as those from the OS gazetteer), if each took five minutes to handle that would mean a challenging 347 person-days of effort. Reducing this to two minutes would result in a more achievable 139 person-days. A group of student editors were trained and mass-editing sessions run over several days, with a small number of expert editors dealing with complex edits which were ‘escalated’ by the students. A total of 140 person-days were expended, comprising 98 person-days of mass-editing and the remainder dealing with the more complex interpretations or other problems. Data layers available for feature classification includes google and bing imagery, together with historical and contemporary maps (eg. OS Open Data).

## **3. Results**

A Definitive Gazetteer for Scotland was created, and this is certainly more useful than either of its components. A ten-point geofeature classification was enriched to 143; rather more than required by INSPIRE. While note has been taken of Linda Hill's ADL work, the geofeature classification was tailored to Scotland eg. lochs, the scattered nature ofcrofting settlements.

Standards had to be adopted; for example water bodies were spatially located by a point which approximated its centre while rivers were spatially located at their termination and other liner features by a random point along their length. The former gives a useful spatial reference, the latter in many cases does not. Within the more sophisticated data model of the GfS, linear features can be better located by the entities to which they relate, for example the Union Canal can be approximately defined by the settlements through which it passes.

Quality checking suggests that 95% of points were located to 100m or better, and 5% located to 20m or better. More than 90% of features are classified correctly, on the basis of the evidence available – any meaningful checking on-the-ground was clearly not possible as part of this project. An obvious problem is that some areas of Scotland are

covered by neither google nor bing imagery, other problems relate to uncertainty of interpretation – it may be difficult to determine when a farm becomes a settlement, which is based on the development of farm cottages or a steading for non-agriculturally related residential purposes.

## 4. Conclusions

The project was successful. It identified the shortcomings of the OS 1:50K gazetteer, which is being widely misused as a gazetteer despite never being designed for that purpose. With a modest budget, the project illustrated how data integration and improvement could give rise to a Definitive Gazetteer for Scotland in compliance with INSPIRE requirements.

In addition to the importance of a place-name gazetteer to meet the requirements of an INSPIRE geographical names layer, the CYGnus project (Gittings & Carter, 2009) illustrated the central importance of a good gazetteer to enable data integration, dissemination, re-use and visualisation across a range of organisations, including national and local government, academia, and the museum and library sector. On this basis it is crucial that any gazetteer service is freely accessible to any application which needs to use it.

The importance of maintenance and the need for some form of 'authority' within Scotland, which would make decisions on how these gazetteers move into the future, needs to be considered.

## References

Hill, L.L. (2006) *Georeferencing: The Geographic Associations of Information*, MIT Press, Cambridge, MA

Barber, C. 2010 *The Places of Scotland: Assessing the value of a definitive place name gazetteer to Scotland*. Unpublished MSc dissertation. University of Edinburgh.

Gittings, B.M. (2009) Reflections on Forty Years of Geographical Information in Scotland: 'Standardisation, Integration and Representation', *Scottish Geographical Journal* **125**(1) p.78-94

Gittings, B.M. and Carter, N. (2009) Getting Beyond The Silos, *Geoconnexionuk* **7**(1) p.14-16

## Acknowledgements

The considerable efforts of Panos Terzis in bringing this project to fruition are acknowledged, along with the work of 32 student editors. Funding was provided by the Scottish Government, and the author is particularly grateful to Dr. Cameron Easton in this regard.

