STUDIA GEO HISTORICA

NR 06|2018 ROCZNIK HISTORYCZNO-GEOGRAFICZNY

Instytut Historii PAN • Polskie Towarzystwo Historyczne

Contents

A note from the editors *5*

International Conference of Historical Geographers

Diogo de Carvalho Cabral
Landscape and Letterscape in Early Colonial Brazil *7*

Karen M. Morin
Prisoners and Animals: An Historical Carceral Geography *28*

Humphrey Southall
Spaces, Places, Features and Units: Web-Enabling Historical Geography *39*

Articles and theses

Agnieszka Latocha, Dominik Nowakowski
Sources and Methods Used in the Reconstruction of Settlement Networks: An Archaeological and Geographical Perspective on the Example of Silesia *55*

Marek Sobczyński
Polish Geographical Survey of Borders until 1945 .... *70*

Roman Czaja
Historical Atlas of Polish Towns. Between Source Edition and the Cartographic Presentation of Research on the History of Towns *80*

Piotr Guzowski
Economic Geography of Poland in the Period of Growth (14–16th Century) in the Polish Historical Study *90*

Dzmitry Vitsko
Historical Cartography in Belarus (Studies on the Administrative Division, Borders and Population of the Grand Duchy of Lithuania) *110*

Henryk Rutkowski
A Few Comments on the Historical Borders in Poland *124*

Bogumił Szady
Historical Cartography of Confessions and Religions in the Polish-Lithuanian Commonwealth: Between Historical Geography and Spatial History *136*

Tomasz Figlus
The Past and Present of Historical Morphology of Rural and Urban Forms in Poland *149*

Maryana Dolynska
Some Principles of Interdisciplinary Investigation
for Recreating the Historical Topography of Urban Spaces *169*

Atlas Fontium

Atlas Fontium
(Marek Słoń) *185*

Tomasz Panecki, Tomasz Związek, Grzegorz Myrda
The Spatial Database Development for Research on Settlements
and Afforestation in Nowy Tomyśl Plain in Early Modern Times *186*

Marek Słoń, Urszula Zachara-Związek (eds.)
The Court Records of Wschowa (1495–1526). Digital Edition *206*

Reviews and discussions

Deutscher Historischer Städteatlas, Nr. 5: Dortmund,
Hrsg. Peter Johanek, Jürgen Lafrenz, Thomas Tippach, Münster 2018
(Michał Słomski) *221*

* * *

Guidelines for Authors *229*

Spaces, Places, Features and Units: Web-Enabling Historical Geography

Humphrey Southall

Introduction

The last thirty years have seen a revolution in the fabric of knowledge, but many academics have spent most of their time in denial. This denial often goes beyond the changes in general practice to denying, sometimes even to themselves, how much their own practice has evolved. I am of course speaking of the creation of the world wide web, launched only in 1991[1]. Of course, we all accept changes to how we submit our research outputs, and to where we read them, but we still focus on creating traditional books and, especially, scholarly articles. We may be using new media, but through our standardisation on the PDF format we try to make our outputs resemble old paper media as much as possible.

This paper argues for a different approach, giving equal weight to new kinds of output which can exist only in digital form. Research using geographical information systems (GIS) and database technologies has created sizeable assemblies of information, intricate masses of "local knowledge". These are largely wasted if the only outputs are research papers presenting statistical summaries. Instead, we need to return to, but build upon, an older tradition in both history and historical geography of publishing annotated "editions" of original sources; "build upon" because even when it existed, the resulting books could only be found in very specialised libraries and presented single sources in isolation. Via the web, complex assemblies of multiple sources can be made accessible not only to academic researchers but also to the general public. In particular, we can make the local knowledge accessible to the people of the relevant localities.

The next section surveys the changing incentives facing British academics in recent years, and then argues that historical geography has a large potential for achieving non-academic impact and public engagement, but only if we make our detailed research accessible via the web. The remainder of the paper explores how best to do this: firstly by developing a critique of "traditional" geographical information systems-based (GIS) approaches; then by discussing how best to implement an alternative geo-semantic approach emphasising place versus space; and finally by outlining how this approach has been adopted within the Great Britain Historical GIS, and its associated web site "A Vision of Britain through Time".

Achieving impact and engagement via the web

Although similar forces have arguably affected academic life in many countries, the author can write only about the British experience. Starting in 1986, a system of periodic Research Assessment Exercises was introduced to allocate funding not for individual research projects but for permanent posts, so that institutions with good gradings would have a high ratio of staff to students, and therefore provide individual academic staff with more time for research. These gradings were decided by subject-specific panels and based mainly on a qualitative assessment of a small number of publications per member of

1 *History of the Web* (https://webfoundation.org/about/vision/history-of-the--web/, access: May 2, 2018).

staff; two publications in 1986. Procedures varied between subject panels, and as a result, a general problem that emerged was that the only public information available was the general guidance given to the panels by the funding councils, and what really mattered to individual researchers was what their heads of department thought most mattered to the panels. In practice, there can be no doubt that in geography, both physical and human, the RAEs led to a strong prioritisation of articles in refereed journals, and discouraged experimentation with new forms of digital output.

The last RAE was in 2008. The Research Excellence Framework (REF) introduced in 2014 differed primarily by including, alongside an expert review of research outputs, an assessment of the non-academic "impact" of research. Panels were instructed to "assess the »reach and significance« of impacts on the economy, society and/or culture that were underpinned by excellent research"[2], based on assessing Impact Case Studies (ICS), and this assessment contributed 20% of the overall grading. Although impact could be economic, social or cultural, economic impact was generally straightforward to demonstrate; social impact could be demonstrated by showing that research had shaped government policy; but measuring cultural impact was clearly problematic, and the Arts and Humanities Research Council launched a "Cultural Value Project" to essentially research this[3].

The Great Britain Historical GIS project's site "A Vision of Britain through Time" (VoB), further described below, was submitted to the REF geography panel in an ICS claiming cultural impact[4]. However, it was poorly rated, being judged to show "reach", via usage data similar to figure 2, but not "significance", which was defined as what the research had enabled non-academic users to do. However, the University of Hertfordshire's ICS on "Old Bailey Online"[5], another UK academic historical web site reaching a wide audience, was not dissimilar and was given the highest possible 4* rating by the history panel[6]. One lesson learned from this process is probably that it is easier to argue "cultural impact" to historians than to a broad group of human and physical geographers, but another is that we cannot simply count our audiences but need to interact with them and learn how exactly they benefit from our work.

Impact is here to stay. It has already been decided that the next REF in 2021 will increase the weighting for impact from 20 to 25 per cent[7]. Further the UK research councils now require funding proposals to include a "Pathways to Impact" statement. The author is a frequent reviewer of proposals to the UK's Arts and Humanities Research Council, and these statements often propose very substantial commitments of time and money, for example for hiring venues for a publicity roadshow, or making cartoon films with schoolchildren loosely linked to a project's literary theme. They almost always propose a project web site, but often just for publicity. Even when academics cannot demonstrate impact, universities expect "public engagement", often locally focused, and the UK research councils fund the National Co-ordinating Centre for Public Engagement[8]. Unlike impact, there are no direct financial rewards for demonstrating engagement, but

2 UK Higher Education Funding Councils, Circular REF 02.2011, *Assessment Framework and Guidance on Submissions*, Bristol 2011, p. 6.

3 Arts and Humanities Research Council, *Cultural Value Project: Open Funding Call*, 2013.

4 University of Portsmouth, *Enabling Access to Local Historical Information for Everywhere* (http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=15992, access: April 30, 2018).

5 University of Hertfordshire, *The Old Bailey Online: Democratising Access to Social History* (http://impact.ref.ac.uk/CaseStudies/Results.aspx?HEI=59, access: April 30, 2018).

6 http://results.ref.ac.uk/DownloadFile/AllResults/xlsx (access: April 30, 2018).

7 UK Higher Education Funding Councils, Circular REF 2017.01, Initial decisions on the Research Excellence Framework 2021 (http://www.ref.ac.uk/media/ref,2021/downloads/REF2017_01.pdf, access: April 30, 2018).

8 http://www.publicengagement.ac.uk/ (access: April 30, 2018).

a university's raised profile benefits recruitment, and a popular web site can be its own reward: over the last nine years, VoB has generated over €135,000 in advertising income, easily covering operating costs and contributing to staff costs.

Provided it is made accessible via the web, historical geography research can be highly engaging, as demonstrated by the GB Historical GIS project's recent GB1900 project, a collaboration with the National Libraries of Scotland and of Wales, and with over a thousand volunteers who were asked to transcribe essentially all the text appearing on a complete set of 1:10,650 maps covering Great Britain *circa* 1900, creating a data set containing approximately 2.6 m coordinates and text strings, which will be in some sense the largest specifically historical gazetteer ever constructed[9]. Partly to prepare for a future ICS, all volunteers were asked to complete an online questionnaire, and detailed telephone interviews were carried out with six of the most active volunteers, who had been spending 15 to 20 hours per week on the project. While similar surveys of volunteers on "citizen science" projects such as Galaxy Zoo show their main motivation was to feel they were contributing to science[10], GB1900 volunteers were motivated mainly by their fascination with the maps, especially pertaining to their home areas, and other areas which had special meaning to them: "I love looking at maps; I love looking at modern maps, but equally, I like looking at old maps, and sort of thinking how things have changed since then"[11].

Although this is probably the largest application of "citizen science" to British historical geography, in the 1970s the ESRC Cambridge Group for the History of Population and Social Structure made very extensive offline use of a similar network of committed volunteers to transcribe parish registers to create a definitive study of long-run population change[12], and the first full transcription of the individual returns of a British census, for 1881, was coordinated by the Genealogical Society of Utah but mainly carried out by family historians around Britain[13]. Even where the nature of the source material makes it hard for the public to contribute to the research, the detailed outputs of other large historical geography projects would be very "engaging" if made accessible by locality on the web; for example, Darby's interpretation of Domesday[14].

Geospatial versus geo-semantic approaches

It is frequently assumed that "putting historical geography on the web" must mean employing geographical information systems technology, and innumerable funding proposals promise "interactive maps", while saying very little about how this will be achieved. Including GIS functionality within web sites poses significant technical challenges, usually adding considerable complexity to the required software and, because of the geometric calculations required, imposing substantial additional computational loads on the web server. However, what follows argues not that a conventional GIS-based approach is technically difficult but that it is undesirable.

9 H.R. Southall et al., *GB1900: Engaging the Public in Very Large Scale Gazetteer Construction from the Ordnance Survey "County Series" 1:10,560 Mapping of Great Britain*, "Journal of Map & Geography Libraries", 13, 2017, p. 7–28.

10 E.g. M. Raddick et al., *Galaxy Zoo: Motivations of Citizen Scientists*, "Astronomy Education Review", 12, 2013 (http://access.portico.org/stable?au=pgg3ztfcv7h, access: April 30, 2018).

11 P. Aucott, H. Southall, C. Ekinsmyth, *Citizen Science through Old Maps: Volunteer Motivations in the GB1900 Gazetteer-building Project*, submitted to "Historical Methods" [in print].

12 E.A. Wrigley, R.S. Schofield, *The Population History of England 1541–1871*, Cambridge 1981.

13 M. Woollard, *The Classification of Occupations in the 1881 Census of England and Wales*, "History and Computing", 10, 2010, p. 17–36.

14 Not so much his summary volume as the regional volumes, such as H.C. Darby, I.S. Maxwell, *The Domesday Geography of Northern England*, Cambridge 1962, and H.C. Darby, G.R. Versey, *The Domesday Geography of England*, Cambridge 1975.

Mainstream GIS takes a **geospatial** approach which distinguishes between spatial information, meaning geographical coordinates, and attribute data. The coordinates define **features** which are recorded as points, lines or polygons, which are then assigned attributes. In other words, the spatial information provides the basic framework, to which other kinds of information are attached. Relationships between features are recorded through their locations in two or three dimensions[15]. Note that traditional GIS data models do not include a time dimension, but there is no great conceptual difficulty in adding time as a third or fourth dimension.

The alternative is a **geo-semantic** approach, using "semantics" conventionally to mean the study of named entities and their relationships. This approach applies the "knowledge organisation systems" developed by information scientists to geographical knowledge[16]. The simplest knowledge organisation system is a **word list**, or controlled vocabulary; for example, when transcribing historical records we may need to limit a particular column to containing the names of one of the counties of England. Most commonly this will simply be arranged alphabetically. The next level of complexity is a **thesaurus**, in which entities are organised hierarchically, narrower terms being instances of broader terms, and preferred terms are distinguished from synonyms. For example, in the Alexandria Digital Library's (ADL) Gazetteer Feature Type Thesaurus, *fortifications* is the preferred term which should be used instead of *castles*, *forts* and *redoubts*, and is a narrower term within *manmade features*[17].

In a traditional paper-based thesaurus, each narrower term can usually be within only one broader term, but a computerised thesaurus can be **polyhierarchic**; for example, if we construct a list of cities as a polyhierarchic thesaurus, *Warsaw* can be an instance of both Polish cities and capital cities; but the only possible relationship remains "instance of", or "is within", a limitation which is removed in **ontologies**. In philosophy, an ontology is a theory of existence, but in information science it means 'a set of concepts and categories in a subject area or domain that shows their properties and the relations between them'[18]. Rather than have preferred terms and synonyms, an ontology aims to identify the underlying entities which then have multiple names associated with them, and in an ontology there can be different types of relationships. As with a GIS, a time dimension is not inherent in the approach, but there is no great difficulty in including dates of existence for both entities and relationships.

To most geographers, the geospatial approach fits in well with how they think about the world: geographers like maps, and are very experienced in using them. However, expressing geographical knowledge semantically has three major advantages.

Firstly, we need to recognise that a geographical training warps the mind: we are not normal[19]. Normal people understand the world not as sets of coordinates, but as named places arranged more or less in hierarchies: my house, within my street, within my town, within my country. The author is very forcefully reminded of this in his walking club, within which most members rely entirely on written directions rather than maps to follow

15 P. Longley et al., *Geographical Information Systems and Science*, Chichester 2005 (chapter 3: *Representing Geography*).

16 G. Hodge, *Systems of Knowledge Organization for Digital Libraries. Beyond Traditional Authority Files.* Washington DC 2000 (http://www.clir.org/pubs/reports/pub91, access: April 30, 2018).

17 http://legacy.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver100301/ (access: April 30, 2018).

18 https://en.oxforddictionaries.com/definition/ontology (access: April 30, 2018).

19 For a study showing that economic training warps the mind, see: R. Frank, T. Gilovich, D. Regan, *Does Studying Economics Inhibit Cooperation?*, "Journal of Economic Perspectives", 7, 1993, p. 159–171.

their country walks, despite the repeated problems they then have with correcting their route after a mistake. Two academic studies draw essentially this conclusion from Australian aboriginal culture[20] and the planning of delivery routes by the U.S. postal service[21]. One very specific implication is that "geographical querying" for most people means searching for a place name, not specifying a bounding box.

Secondly, geospatial approaches work well provided we have accurate mapping, but as we go back in history this becomes more and more questionable. Take the example of English villages and the parish areas which surround them. Presently and as far back as the 1970s, their boundaries are known with certainty and precision, and available in digital form from the government; back to the 1880s, their boundaries appear on paper maps from the national mapping agency, but the only digital versions were created by the Great Britain historical GIS project and are accurate only to 100 or 200 m; before that, parishes were defined by the church rather than the government and their boundaries were often contested, partly because land owners could avoid church taxes by claiming to be "extra-parochial"; and finally, as we move back into medieval times, parishes did not necessarily have well-defined boundaries, instead ending in woods, marshes or moorland, and we sometimes find parishes which no longer exist listed in documents, which may or may not correspond with nameless abandoned villages recorded by archaeologists[22]. A further problem is that what maps existed are not topographically accurate. Faced by this uncertainty, the geospatial approach forces spurious precision on us. It is far better to use a geo-sematic approach which enables us to specify geographical relationships such as *near*, *within*, *part of*, or *south of*.

Thirdly and most practically, this paper's focus is on presenting geographical knowledge on the web, and the web is a semantic rather than a geospatial structure: it consists of named entities, meaning web pages (generally with a title), connected by hyperlinks. This has been made very explicit in the active promotion of the semantic web by Berners-Lee and the W3C[23], but applies also to the common or garden web we all use. Repeated attempts have been made to create true cyberspaces within the web, through technologies such as VRML (Virtual Reality Modelling Language) and SVG (Scalar Vector Graphics), and "walled gardens" such as *Second Life*, but these have never been widely adopted: **cyber-space** simply is not.

Normal people prefer a web consisting of clearly separate pages, explicitly linked, and search engine have enabled this web to grow to great size but remain navigable. A central problem with GIS-enabled web sites is that they are impenetrable to search engines, for two reasons. Firstly, web searches are made possible by Googlebots and similar programs, which systematically follow hyperlinks then index the text on the pages they find, but they cannot reach web pages which are accessible only by completing a form; and web GISs are essentially large forms embedded within pages. Note that this means that the content of database-driven sites is similarly inaccessible. Secondly, bots index text, not graphics, so they can do little with pages containing maps. The original funding for VoB, as described below, required adherence to detailed rules for "accessibility", which meant that

20 D. Turnbull, *Maps Are Territories, Science Is an Atlas*, Chicago 1989.

21 M.R. Curry, *Toward a Geography of a World without Maps: Lessons from Ptolemy and Postal Codes*, "Annals of the Association of American Geographers", 95, 2005, p. 680–691.

22 N.J.G. Pounds, *A History of the English Parish*, Cambridge 2000.

23 T. Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web*, "Scientific American", 284, 2001, p. 34–43.

it had to be usable by people who were blind, using not conventional web browsers but screen readers, programs which read the text on pages and tell the user what links can be followed. Meeting these rules was a difficult challenge for a large geographical site, but a site which works well with screen readers will necessarily also be easily indexed by Googlebots[24].

What kinds of entity: places, features and units

The previous section argued for a primarily textual digital representation of historical-geographical knowledge rather than computerised maps, and so, seemingly, for online systems which are essentially gazetteers. However, this section argues that mainstream digital gazetteers are not appropriate.

Although the nineteenth century saw the compilation of vast multi-volume descriptive gazetteers, several of which have been computerised as components of VoB, currently available digital gazetteers derive not from these but are, overwhelmingly, a by-product of the computerisation of topographic maps. They consequently contain just four items of information for each entry: an identifying number; a coordinate; a text string holding a name or, in the case of certain bi-lingual areas, two names; and a "feature type" derived from the map symbology. This description applies to most gazetteers available from national mapping agencies but in particular to the two very large gazetteers of the United States and of the rest of the world, created by the U.S. Board on Geographic Names and the National Geospatial Intelligence Agency[25]. Because those gazetteers are freely downloadable and in the public domain, they provide the main content of most online gazetteer datasets, including the very widely used Geonames system. Due to their widespread use, existing data standards for digital gazetteers, notably those developed by the Open Geospatial Consortium and the Alexandria Digital Library, have been designed around them[26].

Because such gazetteers derive from topographic maps, most of their **features** exist in the physical landscape, so most "types" in the ADL Feature Type Thesaurus fall within the broad terms *manmade features*, *hydrographic features* and *physiographic features*. However historians and historical geographers work primarily with written texts and deal mainly with geographical entities which do not exist in the physical landscape: they cannot be touched. These are **units**, meaning administrative areas which often also serve as statistical reporting areas, and **places**. Table 1 summarises and contrasts their key characteristics.

Administrative areas can be as large as the European Union, or cover just a single village. Besides their obvious importance to political and administrative historians, their use in statistics has made reconstructing their historical boundaries a central concern of historical GIS, and their use by historical censuses and vital registration systems has made them important for very large numbers of family historians. While boundary lines usually follow physical features, it is generally impossible to identify boundaries from the landscape, instead requiring documentary sources. They are included in standard feature typing, but only in generalised form: ADL identifies within "political areas" just "countries" and then "countries, 1st order divisions" through to "countries, 4th order divisions". This may work for the United States, divided

24 A. Walter, *Building Findable Websites: Web Standards, SEO, and beyond*, Berkeley 2008.

25 M.R. Fournier, *Standardizing Names Nationally: The Work of the U.S. Board on Geographic Names*, in: *Placing Names: Enriching and Integrating Gazetteers*, ed. M.L. Berman, R. Mostern, H.R. Southall, Indianapolis 2016, p. 163–173.

26 L. Hill, *Gazetteers and Gazetteer Services*, in: *Georeferencing: The Geographic Associations of Information*, Cambridge MA 2006, p. 91–154.

Table 1. Types of gazetteer entity

Gazetteer type	Landscape features	Administrative units	Places
Typed	yes	yes	no
Visible	yes	no	no
Defined by	inclusion on topographic maps	legal establishment as corporate bodies	shared perception; mention in texts and discourse – 'social tagging'
Defined as	points (mostly)	legally defined polygons	(mostly) fuzzy polygons

mainly into states, then counties and sometimes townships, but elsewhere many more kinds of units exist. One example of the resulting problems is that the Geonames system classifies modern British parishes as either third or fourth order units, based only on whether they are within a Unitary Authority, or within a district which is part of a county. Historically, it is clearly essential that we record the actual legal nature of each unit, rather than forcing it into a grossly over-simplified framework. The Great Britain Historical GIS, as described below, currently identifies 200 kinds of modern and historical units in a complex extensible typology.

While the geographical names in statistical reports and many other types of official documents mostly refer to units, with less formal historical documents we can usually identify which words are geographical names with certainty, but it is impossible to associate them unambiguously with either features or units. Here we need catalogues of completely unclassified "places", defined neither legally nor through physical existence but simply through some group of people agreeing that a particular name is associated with a certain locality. This is essentially the approach of the English Place Name Survey (EPNS)[27], and follows naturally from a historical approach: over long periods of time, landscape features often prove ephemeral while place names endure and are associated with a succession of different features: there is no ford in Oxford any more, while there have been many different administrative units associated with the city[28].

One continuing issue is how best to construct such spinal gazetteers, given that gazetteers such as Geonames often include a multiplicity of similarly named features in the same locality. Starting from scratch is a daunting prospect, especially given that the EPNS began in 1923 but have yet to complete many counties. In some senses, a spinal gazetteer consists simply of approximate coordinates and identifiers, as all place names come from specific dated sources and hang off the spine, so the spine is ahistorical. The PastPlace project experimented with deriving a spinal gazetteer from Wikidata, a formalisation of Wikipedia. This worked better than using Geonames, as Wikidata allows a single entity to be both a "populated place" and a capital, for example, but there were still many entities with adjacent coordinates and similar names: *Portsmouth*, *Portsmouth Cathedral*, *Portsmouth University*[29]. The next section includes the construction of a spinal gazetteer for Britain.

Case study: A Vision of Britain through Time

Thus far, this paper has advocated a general approach but given few practical details. This final section briefly describes how

27 P. Ell, L. Hughes, H.R. Southall, *Digitally Exposing the Place Names of England and Wales*, in: *Placing Names*, p. 146–162.

28 H.R. Southall, R. Mostern, M.L. Berman, *On Historical Gazetteers*, "International Journal of Humanities and Arts Computing", 5, 2011, p. 127–145.

29 H. Southall, P. Aucott, M. Stoner, *PastPlace Linked Data Historical Gazetteer*, in: *Comprehensive Geographic Information Systems: GIS for socio-economic Applications*, ed. B. Huang, K. Cao, Oxford 2017, p. 9.

the Great Britain Historical GIS (GBH GIS) was rebuilt to support the public web site "A Vision of Britain through Time", funded by the UK National Lottery, and then presents data on its "findability" and usage. A much fuller account of the system's information architecture has been published elsewhere[30].

The original GBH GIS was relatively conventional, holding digital boundary data in ArcGIS and statistics in hundreds of separate tables in an Oracle relational database. The key innovation was the addition of a time dimension, but this required large amounts of custom programming using Arc Macro Language, which is no longer supported by current versions of ArcGIS[31]. Although the main reason for moving away from this architecture was to support the web site, the time-handling worked well with 600–1,800 districts but not with 15,000 parishes; linking statistics to polygons simply using area-names increasingly required *ad hoc* "name standardisation" tables; and requiring polygons as the core framework for other information was problematic when statistical tables could be computerised in days but boundary mapping took years.

The decision was therefore taken to create a quite new core framework, designed not as a GIS but as an ontology. The absolute heart is a list of what **units** existed, each identified by a number and having a **type** as the only required attribute, but including dates of creation and abolition if available. All names are held in a separate table: every unit must have at least one **name**, but can have any number, and each name is assigned to a **language**. A third table records **status**, holding most of the finer detail of administrative geographies, and a fourth holds **relationships** between units, the commonest being "IsPartOf" but also including "AdministeredBy", "SucceededBy" and boundary changes such as "ReducedToCreate". Names, statuses and relationships can all be dated, using **date objects** which can hold anything from a precise calendar date to "Around the time of Edward II", and every row in these tables must be linked to a central table of **authorities**.

The Administrative Unit Ontology (AUO) behind the current VoB defines 87,543 units with 141,887 names, linked by 267,285 relationships, so this may well be the largest ontology of legally-defined geographical areas. Three conceptually similar systems are the U.N. Food and Agriculture Organisation's Geopolitical Ontology[32], the Ordnance Survey Linked Data ontology of modern British administrative geography[33], and a catalogue of modern and historical Finnish units[34]. The system was originally populated mainly from existing British reference works[35], cross-checked against and extended from transcriptions from statistical reports. Under the European Union QVIZ project, it was extended to identify all nation-states existing within Europe since the Congress of Vienna in 1815, with detail to parish level and below for Estonia and Sweden[36]. More recently, all

30 For the current architecture, see: H.R. Southall, *Rebuilding the Great Britain Historical GIS. Part 1: Building an Indefinitely Scalable Statistical Database*, "Historical Methods", 44, 2011, p. 149–159; idem, *Rebuilding the Great Britain Historical GIS. Part 2: A Geo-spatial Ontology of Administrative Units*, "Historical Methods", 45, 2012, p. 119–134; idem, *Rebuilding the Great Britain Historical GIS. Part 3: Integrating Qualitative Content for a Sense of Place*, "Historical Methods", 47, 2014, p. 31–44.

31 I.N. Gregory, H.R. Southall, *Putting the Past in Its Place: The Great Britain Historical GIS*, in: *Innovations in GIS 5: Selected Papers from the Fifth National Conference on GIS Research UK*, ed. S. Carver, London 1998, p. 210–221.

32 http://www.fao.org/countryprofiles/geoinfo/en/ (access: April 30, 2018).

33 http://data.ordnancesurvey.co.uk/datasets/os-linked-data (access: April 30, 2018).

34 T. Kauppinen, J. Väätäinen, E. Hyvönen, *Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal*, in: *Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, LNCS 5021*, ed. S. Bechhofer et al., Tenerife 2008, p. 110–123.

35 F. Youngs, *Guide to the Local Administrative Units of England*, vol. 1–2, London 1979–1991; M. Richards, *Welsh Administrative and Territorial Units*, Cardiff 1969; R. Cheffins, *Parliamentary Constituencies and Their Registers since 1832*, London 1998.

36 P.J. Aucott, A. von Lünen, H.R. Southall, *Exposing the History of Europe: The Creation of a Structure to Enable Time-spatial Searching of Historical Resources within a European Framework*, "OCLC Systems & Services: International digital library perspectives", 25, 2009, p. 270–286.

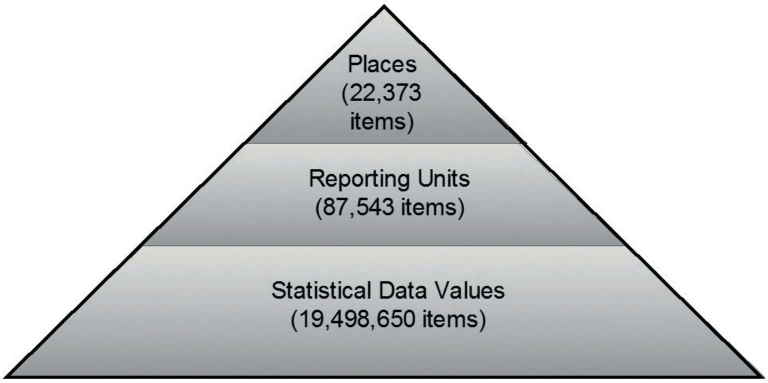

Fig. 1. GB Historical GIS conceptual overview.
Source: own elaboration

the member states of the United Nations have been added, with some information on past colonial relationships.

As figure 1 shows, the AUO is directly linked to a much larger collection of nearly 20 million statistical data values, all held in a single column of one very large table which also includes a unit identifier for every value, dates and links to a second ontology defining statistical meanings[37]. Another table holds boundary polygons if they are available, and although only 47 per cent of units have a linked polygon, 87 per cent of the data values are for these units with polygons, and most other units with data have an approximate point coordinate, enabling almost all of the statistical content to be mapped. If and when new research enables additional polygons to be added to the system, data held for the relevant units automatically becomes fully mappable.

The system also includes a second much simpler gazetteer of places, which exists for two reasons. Firstly, while the original GBH GIS was concerned purely with statistics and boundaries, National Lottery-funding covered the computerisation of four key British travel narratives by William Cobbett, Daniel Defoe, Celia Fiennes and Arthur Young, a collection which has since been greatly extended to create the largest online collection of historical British travel writing. A key goal was to make the writers' accounts of particular areas directly accessible to people interested in those areas by marking up geographical names within the narratives to include references to a gazetteer. VoB uses this mark-up to turn the names into hyperlinks leading from the narratives to pages about localities, to provide links to all the mentions by different travellers of a given locality, and to map the places mentioned in a given text. However, when Daniel Defoe visited Portsmouth, he was not visiting the ancient borough, the nineteenth century Poor Law Union or the modern Unitary Authority, so a different kind of gazetteer was needed.

Secondly, a very early version of the web site contained only the AUO, but usability testing showed that users found the large number of units named after a given settlement confusing, especially where there were several settlements with the same name. For example, the AUO contains 51 units with the name *Newport*, which include eleven units named after a market town in Shropshire, ten for the industrial city in Monmouthshire and ten for the Isle of Wight's capital. By grouping these together into "places",

37 H.R. Southall, *Visualization, Data Sharing and Metadata*, in: *Geographical Visualization: Concepts, Tools and Applications*, ed. M. Dodge, M. McDerby, M. Turner, Chichester 2008, p. 259–275.

users can first select from a set of thirteen Newports with clearly separate locations, and then select a particular unit based on what area around the place it covers, and what data is available for it. The initial set of places was generated algorithmically from the AUO, but it has since been extensively revised and extended manually, partly to improve coverage of places mentioned by travel writers. However, the general aim is to limit it to the main places where people live, and to minimize ambiguity. Instances of place--names appearing in statistical reports, descriptive gazetteers, travel writing, and on maps have been linked to this central spine, and in most cases the site provides links from places to each source, thus assembling a naming history.

This is the barest outline of the system, but how does it work in practice? As argued above, a central aim was to make the site findable via search engines, so that users interested in the history of a particular locality would be served whether or not they had heard of the site. Firstly, data from Google Analytics shows that 75% of all visitors to the web site during April 2018 arrived via a search engine, and 66% arrived specifically from a Google service. For the latter, Analytics also provides data on the search terms used, and of the top 200 search terms 89% include geographical names; perusing the top thousand, most are simply a place name, or add "history of" or the name of a county.

That shows the site is reliant on search engines, but does not show how well it performs with them. Table 2 shows the result of searching Google UK for information on the history of each of the towns and main villages of the author's home county. Few users look beyond the first page of ten results, and most pick one of the first five results, which can usually be seen without scrolling. For 105 places (56%) the first result was a Wikipedia page, and for 46 (25%) it was VoB. For 14, the first result was a local non-commercial site, generally created by either the parish council or a local history society; 3 were local commercial; and 4 national non-commercial other than the four sites listed in the table. However, the four sites listed in the table, uniquely, appear in the results for almost every place: unlike any traditional encyclopaedia, Wikipedia contains an article for almost every village; VoB defines a place for every parish; the University of London's British History Online contains the text of the systematic **Victoria County Histories**; and Family Search was created for genealogists by the Mormon church. It would be difficult for VoB to systematically out-rank Wikipedia, but the table's summary statistics show that it out-performs Wikipedia by being more complete while generally coming second when does not come first. Note that none of these four sites are commercial, and that GIS-driven sites such as Geonames are completely absent from these results.

Google Analytics also provides data on where within the site each user first arrives, their "landing page": again analysing April 2018, 962 of the top 1,000 such pages are "place pages", essentially home pages for each town or village, while the site's main home page is only the 24th ranked landing page. This is designed behaviour, reflecting the addition of an algorithmically constructed hierarchy of places, generated from locations and the maximum population held for any associated unit,[38] and the addition of "no follow" attributes to many hyperlinks within the site, so that Googlebots see it as a simpler hierarchy.

Figure 2 shows the number of visitors to the site each month, almost continuously since its launch in 2004, including a major re-launch in 2009. Steady enhancements to content arguably explain

38 For example, the top level of "places", below "Britain", are Birmingham, Bristol, Edinburgh, Glasgow, Hull, Liverpool, Leeds, London, Manchester and Sheffield; places below Manchester include Blackburn, Bolton, Burnley and Colne Valley; and so on.

*Table 2. Rankings of four popular web sites in Google search results for "history of »placename« Herefordshire"**

Rank	Wikipedia	Vision of Britain	British History Online	Family Search
1	105	46	12	2
2	20	66	28	18
3	22	38	31	30
4	10	13	40	36
5	6	10	24	33
6	2	2	11	18
7	1	5	7	12
8	0	1	1	12
9	0	1	3	3
10	3	1	2	2
Not on 1st page	17	3	27	20
% in 1st 5 links	87.6	93.0	72.6	64.0
% on 1st page:	90.9	98.4	85.5	89.2

* The 186 place names used were those of the Ancient Parishes of the county, with the exception of the three parishes within the county capital, each named after a saint, which were replaced by "Hereford". Results are included whether or not they relate to the place searched for, although most clearly do. Searches carried out on April 28, 2018.

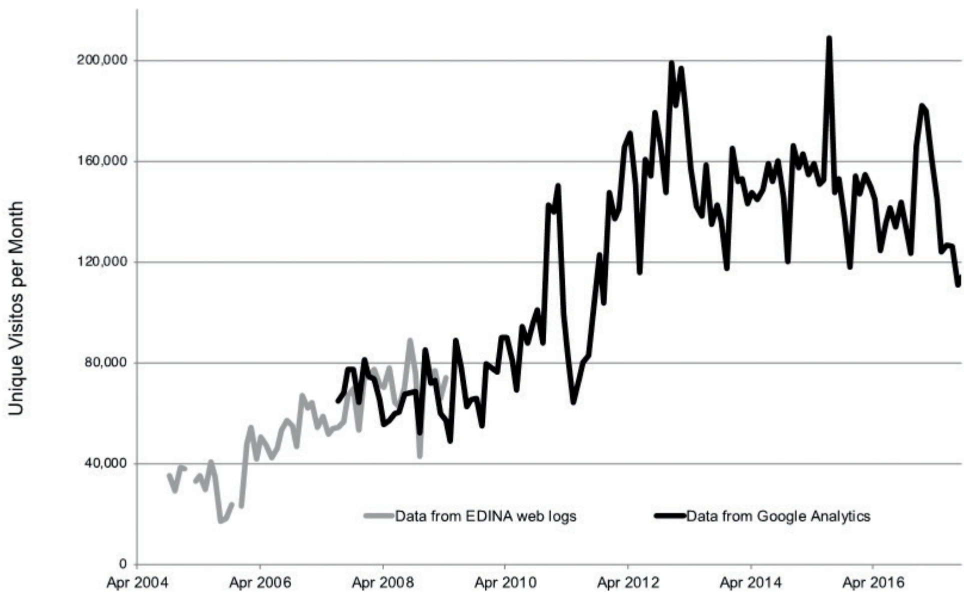

Fig. 2. Usage of "A Vision of Britain through Time", 2004–2018.
Source: own elaboration

growing use from 2004 to 2012. Since then, usage seems to be declining very slowly, possibly reflecting a need to improve the site's usability on mobile devices. Even so, the site was visited by 1,494,096 over 2017, and has roughly the same number of visitors as the main University of Portsmouth site. Note that many academic sites are unable to provide such usage data, but being funded through digital library programmes meant that VoB was required to supply it. These programmes also provided access to expert advice and specialised staff training.

If there is a problem with VoB, it is that its success as a popular web site for local historians means that its underlying data structures are taken less seriously by academic researchers. In particular, most research using historical statistics is limited by the data being divided into many separate data sets. When the data comes from censuses, data for different dates is often structured differently; for example, in the U.S. National Historical GIS almost all data sets are for single years[39]. Conversely, within VoB all statistics are in a single table, always with a date and linked to both a geo-spatial ontology and a statistical domain ontology. Although the system has been used extensively for research by the project team, especially in health geography[40], developing the potential for analysing long-term geographical change, for example by interfacing with analytic engines such as the Geographical Analysis Machine[41], requires collaborators with different expertise.

Some potential for systematic automated analysis has been created by adding applications programming interfaces, including a simple linked data gazetteer search API (applications programming interface), enabling software running elsewhere such as archival record management systems to use VoB as a geographical name authority[42], and a partial implementation of the W3C Data Cube Vocabulary as a means of systematically mining the statistical content[43]. Both these APIs follow linked data standards with the overall aim of making this large geographical resource a hub within the semantic web[44]. This work requires large scale funding before it could be relied on as a historical spatial data infrastructure, but another API, a Web Map Server accessing historical mapping, is used to provide base mapping to the *Historical Gazetteer of England's Place-Names* (www.placenames.org.uk) which accesses the incomplete English Place Names Survey. VoB is already a reliable online reference resource for humans, and one aspect of this is that ever since launch the site has had a public policy about which pages will be held constant and how they should be referenced. Considerable effort has gone into creating a system of stable and easily cited web addresses, Uniform Resource Identifiers[45], like this one for the Portsmouth place page: http://www.visionofbritain.org.uk/place/429.

39 https://www.nhgis.org/ (access: May 2, 2018).

40 H.R. Southall, *Enhancing Life-courses: Using GIS to Construct "New" Aggregate and Individual-level Data on Health and Society in Twentieth Century Britain*, in: *The Routledge Companion to Spatial History*, ed. I.N. Gregory, D. Debats, D. Lafreniere, Oxford 2018, p. 76–91; D. Phillips et al., *Evaluating the Long-term Consequences of Air Pollution in Early Life: Geographical Correlations between Coal Consumption in 1951/1952 and Current Mortality in England and Wales*, "BMJ Open", 8, 2018, p. 1–12.

41 http://www.ccg.leeds.ac.uk/software/gam/ (access: May 2, 2018).

42 H.R. Southall, P. Aucott, M. Stoner, *PastPlace*.

43 World Wide Web Consortium, *The RDF Data Cube Vocabulary*, Cambridge MA 2014 (https://www.w3.org/TR/vocab-data-cube/, access: May 2, 2018); H.R. Southall, M.J. Stoner, *Creating a Spatio-temporal "Data Feed" API for a Large and Diverse Library of Historical Statistics for Areas within Britain*, paper presented to GIS Research UK 2015 (http://leeds.gisruk.org/abstracts/GISRUK2015_submission_122.pdf, access: May 2, 2018).

44 T. Heath, C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, San Rafael CA 2011; G. Hart, C. Dolbear, *Linked Data: A Geographical Perspective*, Boca Raton FL 2013.

45 Chief Technology Officer Council, *Designing URI Sets for the UK Public Sector*, London 2010 (https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector, access: May 2, 2018).

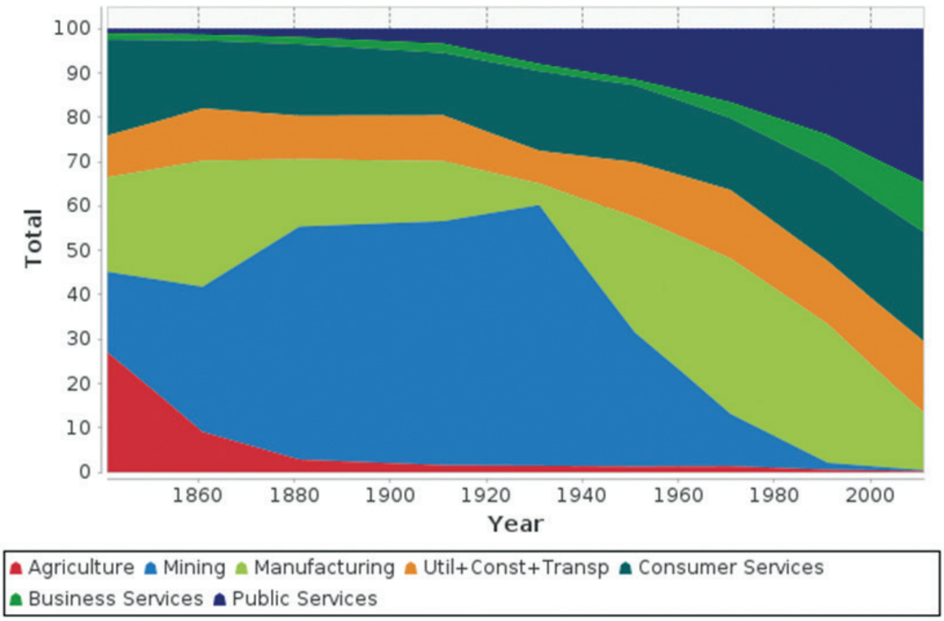

Fig. 3. From "A Vision of Britain through Time": The changing industrial structure of Rhondda Cynon Taff district, South Wales 1841–2011.
Source: own elaboration

Conclusion

This paper began by explaining how the UK Research Assessments since 1986 had initially led to an extreme focus on conventional academic journal articles, but more recently have promoted wider public impact and engagement. The web enables historical geographers to reach out to a wide audience at relatively low cost, and in particular to present numerous place histories each to the relevant locality. Both the large scale use of the "Vision of Britain" site and the commitment made by GB1900 volunteers show the degree to which the public will engage with the materials of historical geography, if presented in the right way.

This paper is mainly concerned with policy and mechanisms, but figure 3 is included to illustrate the actual place histories made available by VoB. It draws on data from nine separate British censuses to show how one particular modern local authority area in South Wales has changed over 170 years. Making data for diverse kinds of historical districts fit into modern boundaries required complex calculations, starting with the redistricting procedure, which first reassigns district-level data to the much more detailed parish-level geography pro rata to population, and then assigns the parish-level estimates to modern units by overlaying the two sets of digital boundaries. The data also had to be reassigned from a variety of different occupational and industrial classifications to a simplified version of the 2007 Standard Industrial Classification.

This has been done not just for Rhondda Cynon Taff, but for each of the 380 districts reported on by the 2011 census. Maps of the distribution of particular sectors in specific years are certainly available on VoB, and a paper analysing overall trends in industrial diversity is in preparation. However, it is local time series graphs such as this which tell the clearest story, in this case about the rise and decline of coal mining. They therefore mean most to the public; and only on the web can we publish all 380. The personal view of this author is that one of Britain's largest current problems is a collective failure to understand how, as a country, we make our living, and how this is changing. When keeping in mind this perspective, making data on historical change widely available and understandable is not only an issue of meeting official targets for "impact" or "public engagement", it is simply necessary. ■

Bibliography

Arts and Humanities Research Council, *Cultural Value Project: Open Funding Call*, 2013.

Aucott P.J., Lünen A. von, Southall H.R., *Exposing the History of Europe: The Creation of a Structure to Enable Time-spatial Searching of Historical Resources within a European Framework*, "OCLC Systems & Services: International digital library perspectives", 25, 2009, p. 270–286.

Aucott P., Southall H.R., Ekinsmyth C., *Citizen Science through Old Maps: Volunteer Motivations in the GB1900 Gazetteer-building Project*, "Historical Methods" [in print].

Berners-Lee T., Hendler J., Lassila O., *The Semantic Web*, "Scientific American", 284, 2001, p. 34–43.

Cheffins R., *Parliamentary Constituencies and their Registers since 1832*, London 1998.

Chief Technology Officer Council, *Designing URI Sets for the UK Public Sector*, London 2010 (https://www.gov.uk/government/publications/designing--uri-sets-for-the-uk-public-sector, access: May 2, 2018).

Curry M.R., *Toward a Geography of a World without Maps: Lessons from Ptolemy and Postal Codes*, "Annals of the Association of American Geographers", 95, 2005, p. 680–691.

Darby H.C., Maxwell I.S., *The Domesday Geography of Northern England*, Cambridge 1962.

Darby H.C., Versey G.R., *The Domesday Geography of England*, Cambridge 1975.

Ell P., Hughes L., Southall H.R., *Digitally Exposing the Place Names of England and Wales*, in: *Placing Names: Enriching and Integrating Gazetteers*, ed. M.L. Berman, R. Mostern, H.R. Southall, Indianapolis 2016, p. 146–162.

Fournier M.R., *Standardizing Names Nationally: The Work of the U.S. Board on Geographic Names*, in: *Placing Names: Enriching and Integrating Gazetteers*, ed. M.L. Berman, R. Mostern, H.R. Southall, Indianapolis 2016, p. 163–173.

Frank R., Gilovich T., Regan D., *Does Studying Economics Inhibit Cooperation?*, "Journal of Economic Perspectives", 7, 1993, p. 159–171.

Gregory I.N., Southall H.R., *Putting the Past in Its Place: The Great Britain Historical GIS*, in: *Innovations in GIS 5: Selected Papers from the Fifth National Conference on GIS Research UK,* ed. S. Carver, London 1998, p. 210–221.

Hart G., Dolbear C., *Linked Data: A Geographical Perspective*, Boca Raton FL 2013.

Heath T., Bizer C., *Linked Data: Evolving the Web into a Global Data Space*, San Rafael CA 2011.

Hill L., *Gazetteers and Gazetteer Services*, in: *Georeferencing: The Geographic Associations of Information*, Cambridge MA 2006, p. 91–154.

Hodge G., *Systems of Knowledge Organization for Digital Libraries. Beyond Traditional Authority Files*, Washington DC 2000 (http://www.clir.org/pubs/reports/pub91, access: April 30, 2018).

Kauppinen T., Väätäinen J., Hyvönen E., *Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal*, in: *Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, LNCS 5021*, ed. S. Bechhofer et al., Tenerife 2008, p. 110–123.

Longley P. et al., *Geographical Information Systems and Science*, Chichester 2005.

Phillips D. et al., *Evaluating the Long-term Consequences of Air Pollution in Early Life: Geographical Correlations between Coal Consumption in 1951/1952 and Current Mortality in England and Wales*, "BMJ Open", 8, 2018, p. 1–12.

Pounds N.J.G., *A History of the English Parish*, Cambridge 2000.

Raddick M. et al., *Galaxy Zoo: Motivations of Citizen Scientists*, "Astronomy Education Review", 12, 2013 (http://access.portico.org/stable?au=pgg3ztf-cv7h, access: April 30, 2018).

Richards M., *Welsh Administrative and Territorial Units*, Cardiff 1969.

Southall H.R., *Enhancing Life-courses: Using GIS to Construct "New" Aggregate and Individual-level Data on Health and Society in Twentieth Century Britain*, in: *The Routledge Companion to Spatial His-*

tory, ed. I.N. Gregory, D. Debats, D. Lafreniere, Oxford 2018, p. 76–91.

Southall H.R., *Rebuilding the Great Britain Historical GIS. Part 1: Building an Indefinitely Scalable Statistical Database*, “Historical Methods”, 44, 2011, p. 149–159.

Southall H.R., *Rebuilding the Great Britain Historical GIS. Part 2: A Geo-spatial Ontology of Administrative Units*, “Historical Methods”, 45, 2012, p. 119–134.

Southall H.R., *Rebuilding the Great Britain Historical GIS. Part 3: Integrating Qualitative Content for a Sense of Place*, “Historical Methods”, 47, 2014, p. 31–44.

Southall H.R., *Visualization, Data Sharing and Metadata*, in: *Geographical Visualization: Concepts, Tools and Applications*, ed. M. Dodge, M. McDerby, M. Turner, Chichester 2008, p. 259–275.

Southall H.R. et al., *GB1900: Engaging the Public in Very Large Scale Gazetteer Construction from the Ordnance Survey “County Series” 1:10,560 Mapping of Great Britain*, “Journal of Map & Geography Libraries”, 13, 2017, p. 7–28.

Southall H.R., Aucott P., Stoner M., *PastPlace Linked Data Historical Gazetteer*, in: *Comprehensive Geographic Information Systems: GIS for Socio-economic Applications*, ed. B. Huang, K. Cao, Oxford 2017, p. 9.

Southall H.R., Mostern R., Berman M.L., *On Historical Gazetteers*, “International Journal of Humanities and Arts Computing”, 5, 2011, p. 127–145.

Southall H.R., Stoner M.J., *Creating a Spatio-temporal “Data Feed” API for a Large and Diverse Library of Historical Statistics for Areas within Britain*, paper presented to GIS Research UK 2015 (http://leeds.gisruk.org/abstracts/GISRUK2015_submission_122.pdf, access: May 2, 2018).

Turnbull D., *Maps are Territories, Science is an Atlas*, Chicago 1989.

UK Higher Education Funding Councils, Circular REF 02.2011, *Assessment framework and guidance on submissions*, Bristol 2011, p. 6.

UK Higher Education Funding Councils, Circular REF 2017.01, Initial decisions on the Research Excellence Framework 2021 (http://www.ref.ac.uk/media/ref,2021/downloads/REF2017_01.pdf, access: April 30, 2018).

University of Hertfordshire, *The Old Bailey Online: Democratising Access to Social History* (http://impact.ref.ac.uk/CaseStudies/Results.aspx?HEI=59, access: April 30, 2018).

University of Portsmouth, *Enabling Access to Local Historical Information for Everywhere* (http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=15992, access: April 30, 2018).

Walter A., *Building Findable Websites: Web Standards, SEO, and beyond*, Berkeley 2008.

Woollard M., *The Classification of Occupations in the 1881 Census of England and Wales*, “History and Computing”, 10, 2010, p. 17–36.

World Wide Web Consortium, *The RDF Data Cube Vocabulary*, Cambridge MA 2014 (https://www.w3.org/TR/vocab-data-cube/, access: May 2, 2018).

World Wide Web Foundation, *History of the Web* (https://webfoundation.org/about/vision/history-of-the-web/, access: May 2, 2018).

Wrigley E.A., Schofield R.S., *The Population History of England 1541–1871*, Cambridge 1981.

Youngs F., *Guide to the Local Administrative Units of England*, vol. 1–2, London 1979–1991. ■

Summary

Although the UK research assessment system initially mainly prioritised traditional academic outputs, more recently it has also promoted non-academic impact and engagement, which historical geography research is well suited to achieve provided detailed local knowledge can be widely disseminated, which is only possible via the world wide web. However, this is best done not via online GIS (geographical information systems) but through geo-semantic systems, supporting web pages about named entities linked by explicit relationships. Although the resulting systems are in some senses gazetteers, they differ from most existing gazetteers by focusing not on landscape features but on administrative units defined in law, and on "places" which are defined through the study of place naming. The final section of the paper briefly describes the Great Britain historical GIS and the web site based on it, "A Vision of Britain through Time", and then presents data on how it performs in place name-based Google searches. ■

Keywords: Gazetteers; geo-semantics; Historical GIS; impact and engagement; place versus space

Humphrey Southall – Professor of Historical Geography at the University of Portsmouth, U.K.
His historical research concerns the development of the north-south divide within Britain,
measured mainly through statistics and including demographic, economic, cultural and political dimensions.
For the last 25 years, he has led the construction of the Great Britain Historical GIS
(geographical information system), and its associated web site "A Vision of Britain through Time"
(e-mail: Humphrey.Southall@port.ac.uk)