

Enriching and Integrating Gazetteers

Enriching and Integrating Gazetteers



**Edited by Merrick Lex Berman,
Ruth Mostern,
and Humphrey Southall**

PLACING NAMES

Enriching and Integrating Gazetteers

Edited by Merrick Lex Berman,
Ruth Mostern, and Humphrey Southall

Indiana University Press

Bloomington and Indianapolis

This book is a publication of

Indiana University Press
Office of Scholarly Publishing
Herman B Wells Library 350
1320 East 10th Street
Bloomington, Indiana 47405 USA

iupress.indiana.edu

© 2016 by Indiana University Press

All rights reserved

No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage and retrieval system, without permission in writing from the publisher. The Association of American University Presses' Resolution on Permissions constitutes the only exception to this prohibition.

⊗ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

Manufactured in the United States of America

Cataloging information is available from the Library of Congress.

978-0-253-02244-8 (cloth)
978-0-253-02256-1 (ebook)

1 2 3 4 5 21 20 19 18 17 16

PLACING NAMES



THE SPATIAL HUMANITIES

David J. Bodenhamer, John Corrigan, and Trevor M. Harris, editors

Imagined Landscapes: Geovisualizing Australian Spatial Narratives
Jane Stadler, Peta Mitchell, and Stephen Carleton

Deep Maps and Spatial Narratives
Edited by David J. Bodenhamer, John Corrigan, and Trevor
M. Harris

Geographies of the Holocaust
Edited by Anne Kelly Knowles, Tim Cole, and Alberto Giordano

Toward Spatial Humanities: Historical GIS and Spatial History
Edited by Ian N. Gregory and Alistair Geddes

*Troubled Geographies: A Spatial History of Religion and Society
in Ireland*
Ian N. Gregory, Niall A. Cunningham, C. D. Lloyd,
Ian G. Shuttleworth, and Paul S. Ell

Locating the Moving Image: New Approaches to Film and Place
Edited by Julia Hallam and Les Roberts

The Spatial Humanities: GIS and the Future of Humanities Scholarship
Edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris

Preface

THE EBB AND flow of social life takes place in space. People are always somewhere. Governments claim territory; commerce moves goods between places. But social life also unfolds along the axis of time. We live in the present, but the organization of places and spaces in which we live change over time; taken in aggregate the social organization of space is remarkably unstable. Places are created but they disappear. Places have names, but those names are not constant, and they are not singular. Named places have locations, but those locations do not stay fixed and the territory they encompass can expand and it can shrink.

We often act on the assumption that the situation at the current moment will last forever. In daily experience we know where we are going and we accommodate the changes in names, locations, and boundaries of the spaces in which we live. The relevance of these changes to us varies depending upon our interests. Although the tax collector pays attention to changes in territory that affect the tax base and the taxee needs to pay attention to the entity to which taxes are owed, few are aware of all the changes that are taking place in the spatial organization of their surroundings. And if one asks not only what has changed but also when it changed, personal experience and memory prove sorely inadequate. And the further into the past we go the harder it is to identify changes. This is of course, why we have authoritative reference works such as gazetteers.

The chapters in this volume are inspired by the view that contemporary gazetteers, whether produced by governments or private efforts, have given scant attention to time as an attribute of space, or to the complex historical and cultural contexts within which places have been named. Taken together they call for enriched gazetteers, consider various models, and propose paths forward. An enriched gazetteer, as this book frames it, is one that includes the best possible information on when changes in place names took place. But this is not quite so simple. Just as the representation of space depends upon scale, so does the identification of points in time. Did it happen on this day? In this year? In this century? What are the many names that a place has had over time? And there are names that refer to imagined places, or places for which the location in time and space is extremely uncertain.

Thus, the complexities of tracking historical place names goes beyond adding a "time element," such as recording sequences of names used over time for a given feature or locality. From the perspective of historical research, each instance of a historical place name has its provenance: we need to know not just

when a given name was used but by whom, and dates appear as part of the documentation of sources. Linguistic place name researchers can extract additional meanings from the earliest forms, for example revealing past landscapes by identifying names associated with woodlands, while archivists and librarians need “place name authorities,” which identify preferred forms among many variants. Accuracy in gazetteers must be equally dependent on the attestation of names, dates, and documentation of sources as it is dependent on locations.

The case for enriched and integrated gazetteers is easy for geographically minded historians to make. They want to know where events happened, the boundaries of a district a population or tax record applied to, what places were linked by roads and railways, and so on. Historical geographic information systems such as those for Great Britain, China, Germany, the United States, and a number of other national territories, have shown that such information can be systematically collected and organized.

A “historical gazetteer” has value to historians, but there is a reason to use the more inclusive term “enriched gazetteer.” This volume begins to answer the question of how agencies and private organizations can integrate change over time into the current stream of toponymic authority systems, and how multiple gazetteers can be used together. In a digital environment the focus is often on staying up to date; there may be no paper records to archive. Our present is the future’s past; without keeping track of changes that are now taking place, we lose the ability to carry out a longitudinal analysis of how our social and environmental contexts have changed. This is not a matter for historians alone. Planning cannot afford to ignore the trajectories of development. This volume, the first book-length work to address these matters, is therefore highly welcome.

Many of the chapters in this volume began as papers for a special three-day track on “Gazetteers for Space-Time System Integration” held at the 2010 Annual Meeting of the Association of American Geographers. There were further sessions at the 2011 meeting. The many participants from universities, government agencies, and industry in those sessions have informed the work here and most papers originate from those panels. As the organizer of that event, I am most gratified to see that this volume has come to fruition.

Peter K. Bol

Acknowledgments

THE EDITORS WISH to thank the Association of American Geographers (AAG), and their Executive Director Doug Richardson, as well as the Harvard Center for Geographic Analysis, and its then-Director Peter Bol, for sponsoring the two workshops at the AAG's 2010 and 2011 Annual Meetings, where first versions of most of these chapters were presented. We also wish to thank Megan Overbey and Wendy Guan for helping organize these meetings, and the many individuals who participated in them.

We also wish to thank Indiana University Press for their patience and assistance in shepherding this book through several years of writing and production. In particular, David J. Bodenhamer offered prompt and excellent advice, intellectual direction, and logistical support at every stage of the process, as did the other editors of the Spatial Humanities series, John Corrigan and Trevor M. Harris. Janice E. Frisch and Raina Polivka have been our calm and expert guides at the press. Karen Kemp reviewed the manuscript with enthusiasm and offered suggestions that considerably improved the project.

Linda Hill was the person who first and most rigorously engaged the three of us in conversations about gazetteers more than ten years ago, and her advocacy for the importance of well-structured gazetteers continues to inspire our thinking. We three editors first met one another, as well as a number of the people who are named in these acknowledgments and whose chapters appear in this book, through meetings of the Electronic Cultural Atlas Initiative in the 2000s, and we warmly acknowledge ECAI Director Lewis Lancaster and Codirector Michael Buckland for organizing an important series of conferences about the spatial humanities.

We gratefully acknowledge the support of the UK Arts and Humanities Research Council for Humphrey Southall's work on this book, as part of their Digital Transformations Big Data program (award AH/L01002X/1).

Finally, we would like to thank one another and all the authors who appear in this book for the several years of excellent and collegial collaboration and stimulating conversation that have now come to fruition.

Contents

Preface/Peter K. Bol	vii
Acknowledgments	ix
Introduction	
<i>Merrick Lex Berman, Ruth Mostern, and Humphrey Southall</i>	1

Section I: What Is a Gazetteer?

1 Gazetteers Past: Placing Names from Antiquity to the Internet <i>Ruth Mostern and Humphrey Southall</i>	15
2 Gazetteers Present: Spatial Science and Volunteered Geographical Information <i>Michael F. Goodchild</i>	26
3 Gazetteers Global: United Nations Geographical Name Standardization <i>Helen Kerfoot</i>	35
4 Gazetteers Enriched: A Conceptual Basis for Linking Gazetteers with Other Kinds of Information <i>Ryan Shaw</i>	51

Section II: Using Gazetteers in Combination

5 International Standards for Gazetteer Data Structures <i>Raj Singh</i>	67
6 Place, Period, and Setting for Linked Data Gazetteers <i>Karl Grossner, Krzysztof Janowicz, and Carsten Keßler</i>	80
7 The Pleiades Gazetteer and the Pelagios Project <i>Rainer Simon, Leif Isaksen, Elton Barker, and Pau de Soto Cañamares</i>	97
8 Historical Gazetteer System Integration: CHGIS, Regnum Francorum, and GeoNames <i>Merrick Lex Berman, Johan Åhlfeldt, and Marc Wick</i>	110

Section III: Exemplars

- 9 Building a Gazetteer for Early Modern London, 1550–1650
Janelle Jenstad 129
- 10 Digitally Exposing the Place Names of England and Wales
Paul Ell, Lorna Hughes, and Humphrey Southall 146
- 11 Standardizing Names Nationally: The Work of the US Board
on Geographic Names
Michael R. Fournier 163
- 12 The Yeosi Project: Finding a Place in Northeast Asia
through History
Youcheol Kim, Byungnam Yoon, Jonghyuk Kim, and Hyunjong Kim 174

Section IV: Doing History with Gazetteers

- 13 Mapping Religious Geographies in Chinese Muslim Society
Mark Henderson and Karl Ryavec 189
- 14 Core-Periphery Structure of the Nobi Region, Central
Japan, with Reference to the Work of G. William Skinner
Tsunetoshi Mizoguchi 200
- 15 Gazetteer GIS and the Study of Taiwan Local Society and
Its Transition
Pi-ling Pai and I-Chun Fan 217
- References 231
- List of Contributors 251
- Index 255

Introduction

Merrick Lex Berman, Ruth Mostern, and
Humphrey Southall

GAZETTEERS, SPECIALIZED INDEXES of place names, are invaluable works of reference and infrastructure for almost all spatial humanities projects. Since the 1990s, humanists, geographers, and information scientists have recognized the importance of standardizing and integrating these sources. After all, gazetteers are the basis for much of the spatial search and visualization that specialists and the public have come to take for granted. Unfortunately, most current large gazetteers are byproducts of topographic mapping that often consist simply of place names and coordinates. They do not adequately support much of the local, vernacular, historical, and cultural scholarship that the spatial humanities requires.

The purpose of this volume is to recognize and bring together new gazetteer research that better serves the needs of the spatial humanities. As the second chapter in the book explains, there was once an earlier tradition of richer, more encyclopedic gazetteers that described places as well as locating them. This book is inspired by that tradition of discursive place making as well as by contemporary approaches to digital data management that have revived the gazetteer and guided its development in recent decades. It is now possible, conceptually and technically, to reunite the two traditions. We have put this book together so that readers can explore the ways in which that work is being done.

The past fifteen years have seen a renewed interest in gazetteers that associate more extensive and descriptive content with place names, which have become a major research focus of the spatial humanities. One line of gazetteer research, largely from the perspective of information science, has focused on designing new frameworks and access mechanisms for existing gazetteer content, most of which was originally created by national mapping agencies, and is now sometimes enhanced through crowdsourcing. Meanwhile, humanists have been active in the field and the archives, researching places and place names, vernacular and historical, and designing computer systems around that content. That is the work that we are showcasing here.

In recent years, a series of international meetings brought together scientists and scholars from both gazetteer research traditions to collaborate with one

another. The largest of those meetings, cosponsored by the Association of American Geographers (AAG) and the Harvard University Center for Geographic Analysis, were held at the 2010 AAG annual meeting in Washington, DC, and the 2011 AAG annual meeting in Seattle. Together they featured over a hundred authors. This book is organized around the most substantial gazetteer research presented at those workshops, supplemented by articles from other leaders in the field whom we invited to submit papers to this volume. Although there are people doing extraordinary work conceptualizing, developing and integrating gazetteers and performing research based upon gazetteer content, little of that work has been published. As a result, it is difficult for members of the spatial humanities gazetteer community to find and cite one another, to train students, or to identify best practices. We hope that this book will fill that gap, and that it will be useful to gazetteer developers, spatial humanists, and anyone seeking an overview of this important and emerging field.

Gazetteers Defined

The discussion that follows is intended as an admiring tribute to earlier work on gazetteer standards development, which remains significant to this day. However, it is also a critical analysis that explains why we believe that the new research presented in this book, which comes from a humanistic and semantic web perspective, is such an important new direction in the field.

Most simply, a gazetteer is a list of places. As Raj Singh's chapter in this volume explains, standards organizations and librarians have long worked on digital gazetteer standards. Foundational work by the International Standards Organization defines two alternative conceptual schemas for digital gazetteers. ISO 19111 covers "spatial referencing by coordinates"¹ and ISO 19112 covers "spatial referencing by geographic identifiers."² The second of these allows for gazetteers that locate one set of named entities relative to a second set of named entities, by being "within," "between," or "adjacent" to them. Coordinates are not required, and a simple list of geographic names would meet the standard if all were contained within some territory.

The ISO defines conceptual reference models, while two other standards define precise formats for digital gazetteers. The Gazetteer Service Profile of the Open Geospatial Consortium's Web Feature Service standard, WFS(G), enables geographical features to be selected from a spatial database based on feature's names, location and type.³ The Alexandria Digital Library (ADL), in work led by Linda Hill, defined a Gazetteer Content Standard,⁴ a Gazetteer Service Protocol describing how another computer can query an ADL-compliant gazetteer over the internet,⁵ and a Feature Type Thesaurus.⁶ This foundational work brought gazetteer design into the computer era and made it possible to efficiently build

standardized gazetteers and integrate them with one another and with other digital resources.

ADL standards development has a large impact on the digital humanities research presented in this volume. The following description and discussion is organized around ADL's three core gazetteer elements. Each of these—feature name, classification, and spatial location—can launch a research agenda for enriched and humanistic gazetteer developers. By examining these elements, we aim to define what a gazetteer is, and also to introduce the challenges that inspire this book. The present volume introduces many domain-specific exemplars that diverge from the ADL standard. It also proposes the development of Linked Open Data standards as a way to link gazetteers such that related concepts can be clustered in a less rigid way than the ADL standard requires.

First, the ADL standard requires that each feature include a name. In practice, each feature also has a unique identifier, usually a number. Since the same name may refer to many different places, it is hard to be sure exactly what qualifier makes a place unique. “Boston” unqualified most probably identifies the capital of Massachusetts, but there is a “Boston” in every continent except Antarctica and in over twenty US states. “Boston, England” most probably refers to the town in Lincolnshire, but there is another in Yorkshire. Further, the same place can have many different names, distinguished partly by historical period and by language but often used interchangeably. The ADL Gazetteer Content Standard allows for the inclusion of “alternate” as well as “preferred” names, but giving primacy to one particular form is often deeply problematic. This vast ambiguity of place names greatly complicates gazetteer construction.

Second, in the ADL standard, every entry is also classified by being assigned to a type, and the ADL Feature Type Thesaurus specifies 210 preferred terms. However, the thesaurus is language specific and also culturally specific. It is defined only in English and is in practice closely based on the set of features symbolized on US Geological Survey maps, including terms such as “court houses” and “capitol buildings.” Given that almost everything has a location, and can therefore be classed as a geographical feature, any particular feature-type thesaurus can be criticized for being incomplete or lacking detail. However, and especially when working with historical events rather than landscape features, the priority may be to bring together the different associations of the same named location, not carefully separate them. In particular, a fixed typology leads almost inevitably to a simplistic treatment of administrative units. For example, the ADL system starts with “countries, 1st order divisions,” and goes down to “4th order divisions.” It adds “statistical areas” to cover almost everything else. This does not effectively reflect the actual political geography of any nation or empire.

Third, every ADL entry must include a spatial location. One nonissue is how to represent these: any global gazetteer, or gazetteers for continental states like

the United States, will usually hold locations using the World Geodetic System or WGS 84, meaning decimal degrees of latitude and longitude; many smaller countries define their own local coordinate system based on a local “flattening” of the globe, such as Britain’s OSGB national grid; but converting between these is very well understood. Less clear is whether to represent everything as points; the ADL content standard allows for entries to be alternatively represented by lines or polygons. However, historical and cultural gazetteers may include many names that certainly refer to places, but ones for which we have only very limited information about location. The most obvious examples are semi- or wholly mythic places such as Avalon, Mount Meru, or Eldorado, but historical sources such as taxation lists contain plenty of prosaic examples as well.

Behind these issues, and a key question for many of the chapters which follow, is the issue of space versus place. Cultural geographer Yi-fu Tuan, who has spent his career explaining how the concept of place emerges from lived experience, has famously asserted that the number of places in the world, that is the locations that someone cares enough about to describe or name, is potentially infinite.⁷ The enriched gazetteers that this volume propounds are intended to more effectively model the kind of cultural, vernacular and perspectival idea of place that Tuan alludes to than current gazetteers are able to do.

Because existing modern gazetteers derive from topographic mapping, they are grounded in a geospatial approach that prioritizes physical features which exist in the landscape, and whose locations and types are often far more significant than their names. By contrast, humanities researchers and cultural heritage workers commonly start, not with the landscape but with texts, and consequently with places that are largely defined through their names and associated discursive and cultural information about them; places that change over time, are contested among many constituencies, and rarely have one authoritative descriptor. This calls for a very different approach to gazetteer building than the ADL approach propounds, one that is geosemantic rather than geospatial, and one focused on systematically identifying all possible named places. Over time, many events will happen in any locale of much significance, and various features of differing ephemerality will be named after it, but the place itself cannot be fundamentally associated with any one feature type, and its location, if known at all, will be hazy.

Gazetteers Enriched

So, the purpose of this book is to make the case for enriching gazetteers beyond the ADL framework, and especially to do so from the perspective of historical and cultural scholarship, global comparison, and spatial humanities agendas. The chapters we have assembled specify at least four directions for enriching

gazetteers. Each one allows place name lists to be integrated with other kinds of information. The first is to identify and link gazetteers to one another and to other resources that are themselves rich in place names. The second is to provide a framework for integrating gazetteers with information from specific domains, such as historical statistics or tables of historical events. The third is to add a time dimension to gazetteers through both dates and named time periods. The fourth is to open gazetteers to user comment and annotation. Through such extensions the gazetteer itself becomes more clearly a work of scholarship. All of these directions suggest a gazetteer development agenda that is quite different from that specified by ADL, and which raises new challenges of design and infrastructure.

First, the ADL and other gazetteer data standards described here do not meet the usual expectations of scholars because no information is included on sources. Each entry is an unsupported assertion that a certain name is associated with a particular location. This may be unproblematic when the gazetteer records current names and is authored by a national mapping agency, but poses clear problems with crowdsourced gazetteers, most of whose contributions come inevitably from visitors to areas rather than residents, and is nonsense once we start assembling variant names from diverse sources. As Michael Fournier from the US Board on Geographic Names explains in his chapter in this volume, even official agencies work with vast quantities of documentation when they assign place names.

One example of a gazetteer-based information system that is enriched according to the principles we are laying out here is the Pelagios project, described in chapter 7. It focuses on gathering name attestations, each of which links a place, identified by the web address of its entry in an existing online gazetteer, with a geospatial document, meaning an old map, historical gazetteer, or itinerary in which the place is named. The source is also identified by a web address, and where possible this should link, not to a catalogue or bibliography entry, but rather to a digital copy of the source itself. Here the potential is to create, not just a gazetteer with variant names, but an integrated corpus of geographical knowledge—a specialized kind of knowledge organization system, as Ryan Shaw puts it in his chapter in this volume, such that places in the online gazetteer directly hyperlink to textual and cartographic representations of those places, making the gazetteer a name index to the map collections.⁸

Similarly, the Great Britain Historical Geographic Information System (GIS) and its web site, *A Vision of Britain through Time*, incorporates a large collection of historical travel writing with a two-way relationship with the gazetteer: variant place names have been harvested from the writers, but the gazetteer's place pages then provide access to place-references within the writing.⁹ Systematically linking to source definitions incorporates dates of publication or attestation into

the gazetteer, creating a naming history. This approach is described in detail in a number of the chapters in this book.

Second, the gazetteer can become an organizing framework to which a wide variety of other information is linked. Some of the clearest potentials are with statistics, especially historical statistics: census reports, taxation listings and many other sources consist mainly of columns of numbers but usually preceded by what are clearly geographical names. However, these are usually not the “places” that appear on topographic maps and in travel writing but some kind of *area*, sometimes defined for statistical reporting but often with some administrative function.

With modern data we would expect to work, not with a gazetteer, but with a Geographical Information System, and the relevant digital boundaries can be downloaded from national mapping agencies or statistical offices. However, as we move deeper into the past boundary data become progressively more problematic: more than forty or fifty years back, or more recently in less developed countries, “born digital” boundaries will not exist and the researcher will need to digitize paper maps; more than a hundred to two hundred years back, there will be no systematic paper boundary maps; going further back, settlements did not necessarily adjoin, instead trailing off into uninhabited areas, so boundaries were not necessarily defined at all; and we will encounter some areas, possibly associated with abandoned settlements, which cannot be located at all.

Somewhere in this progression even a basic gazetteer of named places linked to point coordinates has a role, and even where boundary maps exist, the cost in money or the researcher’s time may make a quicker and simpler approach preferable: a current small project to map intra-urban data from the British 1961 census will be partly gazetteer-based for cost reasons. Further, the previous section described only a minimal implementation of the ADL Gazetteer Content Standard, and a fuller implementation can include, not just a point location, but also dated sequences of boundary polygons and explicit hierarchic relationships, although as noted previously the ADL Feature Type Thesaurus does not enable us to distinguish between the multiplicity of kinds of “district” existing at roughly the same spatial scale but in different periods and with different functions. Of course, once we add a large number of polygons to a “gazetteer” it starts to become a GIS, and we can hold it only in a database with object extensions.

Although situations where we cannot locate statistical data more precisely than with a point coordinate taken from a conventional gazetteer, we should generally avoid directly linking the data to the place record: almost always, the data actually relate to a currently unknown polygon, which further research may be able to reconstruct, or possibly simulate. In most cases, we have some knowledge of the kind of area a dataset covers: settlements plus some wide hinterland or “market catchment”; settlements plus the areas the inhabitants farm;

just the built-up area; or, where the data relate to specific corporate bodies or are essentially decreed by particular corporate bodies, particular buildings. It would be very poor scholarship to associate all these different scenarios simply with “places.”

The Great Britain Historical GIS is the clearest example of indirectly linking statistical data to a gazetteer. The highest level consists of about 20,000 “places,” untyped but mostly towns and villages, and all with point coordinates. About 75,000 “units” exist below the places, with a very detailed and extensible typology; these can have detailed boundary polygons, but are required to have only a hierarchic relationship, enabling them to be added to the system based only on their inclusion in a statistical listing. Finally, at the base, each of c. 14m statistical data values are linked to the units. The resulting system has all the usual capabilities of a GIS for twentieth-century data, but can also hold data from the 1334 Lay Substdy for *Vills* located only as points, or not at all. It also includes *Urban Labour Markets* as a unit type, holding nineteenth and twentieth century unemployment and wage rate data which are fundamentally point-based.¹⁰ Although at heart a gazetteer, the resulting system describes places as well as locating them, and one key advantage over a conventional GIS is its ability to present statistics as time series as well as map them. This is of course a single monolithic computer system, but again the larger potential is for gazetteers to serve as knowledge hubs for data distributed over the internet—an infinite knowledge space.

An enriched gazetteer must go farther than modeling just modern places and their historical names. In addition it must explicitly represent change over time. Thus, one research and development challenge is to incorporate dates and time periods from many modern and historical global systems. Consequent issues include the standardization of dates derived from different calendar systems, the common usage of named time periods in historical sources, and the problem of temporal scale. All of these historical or temporal modeling challenges are unavoidable in the construction of historical gazetteers, and here we survey cutting edge implementations.

There are, for example, several possible methods for standardizing the dates being recorded in the world’s many historical calendar systems, but very few have actually been implemented. One successful case integrated the calendars of China, Japan, and Korea by harmonizing those related but unsynchronized calendar systems based on Julian Day Numbers.¹¹ The consistency of the Julian Day Numbers enables the indexing of any cyclical binary day from the Chinese calendar, as well as looking up any corresponding day in the Japanese or Korean calendars. Mapping particular cyclical binary days to Julian Day Numbers (JDN) also provides a method for defining the time spans of named time periods. We could, theoretically, make use of the same mapping to Julian Day Numbers to construct a timeline for all the various calendars in human history, but this