The cover image is of a torso of Pothos (Roman 1st century BC – 1st century AD) in the Museu Calouste Gulbenkian, Lisbon, Portugal. Photo © Simon Mahony 2013. All rights reserved.
This volume is dedicated to the memory of two people whose untimely death marks a great loss, both personally and to our communities.

Elaine Matthews (died 26 June 2011): one of our esteemed contributors, ambassador and advocate of the Digital Humanities and the place there for Classics, thank you for all your many contributions to scholarship, to this volume, and your generous words on the cover of the earlier *Digital Classicist* (Ashgate 2010) volume.

Gerhard Brey (1954-2012): a valued friend, colleague, and collaborator with whom we shared intellectual ideas as well as coffee and biscuits. Gerhard was always willing to seek out new areas of ‘interest’ and so could be willingly called upon to review chapters in this and the earlier Ashgate volume.
# TABLE OF CONTENTS

Acknowledgements ix  
Abstracts xi  
Abbreviations xv  
Introduction 1  

## Modelling

Andrew Bevan  
*Travel and interaction in the Greek and Roman world. A review of some computational modelling approaches* 3  

Vince Gaffney, Phil Murgatroyd, Bart Craenen, and Georgios Theodoropoulos  
*‘Only individuals’: moving the Byzantine army to Manzikert* 25  

## Texts

Elton Barker, Leif Isaksen, Nick Rabinowitz, Stefan Bouzarovski, and Chris Pelling  
*On using digital resources for the study of an ancient text: the case of Herodotus’ Histories* 45  

Marco Büchler, Annette Geßner, Monica Berti, and Thomas Eckart  
*Measuring the influence of a work by text re-use* 63  

Tobias Blanke, Mark Hedges, and Shrija Rajbhandari  
*Towards a virtual data centre for Classics* 81  

Ryan Baumann  
*The Son of Suda On-line* 91  

## Infrastructure

Elaine Matthews and Sebastian Rahtz  
*The Lexicon of Greek Personal Names and classical web services* 107  

Simon Mahony:  
*HumSlides on Flickr: using an online community platform to host and enhance an image collection* 125  

Valentina Ascuietti and Stuart Dunn  
*Connecting the Classics: a case study of Collective Intelligence in Classical Studies* 147  

Index 161
ACKNOWLEDGEMENTS

The editors would like to thank the Institute for Classical Studies, and specially the Deputy Director and Administrator Olga Krzyszkowska, for their continued support and generosity in hosting and supporting the Digital Classicist seminars. Thanks are also due to all the members of our community who have presented papers at our seminars and conference panels as well as all those who have come along to listen and support these events.

We are grateful to the following scholars for comments, criticism, and advice on individual chapters in this volume; it is through their input that we are able to ensure the high quality of the final work: Chris Blackwell; Gabriel Bodard; John Bodel; Kalina Bontcheva; Gerhard Brey; Hugh Cayless; Graeme Earl; Michael Fulford; Sebastian Heath; Tim Hill; Kathryn Piquette; Dot Porter; Julian Richards; Matteo Romanello; Charlotte Roueché; Melissa Terras; Notis Toufexis; Charlotte Tupman; Michelle Wienhold.
ABSTRACTS

Andrew Bevan  
*Travel and interaction in the Greek and Roman World. A review of some computational modelling approaches*  
pp. 3-24

Inferring dynamic past behaviours from the static archaeological record is always a challenge, but computational and quantitative techniques can be helpful. In particular, they can provide useful insight on patterns of movement and interaction, by better characterising existing archaeological evidence, suggesting simple models of mobile decision-making or proposing expected patterns against which the observed record can be compared. This paper reviews the range of modelling options now available for understanding the movement and interaction behind the archaeological and historical record. There are increasing opportunities not only to pick and choose between different modelling approaches, but also to integrate them in a more theoretically and practically satisfactory way.

Vince Gaffney, Phil Murgatroyd, Bart Craenen, and Georgios Theodoropoulos  
*'Only individuals': moving the Byzantine army to Manzikert*  
pp. 25-43

Traditionally, history has frequently emphasized the role of the ‘Great Man or Woman’, who may achieve greatness, or notoriety, through the consequences of their decisions. More problematic is the historical treatment of the mass of the population. Agent-based modelling is a computer simulation technique that can not only help identify key interactions that contribute to large scale patterns but also add detail to our understanding of the effects of all contributors to a system, not just those at the top. The Medieval Warfare on the Grid project has been using agent-based models to examine the march of the Byzantine army across Anatolia to Manzikert in AD 1071. This article describes the movement model used to simulate the army and the historical sources on which it was based. It also explains why novel route planning algorithms were required in order to surmount problems with standard solutions.

Elton Barker, Leif Isaksen, Nick Rabinowitz, Stefan Bouzarovski, and Chris Pelling  
*On using digital resources for the study of an ancient text: the case of Herodotus’s ‘Histories’*  
pp. 45-62

Involving the collaboration of researchers from Classics, Geography, and Archaeological Computing, and supported by funding from the AHRC, Hestia aims to enrich contemporary discussions of space by developing an innovative methodology for the study of an ancient narrative, Herodotus’s *Histories*. Using the latest digital technology in combination with close textual study, we investigate the geographical concepts through which Herodotus describes the conflict between Greeks and Persians. Our findings nuance the customary
topographical vision of an east versus west polarity by drawing attention to the topological
network culture that criss-crosses the two, and develop the means of bringing that world to a
mass audience via the internet.

In this chapter we discuss three main digital aspects to the project: the data capture of
place-names in Herodotus; their visualization and dissemination using the web-mapping
technologies of GIS, Google Earth, and Timemap; and the interrogation of the relationships
that Herodotus draws between different geographical concepts using the digital resources at
our disposal. Our concern will be to set out in some detail the digital basis to our
methodology and the technologies that we have been exploiting, as well as the problems that
we have encountered, in the hope of contributing not only to a more complex picture of
space in Herodotus but also to a basis for future digital projects across the Humanities that
spatially visualize large text-based corpora. With this in mind we end with a brief discussion
of some of the ways in which this study is being developed, with assistance from research
grants from the Google Digital Humanities Awards Program and JISC.

Marco Büchler, Annette Geßner, Monica Berti, and Thomas Eckart
Measuring the Influence of a Work by Text Re-Use
pp. 63-79

Over the centuries an incredible amount of ancient Greek texts have been written. Some of
these texts still exist today whereas other works are lost or are available only as fragments.
Without considering intentional destruction, one major question remains: why did some
texts remain and others get lost? The aim of this chapter is to investigate this topic by trying
to determine the influence of certain ancient Greek works through detecting text re-use of
these works. Text re-use measures if and how an author quotes other authors and in this
chapter we differentiate between re-use coverage and re-use temperature.

Tobias Blanke, Mark Hedges, and Shrija Rajbhandari
Towards a virtual data centre for Classics
pp. 81-90

A wide variety of digital resources have been created by researchers in the Classics. These
tend to focus on specific topics that reflect the interests of their creators; nevertheless they
are of utility for a much broader range of research, and would be more so if they could be
linked up in a way that allowed them to be explored as a single data landscape. However,
while the resources may be reusable, the variety of data representations and formats used
militates against such an integrated view. We describe two case studies that address this
issue of interoperability by creating virtual resources that are independent of the underlying
data structures and storage systems, thus allowing heterogeneous resources to be treated in a
common fashion while respecting the integrity of the existing data representations.

Ryan Baumann
The ‘Son of Suda On-line’
pp. 91-106

The Son of Suda On-Line (SoSOL) represents the first steps towards a collaborative,
editorially-controlled, online editor for the Duke Databank of Documentary Papyri (DDbDP).
Funded by the Andrew W. Mellon Foundation’s Integrating Digital Papyrology Phase 2
(IDP2), SoSOL provides a strongly version-controlled front-end for editing and reviewing papyrological texts marked up in EpiDoc XML.

Elaine Matthews and Sebastian Rahtz

*The Lexicon of Greek Personal Names and classical web services* pp. 107-24

This chapter documents the data resources of the long-term classical research project, *The Lexicon of Greek Personal Names* (LGPN), published in six volumes since 1987. It explains and demonstrates the web interfaces and services which now make available online the bulk of the LGPN, providing both powerful searching tools for scholars and an interface to allow other systems to link to LGPN data. Making the data available online provides direct, unmediated access to the material and supports exploitation of the data for further research both individual and collaborative.

We describe the work that went into creating the Lexicon, detail the granularity of the data structures, and explain the history of the project's record management. We then move onto the work undertaken in recent years to provide an archival XML-based format for the Lexicon's long-term preservation, and show how this has allowed us to build new web services, including exposure of Resource Description Framework (RDF) metadata, using the ontology of the CIDOC Conceptual Reference Model (CRM) ontology for semantic web applications.¹

Simon Mahony:  *HumSlides on Flickr: using an online community platform to host and enhance an image collection* 125-46

Moving a teaching and research image collection from an analogue to a digital medium for delivery brings with it many advantages but at the same time it also presents many new problems and ones probably not previously considered. This chapter discusses the move of a departmental slide collection, firstly to a proprietary in-house format, and then subsequently to the online community platform Flickr. It draws on the experience and model of the Library of Congress in partnership with Flickr and *The Commons*, as well as initiatives at Oxford and at New York University, and in doing so critically analyses and evaluates the possibilities for the future development of this collection. It asks why this collection is not currently being used to its potential and examines how the development of a user community would help to enrich the collection and ensure long term sustainability and future growth.

¹ CIDOC CRM is an ISO standard (21127:2006) that ‘provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation’ <http://www.cidoc-crm.org/>.
Valentina Asciutti and Stuart Dunn

*Connecting the Classics: a case study of Collective Intelligence in Classical Studies* pp. 147-60

One of the great potentials of the internet is its capacity to aggregate and unify information from diverse sources. Information in the Classics, and data generated by classicists, is inherently fragmented, and organized according to different standards. This paper describes a project at King’s College London which sought to provide a set of aggregating services to humanities scholars. www.arts-humanities.net provides a platform, a library, and a taxonomy to organize and present data: we describe its facilities for supporting a multi-source dataset tracing the paths of Romano-British inscriptions, both in space and conceptually. Itinerant geographies of metrical versus text inscriptions are discussed, including how these can be published in a variety of non-conventional platforms, such as Twitter. We argue that, in the future, these platforms will come to play a critical role in the wider scholarly discourse of the Classics.
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABM</td>
<td>Agent-based modelling</td>
</tr>
<tr>
<td>ADS</td>
<td>Archaeology Data Service</td>
</tr>
<tr>
<td>AHRC</td>
<td>Arts and Humanities Research Council</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>APIS</td>
<td>Advanced Papyrological Information System</td>
</tr>
<tr>
<td>AWIB</td>
<td>Ancient World Image Bank</td>
</tr>
<tr>
<td>CC</td>
<td>Creative Commons</td>
</tr>
<tr>
<td>CI</td>
<td>Collective Intelligence</td>
</tr>
<tr>
<td>CIDOC</td>
<td>International Council of Museums</td>
</tr>
<tr>
<td>CRM-CIDOC</td>
<td>CIDOC Conceptual Reference Model</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-separated data fields</td>
</tr>
<tr>
<td>DANS</td>
<td>Data Archiving and Networked Services</td>
</tr>
<tr>
<td>DARIAH</td>
<td>Digital Research Infrastructure for the Arts and Humanities</td>
</tr>
<tr>
<td>DDbpDP</td>
<td>Duke Databank of Documentary Papyri</td>
</tr>
<tr>
<td>DPI</td>
<td>Dots per inch</td>
</tr>
<tr>
<td>DVCS</td>
<td>Distributed Version Control Systems</td>
</tr>
<tr>
<td>EDM</td>
<td>Europeana Data Model</td>
</tr>
<tr>
<td>GAP</td>
<td>Google Ancient Places</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographical Information Systems</td>
</tr>
<tr>
<td>HEA</td>
<td>Higher Education Academy</td>
</tr>
<tr>
<td>HEFCE</td>
<td>Higher Education Funding Council for England</td>
</tr>
<tr>
<td>HESTIA</td>
<td>Herodotus Encoded Space-Text-Image Archive</td>
</tr>
<tr>
<td>HGV</td>
<td><em>Heidelber</em> <em>Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens</em></td>
</tr>
<tr>
<td>IAph</td>
<td><em>Inscriptions of Aphrodisias</em></td>
</tr>
<tr>
<td>IDP</td>
<td>Integrating Digital Papyrology</td>
</tr>
<tr>
<td>ISAW</td>
<td>Institute for the Study of the Ancient World</td>
</tr>
<tr>
<td>JDI</td>
<td>Image Digitization Initiative</td>
</tr>
<tr>
<td>JISC</td>
<td>Joint Information Systems Committee</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>KML</td>
<td>Keyhole Markup Language</td>
</tr>
<tr>
<td>LaQuAT</td>
<td>Linking and Querying Ancient Texts</td>
</tr>
<tr>
<td>LCCW</td>
<td>Longest Common Consecutive Words</td>
</tr>
<tr>
<td>LGPN</td>
<td><em>The Lexicon of Greek Personal Names</em></td>
</tr>
<tr>
<td>LoC</td>
<td>The Library of Congress</td>
</tr>
<tr>
<td>MDID</td>
<td>Madison Digital Image Database</td>
</tr>
<tr>
<td>OAI-ORE</td>
<td>Open Archives Initiative Object Reuse and Exchange</td>
</tr>
<tr>
<td>OER</td>
<td>Open Education Resources</td>
</tr>
<tr>
<td>OGSAG-DAI</td>
<td>Open Grid Service Architecture–Data Access and Integration</td>
</tr>
<tr>
<td>OGSAG-DQP</td>
<td>Distributed Query Processing</td>
</tr>
<tr>
<td>PDF</td>
<td>Portable Document Format</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>PN</td>
<td>Papyrological Navigator</td>
</tr>
<tr>
<td>PRM</td>
<td>Probabilistic Road Map</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>RIB</td>
<td>Roman Inscriptions of Britain</td>
</tr>
<tr>
<td>SGML</td>
<td>Standard Generalized Markup Language</td>
</tr>
<tr>
<td>SOL</td>
<td>Suda On-Line</td>
</tr>
<tr>
<td>SoSOL</td>
<td>Son of Suda On-Line</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>SVG</td>
<td>Scalable Vector Graphic</td>
</tr>
<tr>
<td>TEI</td>
<td>Text Encoding Initiative</td>
</tr>
<tr>
<td>TLG</td>
<td>Thesaurus Linguae Graecae</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifiers</td>
</tr>
<tr>
<td>V&amp;A</td>
<td>Victoria and Albert Museum</td>
</tr>
<tr>
<td>VRE</td>
<td>Virtual Research Environment</td>
</tr>
<tr>
<td>WFS</td>
<td>Web Feature Service</td>
</tr>
<tr>
<td>WMS</td>
<td>Web Map Service</td>
</tr>
<tr>
<td>WYSIWYG</td>
<td>What-You-See-Is-What-You-Get</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
THE LEXICON OF GREEK PERSONAL NAMES
AND CLASSICAL WEB SERVICES

ELAINE MATTHEWS1 AND SEBASTIAN RAHTZ

Introduction

The *Lexicon of Greek Personal Names* was established in 1972 as a Major Research Project of the British Academy. The overall objective of the LGPN project is to create a comprehensive and authoritative record of the names of all individuals attested in Greek (or with Greek names attested in Latin) in the ancient Greek-speaking world, and so provide the classical research community worldwide with a unique and fundamental resource for the study of all aspects of the ancient Greek world.

Research publications about the *Lexicon* provide a pointer to the range of research fields which the LGPN can illuminate, and, in some instances, makes possible for the first time, including linguistics, the history of religion, historiography and literary history, demographic studies, and above all, cultural interaction. In practice, of course, there is no limit, nor should there be, to the uses researchers will make of the LGPN’s material.

LGPN is internationally recognized as a resource which has transformed the basis on which names may be studied and used. It has done so to date primarily through its publications; so far, over a quarter of a million individuals sharing over 35,000 names have been published in six regional volumes:

1. I, Aegean Islands, Cyprus, Cyrenaica (1987);
2. II, Attica (1994);
3. IIIA, Peloponnese, W. Greece, Sicily, Magna Graecia (1997);
4. IIIB, Central Greece (2000);
5. IV, Macedonia, Thrace, Northern Regions of the Black Sea (2005);
6. VA, Coastal Asia Minor: Pontos to Ionia (2010);

with at least two more in preparation:

- VB, Coastal Asia Minor from Caria to Cilicia (due 2012);
- VC, Inland Asia Minor;

---

1 Elaine Matthews was unfortunately unable to complete her planned enhancements to this paper before her untimely death in June 2011, so there is less critical engagement with other prosopography and person databases than had been planned.

(and, if the project funding continues, extending to a second tranche of work on Palestine, Syria, and the Trans-Euphratic Regions, as well as possible work on Egypt, for which some material has been collected).3

The regional basis on which the Lexicon research and publication has been undertaken has meant that use of the collection as a whole has been relatively limited. The project offers summary data online, which provides the numbers of hits per name, and allows the reader to establish, for example, that the name ʿΑβάσκαντος is attested 147 times.4 However, the other data about those 147 uses of the name have remained on paper. This chapter attempts to unlock some of that information, and show how it can be accessed in a variety of ways.

Lexicon data categories

The key to understanding the LGPN records is the set of data categories established at the start of the project. It was decided then to record the following pieces of information:

1. Normalized primary name form;
2. Sex of person named;
3. Place where the person belonged;
4. Date of the attestation (which can vary wildly in precision);
5. Bibliographical references;
6. Assorted other data. This may include placename variations e.g. alternative places of citizenship; name variants (orthography, dialect), corrections etc; parent/child relationships to other people; status or profession; and editorial corrections/alterations to the record.

The initial data collection was made on paper slips (an example is shown in Figure 1) on the basis of scholars familiar with the region reading primary and secondary sources.

The initial phase of work consisted solely of data collection, but by the late 1970s it was beginning to be recognized that conventional typesetting of the desired publication was likely to be prohibitively expensive, and that managing the entries on a computer should allow for delivery of camera-ready copy to the publisher. Given the technology of the time, it was not clear how this would be achieved, but a text format was agreed which provided the minimal distinctions. In this compact text form of the Lexicon data, a record looks like this, with six fields of information separated by @ characters:

Nani1s @ f @ Athens? @ F4B @ +IG II<2> 12229 @ (%_Na!nei1s!)

This asserts that there is a record of a person called Νανίς, a woman, probably from Athens, in the first half of the fourth century BC, with a bibliographical reference (publication of an inscription). The transliteration is the Lexicon’s own internal system (Table 1): numbers are used to indicate accents and breathings, except for 6 and 7 (used to

3 For more details see the LGPN state of preparation page: <http://www.lgpn.ox.ac.uk/publications/stateprep.html#4prep>.

indicate archaic characters); the number 4 is used to indicate the editorial dot under an unclear reading). The name is normalized from the attested form which has the syllable νείς.

We should note that this is not guaranteed to be different from the person in another record with the name Νανίς, but the Lexicon believes it is a distinct individual. Any other inscriptions which mention the same person will be conflated with this record. The bibliographical reference is usually, but not always, exhaustive. If the person is very well attested, another reference work, such as an encyclopaedia, will be cited, where the full references can be found.

The relationship to the place name is almost always place of birth, usually an ancient city or region, though the name of the modern find-spot may be given where the ancient site cannot be identified.

<table>
<thead>
<tr>
<th>6 and 7</th>
<th>6=F 5=h</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>*A=A *A1=A *A3=A A=A</td>
</tr>
<tr>
<td></td>
<td>A'=A' A1'=A' A13'=A' A'o=A'o A'u=A'u A'u1=A'u3=A'o A1'=A' A14'=A'</td>
</tr>
<tr>
<td></td>
<td>A3'=A A4'=A</td>
</tr>
<tr>
<td>B</td>
<td>B=B</td>
</tr>
<tr>
<td>C</td>
<td>C=X C4=X</td>
</tr>
<tr>
<td>D</td>
<td>D=Δ D4=Δ</td>
</tr>
<tr>
<td>E</td>
<td>*E'=E *E1=Δ E=E E'O=Δ E1=E1 E13=E13 E0'=E0' E01'=E01' E0'=E0' E01'=E01' E03'=E03'</td>
</tr>
<tr>
<td></td>
<td>E1=Δ E14'=E E18'=E E4'=E</td>
</tr>
<tr>
<td>F</td>
<td>F'=Φ F4=Φ</td>
</tr>
<tr>
<td>G</td>
<td>G'=Γ G4=Γ</td>
</tr>
<tr>
<td>H</td>
<td>*H=H *H1=H *H3=H H=H H1=H H3=H H4=H</td>
</tr>
<tr>
<td>I</td>
<td>*I=I *I1=I I1=I I13=I3=I</td>
</tr>
<tr>
<td>K</td>
<td>K=K K4=K</td>
</tr>
<tr>
<td>L</td>
<td>L=Λ L4=Λ</td>
</tr>
</tbody>
</table>
Table 1: LGPN transliteration

Some more complex examples of the Lexicon text markup format:

<table>
<thead>
<tr>
<th>M</th>
<th>M=M M4=M</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>N=N N4=N</td>
</tr>
<tr>
<td>O</td>
<td>*O=O *O1=O O=O O1=O O4=O</td>
</tr>
<tr>
<td>P</td>
<td>P=P P4=P</td>
</tr>
<tr>
<td>Q</td>
<td>Q=Θ Q4=Θ</td>
</tr>
<tr>
<td>R</td>
<td>*R=P R=P</td>
</tr>
<tr>
<td>S</td>
<td>S=Σ S4=Σ</td>
</tr>
<tr>
<td>T</td>
<td>T=T T4=T</td>
</tr>
<tr>
<td>U</td>
<td>*U=Y *U1=Y U=U Y U1=U U14=Y</td>
</tr>
<tr>
<td>W</td>
<td>*W=Ω *W1=Ω W=Ω W1=Ω W3=Ω</td>
</tr>
</tbody>
</table>

Table 1: LGPN transliteration

Some more complex examples of the Lexicon text markup format:
show how the last field is very considerably overloaded with its own fields of information, especially in the ‘final bracket’ (beloved of generations of Lexicon staff), with subfields separated by ‘;’ characters, and within those by commas and many other editorial conventions (e.g. marking of italics with +, and Greek with %). This is markup rather characteristic of its time, managed within the project over 30 years of work.

Critically, the records shown above provide no unique identifier for a Lexicon record, a problem which may have a considerable impact on future work.

Figure 2: Lexicon typeset output

Figure 2 shows the three-column typeset output which was designed for the Lexicon in the early 1980s and has remained more or less consistent for all the printed volumes to date.

The Lexicon’s IT history and current status

The Lexicon has lived through all the generations of humanities computing, in each period espousing the technology of the moment where possible.

During the 1970s, the project had embraced digital storage, and had started transferring data on cards to files using a locally-written flat database (Famulus); there was no method of retrieval, and only an outline plan for producing camera-ready copy by using a pen plotter. By the start of the 1980s, however, the project had to confront the problem of selective retrieval, output, and checking the integrity of data. A period of intensive examination of the data so far input meant that by the mid-1980s, the Lexicon was loading material into a network database (IDMS), had retrieval programs written in FORTRAN, and was able to typeset pages using procedural markup on a Monotype Lasercomp. The database was subsequently converted to a relational model (using Ingres), retrieval programs were rewritten in Pascal and C, and the typesetting switched to producing PDF using TeX. The most significant landmark in the project’s IT history was the design and implementation of the database structure to reflect and provide access to all the research components of an LGPN record, which in the publication books are simply
presented as text, for example chronological and topographical data, and socially relevant data such as statuses and relationships.\(^5\)

The database design (an early draft is shown in Figure 3) was crucial in imposing consistency of format on complex evidence, and the published volumes have all been generated from it in camera-ready form,\(^6\) but it has not been opened up for general research. In the present century, staff continue to use text files with markup as their main editorial tool, backed by a relational database.

Figure 3 Database original schematic

The increasing requirement for data models which emphasize collaboration with other projects, and concerns over sustainability, caused the *Lexicon* to initiate a project in 2005 to remodel the data so that it could be represented in XML conformant with the guidelines of the Text Encoding Initiative (TEI).\(^7\) The aim was to set up an IT infrastructure to support the future maintenance and preservation of irreplaceable research data, and provide direct online access to all the data, thus opening up the full potential of the *LGPN* data to researchers. This included the integration of data from all published (and to be published) volumes into one resource. The intention was that the *LGPN* play as significant a part in the e-research environment as it had played in traditional scholarship, and take a lead in setting standards for encoding names in documents and achieving interoperability with online material worldwide.

---


\(^6\) The first volume was typeset by database routines generating code for a Monotype Lasercomp; subsequent volumes utilized a complex relational database retrieval which extracted and transformed all the fields into TeX markup.

The technical components of the XML phase of the LGPN which helped the Lexicon enhance its publishing and interchange capability consisted of several phases. First, there was the definition of an XML schema, as a customization of the TEI, for an archival form of the Lexicon data suitable for repositories. Existing database retrieval routines were then adapted to output XML conformant to the agreed schema. This allowed the project to provide a simple forms-based interface for searching the database, and delivery of results in XML against the TEI schema; with the option of transforming that to other modern web delivery formats such as HTML for web pages, JSON (data optimized for consumption by Javascript in web pages) or RDF (for use in semantic web applications).

The LGPN XML work coincided with, and stimulated, a major revision of the TEI module relating to names and dates in 2006/07. The work done then to model persons, places, and organizations as first class objects, allowed the Lexicon schema to be a conformant pure subset of the TEI.

The major data categories present in the Lexicon map cleanly to TEI elements as follows:

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>TEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>&lt;nym&gt;</td>
</tr>
<tr>
<td>Person</td>
<td>&lt;person&gt;</td>
</tr>
<tr>
<td>Name form</td>
<td>&lt;persName&gt;</td>
</tr>
<tr>
<td>Sex</td>
<td>&lt;sex&gt;</td>
</tr>
<tr>
<td>Date</td>
<td>Formally, notBefore and notAfter attributes on &lt;birth&gt;; informally in &lt;floruit&gt;</td>
</tr>
<tr>
<td>Status</td>
<td>&lt;socccStatus&gt;</td>
</tr>
<tr>
<td>Reference</td>
<td>&lt;bibl&gt;</td>
</tr>
</tbody>
</table>

The TEI format designed for the Lexicon makes use mainly of the <person> element to contain a Lexicon record.

After transformation, the new long-form XML record for the first Lexicon data example given above looks like this:

```xml
<person xml:id="V2-60057">
  <sex value="2"/>
  <persName type="full" nymRef="#Nani1s">Νανίς</persName>
  <birth notAfter="-0350" notBefore="-0399">
    <placeName key="Athens" cert="?">Athens</placeName>
  </birth>
  <floruit>f.iv BC</floruit>
  <persName type="namevariant" xml:lang="grc">Να<seg type="orth">νείς</seg></persName>
  <bibl>
    <title>IG II<hi rend="sup">2</hi> 12229</title>
  </bibl>
</person>
```

The issue of ‘fuzzy’ dates has been dealt with by storing the human-readable string provided by the editorial compiler (in this case f.iv BC, meaning ‘first half of the fourth century BC’) as the content of the <floruit> element, but mapping it onto absolute year

8 Via a set of ad hoc Perl cleaning scripts, and XSLT transformations.
range using the notAfter and notBefore attributes of <birth>. This allows for sorting and
date range searching of the records.

Names themselves are stored in a set of <nym> records, providing for the name as a
first class object distinct from the <person>. For example, the name in our example above
looks like this:

```
<listNym>
  <nym xml:id="nNani1s">
    <form xml:lang="el-grc">Νανίς</form>
    <form xml:lang="el-grc-x-noaccents">Νανις</form>
    <form xml:lang="el-grc-x-lgpn">Nani1s</form>
    <form xml:lang="el-grc-x-lgpnnoaccents">Nanis</form>
    <form xml:lang="el-grc-x-perseus">*nani/s</form>
  </nym>
</listNym>
```

Variants of the main name in different encodings, with and without accents, are stored
as well in order to make implementation of searching easier.

One of the more complicated examples shown earlier (lexical data categories
Θεόφραστος) in short form exposes some of the less ideal uses of the TEI markup:

```
<person n="2-23" xml:id="V2-33229">
  <sex value="1"/>
  <persName type="main" nymRef="#nQeo1frastos">Θεόφραστος</persName>
  <birth notAfter="-0050" notBefore="-0099">
    <placeName key="LGPN_20500">Hagnous</placeName>
  </birth>
  <floruit>f.i BC</floruit>
  <state key="#relationship">
    <label>s.<persName type="relationship" xml:lang="el-grc" nymRef="#nQemistoklh3s">Θεμιστοκλῆς I</persName>
  </label>
  <state key="#relationship">
    <label>
      <persName type="relationship" xml:lang="el-grc" nymRef="#nAke1stion">Ἀκέστιον</persName>
    </label>
  </state>
  <state key="#relationship">
    <label>
      <persName type="relationship" xml:lang="el-grc" nymRef="#nQemistoklh3s">Θεμιστοκλῆς II</persName>
    </label>
  </state>
  <state key="#relationship">
    <label>
      <persName type="relationship" xml:lang="el-grc" nymRef="#nSofoklh3s">Σοφοκλῆς</persName>
    </label>
  </state>
  <bibl>Paus. i 37. 1</bibl>
```

Offprint from BICS Supplement-122 © The Institute of Classical Studies University of London 2013
It will be clear to the experienced TEI user that the <bibl> records could usefully be properly structured, and that the use of <state> to model relationship claims is not as good as a proper <relation> element would be. Unfortunately, the assertion in the Lexicon that ‘Θεόφραστος is the father of Θεμιστοκλῆς II’ does not permit us to automatically identify which record ‘Θεμιστοκλῆς II’ applies to (although this sort of record works well in print where a human can extrapolate).

Places are modelled using the TEI <place> element, pointed to by the key attribute on <placeName>. This allows us to maintain a single hierarchical <listPlace> containing all place names used by the Lexicon. Thus:

```xml
<listPlace>
  <place type="region" xml:id="LGPN_33014">
    <region>Achaia</region>
  </place>
  <place type="settlement" xml:id="LGPN_33915">
    <placeName>Aiga</placeName>
  </place>
  <place type="settlement" xml:id="LGPN_33003">
    <placeName>Aigeira</placeName>
  </place>
  <place type="settlement" xml:id="LGPN_33917">
    <placeName>Aigeira (Hyperesia)</placeName>
  </place>
  <place type="settlement" xml:id="LGPN_33004">
    <placeName>Aigion</placeName>
  </place>
</listPlace>
```
The *Lexicon* has defined five levels of data interchange as a result of the XML work:

1. **Character interchange**: ASCII text version of data separately from the binary format used by any database system. This was the minimal form of interchange supported in the initial stages of the project.

2. **Character encoding**: The *Lexicon* defined its own transliteration for Greek, independently of, for example, TLG betacode, and continues to use it for internal purposes. It is a happy accident that the characters used in the transliteration allow the names to be used in human-readable URLs.

3. **Standardized structural markup**: Data relationships follow the schema defined in 1983. The hierarchical and network structure used in project databases are maintained in the XML records.

4. **Standardized semantic markup**: The XML representation of the *Lexicon* is aligned with the vocabulary and semantics for XML elements of the Text Encoding Initiative. The TEI elements are themselves in the process of alignment with the CIDOC CRM, allowing even wider understanding and a serious ontology within this field.

5. **Information linking**: For most categories of data (name, sex, data, bibliography), the *Lexicon* can be fully linked to comparable data. Places information is more complex and will be addressed later in this chapter.

One of the important additions to the new LGPN XML representation is the exposure of an ID for each record, to allow the project to offer a permanent URL. The IDs are of the form `volume number-person number`, for example V2-1030. Another form of identifier available is the sequence number for each name, as shown in the published volumes, for example ‘Ἀρχίτιμος 10’. For users of the books, this is the only way they can refer to an entry, but such usage raises the very considerable problem of providing updates and additions for published material. The *Lexicon* has not yet resolved this issue.

**Dealing with place names**

There are approximately 3000 place names referred to in the *Lexicon*, in data which was collected long before it occurred to anyone to plot name occurrences on maps. The places are managed in a three-level geographical/political hierarchy (region, settlement, and deme) with occasional granularity down to the quasi-geographical tribe. In order to provide the map display described in the next section, we need to establish, at a minimum, a latitude and longitude for each place. This leaves aside, for the present, the issue of what point to use for a large place like Athens (the geographic centre? the Parthenon?) and the problem of variable size of settlements over time. To locate all 3000 places in the *Lexicon* from scratch is a considerable task, bearing in mind that:

1. the common geo-gazetteers (*e.g.* GeoNames) do not include tiny villages in northern Greece where the name given in the secondary literature from which the *Lexicon* derives may be an idiosyncratic transcription, and an older name;

2. the precision of the location may only be regional (*e.g.* Crete);

3. the recorded name may be ambiguous when checked against modern atlases.

---

9 Described at the CIDOC CRM Home page: <http://www.cidoc-crm.org>; compare this with the TEI Ontologies SIG: <http://www.tei-c.org/Activities/SIG/Ontologies>.

Desirable though it would be to revisit all the sites with a GPS, the Lexicon method is, in practice, three-fold. Firstly, place names which are unambiguous, and do appear in the modern gazetteers (e.g. Athens), are matched with a latitude and longitude quickly. Secondly, places which can be located in the Barrington atlas\(^\text{11}\) can be given an intermediate record of name, page number and grid reference, in the knowledge that we will be able to use the work of the Pleiades project,\(^\text{12}\) which is gradually digitizing all the material from Barrington, to resolve an identifier like akraiphiai-55-e4 (this place is in fact http://pleiades.stoa.org/places/540617\(^\text{13}\)). Finally, the Lexicon is a partner in the larger CLAROS\(^\text{14}\) project, which brings together a set of classical art resources, including a large set of common place names. As each partner geolocates places, the Lexicon can share the data.

A more complete place record may now be shown, enhanced with latitude and longitude, modern place names, and reference to the Barrington atlas.

At the time of writing, only about half the Lexicon placenames have been fully geolocated and/or linked to other gazetteers.

The Lexicon online

With the conversion to XML available, it is now possible to deliver 250,000 published records in a single new interface. The service offers a fairly conventional form-based interface to allow users to search by any of the available data fields: name (in various transliterations), date, place, florisit, and status.\(^\text{15}\) An initial form showing name (with on-screen keyboard for Greek) and date (Figure 4) can be expanded to cover the other fields (Figure 5). Names and places can also be picked from pre-built summary lists (Figures 6 and 7). The fields are simply additive; the results must satisfy all criteria at once, thus not allowing for disjoint queries such as ‘names from Cyprus or Messenia’, or ‘names ending in ιμος from the fourth century AD or fourth century BC’.


\(^{12}\) Pleiades, a gazetter of ancient places: <http://pleiades.stoa.org/>.

\(^{13}\) Thanks to help from Tom Elliott, we were able to make a trial digitization of c.60 places ourselves, ahead of Pleiades schedule, which helped us refine the Lexicon workings.

\(^{14}\) Claros: <http://www.clarosnet.org/>.

\(^{15}\) LGPN search interface: <http://www.lgpn.ox.ac.uk/database/lgpn.php>.
Figure 4: LGPN online simple search form

Figure 5: LGPN online advanced search form

Figure 6: LGPN online, pick lists for names
The results can be returned in a variety of ways. For normal browsing, the default is to return simply the number of hits, in order to avoid unnecessary data transfer. Alternatively, a tabular display is provided, with sortable columns and narrowing via a search box (Figure 7). Results are batched, but queries which would result in more than 10,000 records to display are not permitted. For those interested in the geographical spread of results, the Map tab utilizes Google Maps to offer simple point display (Figure 8). It should be noted, however, that not all of the places known to the Lexicon have been geolocated, as discussed in the previous section.

Figure 7: LGPN online, pick lists for places

Figure 8: LGPN online result table

_requests for data formats, rather than web pages for human consumption, are not limited by size._
The results of this web form-based searching can be exploited further using a variety of other data formats. The underlying TEI XML can be returned for those wishing to perform their own transformations, JSON data can be used for web-based visualization, the KML (Keyhole Markup Language) format of XML for Google Earth and Google Maps can be downloaded, and a CSV (comma-separated data fields) summary can be imported into a spreadsheet for data exploration.

The XML RDF format is a conversion of the TEI markup to the CIDOC CRM ontology for interchange and cross-searching with other data sets. Making Lexicon data available against this formal ontology, mapping Lexicon data categories onto well-defined concepts, allows it to be ingested immediately into semantic web databases, and start to participate in the universe of open-linked data. This is intended to be used in computer-to-computer interaction, using a standardized query language (SPARQL).\textsuperscript{17}

The CLAROS project is an example of using the Lexicon RDF feed;\textsuperscript{18} it provides an aggregating searchable cache across large classical art history databases, but also includes the Lexicon data. This can allow, for example, the formulation of queries which combine

\textsuperscript{17} SPARQL is a W3C recommended query language for RDF.

\textsuperscript{18} Claros: <http://www.clarosnet.org>.
questions about vase forms with names of people, linked by place name. Europeana is another example of a large cache based on aggregating RDF data against the CRM.19

The online search is implemented using a simple read-only relational database which contains chunks of TEI XML (i.e. a <person>), and a set of extracted index terms. This allows for efficient sorting and searching. The TEI XML fragments are taken from the database and converted to the appropriate output format (HTML, JSON, KML, etc.) using XSLT transformations. An initial implementation using an XML database (eXist)20 was unable at that time to support very large result sets and complex additive queries with a sufficiently good response time. Many other systems could be used to provide the same service in future.

The web pages for use by classical scholars are only one way in which the data can be accessed. They are also available through persistent and predictable URLs. We try to follow here some of the modern guidance about providing 'cool URIs'.21 Lexicon URLs take the form: http://www.lgpn.ox.ac.uk/type/query/format, where type is one of:

1. batlas: grid square, page and name in Barrington atlas;
2. date: date in the form yeartoyear;
3. floruit: date range, using Lexicon conventions;
4. id: Lexicon person ID;
5. lexname: transliterated name (lexnamenoaccents: without accents);
6. n: a combination of volume-name in accented Greek-publication number;
7. name: name in UTF-8 Greek (namenoaccents: without accents);
8. place: place name from Lexicon authority list;22
9. placecode: Lexicon internal code for place ;
10. region: geographic region, using Lexicon names;
11. status: status from Lexicon list;23

…and format is one of:

1. csv: comma-separated values in table;
2. exhibitdata: JSON code suitable for consuming by Simile Exhibit;24
3. html: human-readable web page;
4. json: Javascript JSON data format;
5. kml: KML for display in Google Earth or Maps;
6. rdf: RDF XML;

20 eXist, an Open Source database: <http://exist.sourceforge.net>.
21 For more discussion of this topic see W3C, Cool URIs don’t change: <http://www.w3.org/Provider/Style/URI>.
22 LGPN place name authority list: <http://www.lgpn.ox.ac.uk/place>.
23 LGPN status and occupations list: <http://www.lgpn.ox.ac.uk/status>.
The Digital Classicist 2013

7. summaryjson: Javascript JSON showing summary detail of date range and count for name;
8. xml: TEI XML.

A selection of examples is given in Table 2.

<table>
<thead>
<tr>
<th>Search by name, return HTML</th>
<th>/lexnamenoaccents/Paramonos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search by name, return XML</td>
<td>/lexname/Paramonos/xml</td>
</tr>
<tr>
<td>Search by name in accented Greek, return KML</td>
<td>/name/Παράμονος/kml</td>
</tr>
<tr>
<td>Search by Lexicon ID, return TEI XML</td>
<td>/id/V1-2697/xml</td>
</tr>
<tr>
<td>Search by status, return JSON</td>
<td>/status/potter/json</td>
</tr>
<tr>
<td>Search by date, return RDF</td>
<td>/date/250to265/rdf</td>
</tr>
<tr>
<td>Search by place name, return comma-separated data</td>
<td>/place/Aloros/csv</td>
</tr>
<tr>
<td>Search for 1st Αργόφιλος in volume 3a</td>
<td>/v/3a-Αργόφιλος-1</td>
</tr>
</tbody>
</table>

Table 2: Lexicon persistent URL patterns

With Lexicon data available in a standardized way, what sort of services can be built on top of it? One example is a ‘name decorator service’. If we have web pages showing Greek inscriptions, and the names are identified in some way, we can run over the page, look up each name in the LGPN, and enhance the page. Thus the HTML may have markup like this:

```
...κὲ παντευλόγ [...ἱον<br/>εις ἄπονθησιν<br/>της πλήθη ἔκτισαν<br/>ἐξ ἱδίων μνῆμα<br/>

<abbr class="lgpn" title="Ἰαηλ">Ιαηλ</abbr> προστάτης
<abbr class="lgpn" title="Ἰωσούα">Ἰωσούᾳ</abbr> ἄρχοντι
<abbr class="lgpn" title="Παλατῖνος">Παλατῖνος</abbr> σὺν ὑἱῷ ...</div>
```

in which names are identified using an HTML class attribute. The Javascript in greeknames.js will take care of the lookup by making a series of requests to, for example, http://www.lgpn.ox.ac.uk/name/Θεόδοτος/summaryjson which returns a record like this:

```
{
  "query": "Θεόδοτος",
  "id": "Qeodotos",
  "name": "Θεόδοτος",
  "notBefore": -500,
  "notAfter": 999,
  "number": 393,
  "firstChar": "Θ"
}
```

containing the information about the date range and number of occurrences of the name. This can then be used to add a popup on the name, in which unknown names have a red underline and known names have a green underline. This complete example (facing page) is available at <http://clas-lgpn2.classics.ox.ac.uk/Demo/iAph110055.html>, which includes a link to the XML source.

25 We are grateful to Gabriel Bodard at King’s College London for this example from the Aphrodisias inscriptions. The HTML was created by transforming the project’s TEI XML markup.
Conclusions

We believe that this story of the Lexicon of Greek Personal Names illustrates four points. Firstly, the conceptual data model of the 1970s has survived the test of time; it has gone through many completely unforeseen changes and challenges, but has required no serious rethinking. Secondly, the Lexicon experience shows that the modern web techniques of machine/machine interchange, and rich exploratory tools, can be retrofitted effectively to older projects. Thirdly, we believe that the extra abilities added to interoperability by the adoption of open standards and linked data are crucial to the future of research data like the Lexicon.

Finally, we hope that there are entirely new academic questions waiting to be answered by this version of the Lexicon data, as well as uses we have not yet imagined.

Acknowledgements

We are grateful to Tom Elliott at the Institute for the Study of the Ancient World, NYU, for assistance during the work that led to this chapter and for essential help with locating places and ideas about URLs; Janet McKnight, for Javascript programming and geolocating; Gabriel Bodard, for forcing us to complete this work by giving us a deadline for the seminar paper on which this chapter is based; and Donna Kurtz and the Beazley Archive at Oxford for stimulating ideas about interchange.

Elaine Matthews (late of All Souls College, Oxford)
Sebastian Rahtz (University of Oxford) sebastian.rahtz@it.ox.ac.uk
Bibliography


Fraser, P. M., Greek ethnic terminology (Oxford 2009).


