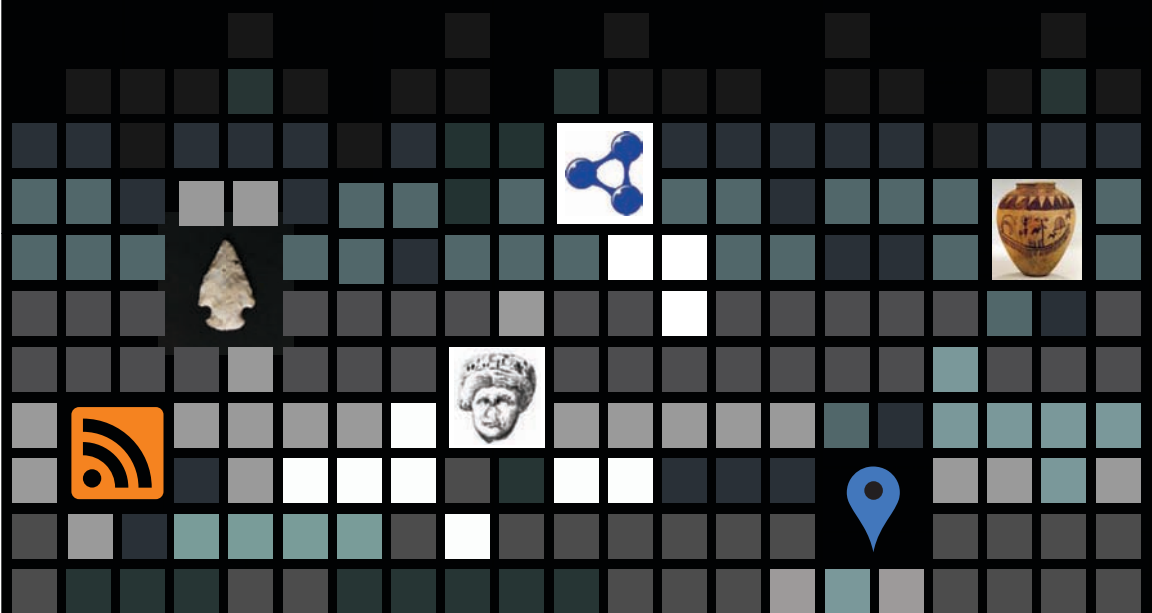
ARCHAEOLOGY 2.0

new approaches to communication & collaboration



edited by Eric C. Kansa, Sarah Whitcher Kansa, & Ethan Watrall

Archaeology 2.0: New Approaches to Communication and Collaboration

Cotsen Digital Archaeology Series

Volume 1. *Archaeology 2.0: New Approaches to Communication and Collaboration*, Eric C. Kansa, Sarah Whitcher Kansa, and Ethan Watrall (Editors)

Archaeology 2.0:
New Approaches to
Communication and Collaboration

Edited by

Eric C. Kansa,

Sarah Whitcher Kansa, and

Ethan Watrall

Cotsen Digital Archaeology 1

THE COTSEN INSTITUTE OF ARCHAEOLOGY PRESS is the publishing unit of the Cotsen Institute of Archaeology at UCLA. The Cotsen Institute is a premier research organization dedicated to the creation, dissemination, and conservation of archaeological knowledge and heritage. It is home to both the Interdepartmental Archaeology Graduate Program and the UCLA/Getty Master's Program in the Conservation of Archaeological and Ethnographic Materials. The Cotsen Institute provides a forum for innovative faculty research, graduate education, and public programs at UCLA in an effort to positively impact the academic, local and global communities. Established in 1973, the Cotsen Institute is at the forefront of archaeological research, education, conservation and publication and is an active contributor to interdisciplinary research at UCLA.

The Cotsen Institute Press specializes in producing high-quality academic volumes in several different series, including Monographs, World Heritage and Monuments, Cotsen Advanced Seminars, and Ideas, Debates and Perspectives. The Press is committed to making the fruits of archaeological research accessible to professionals, scholars, students, and the general public. We are able to do this through the generosity of Lloyd E. Cotsen, longtime Institute volunteer and benefactor, who has provided an endowment that allows us to subsidize our publishing program and produce superb volumes at an affordable price. Publishing in nine different series, our award-winning archaeological publications receive critical acclaim in both the academic and popular communities.

THE COTSEN INSTITUTE OF ARCHAEOLOGY AT UCLA

Charles Stanish, Director	Willeke Wendrich, Editorial Director
Gregory Areshian, Assistant Director	Julie Nemer, Publications Manager

EDITORIAL ADVISORY BOARD OF THE COTSEN INSTITUTE OF ARCHAEOLOGY

Willeke Wendrich	Area Editor for Egypt, North, and East Africa
Christopher Donnan	Area Editor for South and Central America
Jeanne E. Arnold	Area Editor for North America
Jeffrey P. Brantingham	Area Editor for the Paleolithic and Environmental Archaeology
Aaron Burke	Area Editor for Southwestern Asia
Lothar Von Falkenhausen	Area Editor for East and South Asia and Archaeological Theory
Sarah Morris	Area Editor for the Classical World
John Papadopoulos	Area Editor for the Mediterranean Region
Ex-Officio Members:	Charles Stanish, Gregory E. Areshian, and Julie Nemer
External Members:	Kusimba Chapurukha, Joyce Marcus, A. Colin Renfrew, and John Yellen

Cover design by Ethan Watrall. Editorial and production services: Leyba Associates, Santa Fe, NM

Library of Congress Cataloging-in-Publication Data

Archaeology 2.0 : new approaches to communication and collaboration / edited by Eric C. Kansa, Sarah Whitcher Kansa, and Ethan Watrall.

p. cm. — (Cotsen digital archaeology ; 1)

Outcome of a session held at the 2008 meeting of the Society for American Archaeology (SAA) in Vancouver, British Columbia.

Includes bibliographical references and index.

ISBN-13: 978-1-931745-85-7 (pbk.) ISBN-10: 1-931745-85-4 (pbk.) ISBN- 978-1-938770-65-4 (eBook)

1. Archaeology—Computer network resources. 2. Web 2.0. 3. Archaeology—Data processing. 4. Archaeology—Methodology. 5. Archaeology—Information technology. 6. Digital electronics. 7. Internet. I. Kansa, Eric Christopher. II. Kansa, Sarah Whitcher. III. Watrall, Ethan. IV. Society for American Archaeology. Meeting (73rd : 2008 : Vancouver, B.C.)

CC80.4.A66 2011

930.1028—dc23

2011030384

© Copyright 2011. Regents of the University of California.

This volume carries a Creative Commons BY-SA (By Attribution, Share Alike, http://creativecommons.org/licenses/by-sa/3.0/) license. In short, this means that others can freely distribute, remix, and build upon the contents of this volume, provided two very important conditions are met: the original author receives proper attribution (especially citation) and all subsequent works carry the same license. We chose a Creative Commons license primarily because of our deep concerns in the sustainability of sharply escalating costs in scholarly publishing. These costs make it increasingly difficult for educational institutions, our colleagues in commercial archaeology, students, and members of the interested public to (legally) obtain peer-review publications. Please note that the Creative Commons BY-SA license allows for commercial use, as well as free distribution both inside and outside of the Academy. Permissions for commercial reuse does not, however, mean commercial appropriation. The "copyleft" philosophy embodied by this license enables this work to move in many contexts, but any adaptation or enhancement of this work must be shared back, openly, with the community. Finally, because this license requires proper attribution in any subsequent duplication or adaptation, we hope this volume helps build exposure and recognition for our contributions, and that our colleagues follow in this example. With enough accessible and open data ("data" that includes content like this book), we open up more opportunities for text-mining, tagging, aggregating, linking, visualizing, and hopefully better understanding.



Contents

List of Figures vii
List of Tables viii
Volume Editors ix
List of Contributors xi
Preface xiii
Acknowledgments xiv

Introduction: New Directions for the Digital Past
Eric C. Kansa 1

Section I: A Web of Archaeological Data: Infrastructure, Services, and Interoperability 27

Chapter 1: The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing
Julian Richards, Stuart Jeffrey, Stewart Waller, Fabio Ciravegna, Sam Chapman, and Ziqi Zhang 31

Chapter 2: Toward a Do-It-Yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing
Eric C. Kansa and Sarah Whitcher Kansa 57

Section II: The Technical and Theoretical Context of Archaeology on the Web 93

Chapter 3: Poor Relatives or Favorite Uncles? Cyberinfrastructure and Web 2.0: A Critical Comparison for Archaeological Research
Stuart Dunn 95

Chapter 4: Archaeological Knowledge Production and Dissemination in the Digital Age
Robin Boast and Peter Biehl 119

Section III: Archaeological Data Management and Collaboration 157

Chapter 5: Creating a Virtual Research Environment for Archaeology
Michael Rains 159

Chapter 6: iAKS: A Web 2.0 Archaeological Knowledge Management System
Ethan Watrall 171

CHAPTER 7: User-Generated Content in Zooarchaeology: Exploring the "Middle Space" of Scholarly Communication
Sarah Whitcher Kansa and Francis Deblauwe 185

SECTION IV: Sustainability, Quality, and Access 207

CHAPTER 8:UCLA Encyclopedia of Egyptology, Archaeological Data, and Web 2.0
Willeke Wendrich 211

CHAPTER 9: Open Access for Archaeological Literature: A Manager's Perspective
Jingfeng Xia 233

CHAPTER 10: What Are Our Critical Data-Preservation Needs?
Harrison Eiteljorg 251

CONCLUSION: *Web 2.0 and Beyond*, or On the Web, Nobody Knows You're an Archaeologist
W. Fredrick Limp 265

List of Figures

Figure 1.1. A screen-shot of the prototype faceted classification browsing tree for the Archaeotools project 40
Figure 1.2. Archaeotools process architecture 43
Figure 1.3. Example annotations from a gray literature report showing title, author, and published date 46
Figure 1.4. A screen-shot showing a "tagstract" of indexing terms for a gray literature report 49
Figure 2.1. Open Context's "faceted browse" tool, showing filtered results . .62
Figure 2.2. Example of a project overview in Open Context 63
Figure 2.3. Overview of search strings that linked to Open Context content over a three-month period 66
Figure 2.4. Open Context uses Dublin Core metadata to generate bibliographic citation information and a stable URL for each item (or tagged set of items) 73
Figure 2.5. Illustrations of RESTful services in Open Context 78
Figure 2.6. Web services demonstration using Google Earth to visualize Open Context KML data, showing species ratios at different sites in the Near East 80
Figure 3.1. Thumbnail images on the Google Maps interface form a clear linear representation of Hadrian's Wall 110
Figures 4.1–4.4. Multimedia applications in photo, video and excavation documentation and digital reconstructions and visualizations 130–133
Figures 4.5–4.6. Webcam and videostreaming and its practical application on site 135–136
Figure 4.7. Screen-shot of the experimental Blobgects interface (note the tag cloud and recent comments displayed on the right side) . . .139
Figure 4.8. Screen-shot of the control Blobgects interface (note the lack of the tag cloud and comments) 140
Figure 4.9. Comparison of Zuni comments (left) and MAA catalog records (right) for four sample objects, grouped according to metadata field or general category 146
Figure 5.1. A simple diagram showing stratigraphic units (contexts) and the stratigraphic links between them 162
Figure 5.2. Structure diagram with added image, plan, and document 162
Figure 5.3. Researching a coin assemblage within a virtual research domain (VRD) 163

Figure 5.4. Online collaborative document editing within a virtual research domain 164
Figure 6.1. The iAKS One Client/One Server model 176
Figure 6.2. The iAKS Many Clients/One Server model 177
Figure 6.3. The iAKS Many Clients/Many Projects/One Server model 178
Figure 7.1. Custom feeds of zooarchaeology-related data flowing from Open Context into BoneCommons 190
Figure 7.2. Diagram showing the flow of information between the ZOOARCH email list and BoneCommons. 191
Figure 7.3. Screen-shot of the Zooarchaeology Social Network 193
Figure 7.4. A search in Google Insights for Search shows that archaeology-related domains make up a smaller proportion of the Web over time . . .196
Figure 7.5. A Yippee (formerly Clusty) search for domain names shows an increase in .com at the expense of .edu 197
Figure 7.6. A Google Blogs search for "archaeology," showing the relative strength of different domains 197
Figure 7.7. The "long tail" of 1,132 searches leading to BoneCommons (searches from April to June, 2010) 202
Figure 8.1. The interface of the UEE Open Version in eScholarship (http://escholarship.org/uc/nelc_uee) 216
Figure 8.2. The interface of the Uee Full Version 218
Figure 8.3. The UEE time map, overview 220
Figure 8.4. The UEE time map, zoomed in to the site of Tanis in the Eastern Nile Delta 223
Figure 9.1. A sample web interface from arXiv 237

LIST OF TABLES

Table 1.1.Ontology mismatches between record entries and the appropriate thesauri after initial analysis 41
Table 1.2."Actual" identifications for 906 gray literature reports 48
Table 1.3. "Reference" identifications for 906 gray literature reports 48
Table 1.4. "Actual" identifications for 3991 PSAS papers 51
Table 1.5. "Reference" identifications for 3991 PSAS papers 51
Table 2.1. Links to examples from Open Context discussed in the text 60
Table 8.1. Feature types for the UEE time map 221

Volume Editors

Eric Kansa

Eric Kansa leads development of Open Context (http://opencontext.org). His research interests explore Web architecture, service design, and how these issues relate to the social and professional context of the digital humanities. Eric also researches policy issues relating to intellectual property, including text-mining and cultural property concerns. He actively participates in a number of Open Science, Open Government, cyberinfrastructure, text-mining, and scholarly user needs initiatives. Over the past three years, he has taught and practiced project management and information service design in the UC Berkeley School of Information's Clinic program, and his ongoing work at UC Berkeley includes participation on Project Bamboo, a humanities and social science cyberinfrastructure initiative. Eric has been a principal investigator and co-investigator on projects funded by the William and Flora Hewlett Foundation, the NEH and IMLS, Hewlett-Packard, Sunlight Foundation, Google, and the Alfred P. Sloan Foundation.

Sarah Whitcher Kansa

Sarah Whitcher Kansa is Executive Director of the Alexandria Archive Institute (http://alexandriaarchive.org), where she advocates for data sharing and publication in various archaeological and cultural heritage communities. In addition, she is Chief Editor of Open Context (http://opencontext.org), an open access data publication initiative for archaeology and related fields. Sarah has a Ph.D. in archaeology and has spent the past 15 years conducting zooarchaeological research at sites in the Near East. She also manages BoneCommons.org on behalf of the International Council for Archaeozoology (ICAZ), where she currently stands on the Executive and International Committees.

Ethan Watrall

Ethan Watrall is an Assistant Professor in the Department of Anthropology and Associate Director of Matrix: The Center for Humane Arts, Letters & Social Sciences Online (www.matrix.msu.edu) at Michigan State University. In addition, Ethan is Director of the Cultural Heritage Informatics Initiative and the Cultural Heritage Informatics Fieldschool at Michigan State University (chi.matrix.msu.edu). Ethan's research interests fall in the domain of cultural heritage informatics, with particular (though hardly exclusive) focus on digital archaeology and serious games for cultural heritage learning, outreach, and engagement.

List of Contributors*

* Email addresses are provided for lead authors only

Peter Biehl
380 MFAC Ellicott Complex
SUNY Buffalo
Buffalo, NY 14261

Robin Boast
Museum of Archaeology and Anthropology
University of Cambridge
Downing Street
Cambridge
CB2 3DZ, UK
robin.boast@maa.cam.ac.uk

Sam Chapman
Knowledge Now Limited
Aizlewood's Mill
Nursery Street
Sheffield
S3 8GG, UK

Fabio Ciravegna
Web Intelligence Technologies Lab
Organisations, Information and Knowledge (OAK) Group
Department of Computer Science
University of Sheffield
S1 4DP, UK

Francis Deblauwe
The Alexandria Archive Institute
125 El Verano Way
San Francisco, CA 94127

Stuart Dunn
Centre for e-Research, King's College London
26-29 Drury Lane
London, WC2B 5RL, UK
stuart.dunn@kcl.ac.uk

Harrison Eiteljorg
P. O. Box 60
Bryn Mawr, PA 19010
nickeiteljorg@mac.com

Stuart Jeffrey
Archaeology Data Service
Department of Archaeology
The King's Manor
University of York
Y01 7EP, UK

Eric Kansa
School of Information
University of California, Berkeley
211 South Hall
Berkeley, CA 94720-4600
ekansa@ischool.berkeley.edu

Sarah Whitcher Kansa
The Alexandria Archive Institute
125 El Verano Way
San Francisco, CA 94127
skansa@alexandriaarchive.org

Fredrick Limp
University of Arkansas
304 JBHT
Fayetteville, AR 72701
fred@cast.uark.edu

Michael Rains
York Archaeological Trust
admin@yorkarchaeology.co.uk

Julian Richards
Archaeology Data Service
Department of Archaeology
The King's Manor
University of York
Y01 7EP, UK
jdr1@york.ac.uk

Stewart Waller
Archaeology Data Service
Department of Archaeology
The King's Manor
University of York
Y01 7EP, UK

Ethan Watrall
Department of Anthropology
MATRIX: The Center for the Humane Arts, Letters, and Social Sciences Online
418 Natural Sciences
Michigan State University
East Lansing, MI 48824
watrall@msu.edu

Willeke Wendrich
UCLA, Cotsen Institute of Archaeology
308 Charles E Young Dr. North
A210 Fowler Building/Box 951510
Los Angeles, CA 90095-1510
wendrich@humnet.ucla.edu

Jingfeng Xia
755 W. Michigan Street, UL 3100N
SLIS - IUPUI
Indianapolis, IN 46202
xiaji@iupui.edu

Ziqi Zhang
Web Intelligence Technologies Lab
Organisations, Information and Knowledge (OAK) Group
Department of Computer Science
University of Sheffield
S1 4DP, UK

Preface

This volume is the outcome of a session held at the 2008 meeting of the Society for American Archaeology (SAA) in Vancouver, British Columbia. The session carried the title "Web 2.0 and Beyond: New Tools for Archaeological Collaboration and Communication" and was organized by the editors. The majority of the chapters in this volume are based on presentations given in that session. However, in order to keep up with the sometimes fast pace of technological change, all contributors made significant updates to their chapters in mid-2010 to make the information as current as possible. In addition to the conference papers, the editors solicited two additional chapters (Chapters 7 and 10). These two chapters bring perspectives that we feel offer a more complete picture of the current uses and challenges of new Web-based technologies in archaeological communication.

This volume is somewhat experimental, being the first volume published by the Cotsen Institute of Archaeology Press in their new Cotsen Digital Archaeology (CDA) series, which uses the University of California's eScholarship framework for digital publishing. Through eScholarship, the contents of this volume are available online for free, though to accommodate the expectations of traditional print publication, print-on-demand copies are also available for a fee. Publication of this volume, open access, in a primary electronic medium, seems well suited for its subject matter. It will be interesting to see if open access venues become more accepted and expected as scholarly communications continue to evolve.

Finally, readers should be aware that this volume does not represent a comprehensive overview of digital archaeology. Because this volume originated from a Society for American Archaeology conference, it does not explore digital initiatives outside of North America or the United Kingdom. Thus, impressive achievements of colleagues in Europe, Asia, and elsewhere are not represented. This book also does not cover the equally impressive work of museum researchers, classicists, and scholars of the ancient Near East and Pacific. While this volume is not comprehensive, the sampling of perspectives presented here can nevertheless inform more general debates over issues of accessibility, sustainability, information quality, and semantics that cross-cut disciplinary and geographic boundaries.

ACKNOWLEDGMENTS

First and foremost, we would like to thank the contributors to this volume for sharing their valuable perspectives. The William and Flora Hewlett Foundation sponsored the session at the 2008 Society for American Archaeology (Vancouver, BC) where the contributions to this volume were first presented. A grant from the National Endowment for the Humanities and the Institute of Museum and Library Services (PK-50072-08) provided a portion of the support needed for the organization and production of this volume. Finally, we are grateful to Carol Leyba for her dedicated editorial work, Willeke Wendrich for her encouragement to publish in the Cotsen Digital Archaeology series, and two anonymous reviewers for their constructive comments on the original version of this volume.

Introduction

New Directions for the Digital Past

Eric C. Kansa

This volume looks at archaeology in the context of the World Wide Web, a communication system that has witnessed over two decades (and counting) of exponential growth. In many ways, the Web represents a revolution in communications and information sharing that rivals in significance the invention of the printing press or the origins of writing. In the past two decades, the Web has come to permeate virtually every aspect of our lives, transforming journalism, the arts, commerce, and the way we socialize.

Who Should Read this Book and Why

> As long as the centuries continue to unfold, the number of books will grow continually, and one can predict that a time will come when it will be almost as difficult to learn anything from books as from the direct study of the whole universe. It will be almost as convenient to search for some bit of truth concealed in nature as it will be to find it hidden away in an immense multitude of bound volumes.
>
> —Denis Diderot, "Encyclopédie" (1755), quoted in Wikipedia

As part of this transformation, the Web's growth means that we live in an information-saturated era. This volume adds minutely to this ever expanding body of data. While the Web increasingly enriches our work and social lives with new information, we face often overwhelming demands on our attention. Not only do we have new sources of news, information, and entertainment to jostle for our attention, we have many new sources of disinformation, spam, propaganda, and plain junk. It is no wonder that many scholars, overloaded with information, lament the passing of bygone days of quiet scholarly contemplation (Harley et al. 2010). No doubt many of the

complaints about today's Web reflect some romanticism about the past. Even with the development of telegraph networks in the nineteenth century, people complained about information overload (Standage 1998:165). While the phenomenon of information overload may not be a new symptom of modernity, even some pioneers of cyberspace worry about the current data deluge and its impact on creativity and deep thinking (Lanier 2010).

Given all these competing demands on our limited attention, this book about archaeology on the Web requires some justification. Instead of merely feeding information overload, we hope this volume will help archaeologists take stock and better understand how the Web is transforming the professional practice of archaeology, just as it transforms professional communications in other disciplines. Reflection on these changes can help us better understand the state of this discipline. Moreover, researchers who study scholarly communications more generally will find this book a useful source of case studies. Archaeology is an inherently multidisciplinary enterprise, with one foot in the humanities and interpretive social sciences and another in the natural sciences. As such, case studies in digital archaeology can help illuminate changing patterns in scholarly communications across a wide array of disciplinary contexts.

Archaeologists will find this book a useful guide in understanding how this revolution in communications technology reverberates across this discipline. Many of the contributions describe technologies, user interface designs, and organizational practices that attempt to mitigate some of the problems associated with the Web, especially information overload and disinformation. Contributions to this book also explore how the Web can be used to transform archaeological communications into forms that are more open, inclusive, and participatory. Some discussions focus on ways that Web-based systems can make archaeological knowledge production more open and transparent, while others focus on the challenges of archiving and preserving digital data. Finally, some chapters describe case study examples of digital projects. Sharing these experiences can provide useful guidance for other researchers wanting to create and apply technology to archaeology.

LOOKING BACK AT WEB 2.0

The book is loosely themed on so-called Web 2.0 approaches to these issues. Most contributions presented here derive from presentations given at the

2008 Society for American Archaeology conference in Vancouver, British Columbia. However, as is typical of many scholarly publishing cycles, transitioning conference presentations to a book form has taken about three years. Thus, many of the new technologies and perspectives presented at the 2008 conference are no longer so new. Nevertheless, the slower pace of academic publishing does offer some advantages. The lengthier review and revision cycle gave many of our contributors some added perspective and a chance for more nuanced reflection. Thus, many discussions of Web 2.0 presented in this volume have an element of retrospection.

Originally coined in 2004 by Tim O'Reilly, founder and CEO of O'Reilly Media, the term "Web 2.0" describes social networking systems, blogs, and other Web-based platforms emphasizing collaboration and sharing, rather than the unidirectional flow of information in a traditional Web 1.0 architecture (O'Reilly 2005). Many archaeologists have embraced Web 2.0 tools and technologies, allowing them to integrate different bodies of content and develop new tools and interfaces for peer-to-peer communication and collaboration. Ultimately, these new tools and platforms allow opportunities for research and public participation in archaeology.

However, "Web 2.0" is fast becoming a clichéd and obsolete expression. It stems from the revival of investment and interest in the Web following the dot-com collapse at the turn of the century. Now, in the aftermath of another and far graver financial collapse, many of the startups branded as Web 2.0 will likely fail. The term "Web 2.0" soon may be consigned to history.

The Promise of Web 2.0

As archaeologists accustomed to dealing with "deep time," it makes sense to consider the Web's impact on the discipline with a longer time horizon than is typical of most discussions of Web 2.0. While much about Web 2.0 will be a passing trend, the term still points to perspectives and developments likely to have lasting value and significance. In general, the term evokes designs and services emphasizing user interaction and the user as a source of extra value. This value can take multiple forms, including (but not limited to):

- *User-generated content:* Many Web 2.0 systems provide users with platforms for sharing and publishing content. These range from images to videos, and from essays to short, 140-character messages in microblogging services like Twitter.

- *Crowd-sourced classification:* Many Web 2.0 systems provide mechanisms for users to create and share metadata ("information about information") that can describe content to facilitate search and retrieval. Tagging and folksonomy (informal classification) systems provide important information retrieval services.

- *Remixable data*: Providing Web-based data in ways that can be easily manipulated by software represents another common theme of Web 2.0. Web services, often called "application program interfaces" (APIs), publish data on the Web in formats intended for use by third-party software. The intent behind an API is often to "crowd-source" interesting and useful software applications based on content. Google Maps[1] is currently an excellent example. Google provides mapping data and tools to third-party web developers. These developers install Google Maps on their sites, and in the process, Google (and its brand) becomes an increasingly ubiquitous feature of the Web.

- *Enhancing and evaluating information quality:* Web 2.0 often has connotations of an anarchic free-for-all, lacking traditional gate-keeping mechanisms to maintain quality. While information quality concerns rightly make researchers skeptical of some Web 2.0 platforms, in other contexts Web 2.0 systems try to promote quality. Sometimes Web 2.0 evaluations of quality are little more than popularity contests. In other cases, experimental scientific journals, such as *PlosOne*,[2] attempt to use Web 2.0–style rating and commenting systems as a new form of "enhanced peer-review."

Leveraging user communities and user interactions in the ways described above will probably continue to feature in future technology developments, in both popular and scholarly media.

As is evident to many of us, the speed of implementation, rate of adoption, and impacts of drawing value from the efforts of users can be highly uneven. Developments in commercial web services, technologies, and tools are transforming the professional and personal lives of archaeologists, often by blurring the boundaries between these lives. The founders of Google,[3] one

[1] http://maps.google.com/
[2] http://plosone.org
[3] http://www.google.com/

of the first and most successful Web 2.0 companies, discovered a powerful method for finding value in the distributed work of millions, with the PageRank algorithm (Brin and Page 1998). Google searches, as well as more specialized services such as Google Scholar,[4] have transformed the way researchers find information (Markey 2007; Yu and Young 2004). Google has also changed the way university libraries serve their patrons, influencing them to adopt single-box "full-text" search and stronger ranking algorithms.

Web 2.0 increases the diversity of content that people can find readily on the Web. Social networking sites, blogs, shared web bookmarks (Delicious.com[5] and the like), ratings systems (Digg[6]), and shared media (Flickr[7] and YouTube[8]) have expanded the range of media that people publish online. However, trends impact scholarly communications in slower and more tentative ways. While the peer-reviewed, journal-published paper is still the main currency of professional research communication, an expanding number of research-themed blogs make less formally published material available. More archaeologists now publish images to Flickr, share presentations on SlideShare,[9] and participate in open access publication (usually through self-archiving).

While Web 2.0's impact is far reaching, it does seem to have limits. Web 2.0 platforms and services mainly facilitate informal communications among archaeologists. Web 2.0 systems are simple to use, fast, and geared to content that requires relatively minimal investment to create. Archaeologists tend not to use Web 2.0 platforms as the primary dissemination channel for forms of content that take a great deal of effort and expertise to create. In this light, data sets and sophisticated scholarly manuscripts see less circulation in Web 2.0 channels.

Finding Web 2.0 Solutions for Primary Data?

If related data and documents can be linked together in a scholarly information infrastructure, creative new forms of data- and information-

[4] http://scholar.google.com/schhp?hl=en&tab=ws
[5] http://www.delicious.com/
[6] http://digg.com/
[7] http://www.flickr.com/
[8] http://www.youtube.com/
[9] http://www.slideshare.net/

> intensive, distributed, collaborative, multidisciplinary research and learning become possible. Data are outputs of research, inputs to scholarly publications, and inputs to subsequent research and learning. Thus they are the foundation of scholarship (Borgman 2007: 115).

Borgman describes a desired goal for so-called cyberinfrastructure systems that support research through more efficient communication and data preservation. Unfortunately, most popular Web 2.0 tools and services cannot deal with the complexities required of such a system. Flickr, Google Docs,[10] and other applications make certain aspects of archaeological fieldwork convenient to publish online, but they cannot describe key contextual information in a precise, consistent, and machine-readable way. For instance, contextual relationships are difficult to precisely describe in Flickr's annotation (tagging) system, limiting Flickr's usefulness in publishing images from an archaeological excavation. Similarly, Google Docs, ManyEyes,[11] Swivel.com (now defunct), and other online systems for sharing tabular (structured) data limit dissemination of structured data to data structures that can be represented on single tables. This is sufficient for archaeological data sets like individual zooarchaeological analyses, but it does not work well for publishing data on complex, multidisciplinary archaeological projects involving data sets generated by several different specialists and describing complex contextual relations.

Most Web 2.0 systems are simple to use and place minimal requirements on end users to prepare and describe content. Users generally find it easy to retrieve relevant pictures or videos by searching content indexed by user-generated tags. Contributing (publishing) to a Web 2.0 system can be quite easy, because the content published is generally simple and described with limited and informal metadata, such as user-generated tags. So, Web 2.0 systems sufficiently serve many popular needs and applications. However, they have not been widely used in academic communities for content central to research. To respond to that problem, this volume presents recent advances in archaeological data sharing and explores how Web 2.0 services affect communication and collaboration in archaeology.

[10] http://docs.google.com/
[11] http://manyeyes.alphaworks.ibm.com/manyeyes/

OVERVIEW OF CONTRIBUTIONS

The chapters in this volume illustrate the possibilities and limitations of the Web in meeting the specialized needs and requirements of professional researchers. Because Web 2.0 is best understood loosely as a zeitgeist or even as a marketing term, contributors define and discuss Web 2.0 from their own perspectives. The chapters address various semantic, intellectual-property, technical, social, and professional challenges of networking archaeological information. They present different perspectives on conceptual, theoretical, and practical approaches to communicating archaeological knowledge with new technologies and platforms. Some have successfully implemented Web 2.0 tools and approaches, while others have rejected such approaches. Issues about information quality, audience, and authority also inform their discussion.

This volume shows how emerging digital forms of archaeological communication differ from traditional paper-based media, and how these differences require examination and rethinking of knowledge production processes. Many example projects in this volume are rich in structured data and multimedia content. Some of this content is generated "in real time" in active field programs and sees little editing or filtering before global dissemination. These projects hope to use the inherent capabilities of Web 2.0 technologies and platforms to make archaeology more collaborative and more transparent. However, they also raise difficult questions about information quality, information overload, intellectual property, and relationships between professional researchers, students, and different public communities.

This book is divided into themed sections to help highlight certain particularly salient points of discussion made by the various contributions. The section themes are summarized as follows:

- *Section I* focuses on information retrieval and information-access approaches, especially centered on gray literature and primary field data. These forms of content have traditionally seen little dissemination.
- *Section II* explores larger conceptual concerns regarding information access and management. The contributions in this section discuss practical as well as theoretical concerns inherent in various design choices for archaeology's computing infrastructure.

- *Section III* presents projects that aim to enhance collaboration in archaeology through various approaches, such as the adaptation and development of certain technologies for mobile field-based collaboration, coordination and data management of field-based researchers and other specialists, and collaboration among the international researcher community.
- *Section IV* addresses scholarly communications issues, with a particular emphasis on concerns over information quality and access in light of sustainability and preservation imperatives.

SECTION I

Chapter 1

In "The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing," Julian Richards and colleagues discuss design innovations in information retrieval and integration of different data services. A key problem for archaeological information sharing is information overload. Standard keyword search systems often retrieve too much irrelevant information or fail to deliver relevant information if keywords are not mapped to synonyms. These problems make keyword searches somewhat unreliable and prone to deliver different results depending on the keyword inputs. However, by applying faceted search mechanisms, researchers gain greater comprehension of an entire corpus of material and can progressively refine searches to obtain specific information relevant to their interests.

Beyond search methodologies, Richards et al. discuss emerging frontiers of archaeological information management. Techniques in natural language processing (NLP) promise to enhance the value of archaeological literature. Automated and semiautomated NLP techniques help address some limitations of Web 2.0 methods in narrow niche professional contexts. How do you crowd-source metadata creation via techniques like social tagging when there is no crowd? The archaeological research community is relatively small, typically has very specialized interests, and may not be especially interested in social collaborative tagging. Thus, NLP offers a viable strategy to generate metadata to improve information retrieval (through faceted search and other techniques) without requiring the action of a (nonexistent) "crowd."

Richards et al.'s discussion of faceted search and NLP techniques for metadata creation is on the cutting edge of archaeological informatics, indicating important features of the landscape of archaeological information retrieval for years to come. As these systems are developed and deployed, they will shape how professionals and members of the public encounter cultural heritage. In other words, our record of the past will be increasingly shaped and organized by algorithms. This trend will be a fascinating topic for future research and critique. Will these algorithms be part of unobserved background processes that rarely see scrutiny? Or will there be discussion and debate about who creates and deploys these algorithms, and for what agenda and purpose? How will different perspectives and agendas be accommodated? Many of the theoretical debates and concerns that shaped the content of archaeological literature may also emerge in the context of automated processes to organize and retrieve that literature.

Chapter 2

In Chapter 2, "Toward a Do-It-Yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing," Eric Kansa and Sarah Whitcher Kansa discuss how techniques of Web 2.0 systems can be applied for research applications. Simple web services delivering machine-readable data can help make archaeological information open and reusable for research, instruction, and creativity. However, fitting new modes of communication and collaboration into traditional research practices poses potentially insurmountable problems with regard to time, recognition, technical challenges, and workflows. These concerns have guided new developments to Open Context,[12] an open source publishing system designed to facilitate sharing, collaboration, and integration of archaeological content.

The Kansas discuss some of the successes and failures of Web 2.0 in Open Context. They discuss how folksonomies provoked some initial curiosity in the system but failed to engage enough users to create useful metadata. In contrast, other aspects of Web 2.0 seem to have greater long-term traction and significance for their project, particularly exposure of machine-readable data through web services. Approaches to design and delivery of

[12] http://opencontext.org/

machine-readable data through web services represent one legacy of Web 2.0 likely to have long-term impact and application for research data sharing. They explore how web services enable data reuse across different applications and collections. When data are made machine-readable, content can be freed from individual silos and used with content from other sources. These may include other archaeological collections or systems supporting data sharing in other disciplines. Content is also freed from a single mode of presentation and visualization. Web services therefore encourage development of new user-interface paradigms and greater flexibility in user interactions with diverse content. Finally, certain technical design perspectives encourage architectures that better support important scholarly conventions, including citation and linking.

Discussing the reluctance of some researchers to share data, they emphasize a strategy that casts data sharing as a form of publication, where many of the conventions for citation and editorial oversight used in narrative publication can be applied. This perspective has increasing traction across many scientific fields, as indicated by recent editorial comments in the journal *Nature* ("Data's Shameful Neglect," 2009). Beyond serving as a useful publication model, narratives also provide context and meaning for archaeological data sharing. Kansa and Whitcher Kansa discuss "tacit knowledge" and the implicit understandings and background required to make sense of archaeological data. They discuss transmission of tacit knowledge via formal classification systems and ontologies, or through social scholarship enabled by Web 2.0 systems. They highlight the need to integrate data publication with narrative and interpretive publication to make shared primary data intelligible and usable by a wider community. This last point is made by many other contributors to this volume.

SECTION II

Chapter 3

Historically, archaeologists became interested in computing and databases to control huge quantities of excavation data. They looked to the computer as a tool to retrieve and analyze information across multiple data sets and excavations to create broad syntheses. Unfortunately, the promise of digitally based meta-analysis has not panned out. The mass of digital material generated by archaeological activity is geographically distributed, fuzzy, incom-

plete, inconsistent, and often hard to access. The resulting complexity deluge presents a whole new set of problems for archaeology. Stuart Dunn's contribution in Chapter 3, "Poor Relatives or Favorite Uncles? Cyberinfrastructure and Web 2.0: A Critical Comparison for Archaeological Research" critically reviews Web 2.0 methods and technologies that address this emerging problem. He explores cyberinfrastructure/e-science and how it relates to Web 2.0 technologies and techniques in archaeology. Dunn divides the process of using archaeological data into collection and harvesting; analysis, integration, and interpretation; and social research. In these three domains, he explores various hallmark technologies and methodologies commonly associated with both Web 2.0 and cyberinfrastructure. In particular, he sees folksonomy as a way to supplement and enhance traditional taxonomies. He advocates a "spade to screen" documentation process, to ensure that methods used to author and create digital objects are transparent and attributable. Dunn concludes that the top-down approach of cyberinfrastructure and the bottom-up approach of Web 2.0 are not two irreconcilable models, but different layers in the same structure.

Chapter 4

In Chapter 4, "Archaeological Knowledge Production and Dissemination in the Digital Age" Robin Boast and Peter Biehl discuss the ways in which different cultural contexts shape information management, retrieval, and use. They explore the diversity of ontologies and classification systems among expert communities and others, including different indigenous communities. Their discussion begins with an exploration of archaeological approaches to knowledge creation, contrasting "classificatory" versus "interpretive" paths. Whatever interpretive approach is taken, tangible cultural heritage becomes embedded in intangible processes that shape understandings of that tangible heritage. They argue that online information systems are contact zones where different understandings collide and inform one another.

Boast and Biehl then discuss how different conceptual systems can inform one another through Web-mediated collaboration. They look at how museums and their educational programs try to bridge understandings between museum experts and various professional communities. However, as they note, museums typically concern themselves only with managing their own, expert-informed classifications and documenting their own collections.

They do little to document how different public communities understand these collections. The perspectives of outsiders, though present and expressed in museum educational performances, rarely end up being recorded or informing experts.

Alternatives to the model of "one-way" broadcasting of museum expert knowledge are emerging. Many of these use Web 2.0 ideas of two-way communication and participation in creating content and sharing ideas. Boast and Biehl's exploration of classification is of interest to those wishing to foster reciprocal information sharing across different community settings. Essentially, they find that categorizing cultural heritage, even in loosely structured and constrained "folksonomy" approaches, is of limited appeal and interest to many people outside of museum professional circles. They find much more interest in digital representations of material culture to support narratives. In other words, opening up museum collections for social tagging resonates less than encouraging storytelling.

This observation has important implications. Concerns over classification and standards for classification dominate thinking about cyberinfrastructure, the Semantic Web (or "Linked Data"), and cultural heritage data sharing. Many discussions of Web 2.0 and folksonomies emphasize classification issues in information sharing. However, while classification is important, it is not the only concern. In some cultural contexts, construction of narratives has greater priority. By looking at how cultural heritage information is used in different contexts, Boast and Biehl highlight the need to move beyond classification to other social uses of information. Their insights help guide future attempts to bridge gaps between museums and other communities and also highlight the importance of narrative even within academic and museum professional circles.

SECTION III

Chapter 5

Archaeological projects are rarely blessed with full-time, permanent staff. In some cases, part-time specialists are employed full time at other museums or institutions, or work as freelance archaeological specialists involved in a wide range of additional projects. In other cases, specialists focus so intently on their particular research interest that they have no real sense of the totality of the project. The result is that specialists often feel isolated or semidetached.

In Chapter 5, "Creating a Virtual Research Environment for Archaeology," Michael Rains discusses VERA (Virtual Environment for Research in Archaeology), a system attempting to address these issues. Funded by JISC (Joint Information Systems Committee), VERA is a Web-based virtual research environment (VRE) collaboratively developed by the University of Reading, University College London, and York Archaeological Trust. It is centered on the Silchester Town Life Project[13] at the University of Reading. This is a large-scale, ongoing excavation of part of the abandoned Roman town of Calleva Attrebatum at Silchester, approximately 80 km west of London. Silchester has used the Integrated Archaeological Database (IADB) as its data management system since the start of the project in 1996. Key aims of the VERA project include improving the flow of information from excavation, through analysis and research, to publication and dissemination, and developing a collaborative working environment involving all members of the project team. At the heart of the VERA interface is one or more interactive graphical representations or visualizations of part of the project database. For example, a particular phase in site development is displayed as a standard archaeological stratigraphy diagram. What is unique about VERA is that additional content from the project database, such as plans of stratigraphic units, photographs, and field notes, can be attached to the diagram. This process adds context to excavation materials and allows all Web-connected stakeholders, regardless of their location, to contribute to the project.

Chapter 6

Ethan Watrall's contribution in Chapter 6, "iAKS: A Web 2.0 Archaeological Knowledge Management System," proposes a system to leverage web services and Web 2.0 technologies. iAKS aims to solve many data collection, storage, and visualization challenges currently faced by archaeologists. Database field tools are difficult to find and are often expensive, complex systems developed on a project-specific basis. iAKS, by contrast, is a flexible design that can be used by many different projects with very different research designs. Most importantly, iAKS can be used in the field and offers different types of service, depending on whether a project has Internet access, a local server, or simply a hard drive. Content created using iAKS is converted into XML,

[13] http://www.silchester.rdg.ac.uk

making it easy to share, integrate with other content, and preserve. This holistic approach, from field-based design and data collection to data sharing and integration via the Internet, makes the iAKS system particularly promising. Most importantly, the architecture approaches Watrall advocates explore new possibilities in mobile computing, where global information systems and infrastructure can come together with handheld, field-ready devices (phones, tablet computers, etc.). These capabilities can make archaeology increasingly "glocal" (simultaneously global and local), as particular finds and contexts observed locally can be related to other digital documentation found on global information networks.

Chapter 7

In their contribution, "User-Generated Content in Zooarchaeology: Exploring the 'Middle Space' of Scholarly Communication," Sarah Whitcher Kansa and Francis Deblauwe review the emerging role of user-generated content in archaeological communications. As discussed above, archaeologists mainly participate in Web 2.0 platforms for less formal types of communication and sharing. This chapter, based on the experiences of people who actively manage and participate in Web 2.0 systems, explores why archaeologists sometimes see a valuable role for Web 2.0 channels.

This chapter makes a number of interesting points with regard to incentives for professional researchers to participate in online social media. Blogging, as a social media platform, has been widely used now for several years. While there are several professionally oriented archaeological blogs, this chapter notes that blogging is still a somewhat niche activity in the discipline. Nevertheless, despite the fact that most blog posts will see little comment, the impacts of blogging may be greater than is immediately apparent: they can help spark interest in a paper, a website, or a grant announcement through the passive engagement of readers, and even through improving the search engine exposure of web resources referenced in a blog post.

Similarly, community portals are commonplace on the Web but are less discussed in calls for disciplinary "cyberinfrastructure." One such portal, BoneCommons,[14] has seen various incarnations since its initial launch in 2006 and now enjoys continual use by the zooarchaeological community.

[14] http://alexandriaarchive.org/bonecommons/