Digital Humanities and Information Visualization: Innovation and Integration

14] Introduction
by Joan Beaudoin and Sarah Buchanan, guest editors

16] The Dynamics of Primary Source and Electronic Resource: The Digital Renaissance and the Post-Reformation Digital Library
by Jordan J. Ballor

20] A Brief Introduction to Data Mining Projects in the Humanities
by Jonathan Hagood

24] Pale Males 2.0: Revisiting a Traditional American Studies Project Using Digital Humanities Tools
by Stephanie Margolin

27] When Computers Read: Literary Analysis and Digital Technology
by Sarah Jones

31] Visualizing Social Connections in the Humanities: Beyond Bibliometrics
by Chris Alen Sula

36] Interactive Visualization for Multilingual Search
by Stan Ruecker, Ali Shiri and Carlos Fiorentino

41] From Records to Data with Viewshare: An Argument, An Interface, A Design
by Jefferson Bailey and Trevor Owens

45] Beyond the Score: Music Visualization and Digital Humanities
by Margaret Lam

The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus
by Leonard Will
The President’s Page in this issue should be required reading for all ASIS&T members. In order to improve the international appeal of ASIS&T, Diane Sonnenwald proposes that we change our name, while leaving the current acronym. The name she suggests is ASsociation for Information Science and Technology. She is opening discussion on this topic at www.quicktopic.com/47/H/bvJhSC8HTs, which will be active through May 1, 2012. The discussion will determine whether a formal proposal is made to the Board, so let your voice be heard.

Our special section abounds with SIG synergy. Experience it first hand! SIG/VIS (Visualization, Images and Sound) and SIG/AH (Arts & Humanities) have combined to create a special section entitled Digital Humanities and Information Visualization: Innovation and Integration. As the editors state, the articles range “from the theoretical to the practical, from those that identify how the mere act of digitizing and providing electronic access to resources can benefit scholarship to those that present highly sophisticated tools and techniques for data analysis and display.” Whereas our previous issue focused on museum informatics (www.asist.org/Bulletin/Feb12/Bulletin_FebMar12_Final.pdf), this one is concerned with the role of visualization to expand research in literature, history, film and music. Together they provide many excellent examples of the use of new technologies to support the humanities and cultural heritage institutions.

Leonard Will has contributed a feature article which provides a guide to the extensive new data model for a thesaurus structure underlying the recently published international standard ISO 25964-1:2011 – Thesauri for Information Retrieval. Changes include, for example, a clear distinction between “term” and “concept.” The article speaks to a core interest of many of our readers.

Have you recently been incommunicado on Facebook? Posted great things that nobody saw? It happened to no less a navigation expert than our associate editor for information architecture, Thom Haller. In his IA Column, Thom discusses how his misinterpretation of an ambiguous menu led to this unhappy state of affairs and how the choices might have been improved.

Interaction is considered to be a good thing these days, so please be interactive and remember to pass along your thoughts on the name change to Diane.
Dear Colleagues,

International organizations, such as ASIS&T, have responsibilities and considerations that local and national organizations do not have. International organizations also provide many benefits to their members that local and national organizations cannot provide. These include the opportunity to learn from – and share expertise and knowledge with – colleagues who have different expertise and knowledge. Organizations can be international in their deeds (what they do) and words (what they say.) Being international in both deeds and name is often vital to organizations such as ours and a strong indicator of our collective embrace of geographic diversity.

Deeds

Recognizing that individuals around the globe face different economic and social conditions, the Board recently unanimously approved a motion to offer a reduced membership fee to individuals residing in countries that the United Nations identifies annually as countries with medium and low human development achievement using the Human Development Index. The UN Human Development Index is updated annually in the UN’s Human Development Report [1] and takes into account economics and “the range of things that people can do or be in life” [2]. The reduced fee is $40.00 per year, which is identical to the current fee for students and members experiencing financial difficulties.
The motion was based on a proposal developed by Daniel Alemneh, SIG/III president, in collaboration with SIG/III officers. Our thanks to Daniel and his colleagues for their hard work in developing the proposal.

**Words**

When I was growing up, common phrases were “actions speak louder than words” and “sticks and stones may break my bones but words will never hurt me.” As I grew older, I learned that contrary to these sayings, words matter. What we name things influences behavior. As Buddha eloquently said, “Whatever words we utter should be chosen with care for people will hear them and be influenced by them for good or ill.” Jean-Paul Sartre also commented, “Words are more treacherous and powerful than we think” [3].

One word in the name of our organization has different implications for members. The word *American* can make it difficult for members outside the United States to get support and receive recognition for belonging to and participating in ASIS&T. For other members, it may reflect a national pride and distinction. The name is a branding that, depending on our various perspectives, can be seen to add prestige or make it challenging to participate in the organization.

Some members have suggested keeping the name and establishing close ties with similar organizations outside the United States. This approach was tried for several years; however, ASIS&T has much greater disciplinary breadth than other information science organizations, and matching counterparts in other countries often don’t exist. Our membership includes practitioners, researchers and institutions, and we focus on the interplay of people, information, technology and social structures. ASIS&T is unique.

Others have suggested it’s too soon to change our name again. As you may know, our organization was founded in 1937 as the American Documentation Institute (ADI). In 1968 the name officially became the American Society for Information Science. In 2000 “and Technology” was added to the name. Yet if we keep the acronym, ASIS&T, might this concern be mitigated? And couldn’t changing the name help show we are not a static organization, but a dynamic, thoughtful, learning organization that embraces geographic diversity?

What might ASIS&T stand for, if not “American Society for Information Science and Technology”? Suggestions range from “Awesome Society for Information Science and Technology” to “Association for Information Science and Technology.” (Truth be told there were a few more suggestions but it’s probably best not to mention the other adjectives beginning with the letter “a” which were suggested...) The word *association* is defined as “an organization of people with a common purpose and having a formal structure...friendship, companionship...connection or combination” [4]. These meanings seem very applicable to ASIS&T. We have a common purpose to “advance the information sciences and related applications of information technology by providing focus, opportunity and support to information professionals and organizations” [5]. There are many friendships among members. We connect ideas and practices through our interactions and sharing at Annual Meetings, workshops, seminars, webinars, papers and other documents.
To change any aspect of our name requires a vote by the membership (which has declined 28% since 2005). Three quarters of the members voting on this type of proposed change is needed to approve the change. I would like to invite all members to comment on the proposal to keep the acronym and change its wording to “ASsociation for Information Science & Technology.” A discussion page has been created for this purpose. To access the discussion page, go to www.quicktopic.com/47/H/bvJVhSC8HTs. The page will be open for comments until May 1, 2012. At that time the Board will consider whether or not to call for a membership vote on this issue.

William Morris remarked, “It took me years to understand that words are often as important as experience, because words make experience last” [3]. Thus I urge all members to thoughtfully listen to, and constructively respect, different perspectives on this issue.

Acknowledgement

My thanks to Linda Smith, Mei-Mei Wu, Dick Hill and Andrew Dillon for their comments on a draft of this column.

Resources Mentioned in the Column


Dr. Gloria Leckie, professor emeritus of the University of Western Ontario (UWO), Canada, delivered the 2012 ASIS&T Lecture Series presentation, sponsored by the School of Information Studies, Charles Sturt University, the recipient of the 2nd ASIS&T Lecture Series award. The speech was delivered at the National Library of Australia, Canberra as this issue of the Bulletin was being prepared.

Dr. Leckie’s talk – From Facebook to Twitter and Into the Cloud: Where is Library and Information Science in our Googleized World – explores the meaning of libraries and information centers in today’s technological world, through a discussion of educational and research-related issues.

ASIS&T created the ASIS&T Lecture Series to promote the progress of information science and technology. Charles Sturt University is the first Australian host for this series.

Dr. Leckie holds a master of library of information science, as well as a master of arts and a PhD in geography. During her professional career, she has worked in a number of different positions as a research librarian in the Canadian government; coordinator for the Cataloguing in Publication Program for Eastern Canada; and in academic roles. Currently, Dr. Leckie continues to work with her doctoral students and gives lectures and workshops. She has also taken on faculty associate position in UWO’s Teaching Support Centre, where her mandate is to examine North American trends in the development and characteristics of professional master’s programs across a wide variety of disciplines; she also advises faculties and departments who wish to develop such programs.

In her speech, Dr. Leckie explored the three faces of library and information science: as an academic discipline, as an educational space for the modern librarian and as a profession that cares about the role of information in society. Each aspect was examined through specific topics, including the changing role and worth of the library, the need to be proactive with technology (e.g., digital preservation projects), how we can (and should) interact with other academic disciplines and the challenges faced by educational institutions in preparing the librarians of the future.

Deadline Extended for History of ASIS&T Pre-Conference

The ASIS&T 75th Anniversary Task Force has made a few alterations to the call for papers for the 2012 pre-conference on the history of ASIS&T and the fields of information science and technology. That’s good news on two fronts: the call is undoubtedly improved, and the deadline for submitting abstracts is extended to March 19. The pre-conference will be held immediately prior to the 2012 Annual Meeting of ASIS&T, October 26-31, 2012, in Baltimore, Maryland.

This conference will explore the 75 year history of ASIS&T and the longer history of information science and technology worldwide. Abstracts of papers (maximum 1,000 words) not previously published or submitted elsewhere are eligible. All submissions will be refereed by a panel of experts and accepted papers will be published as a web-based publication prior to the conference. Submissions by students and authors outside the United States are particularly encouraged.

Among the topics detailed in the revised call for papers are the following: continued on next page
continued from previous page

- development of the foundational ideas and theories of information science and its earlier name, documentation;
- forces that contributed to shaping the research agenda and scope of the field of information science worldwide over the past 75 years and how the field might evolve over the next 75 years;
- major researchers and developers of the ideas, practices and theories in information science and closely related fields of study;
- evolution of the multifaceted and multiple-named field of information science with a focus on what the past says about the future of information science research and development;
- treatment of the origin and evolution of information science ideas and practices specific to an individual country;
- historical contexts of major technological innovations, and the impacts they have had on societies, organizations, governments or individuals worldwide;
- development of ASIS&T (and its predecessors, the American Documentation Institute and the American Society for Information Science) as a professional organization and scholarly society in context with its times;
- role of government (nationally or internationally) in influencing the development of information science and technology, with particular attention to the support of research and development of information policies;
- development of information science education and its relationships with other fields of study.

For more information and questions about this call for papers contact Robert V. Williams at bobwill(at)sc.edu or call 803-777-2324.

Abstracts should be submitted electronically to www.softconf.com/assist2/Hist_Pre_Con/.

---

SIG/USE Starts Anniversary History Project

SIG/Information Needs, Seeking and Use (SIG/USE) is the latest ASIS&T unit to launch a project in celebration of the upcoming 75th anniversary of the organization. In response to a call from the ASIS&T leadership for SIGs to help in the effort to preserve the Society’s history and its historical artifacts, SIG/USE is collecting papers, photos, presentations, links, videos and other memorabilia pertaining to the SIG’s storied past. Sanghee Oh, assistant professor at Florida State University, is leading the project with the SIG/USE wiki at http://siguse.wikidot.com set up to share all appropriate materials. If you have an item or information to contribute or would like to join the team working on the SIG/USE history project, contact the group at siguse_history@yahoo.com.

---

Deadlines Approaching for History Awards Entries

The ASIS&T History Fund, established by the ASIS&T Board of Directors in 2000, supports and encourages research and publication in the history of information science and technology. To that end, the fund sponsors two annual awards: Research Award and Best Paper Award. Nominations for both are due by May 1.

The ASIS&T History Fund Research Award goes to the best research support proposal on a topic relevant to the history of information science and technology. A proposal should state the central topic or question to be researched, qualifications of the researcher, a budget and how the funds will be expended. The award carries a maximum $1000 grant.

The ASIS&T History Fund Best Paper Award, a $500 award, will honor the best paper on a topic relevant to the history of information science and technology. The paper should not have been previously published or submitted to a journal, and it should not exceed 30 pages double-spaced, including notes/references, using APA style manual. If an award is made for 2012, the winner will be expected to present the paper at the 2012 ASIS&T Annual Meeting pre-conference and give first rights of refusal for publication to the Journal of the American Society for Information Science and Technology.

For more information about the awards, contact Lai Ma at lama@indiana.edu.

continued on next page
continued from previous page

The ASIS&T History Fund is supported by donations (including book royalties) from ASIS&T members and others with interest in the history of information science and technology. The Fund Advisory Board encourages donations from anyone interested in supporting historical study of information science and technology.

Members of the ASIS&T History Fund Advisory Board for 2012 are Kathryn La Barre, chair; Sarah Buchanan, chair-elect; Julian Warner, past chair; Michael Buckland; and Samantha Hastings.

News about ASIS&T Members

Linda C. Smith, immediate past president of ASIS&T and professor and associate dean in the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign, is the recipient of the 2012 Service Award from the Association for Library and Information Science Education (ALISE). Linda has served ALISE in a variety of capacities, including board service as secretary-treasurer; director for special interest groups; and vice president, president and past president. She joined ALISE as a doctoral student in 1977 and says the organization has been significant at “every stage in my career.”

Marjorie M. K. Hlava, former ASIS&T president and founder and CEO of Access Innovations, received the 2012 Ann Marie Cunningham award for outstanding service from the National Federation of Advanced Information Services (NFAIS). The honor, named for a former executive director of NFAIS, honors members who go above and beyond the normal expectations of service to the organization. Hlava was cited for what current president Keith MacGregor calls “her years of hard work as chair of the NFAIS Standards Committee” and her years of service in numerous other capacities, such as board member, former president, committee member and editor of meeting papers.

News from ASIS&T SIGs

SIG/Metrics (SIG/MET) is looking for entries in its second student paper contest designed to recognize promising student research relating to the interests of the SIG – the measurement of information production. The contest is open to any full time student regardless of ASIS&T membership. Submissions must be single-author, original and previously unpublished. Among the topics suggested by contest planners are metric-related theory; methods and new techniques; citation and cocitation analysis; altmetrics; webometrics; and research policy. Entry deadline is April 30 to www.easychair.org/conferences/?conf=sigmetricspc2012. Queries can be made to Chaoqun Ni at chni[at]Indiana.edu or Carrie Change at carriehc[at]gmail.com.

News from ASIS&T Chapters

The New England Chapter of ASIS&T (NEASIST) will present Going Mobile: Library Websites, Services and Apps on Mobile Devices, a half-day workshop designed to help libraries better understand their mobile users of today and the future; plan for mobile-friendly content; ensure contextually relevant information; and collaborate with technology partners. The April 23 program will be held at Massachusetts Institute of Technology.

The International Calendar of Information Science Conferences (icisc.neasist.org/) is a nonprofit collaboration between the Special Interest Group/International Information Issues (SIG/III) and the European (ASIST/EC) and New England (NEASIST) chapters of the American Society for Information Science and Technology, with the additional support of Haworth Press.
In Memoriam
Deborah Barreau

Message from Gary Marchionini, dean of the School of Information and Library Science at the University of North Carolina at Chapel Hill

It is with great sadness that I share news that our friend and colleague, Dr. Deborah Barreau, lost her battle with cancer in mid-February. Although some of you may not have had the opportunity to know Deborah, she was a talented and gifted teacher and researcher who specialized in personal information systems, organizational behavior and organizational communication as well as the design, development and use of information systems. She was a student of our school, graduating with her master’s in library science in 1986, before earning her Ph.D. from the University of Maryland at College Park. She was a Tar Heel at heart, receiving her bachelor’s and master’s degrees in sociology from the University of North Carolina at Chapel Hill.

She came to SILS in 2002 from Catholic University where she had been an assistant professor. Deborah was a wonderful addition to our faculty, most recently serving as the Frances Carroll McColl Term Professor. She always put our school above her own ambitions and did many things to make us civil, effective and collaborative. She did this through her wisdom, warm smile and willingness to go the extra step to give us all her time, advice and a sympathetic ear. She was an exceptional teacher, winning the school’s Award for Teaching Excellence twice and the outstanding teacher award from the American Society for Information Science and Technology (ASIS&T). Her involvement with professional associations such as ASIS&T, where she was elected to the Board as director-at-large; Association for Library and Information Science Education; Special Libraries Association; and others, demonstrated her extraordinary commitment to the field.

Deborah was not only a remarkable faculty member and researcher, she was also one of the most compassionate and understanding people I’ve had the opportunity to know. She truly cared about her students and their successes and her fellow faculty members and staff. I have never met anyone as selfless and devoted to others, especially to our students. She has been a great influence in my life, the lives of her students and colleagues and has brought true meaning to the words kindness and integrity. Deborah had a quote on her door that exemplified her beliefs and her life practice: “Everyone is indispensable.” She believed this and lived her life accordingly. Her spirit and inspiration are truly indispensable, and although we will miss her terribly, we are all better for having known her. Her death is an incredible loss to SILS and to the information and library science world.

Our thoughts and prayers are with Deborah’s family as they struggle with this great loss. A memorial tribute to her life will be held on Friday, April 27, 2012, at 2 p.m. in the George Watts Hill Alumni Center Ballroom on the UNC at Chapel Hill campus.

The SILS Award for Teaching Excellence, which Deborah won twice, has been renamed in her honor. Contributions to the Deborah Barreau Award for Teaching Excellence fund can be made by sending a check payable to UNC SILS with a note directing the donation to fund 2954 or online at https://secure.dev.unc.edu/gift/default.aspx?.
ASIS&T Presents Annual Audit

The report of the ASIS&T auditors on the 2011 financial statements is presented on this and the following pages.

AMERICAN SOCIETY FOR INFORMATION SCIENCE & TECHNOLOGY
FINANCIAL STATEMENTS AND INDEPENDENT AUDITORS’ REPORT
SEPTEMBER 30, 2011 AND 2010

INDEPENDENT AUDITORS’ REPORT

Board of Directors:
American Society for Information Science & Technology
Silver Spring, Maryland

We have audited the accompanying statements of financial position of the American Society for Information Science & Technology (a nonprofit organization) as of September 30, 2011 and 2010 and the related statements of activities, functional expenses, and cash flows for the years then ended. These financial statements are the responsibility of the Organization’s management. Our responsibility is to express an opinion on these financial statements based on our audit.

We conducted our audit in accordance with auditing standards generally accepted in the United States of America. Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatements. An audit includes examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing accounting principles used and significant estimates made by management, as well as, evaluating the overall financial statement presentation. We believe our audit provides a reasonable basis for our opinion.

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the American Society for Information Science & Technology as of September 30, 2011 and 2010, and the changes in its net assets and its cash flows for the years then ended, in conformity with accounting principles generally accepted in the United States of America.

Columbia, Maryland
January 12, 2012
### American Society for Information Science & Technology

#### Statement of Functional Expenses

**For the Year Ended September 30, 2011**

<table>
<thead>
<tr>
<th>Category</th>
<th>Membership</th>
<th>Conferences and Meetings</th>
<th>Publications</th>
<th>Continuing Education</th>
<th>Total Program Support</th>
<th>Management and General</th>
<th>Total Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awards and Honors</td>
<td>$17,469</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$17,469</td>
<td>$1,580</td>
<td>$19,049</td>
</tr>
<tr>
<td>Banquet and Hospitality</td>
<td>1,183</td>
<td>281,416</td>
<td>-</td>
<td>-</td>
<td>1,183</td>
<td>300,300</td>
<td>311,483</td>
</tr>
<tr>
<td>Books and Products</td>
<td>7,777</td>
<td>5,858</td>
<td>-</td>
<td>-</td>
<td>13,635</td>
<td>15,411</td>
<td>29,046</td>
</tr>
<tr>
<td>Chapter Direct Expenses</td>
<td>1,415</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15,411</td>
<td>15,411</td>
<td>-</td>
</tr>
<tr>
<td>Decoration and Accreditation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>23,364</td>
<td>23,364</td>
<td>-</td>
</tr>
<tr>
<td>Duplication</td>
<td>1,258</td>
<td>710</td>
<td>63</td>
<td>43</td>
<td>2,063</td>
<td>3,873</td>
<td>5,936</td>
</tr>
<tr>
<td>Editorial</td>
<td>26,000</td>
<td>2,836</td>
<td>43,854</td>
<td>-</td>
<td>51,684</td>
<td>10,684</td>
<td>62,368</td>
</tr>
<tr>
<td>Equipment Rental and Maintenance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16,561</td>
<td>16,561</td>
<td>16,561</td>
<td>16,561</td>
</tr>
<tr>
<td>Exhibits and Conference</td>
<td>3,212</td>
<td>282,994</td>
<td>-</td>
<td>-</td>
<td>3,212</td>
<td>31,985</td>
<td>35,201</td>
</tr>
<tr>
<td>Insurance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3,863</td>
<td>6,295</td>
<td>10,158</td>
</tr>
<tr>
<td>Legal, Accounting, and Auditing</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>118</td>
<td>53,435</td>
<td>53,553</td>
</tr>
<tr>
<td>Other</td>
<td>23,260</td>
<td>57,031</td>
<td>4,885</td>
<td>28,237</td>
<td>93,375</td>
<td>129,875</td>
<td>223,250</td>
</tr>
<tr>
<td>Postage and Delivery</td>
<td>2,073</td>
<td>11,410</td>
<td>877</td>
<td>342</td>
<td>14,352</td>
<td>53,895</td>
<td>58,247</td>
</tr>
<tr>
<td>Printing and Related Expenses</td>
<td>683</td>
<td>22,942</td>
<td>10,009</td>
<td>-</td>
<td>12,970</td>
<td>30,903</td>
<td>43,873</td>
</tr>
<tr>
<td>Proceedings</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9,377</td>
<td>1,957</td>
<td>11,334</td>
</tr>
<tr>
<td>Rent</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>39,902</td>
<td>39,902</td>
<td>39,902</td>
</tr>
<tr>
<td>Supplies, Benefits, and Taxes</td>
<td>30,062</td>
<td>165,812</td>
<td>21,092</td>
<td>22,111</td>
<td>311,477</td>
<td>250,631</td>
<td>562,108</td>
</tr>
<tr>
<td>Taxes</td>
<td>2,567</td>
<td>2,567</td>
<td>-</td>
<td>-</td>
<td>5,134</td>
<td>5,134</td>
<td>5,134</td>
</tr>
<tr>
<td>Travel and Related Expenses</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>25,000</td>
<td>25,000</td>
<td>25,000</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>$140,067</td>
<td>$627,410</td>
<td>$111,229</td>
<td>$22,911</td>
<td>$1,207,270</td>
<td>$467,229</td>
<td>$1,674,509</td>
</tr>
</tbody>
</table>

See Independent Auditors’ Report and Notes to Financial Statements.

---

### American Society for Information Science & Technology

#### Statement of Functional Expenses

**For the Year Ended September 30, 2010**

<table>
<thead>
<tr>
<th>Category</th>
<th>Membership</th>
<th>Conferences and Meetings</th>
<th>Publications</th>
<th>Continuing Education</th>
<th>Total Program Support</th>
<th>Management and General</th>
<th>Total Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awards and Honors</td>
<td>$13,888</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$13,888</td>
<td>$233,737</td>
<td>$247,625</td>
</tr>
<tr>
<td>Banquet and Hospitality</td>
<td>5,054</td>
<td>184,415</td>
<td>-</td>
<td>-</td>
<td>5,054</td>
<td>233,737</td>
<td>238,791</td>
</tr>
<tr>
<td>Books and Products</td>
<td>9,124</td>
<td>-</td>
<td>3,796</td>
<td>-</td>
<td>12,920</td>
<td>12,920</td>
<td>12,920</td>
</tr>
<tr>
<td>Chapter Direct Expense</td>
<td>1,141</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1,141</td>
<td>1,141</td>
<td>1,141</td>
</tr>
<tr>
<td>Decoration and Accreditation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>25,364</td>
<td>25,364</td>
<td>25,364</td>
</tr>
<tr>
<td>Duplication</td>
<td>1,270</td>
<td>2,470</td>
<td>6,459</td>
<td>4,542</td>
<td>6,459</td>
<td>8,991</td>
<td>15,440</td>
</tr>
<tr>
<td>Editorial</td>
<td>4,521</td>
<td>15,922</td>
<td>1,802</td>
<td>509</td>
<td>15,922</td>
<td>27,587</td>
<td>43,514</td>
</tr>
<tr>
<td>Equipment Rental and Maintenance</td>
<td>5,255</td>
<td>23,181</td>
<td>18,896</td>
<td>250</td>
<td>10,745</td>
<td>10,745</td>
<td>10,745</td>
</tr>
<tr>
<td>Exhibits and Conferences</td>
<td>18,043</td>
<td>84,418</td>
<td>-</td>
<td>3,655</td>
<td>3,655</td>
<td>26,052</td>
<td>29,707</td>
</tr>
<tr>
<td>Insurance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5,275</td>
<td>5,275</td>
<td>5,275</td>
</tr>
<tr>
<td>Fines and Penalties</td>
<td>4,444</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Legal, Accounting, and Auditing</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>40,584</td>
<td>40,584</td>
<td>40,584</td>
</tr>
<tr>
<td>Other</td>
<td>36,690</td>
<td>100,941</td>
<td>760</td>
<td>32,693</td>
<td>239,151</td>
<td>267,842</td>
<td>507,003</td>
</tr>
<tr>
<td>Postage and Delivery</td>
<td>4,521</td>
<td>15,922</td>
<td>1,802</td>
<td>509</td>
<td>15,922</td>
<td>27,587</td>
<td>43,514</td>
</tr>
<tr>
<td>Printing and Related Expenses</td>
<td>5,255</td>
<td>23,181</td>
<td>18,896</td>
<td>250</td>
<td>10,745</td>
<td>10,745</td>
<td>10,745</td>
</tr>
<tr>
<td>Proceedings</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7,616</td>
<td>7,616</td>
<td>7,616</td>
<td>7,616</td>
</tr>
<tr>
<td>Supplies, Benefits, and Taxes</td>
<td>78,273</td>
<td>162,482</td>
<td>27,960</td>
<td>7,630</td>
<td>298,157</td>
<td>257,836</td>
<td>555,993</td>
</tr>
<tr>
<td>Taxes</td>
<td>4,400</td>
<td>3,329</td>
<td>-</td>
<td>214</td>
<td>6,729</td>
<td>6,729</td>
<td>6,729</td>
</tr>
<tr>
<td>Telephone</td>
<td>118</td>
<td>2,341</td>
<td>31</td>
<td>1,100</td>
<td>2,483</td>
<td>4,734</td>
<td>7,217</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>$146,089</td>
<td>$269,297</td>
<td>$35,260</td>
<td>$21,827</td>
<td>$390,396</td>
<td>$429,290</td>
<td>$819,686</td>
</tr>
</tbody>
</table>

See Independent Auditors’ Report and Notes to Financial Statements.
# AMERICAN SOCIETY FOR INFORMATION SCIENCE & TECHNOLOGY
## NOTES TO FINANCIAL STATEMENTS
### SEPTEMBER 30, 2011 AND 2010

### PURPOSE OF ORGANIZATION
The American Society for Information Science & Technology (the Society) is a non-profit (501)(c)(3) professional association organized for scientific, literary and educational purposes. The Society is dedicated to the creation, organization, dissemination and application of knowledge concerning information and its transfer. The mission of the Society is to foster and lead the advancement of information science and technology. The Society is headquartered in the Washington, DC area. The Society’s membership base is primarily in North America, but the Society has members throughout the world.

### SUMMARY OF SIGNIFICANT ACCOUNTING POLICIES
**Basis of Accounting:** The Society’s financial statements have been prepared on the accrual basis of accounting and, accordingly, reflect all significant receivables, payables, and other liabilities.

**Basis of Presentation:** The Society is required to report information regarding its financial position and activities according to three classes of net assets: unrestricted net assets, temporarily restricted net assets, and permanently restricted net assets.

There were no temporarily restricted or permanently restricted net assets at September 30, 2011 and 2010.

**Cash and Cash Equivalents:** Cash and cash equivalents includes all highly liquid investments with a maturity of three months or less when purchased to be cash and cash equivalents.

**Allowance for Doubtful Accounts:** Accounts receivables are reported net of an allowance for doubtful accounts. The allowance is based on management’s estimate of the amount of receivables that will actually be collected. The allowance for doubtful accounts is based specifically on the identification of uncollectible accounts and the Society’s historical collection experience. The allowance was $1,179 and $1,179 at September 30, 2011 and 2010, respectively.

**Royalties Receivable:** The Society receives royalties for publications sold during the year. The Society records the receivable and the related revenue for royalties earned as of September 30 each year. An allowance for uncollectible amounts has not been established for royalties receivable as the Society deems all royalties to be collectible.

**Prepaid Expenses:** Prepaid expenses represent costs associated with Society activities paid prior to year-end whose benefits will be realized by the Society as expenses are incurred.

**Inventory:** Inventory, consisting of publications and other items for sale, is recorded at the lower of cost or market using the first-in, first-out (FIFO) method.

**Property and Equipment:** The Society follows the practice of capitalizing, at cost, all expenditures for property and equipment in excess of $500. Property and equipment are depreciated using the straight-line method over the useful lives of three to eight years. Leasehold improvements and software costs are amortized over ten years. The web design costs are amortized over seven years.

**Functional Allocation of Expenses:** The costs of providing the various programs and other activities have been summarized on a functional basis in the Statement of Activities. Accordingly, certain costs have been allocated among the programs and supporting services benefited.

**Deferred Revenue:** The Society’s deferred revenue represents membership dues, conference registration fees and publication subscriptions for which services have not yet been provided.

See Independent Auditors’ Report.

---

### AMERICAN SOCIETY FOR INFORMATION SCIENCE & TECHNOLOGY
## STATEMENTS OF CASH FLOWS
### FOR THE YEARS ENDED SEPTEMBER 30, 2011 AND 2010

<table>
<thead>
<tr>
<th>Description</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash Flows From Operating Activities</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>Change in Net Assets</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>Adjustments to Reconcile Change in Net Assets to Net</td>
<td>$267,020</td>
<td>$178,826</td>
</tr>
<tr>
<td>Cash Provided by Operating Activities:</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>Depreciation and Amortization</td>
<td>23,644</td>
<td>25,067</td>
</tr>
<tr>
<td>Changes in Operating Assets and Liabilities</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>Increase/Decrease in Assets</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>Accounts Receivable</td>
<td>(3,398)</td>
<td>(4,365)</td>
</tr>
<tr>
<td>prepaid Expenses - General</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td>Prepaid Expenses - Conferences</td>
<td>5,235</td>
<td>(10,098)</td>
</tr>
<tr>
<td>Prepaid Expenses - Taxes</td>
<td>(60)</td>
<td>368</td>
</tr>
<tr>
<td>Postage Deposited</td>
<td>1,826</td>
<td>440</td>
</tr>
<tr>
<td>Increase/Decrease in Liabilities</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>Accounts Payable</td>
<td>11,835</td>
<td>(2,253)</td>
</tr>
<tr>
<td>Income Taxes Payable</td>
<td>148</td>
<td>320</td>
</tr>
<tr>
<td>Accrued Expenses</td>
<td>(20,477)</td>
<td>46,678</td>
</tr>
<tr>
<td>Deferred Revenue</td>
<td>(60,871)</td>
<td>104,688</td>
</tr>
<tr>
<td>Net Cash Provided by Operating Activities</td>
<td>190,204</td>
<td>903,068</td>
</tr>
</tbody>
</table>

| Cash Flows From Investing Activities             |          |----------|
| Purchase of Property and Equipment               | (4,846)  | (711)    |
| Net Cash Used in Investing Activities            | (4,846)  | (711)    |

| Net Increase in Cash and Cash Equivalents        | 107,876  | 362,557  |

| Cash and Cash Equivalents, Beginning of Year     | 1,840,147| 1,479,690|
| Cash and Cash Equivalents, End of Year           | $2,029,826| $1,842,147|

<table>
<thead>
<tr>
<th>Cash and Cash Equivalents at September 30 Consolidated of:</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash and Cash Equivalents - Headquarters</td>
<td>$1,906,613</td>
<td>$1,717,873</td>
</tr>
<tr>
<td>Cash and Cash Equivalents - Local Chapters</td>
<td>$102,352</td>
<td>$104,476</td>
</tr>
<tr>
<td>Total</td>
<td>$2,009,825</td>
<td>$1,822,447</td>
</tr>
</tbody>
</table>

### Supplemental Disclosure of Cash Flow Information:

<table>
<thead>
<tr>
<th>Description</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash Paid During the Year for Income Taxes</td>
<td>$1,805</td>
<td>$730</td>
</tr>
</tbody>
</table>

See Independent Auditors’ Report and Notes to Financial Statements.
American Society for Information Science & Technology
Notes to Financial Statements
September 30, 2011 and 2010

Summary of Significant Accounting Policies (Continued)

Use of Estimates: The preparation of financial statements, in conformity with generally accepted accounting principles, requires management to make estimates and assumptions that affect certain reported amounts and disclosures. Accordingly, actual results could differ from those estimates.

Income Taxes: The Society is exempt from federal and state income tax (except taxes on unrelated business income) under Section 501(c)(3) of the Internal Revenue Code. Although the Society is exempt from income tax, it has certain activities considered unrelated to its exempt status that are subject to income tax. The Society has accrued taxes of $464 and $652 for unrelated business income for the years ended September 30, 2011 and 2010, respectively.

Reclassifications: Certain balances in the prior year financial statements have been reclassified for comparative purposes to conform to the presentation in the current year financial statements. These reclassifications have no impact on net income.

Subsequent Events: Management has evaluated subsequent events through January 12, 2011, the date the financial statements were available to be issued.

Note 1. Concentrations of Credit Risk

The Society uses the Certificate of Deposit Account Registry Service (CDARS), administered by the FDIC, which affords large deposits to be split among various U.S. banks to achieve FDIC protection.

Note 2. Property and Equipment

A summary of property and equipment at September 30, 2011, follows:

<table>
<thead>
<tr>
<th>Description</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furniture and Fixtures</td>
<td>8,769</td>
<td>8,769</td>
</tr>
<tr>
<td>Computer and Office Equipment</td>
<td>160,601</td>
<td>162,364</td>
</tr>
<tr>
<td>Leasehold Improvements</td>
<td>11,263</td>
<td>11,383</td>
</tr>
<tr>
<td>Digital Library</td>
<td>106,424</td>
<td>106,424</td>
</tr>
<tr>
<td>Web Design</td>
<td>59,698</td>
<td>59,598</td>
</tr>
<tr>
<td><strong>Total Cost</strong></td>
<td>346,075</td>
<td>348,858</td>
</tr>
</tbody>
</table>

Less: Accumulated Depreciation and Amortization

<table>
<thead>
<tr>
<th>Description</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>(281,261)</td>
<td>(274,900)</td>
<td></td>
</tr>
<tr>
<td><strong>Depreciation and Amortization</strong></td>
<td>$ 54,754</td>
<td>$ 73,952</td>
</tr>
</tbody>
</table>

Depreciation and amortization expense was $23,644 and $29,267 for the years ended September 30, 2011 and 2010, respectively.

See Independent Auditors’ Report.

Note 3. Pension Plan

The Society sponsors a defined contribution retirement plan that operates under section 403(b) of the Internal Revenue Code. The Plan covers all full-time employees and part-time employees with more than 1,000 hours of service. Participation in the Plan begins after completion of twelve months of service. Employees may contribute to the Plans and the Society contributes 5% of the portion of an employee’s salary within the Social Security wage base. The Society’s contributions to the Plans total $21,260 and $22,171 for the years ended September 30, 2011 and 2010, respectively.

Note 4. Board Designated Net Assets

The Board of Directors of the Society has designated net asset balances into the following funds as of September 30:

<table>
<thead>
<tr>
<th>Fund</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve Fund</td>
<td>$233,932</td>
<td>$233,932</td>
</tr>
<tr>
<td>Chapter Development Fund</td>
<td>20,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Local Chapter Funds</td>
<td>123,212</td>
<td>124,274</td>
</tr>
<tr>
<td>Special Interest Group Funds</td>
<td>5,000</td>
<td>5,000</td>
</tr>
<tr>
<td>Special Interest Group Funds</td>
<td>41,266</td>
<td>25,118</td>
</tr>
<tr>
<td>Chapter Project Funds</td>
<td>3,200</td>
<td>3,200</td>
</tr>
<tr>
<td>SIG Digital Scholars Fund</td>
<td>9,780</td>
<td>11,090</td>
</tr>
<tr>
<td>History of Information Science Fund</td>
<td>18,415</td>
<td>17,315</td>
</tr>
<tr>
<td>Scholarship Fund</td>
<td>4,518</td>
<td>4,518</td>
</tr>
<tr>
<td>New Initiatives Fund</td>
<td>13,000</td>
<td>10,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$459,662</td>
<td>$454,717</td>
</tr>
</tbody>
</table>

Note 5. Commitments

Leases: The Society is obligated under non-cancelable lease agreements for office space which will expire in the year 2015. The amounts due under the lease agreements are subject to increases based on the greater of a fixed 4% annual escalation or on a percentage of the change in the consumer price index.

The minimum future rental commitments through the remaining term of the leases are as follows:

<table>
<thead>
<tr>
<th>Year ended September 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
</tr>
<tr>
<td>2013</td>
</tr>
<tr>
<td>2014</td>
</tr>
<tr>
<td>2015</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

Rent expense paid for all operating leases was $29,009 and $47,132 for the years ended September 30, 2011 and 2010 respectively.

Contracted: The Society has a contract with Wiley – Blackwell (John Wiley and Sons, Inc.) in which the Society turns over the publication, distribution, and storage of the Society’s journal. In addition, the Society gave Wiley rights to all gross revenues earned from the journal. In return, the Society will earn royalties from the journal based on a percentage of gross revenues from circulation. The Society is guaranteed a minimum annual royalty of $336,350. The monies are paid by Wiley in advance, on a quarterly basis. As of September 30, 2011 and 2010, royalties of $546,251 and $536,543 respectively, have been earned and recognized under the contract. The Society expects to receive royalty payments in 2017.

The Society has entered into various contracts with hotels for meeting rooms and guest rooms for its meetings to be held in fiscal year ending September 30, 2011. The total values of the contracts are still to be determined. The Society, however, could be liable for a portion of the cost of the rooms not filled if the event is cancelled.

See Independent Auditors’ Report.
Digital Humanities and Information Visualization: Innovation and Integration

by Joan Beaudoin and Sarah Buchanan, guest editors

We are pleased to bring you this special issue of the Bulletin dedicated to the fields of digital humanities and information visualization. As the current chairs of two ASIS&T special interest groups (SIGs), SIG/Arts & Humanities (SIG/AH) and SIG/Visualization, Images and Sound (SIG/VIS), we recognize that the two SIGs have long-shared, common interests related to the access, retrieval and use of visual, textual and auditory information in the online environment. With these shared interests in mind we decided to highlight the remarkable range of current activities within digital humanities and information visualization. The various papers included in this issue represent a cross-section of topics falling within these two areas. Our mutual interests in coordinating this collection of articles would have borne no fruit without the efforts of many individuals. The Bulletin’s dedicated editor, Irene Travis, shared her advice and guidance, for which we are very grateful. We also wish to thank the many authors who responded to our call for papers. These authors bring an engaged perspective to their writing as they share current developments and innovative technologies from the profession.

As the pace of change in the information field continues to quicken, many observers and theorists strive to contextualize and understand the research potential of new technologies. The authors of the papers presented here are no exceptions, with each of their papers discussing how technology has the potential to advance research in their respective areas of expertise. The papers selected for this issue range from the theoretical to the practical, from those that identify how the mere act of digitizing and providing electronic access to resources can benefit scholarship to those that present highly sophisticated tools and techniques for data analysis and display. One aspect holds all of
these papers together: each paper shows a commitment to the development of knowledge within their respective domains.

The issue begins with Jordan Bailor’s paper elucidating our current digital renaissance and how the Post-Reformation Digital Library offers theological researchers a greater scope of resources than was previously available.

Continuing the discussion concerning the advantages of digitizing paper documents into machine-readable text is Jonathan Hagood’s practical survey of the practice of data mining and the tasks it can enable. Hagood discusses how data mining has enabled researchers to investigate patterns and trends that were previously imperceptible. He also summarizes several existing literary research projects and tools useful to this line of research.

The use of tools to analyze materials anew is also discussed by Stephanie Margolin in her paper on how research in the field of American studies might be enhanced through the use of digital tools. Margolin’s paper surveys a range of presentation methods using video and collaborative programs.

Researchers and practitioners in the digital humanities have for several decades made use of literary studies to expand the depth of analysis of textual works. Sarah Jones presents an educational discussion of how computers can enhance the practice of literary analysis and suggests that such technologies blur the distinction between the “authority” and the “learner.”

Jones’ challenge for a re-examination of the analysis performed within a disciple is mirrored by Chris Sula, who draws attention to limited research attention paid to the humanities within the domain of library and information science. Sula notes the limitations of bibliometrics for the study of the humanities and offers a constructive discussion for other types of relationship studies that could benefit future humanities-based and scientific research.

A paper that truly bridges and combines this issue’s two fields of investigation is the discussion by Stan Ruecker, Ali Shiri and Carlos Fiorentino of two visual user interfaces – Searchling and T-Saurus – and their innovative multilingual features. The authors also provide an informed discussion of user behaviors that has broader interest and implications across ASIS&T’s membership.

A similar combination of an interactive tool for the visual display of data can be found in the paper describing the Library of Congress’s recently launched Viewshare.org. This free, open-source toolkit for creating maps, timelines and other interfaces for digital collections is the subject of Jefferson Bailey and Trevor Owens’s article. These authors guide readers through the platform’s design and workflow in addition to providing a case study of the Fulton Street Trade Card Collection interface.

The final paper, by Margaret Lam, incorporates the subject of music informatics into the discussion of the digital humanities. In focusing on the “meaning” of studying music, Lam provides basic guidelines for the development of tools useful for examining music. Lam emphasizes the socio-cultural focus of this vein of research, which is so often lost in purely technical studies.

The digital humanities and information visualization are unique in that they combine the best of humanistic thought and scientific ingenuity to develop innovative means of analyses useful to the creation of knowledge. Much of the research in the library and information science domain has revolved around developing innovative techniques and technologies useful for the analysis of information. Examining the many applications that these techniques and technologies have outside of our own domain is an exciting and instructive endeavor. We hope you will find these articles as engaging as we have, and we encourage you to share your ideas and reactions by joining SIG/AH or SIG/VIS. We promise you will find colleagues sharing these common interests.
The Dynamics of Primary Source and Electronic Resource: The Digital Renaissance and the Post-Reformation Digital Library

by Jordan J. Ballor

Editor’s Summary

The transition from mechanical printing to electronic information dissemination amounts to a digital renaissance, enabling primary source documents to be reborn as electronic resources. For history scholars, this presents opportunities and challenges for preserving centuries-old original texts in a digital environment, supporting downloading and digital access and presenting original research. The Post-Reformation Digital Library (PRDL) serves as a case study, capturing a select set of resources on theology and philosophy of the 15th to 18th centuries from a variety of physical locations and digital libraries. Through digitization, digital access and systematic cataloging, the PRDL – along with the database, social networking and dedicated website it has prompted – are helping to overcome barriers to access. Though a modest initiative, the progress of the PRDL demonstrates possibilities for rediscovering historical materials and making them available for modern scholarly studies.

Keywords

primary literature
analog data
digitization
access to resources
religion
philosophy

W

e are today at the leading edge of a digital revolution, a revolution of information dissemination comparable in scale to that experienced in Europe over 500 years ago. It has become commonplace, if not a cliché, to invoke the Renaissance when discussing historical precedents for the radical implications of technological advances in the information sciences of the last few decades. But the comparison is indeed apt; it may well be that we face an end to the dominance of the mechanically printed text akin to the end of the dominance of the manuscript in the West in the 15th century. Some who herald this transition go so far as to call our current era “the late age of print” [1, p. 93].

When speaking of a digital renaissance in the context of religious history and historical theology, what comes foremost to mind is literally a rebirth of primary source documents into an electronic world. Whereas the Renaissance of the late-medieval and early-modern world focused on bringing to bear the wisdom of the ancient world, including a return to the textual sources of that era, today’s digital renaissance is concerned with bringing to bear the printed sources of bygone times, in this particular case those of the Renaissance and Reformation, in electronic form.

The widespread availability of many of these sources in digital form is a relatively novel phenomenon. An essay on resources in Reformation research published as recently as 2008 could make no mention, for instance, of the vast array of continental sources that were rapidly becoming available from Google Books [2, pp. 25–56]. This example illustrates how quickly things are changing in this digital renaissance. Whereas in the first Renaissance it took decades for printing presses to populate the European continent, the transition to digital sources freely and widely available the world over has taken just a few years.

Digital Humanities and Information Visualization

We are today at the leading edge of a digital revolution, a revolution of information dissemination comparable in scale to that experienced in Europe over 500 years ago. It has become commonplace, if not a cliché, to invoke the Renaissance when discussing historical precedents for the radical implications of technological advances in the information sciences of the last few decades. But the comparison is indeed apt; it may well be that we face an end to the dominance of the mechanically printed text akin to the end of the dominance of the manuscript in the West in the 15th century. Some who herald this transition go so far as to call our current era “the late age of print” [1, p. 93].

When speaking of a digital renaissance in the context of religious history and historical theology, what comes foremost to mind is literally a rebirth of primary source documents into an electronic world. Whereas the Renaissance of the late-medieval and early-modern world focused on bringing to bear the wisdom of the ancient world, including a return to the textual sources of that era, today’s digital renaissance is concerned with bringing to bear the printed sources of bygone times, in this particular case those of the Renaissance and Reformation, in electronic form.

The widespread availability of many of these sources in digital form is a relatively novel phenomenon. An essay on resources in Reformation research published as recently as 2008 could make no mention, for instance, of the vast array of continental sources that were rapidly becoming available from Google Books [2, pp. 25–56]. This example illustrates how quickly things are changing in this digital renaissance. Whereas in the first Renaissance it took decades for printing presses to populate the European continent, the transition to digital sources freely and widely available the world over has taken just a few years.

Jordan J. Ballor is a research fellow at the Acton Institute for the Study of Religion & Liberty, where he serves as executive editor of the Journal of Markets & Morality. He is a founding member of the executive board of the Post-Reformation Digital Library, and his scholarly interests include Reformation studies, church-state relations, theological anthropology, social ethics, theology and economics, and research methodology. He can be reached at jordan.ballor@acton.org.
Digital Bibliography and the Problem of Curation

There are today at least two basic kinds of projects for digital humanities scholars: 1) the presentation and preservation of native analog texts (primary sources) in a digital environment, whether for research use, classroom use or both and 2) the presentation of research (secondary sources) in a born-digital format. It is the former efforts that this essay addresses, that is, the dynamic between “source,” the primary texts and documents of the 16th and 17th centuries, and “resource,” the digital tools to bring these sources to bear in some way, especially in the display of analog texts in digital form and within the context of religious history and historical methodology.

What does this rapid evolution mean concretely for academic research today? Take the case of the bibliography, a basic element of scholarly research. Imagine the class of texts that in the best possible world are relevant to a topic, all those texts that have ever existed, or will ever exist, that relate to what is being researched. Let us call this broadest class of theoretically possible-to-include texts archetypal. This broadest class is the standard for which scholars strive in our bibliographic work.

The class of texts that are actually acquirable for us depends on a number of contextual circumstances, however. Depending on when we live, some texts might not yet have been created or might not yet have been discovered or, worse yet, might have been created but not survived in any concrete form or recorded memory. This narrower class of acquirable texts, different in each context because of any number of variables, we might call ectypal, that bibliography of which we are actually capable, corresponding in a greater or lesser measure to the archetypal bibliography.

Whereas research standards used to depend on factors like travel budgets, archival access and local library holdings, they now also include factors like download ability and digital accessibility. As the church historians and research methodologists James E. Bradley and Richard A. Muller wrote in 1995, “[t]he areas in which students can safely ignore the new methods and source mediums are becoming fewer, and even those scholars working in areas as yet untouched by this technology can still benefit from an exposure to the conceptual elegance of unimpeded research, and exhaustive, near-perfect bibliographies” [3, p. 74]. What the digital renaissance has done is to broaden the class of texts for ectypal (near-perfect) bibliographies to more closely correspond to our hypothetical, archetypal (perfect) bibliography.

But if the reach of our ectypal bibliographies comes increasingly closer to our ideal archetypal bibliographies, the challenge is for our grasp to sufficiently match that reach. This digital renaissance raises serious issues for the orderly accessibility of digital sources. This is, at its core, a problem of curation – the management, organization, preservation and care for and the stewardship of – a set of digital artifacts.

The Post-Reformation Digital Library

The Post-Reformation Digital Library (PRDL; www.prdl.org) is an example of one kind of response to this current situation, a resource intended to provide streamlined access to a specialized set of sources. Apart from its usefulness for a particular area of scholarship (in this case early modern religious history), the PRDL is a case study of what can be done in reaction to this dynamic of source and resource. This library is a model with some application across a number of different disciplines: philosophy, language, literature, medicine, history and so on.

The Post-Reformation Digital Library is a select database of digital books relating to the development of theology and philosophy during the Reformation and post-Reformation/early modern era (the late 15th-18th centuries). Late medieval and patristic works printed and referenced in the early modern era are also included, as well as select modern editions, especially Opera, of relevant authors now available on the web. PRDL spans publicly accessible collections from major research libraries, independent scholarly initiatives and corporate documentation projects. With the proliferation of digital books scattered across various places on the web, it can be difficult for the individual scholar to find or keep track of all the new content that is appearing almost daily. The PRDL is a collaborative effort to organize this content for scholars of early modern theology.

The core of the PRDL project involves the organization of thousands of documents available in digital form from sources including Google Books and the Internet Archive. Also included are the offerings of a range of select digital libraries from Europe and North America, which are beginning to
make digitized forms of their holdings available to the public. The project covers the work of thousands of authors from a wide variety of theological, philosophical and ecclesiastical traditions.

The PRDL had its beginnings in a group of doctoral students associated with the Calvin Theological Seminary, specifically doctoral students in historical theology studying with Dr. Richard Muller. At some point in 2008, through a series of informal discussions as well as disciplinary-focused colloquia, they determined that there was an obvious need to join together to more systematically catalog the various digital sources that were becoming newly available every day. The hope was that collaboration might eliminate redundancies since many students were spending time finding the same works repeatedly and also that such efforts would eliminate lacunae, tapping into respective areas and figures of expertise and focus to cover theological sources in the period more broadly.

The project was also focused on the practical problem faced by many of our colleagues that upon graduation, when heading to other institutions, sometimes domestically but more often abroad, they were not able to easily access the quality of sources made readily available by the Hekman Library and Meeter Center at Calvin College and Seminary. Dr. Muller initiated the effort by providing a document with a finding list of 207 theological works and 76 philosophical works (mostly available from Google Books and based on the bibliography for his four-volume work, *Post-Reformation Reformed Dogmatics*). The effort first began with a large email list of potentially interested scholars who could add to and amend this list as appropriate, but this approach quickly became unwieldy.

Eventually the list was moved to a wiki format, specifically PBWorks (formerly PBWiki), which would allow users to add links to other sources they had found. A small core of active participants took the initiative to arrange the finding list more formally and systematically. The participants focused on a basic arrangement of the authors to be covered. As the various lists began to take shape, discussions began about hosting a public version of the site. Going public necessitated some additional formalization of the work, and so an editorial board was formed and an agreement was negotiated for the Meeter Center to host PRDL.

Once a basic level of coverage in the major areas was achieved, the pages were migrated from the PBWiki site to Calvin College’s LibGuides system, a fairly common web format used by a number of libraries for study guides and bibliographies, as well as a number of other functions. Student workers helped with the migration, and PRDL (http://libguides.calvin.edu/prdl) went public in the fall of 2009.

As of April 1, 2010, PRDL covered roughly 800 authors from a variety of traditions. Given the expertise of the scholars on the executive board, the strength of the PRDL has been its coverage of the Reformed tradition, so that more than half of the 800 authors were Reformed authors from the 16th and 17th centuries. Todd Rester, a PRDL executive board member, rightly called this a veritable “Abbadie to Zwingli” of Reformed theology [4].

Recognizing that there are significant limits on the initial bibliographic form of PRDL, the executive board took steps over the last half of 2010 and the beginning of 2011 to begin to address these shortcomings, resulting in the launch of a new version of PRDL on its own website in the fall of 2011.

The most significant aesthetic and functional difference between the two iterations of the PRDL is that the project has moved from a digital bibliographical format to a database-driven format. This change allows the site to show in real-time the number of authors that are covered as well as the number of works, in addition to allowing increasingly focused and narrow search options. For instance, the front page of PRDL now dynamically updates these stats (currently including coverage of more than 2,600 authors). Extremely complex searches are now also possible by variables including genre, topic, publication information or related biblical text (in the case of biblical commentaries, sermons, disputations and so on).

PRDL has also expanded coverage to include more of the newer digitization efforts, particularly those at major European universities. PRDL also formed an editorial advisory board with scholars from around North America and Europe in the spring of 2011, whose main responsibility is to provide guidance about authors and sources in other theological traditions.

**Source and Resource in History’s Digital Future**

The Post-Reformation Digital Library represents one attempt among
many to come to grips with the challenging dynamic of digital source and resource in research methodology today. It is a project with limited scope, but perhaps for this very reason it represents a fruitful and instructive effort that demonstrates the kinds of things that can be possible for history’s digital future. At one level there are and must continue to be analogous projects creating digital bibliographies for the purposes of other disciplines such as law, politics, literature, philosophy and language.

But at a broader level PRDL represents one possible and hopeful future for digital history. Richard Muller once wrote, “where there is text, there is hope” [5, p. viii]. Given various digital efforts around the world and attempts to organize these resources and make their sources available in a coherent fashion, there is increasing access to the texts of the early modern era. For this reason there is growing hope that we might better understand the “historical Calvin” in Muller’s case, and in the case of the Post-Reformation Digital Library, everyone from the historical Abbadie to the historical Zwingli.

One of the notable characteristics of the Renaissance of the early modern era was the application of critical sources, methods and linguistic tools to the received scholastic method. In a similar way today’s digital renaissance has critically altered research methodology, ushering in an era in which scholarship must make responsible use of the digital sources and resources. As Bradley and Muller write, “The scholar who neglects current technological advances in the manipulation and accessing of sources puts himself or herself in the position of the student who refuses to adopt the methodological advances of the Enlightenment; they become, by definition, precritical” [3, p. 74]. As both the Post-Reformation Digital Library and the ad fontes impulse of the Renaissance show, the first step of today’s digital renaissance is taken with a new and hopeful return “to the sources.”

Resources Mentioned in the Article

A Brief Introduction to Data Mining Projects in the Humanities

by Jonathan Hagood

Imagine having the ability to search the entire canon of Western literature quickly and easily for the use of a specific metaphor, references to a particular place or instances of an exact sequence of words or phrases.

Last year’s publication in the journal Science of research using preliminary results from Google’s book digitization project drew attention to the potential of such data mining for exploring a variety of fields in the social sciences and the humanities [1]. At its simplest, data mining is the process of extracting new knowledge (usually in terms of previously unknown patterns) from sets of data already in existence. For instance, Shakespeare scholars have used data mining techniques to identify patterns of word usage in his plays, the texts of which have already been digitized. Similarly, there is a long history of researchers making use of U.S. census data to identify demographic trends or correlations with other datasets. Data mining is inherently an exercise in quantitative analysis, the results of which are subject to qualitative analyses that link the newly discovered patterns back to particular, representative examples from the original set of data.

In the humanities, data mining necessarily entails an interdisciplinary and collaborative practice because it combines tools, techniques and methodologies from computer science and the humanities. As a consequence, data mining is often associated with the term digital humanities, which includes using cutting edge technology both to present the results of research and to conduct the research itself. Data mining is one example of the latter, and at its best a data mining project involves active collaboration between humanities scholars and information professionals to design and carry out the research program. In addition, because data mining is a relatively recent practice, the research project is often as novel for the computer and information sciences as it is for the humanities. Therefore, data mining projects are driven as much by the information professional as the humanities scholar.

Jonathan Hagood is an assistant professor in the Department of History at Hope College in Holland, Michigan, where his research focuses on science and medicine in Latin America, the history of international nursing and incorporating digital humanities into undergraduate education and research. He can be reached at hagood<at>hope.edu.
One critically important aspect of data mining is negotiating the method’s relationship with traditional research practices within the humanities. Data mining is certainly a promising tool for investigation and data collection. However, data mining also has the potential to influence the formulation of research questions. That is, while projects can certainly use data mining to test pre-existing hypotheses or answer questions already raised by the researcher, humanities scholars are beginning to see the value in using data mining to develop the hypotheses or questions themselves. Returning to the definition of data mining presented above, the phenomenon of data-driven questions largely stems from the fact that data mining identifies previously unknown patterns that necessarily entail the existence of questions that, in a similar way, would have otherwise remained unasked. Therefore, data mining may join critical reading, comparative analysis and other traditional research methodologies in the humanities as tools for initiating and shaping inquiry.

Technical Components of Data Mining

Within the field of computer science, data mining constitutes part of knowledge discovery in databases (KDD) [2]. While there are several multi-stage versions of this process, the simplest is (1) pre-processing, (2) data mining and (3) results validation. As noted above, humanities scholars often simplify the pre-processing stage by making use of existing datasets. In a specific example, historian Sharon Block and computer scientist David Newman have published research based upon data mining article abstracts from widely used databases [3]. Similarly, economic historian Jeremy Atack has discussed the possibilities for researchers available in the Bateman-Foust sample, which is a database of agricultural and population census data from 1850 to 1880 first begun with machine-readable punch cards in the late 1960s and expanded in the 1990s [4]. The increased number of digitization projects for a variety of texts means that humanities scholars are discovering access to a wealth of potential sets of data.

Many different tasks make up the process of data mining itself, which authors and disciplines have categorized, labeled and prioritized differently. For data mining projects in the humanities, the following are the four most important tasks to consider:

- **N-gram Identification.** When analyzing text, data mining projects often search for instances of a given “n-gram,” a sequence of n items that can be either characters or words. For example, a project might search for instances of the word *cosmopolitan* within a set of newspaper articles from the early 1900s. A critical component of such an analysis is identifying the relevant n-grams either *a priori* or during the project. That is, data mining has the potential of revealing n-grams whose significance were previously unknown to the research team.

- **Classification.** With data mining, items within a dataset often need to be assigned to one or more classes or categories. Strictly speaking, this task is part of the pre-processing of the dataset; but in practice early results from data mining often suggest additional classification. A feedback loop is therefore created. For example, a project to classify the public speeches of President Franklin D. Roosevelt might begin with a particular categorization scheme but may discover new themes through the process of data mining.

- **Dependent Modeling.** Ultimately, the goal of most data mining projects is to identify dependency and correlation among variables (be they n-grams or classes). Standard tools of quantitative analysis can then evaluate the relative strength or weakness of discovered dependencies. Also, preliminary analysis of portions of a dataset can lead to hypotheses that the project can then test on the rest of the data. For example, a data mining project may discover that a selection of Romantic poetry exhibits a strong correlation between the use of n-gram X and theme Y and an inverse relationship with theme Z. The project may then turn to a larger sample of work to test these correlations.

- **Clustering.** Once a data mining project identifies n-grams, classes and dependencies, further analysis can reveal sub-populations of the dataset. For example, data mining a body of texts that include novels, short stories and essays may reveal clusters of themes, word usage, etc. that transcend the categories of genre and format. For example, although approaches to rendering clusters visible depend upon the specific project, common methods include the use of word clouds, depicting the relative weight of multiple dependencies visually as in a network.
Diagram, and plotting a pair of variables in two-dimensions while coding the data with a third variable (for example, the number of main characters vs. the average length of monologues might reveal clusters that do or do not correspond directly to genre).

Data mining projects in the humanities can automate each of these tasks to varying degrees depending upon the skills, interests and time available to the particular members of the project team. Also, because the interdisciplinary nature of data mining necessarily involves collaboration between information professionals and humanities scholars, data mining also brings up questions about displaying and communicating data to readers with differing expectations, learning styles and levels of technical literacy. Therefore, the collaborative and interdisciplinary nature of data mining as a practical research exercise means that the discussions that take place within interdisciplinary teams while carrying out the research itself foreshadow the ways in which scholars from diverse disciplines will receive and interpret a project’s published findings. Finally, any survey of the current state of digital humanities research underscores the fact that these interdisciplinary teams frequently include humanities scholars who are pursuing non-traditional and non-tenure-track careers, such as many of the contributors to #alt-academy (http://mediarecords.futureofthebook.org/alt-ac/).

Examples of Data Mining Projects

As the research using Google Books demonstrated, the most efficient way to get a better sense of the potential of data mining is to examine the results of research projects that have made use of data mining methods. Here are some notable examples:

- Literary scholar Eric Gardner constructed a dataset of “subscribers” and “readers” of the Christian Recorder, the principal publication of the African Methodist Episcopal Church, from the lists of acknowledgments published between November 1864 and November 1865 [5]. The set of 834 items included information that allowed Gardner to interrogate the question of readership, particularly how the editors of the Christian Recorder understood the concepts of subscription, dissemination, readers and reading.

- In October 2011, Michael Witmore, the director of the Folger Shakespeare Library, presented early results of a project that data mined excerpts from Shakespeare’s First Folio using software called DocuScope [6]. Through this project, Witmore discovered evidence of variance within vocabulary and syntax among Shakespeare’s comedies, historical plays and tragedies [7]. In particular, the analysis suggests that Othello, which scholars have traditionally categorized as a tragedy, has more in common stylistically with Shakespeare’s comedies.

- Computer scientists David Elson and Kathleen McKeown and literary scholar Nicholas Dames have worked on a project that attempts to recover the social networks within 19th-century novels through data mining of the texts [8]. Their main approach is to identify instances of dialogue between characters, and in this way the project takes an existing database (texts of 19th-century fiction) and extracts a new database (conversations between characters) that is itself then mined for new patterns.

In addition, the National Endowment for the Humanities through its Office of Digital Humanities partnered with the National Science Foundation, Canada’s Social Sciences and Humanities Research Council and the United Kingdom’s Joint Information Systems Committee to create the Digging into Data Challenge [9]. Some of the project teams that have received funding from this program used data mining techniques to analyze digitized music, images and maps, the linguistics of the spoken word and letters written by key Enlightenment figures.

Publicly Available Tools

Most data mining projects are very open about the tools used to undertake the research. Although the outcomes are often the result of significant customization and development, software tools for data mining are available online:

- Google Ngram Viewer (http://books.google.com/ngrams/) allows the user to run searches on Google’s datasets, which are also available for download.
**Machine Learning for Language Toolkit (MALLET)** (http://mallet.cs.umass.edu/) is a downloadable piece of software for document classification, sequence tagging, topic modeling and numerical optimization of sets of textual data.

**The Metadata Offer New Knowledge (MONK) Project** (www.monkproject.org/) provides online access to a suite of textual analysis tools and publicly available texts from American literature and the works of Shakespeare.

**The Stanford Natural Language Processing Group** (http://nlp.stanford.edu/software/) has made portions of its software available for download and incorporation into applications that analyze language.

**Voyeur** (http://hermeneuti.ca/voyeur) is a web-based tool for studying the frequency and distribution of data within user-provided texts.

These examples demonstrate the variety of data mining tools available that perform similar functions yet vary in terms of access (via a web interface or as a download), the coding expertise necessary to make effective use of them, the inclusion of publicly available data and the ability of users to apply these tools to their own datasets.

**Conclusion**

Information professionals and humanities scholars interested in data mining projects should begin by establishing collaborative and interdisciplinary relationships as early as possible in the development of the project. Research interests and questions from both computer science and the humanities can drive either original or ongoing research. Such projects have the potential to re-examine existing assumptions and theories within particular fields and to develop new lines of research.

---

**Resources Mentioned in the Article**


Pale Males 2.0: Revisiting a Traditional American Studies Project Using Digital Humanities Tools

by Stephanie Margolin

EDITOR'S SUMMARY

Though research papers have traditionally been presented as text, digital humanities tools bring new capabilities that can transform a presentation. Moving from a familiar written format to a digital multimedia work can be daunting, involving an unfamiliar work process. Technology options and support are available through a variety of sources, including universities, the American Studies Association’s Digital Humanities Caucus and the Alliance for Digital Humanities Organizations. Choosing a presentation tool that supports desired features, such as interactive conversation, is key. While the process of developing a digital presentation can be initially challenging for an author, reproducing a traditional thesis in a dynamic format makes it more lively and engaging, collaborative and available for broader use.

KEYWORDS

humanities, digital audio files, scholarly publishing, information technology, multimedia, dissertations, digital video files

Stephanie Margolin is a freelance librarian, currently working as a digital librarian for Knowledge in the Public Interest and as an adjunct reference librarian at Hunter College (CUNY). Author of the original "Pale Males" paper, Stephanie is drawn to the new possibilities that digital humanities bring to American studies. She can be reached at smmargolin<at>yahoo.com.

This paper considers how one might use digital humanities tools to re-think an American studies project. With these tools, we can transform a traditional, text-based thesis into a multimedia presentation, which can then be used to generate further thought, discussion and opinions. The original project, titled Pale Males Gone Pell Mell: James Dean, Axl Rose and Adolescent Male Rebellion, was my master’s thesis written in 1992 [1]. The paper examined James Dean as “Jim Stark” in the film Rebel Without a Cause and Axl Rose of the band Guns’n’Roses through the lens of men’s studies. While the central thesis relied heavily on the visual, the format was limited to a written analysis. Using the currently available digital humanities tools, we can add a much-needed visual component to this thesis. At the same time, readers who took a passive role in the past might now be invited to interact with the material, draw their own conclusions and expand their experience.

Getting Started

For American studies scholars interested in adding a digital humanities component to their repertoire, the American Studies Association now includes a Digital Humanities Caucus in which they might participate (www.theasa.net/caucus_digital_humanities/P5). The Alliance for Digital Humanities Organizations (ADHO) (digitalhumanities.org) also offers resources and support for these endeavors. However, the first place to seek help or mentoring might well be on one’s own campus as a growing number of colleges and universities house digital humanities programs. Often these programs offer support that includes development of homegrown technology solutions or customization of extant open source options to meet the particular
needs of scholars and projects. When a digital humanities program is not available, ADHO’s wiki of digital research tools (called DiRT) has been a useful place to go to learn more [2]. Tools discussed below were selected from that site in December 2011. However, the wiki is no longer being updated and is being replaced by Project Bamboo’s site, “Bamboo DiRT” [3].

Moving from a traditional academic presentation such as a written thesis to a presentation using digital humanities tools can transform telling to showing, thus enriching the audience’s experience. Consider the video essay, which has proven itself more effective than mere written analysis. In film critic Jim Emerson’s video essay, “In the cut, Part I: Shots in the Dark (Knight)” Emerson pastes together the key scenes that comprise his thesis about an action sequence in the film that he finds unsatisfactory. He then narrates his argument via voice-over [4]. Already used in cinema studies, such an essay seems an ideal strategy for *Pale Males* and other American studies projects and possibly also a powerful choice for audio and/or video projects in art history, performance studies and/or media studies.

While digital humanities tools provide unique ways with which to interact with material, the work process is often unfamiliar to traditional scholars and can be complex. How, for example, does one construct a video essay? Jim Groom and Alan Levine have created an excellent guide in their “Video Essay Tutorial” as part of their online digital storytelling course. This guide, written in an easy-to-follow, conversational style, provides step-by-step guidance through the process of creating a video essay. The process is complicated and includes finding the desired clips, “ripping” the clips and putting the whole thing together with narration. Groom and Levine’s guide recommends tools for each step of the process, presenting tools available for both PC and Mac, and many of their recommended tools are free [5].

Once the video essay has been created (or at least outlined), the scholar must choose the most appropriate presentation tool, aligned with the stated project goals. For our *Pale Males* project, in addition to adding the multimedia component, we hope to invite reader feedback, in order to make the project more interactive and perhaps build some kind of community around the project. Therefore, we are looking for a presentation platform that can engage the viewers and accommodate feedback and other user interactions.

**Selecting the Right Tools**

As discussed earlier, we might find – or build – an effective presentation platform with the help of our own institution’s digital humanities program. Without such a program, or while searching for a ready-made solution, consider one of the tools mentioned below. As outlined here, each option offers a distinct benefit, but further research and testing is needed to determine the best tool for transforming *Pale Males* into an interactive presentation.

First, we consider Sophie (tagline: “redefines the notion of a book to include rich media, reader feedback & conversation within a networked environment”), which was developed by the Future of the Book project and has been adopted by the Institute for Multimedia Literacy (IML) in the School of Cinematic Studies at the University of Southern California (www.sophieproject.org). In promising to combine rich media and reader conversation, Sophie appears to meet two key goals we’ve set for the *Pale Males* project. However, Sophie can be challenging for viewers to use. They must download the (free) Sophie reader tool in order to view a Sophie project (analogous to, but more complicated than, reading an Adobe PDF file, for example). Authoring appears to be complex, too. While the features are exciting there appears to be a significant learning curve to working with Sophie.

**Pachyderm** (tagline: “multimedia authoring for peanuts”) is somewhat less complex: visitors can view projects directly without downloading additional tools. “Pachyderm pieces are interactive, Flash-based presentations that can include images, sounds, video, and text” [6]. The software was originally developed by the New Media Consortium (NMC). While the group no longer supports Pachyderm, the open-source code is available through SourceForge (sourceforge.net/projects/pachyderm/?source=directory). Projects in the Pachyderm gallery on the NMC site appear somewhat lackluster. There is little evidence of interactivity or space for user comments, begging the question: Is this tool robust enough for the interactive video essay presentation envisioned for *Pale Males*?

**VoiceThread** (tagline: “conversations in the cloud”) is the final option we will consider (http://voicethread.com). Essentially a slide show with additional features, VoiceThread focuses on collaborative conversations,
allowing participants to comment by text or via recorded audio or video. One limitation of note is that VoiceThread (both the free version and the paid version) has file size restrictions: 25 MB and 100 MB, respectively. While VoiceThread does not have the production value of the other two tools, it is the most user-friendly. VoiceThread might be the quickest and easiest way to prototype a project of this sort.

It is exciting to think of how our *Pale Males* project might be adapted and expanded using digital humanities tools. Becoming comfortable with the new digital humanities technology and new work processes will be a significant challenge, though one that also promises rich intellectual rewards. While creating a simple prototype video essay will not be simple, doing so will breathe new life into the project and greatly advance its relevance, inviting the audience to actively see the material and form their own opinions of it.

With the ability to present the evidence such as the film and video clips directly to the audience, there are new possibilities for *Pale Males*. The completed project might be used as a learning module or online lesson in an American studies class where students are invited to share their own thoughts or contribute an essay of their own. Or the project might launch an online collaboration among a community of scholars addressing the material collaboratively. Each of these is far removed from the fate of the first version of *Pale Males* back in 1992.

### Resources Mentioned in the Article

When Computers Read: Literary Analysis and Digital Technology
by Sarah Jones

EDITOR'S SUMMARY
The study of literature is changing in dramatic ways, stimulated by new opportunities that digital technology presents. Data visualization upends the dynamic for literary analysis, focusing not on questions stemming from a critic's personal viewpoint but on revealing and displaying connections between elements of the literary experience. The dominant association between critic and text is downplayed, replaced with associations within the text and between it and its context. The basis of interpretation shifts from reading to seeing, from qualitative analysis to quantitative. The reader's role is transformed, as well, from following the critic's path of thinking to actively exploring a network of multisensory and interdisciplinary information. The distinction between the authoritative presenter/critic and the learner/explorer is blurred. By inviting literary scholars to ask different questions for computational analysis, digital technology and visualization inspire innovative investigations and enable new insights.

KEYWORDS
literature
electronic visualization
quantitative analysis
information technology

Digital technology has encouraged innovation not only in the methods of research but also in the presentation of results. Data visualization turns findings into charts, graphs, maps and exhibits – all forms that make literary interpretation a visually experienced object or event rather than an abstract concept. Lev Manovich, a professor of visual arts, describes the difference between traditional expository writing and data visualization: “What’s interesting about culture is that the categories are continuous. Instead of using these techniques to reduce complexity, to divide data into a few categories, I want to map the complexity” [1, p. 11].

Computational analysis promises the discipline of literature access to the kind of knowledge that was once regarded as the antithesis of the humanities: hard facts. As a result of emerging digital technologies literature can now be studied and presented with techniques that have traditionally been confined to sociology, natural science, neuroscience, history, psychology and linguistics. This paper will address two questions raised by the introduction of digital methods into the humanities: What does it mean for the literary critic? What does it mean for the reader of the resulting literary analysis?

New Formats: From Papers to Puzzles
Imagine, for example, a web representing the conversations among characters in Hamlet. Now imagine an essay describing the structure of these connections. An essay outlining these phenomena in a linear fashion would run the risk of being either confusing or reductive. In addition, the essay’s author would make the subjective personal choice of which connections to describe in further detail. Traditional literary interpretation favors a particular message, point or theme. By contrast, patterns, shapes and trends would be immediately apparent in the visual web. Data webs represent multiple angles, connections and shapes simultaneously. Literary historian
Susan Brown has noted that some areas of the humanities are “rich in dense and complex interlinkages which almost defy explanation in words” [2, p.5]. In a well-crafted visual form the complexity of the data would not have to be sacrificed for the sake of clarity.

Data visualization changes the experience of literary interpretation not just for the critic but for the reader as well. It has already been noted that the traditional analytic essay is the personal commentary of a single reader who is recognized as an expert qualified to interpret texts. While the literary critic draws from a cultural knowledge and intellectual tradition, the interpretation is still very much an explanation of the personal interaction between this privileged reader and the text. Readers of the resulting criticism experience the literary work filtered through the personal interpretation of the theorist. While the readers must apply their own thought processes to the texts, the critic makes many decisions for them, such as which factors and connections within the literary work are worth attention. In many ways the shape of the reader’s literary experience is drawn by the critic.

A reader’s interaction with data visualization is still influenced by a privileged reader, the critic. The critic (in this situation “researcher” may be a more appropriate title) decides which questions to “ask” the computer and then designs the visual representation of the results. The outcome of the data, however, is somewhat unpredictable. The researcher determines the hypothesis and method of the experiment but cannot control the results. It is highly unlikely, for example, that Franco Moretti, a professor at Stanford University and one of the most prominent digital-humanities academics, had anticipated the results of a study he performed on literary genre: that the word the appeared more frequently in Gothic novels [3]. In this situation the literary critic is in the position not of documenting individual experience with the act of reading, but of presenting hard data. It is important to note that this type of analysis is still affected by the researcher’s personal bias—the subject and experiment are, after all, controlled by the critic. However, as will be discussed in greater detail, results presented in the forms offered by humanities computing have the potential to step closer to objectivity.

In reference to a study that mapped the connections among women writers in England, Brown observed that “seeing a visual representation that summarizes a pattern in the data is different from reading the same data as a text ... an interactive visual environment is intended primarily to assist in pattern-finding” [2, p. 3]. Data visualization, then, makes the act of literary interpretation less reading and more seeing. The viewer of data visualization processes literary study differently from the reader of criticism. Brown’s usage of the word interpret suggests a more active role on the part of the reader/viewer. An authoritative critic may be called upon to explain data but ultimately numbers and visuals require more interpretation on the part of the reader than does expository writing. To read a critical essay is to follow the writer along a path of thought; studying a chart is a very different experience, one that belongs more to the reader.

New Questions, New Answers

The domination of the literary canon is weakened by digital literary studies. Computers detect quantity, not quality, which means that scholars must change the nature of the questions they ask about literature. Quantitative data lends itself to inquiries like “Why was this author commercially successful?” rather than “What makes this novel good?” When studying Sir Arthur Conan Doyle, Moretti does not attempt to explain why he believes the Sherlock Holmes stories to be objectively great works of literature. Instead, he asks, “Why did readers of the era prefer Conan Doyle’s writing to that of other detective novelists?” He uses a “tree” to find patterns that reveal correlations between the form of a novel and its commercial success [4]. A more traditional approach would have focused on the canonized work, ignoring the less prestigious writing from which Moretti draws a great deal of his insight.

While humanities computing can allow literary studies to more closely resemble scientific inquiry, it is important to note that digital innovations also cause a structural break from both traditions. The expository essay form has much in common with the scientific method as both are structured to prove, disprove or illuminate a single point. Classifications (such as species in science and genre in literature), causal relationships and definitive conclusions are relied upon heavily in both disciplines. The essay and the experiment are linear forms that begin with a question and end with an answer.

The difference between an insight expressed in writing and one expressed...
visually or numerically raises an interesting question: How do we define and structure knowledge? Is it the understanding of causal relationships or an awareness of a number of interconnected factors at play? According to media theorist Donna Haraway, traditional Western thought has sought to make sense of the world by dividing it into culturally determined categories and hierarchies [5]. We can translate this concept to the study of literature – “themes” or “messages” that critics derive from literary works function in the same way. Due to its demands for summary and conclusion, the linear essay form can be seen as both encouraging and reflecting this kind of hierarchical thought. It prompts the writer to arrange ideas vertically according to perceived importance.

In discussing the organization of human culture, Haraway recommends an alternative to the traditional stratified power structure: a network “suggesting the profusion of spaces and identities and the permeability of boundaries” [5, p.170]. If we apply this recommendation to the hierarchical form of traditional literary study, it seems that humanities computing has the potential to present the products of criticism in the form Haraway recommends. In these forms – a table of figures, a map of connections – knowledge is organized not as a line or a column, but as a tangled web.

New Roles for Readers

New forms of literary analysis create a new kind of reader. As already discussed, digital literary scholarship takes some of the interpretative responsibility from the critic and places it on the reader. In his criticism of the written word, Plato presents the written idea as sterile. That is, it cannot grow and be shaped through interaction in the manner of an idea spoken in conversation. The idea born from interactive discussion, on the other hand, is more provocative [6]. This distinction can be applied to modern forms of scholarship, too. The expository essay speaks in a single static voice while the chart permits divergent interpretations by presenting a sort of conversation among elements in the text. This kind of insight is less about the authority telling the reader how the text functions and more about showing aspects of the work through the presentation of data. This creates a more autonomous reader who must examine a web of findings, decide which connections are significant or interesting, and then, if appropriate, conceive of possible explanations for the pattern. While these explanations may be artificial “wholes,” the emphasis on reader-created interpretations of data leads to a number of diverse viewpoints. This makes the discipline more open to polyvocality, thereby blurring the hierarchical division between the “authority” and the “learner.”

It has already been established that the digital-humanities reader must play a more active role, but what does this responsibility mean in practice? The difference between “active” and “passive” audiences is a much-discussed subject among media theorists. Marshall McLuhan asserts that the manner in which information is presented is just as influential as the information itself. McLuhan divides formats into two categories: “hot” and “cool” media. Hot media are characterized by their tendency to demand the full attention of a single human sense and demand only minor effort on the part of the audience. Cool media contain less and more scattered data and require a more active audience who will connect the dots and fill in the blanks [7]. It is, in essence, the difference between a puzzle and a picture. Hieroglyphs (cool) and letters (hot), for example, differ from each other in that hieroglyphs are pictures representing objects. The reader must figure out the relationships among the images. Words, by contrast, are abstracted from the physical forms they represent and thus a person who understands the phonetic system is provided with all the connections necessary for a complete understanding of the message.

With McLuhan’s ideas in mind, it would seem that the switch from written theory to visual and numerical forms would mark a sea change in sense perception and have an enormous impact on the reader’s experience of literary criticism. Data visualization turns concepts into physical objects and display them in relation to each other. Therefore, it more closely resembles a hieroglyph than it does a written word; it requires more dot-connecting and independent work on the part of the reader than would a traditional essay. The products of machine-based research are not so much read as they are operated; they are not absorbed, but worked with.

What are the implications of literary criticism that demands a more active reader? McLuhan suggests that the tendency of hot media to privilege a single sense leads to a sort of tunnel vision – a specialized and fragmented worldview [7]. This state of affairs bears a striking resemblance to traditional humanities studies in which literature is isolated from other
disciplines and their corresponding senses. With this in mind it makes perfect sense that the cooling-down of literary studies would coincide with an increased interest in work that is interdisciplinary, multisensory and large in scope. When critics and readers must connect numerical data to historical context and literary texts, there is an increased need for exploration in areas such as science, statistics, linguistics and history.

**Conclusion: What’s Next for Critics and Readers?**

Both readers and critics must now interact with literary insight not just through the written word, but as a multisensory and interdisciplinary experience. The digital-humanities readers are active because they must connect the dots between pieces of data; when working with the new media forms produced by the digital humanities, critics, too, must adopt new methods of making connections. Hard data lends itself to questions that are more about the grand scope – the cultural and historical – and less about the personal – the moral and emotional – effect of a text. The digital-humanities critic, then, must focus on connections between texts and texts and between texts and culture; the connection between the critic and the text is deemphasized.

For the critic (or “writer/researcher/designer”), humanities computing encourages movement that breaches the confines that characterize close reading, including personal viewpoint, time constraints, historical situation and disciplinary subject boundaries. While no theory can completely escape the influence of the personal and cultural context of the critic, digital and quantitative methods allow literary analysis to shed some of its subjectivity and move along the continuum towards objectivity. For the reader (or “viewer/player/listener”) of literary analysis, humanities computing can have the opposite effect. The reader must take an active role in assessing the meaning of the data when the results of literary analysis are presented in a form that does not contain a prepackaged conclusion. The products of digital literary study encourage the reader to make a number of individual choices. As a result, the reader’s experience of literary criticism is actually more personal in a computational context.

For both the critic and the reader, the digital humanities provide a new conception of the world of literature. Not only is this world larger – the sheer volume of the material we can access is unprecedented – but it is open to levels of analysis that could never be achieved by human brainpower alone. Hierarchies and themes fade into the background as patterns and networks emerge. These methods simultaneously divide texts into new categories and connect them to each other to form new wholes. As digital innovations progress, literary scholars and their audiences must work through new issues emerging in their discipline. What should critics do when humanities computing produces inexplicable results? How will active readers change the field of literary interpretation? What tools and skills should the critic and the reader acquire to facilitate their interactions with charts and statistics? In the face of hard data, the most important question is this: What kinds of insight are valuable to the study of literature?

---

**Resources Mentioned in the Article**


Visualizing Social Connections in the Humanities: Beyond Bibliometrics

by Chris Alen Sula

Intellectual history and critical self-reflection are distinguishing features long associated with the humanities. The growing movement of digital humanities affords new opportunities for studying both through high-volume, longitudinal datasets on people and publications, as well as advanced algorithmic analyses. In addition, visualization techniques can help render this information in salient ways and open new paths for exploration. This article addresses one intersection of digital humanities and information visualization: the study of social connections among humanists.

The first section reviews previous studies of the structure of the humanities, particularly bibliometrics, and notes the limitations of this approach. The second section discusses several studies that support greater consideration of social connections in the humanities as well as data sources from which such connections may be gathered. Three broad categories of relationships are discussed, including student/teacher ties, departmental colleagues and other relationships, such as conference participants. The final two sections address the prospects for visualizing these connections, most notably in the form of network graphs, and speculate on the larger significance of this social analysis, both for the humanities and for the academy in general.

Bibliometrics and Its Limits for the Humanities

Those who study the structure of academic disciplines have long been interested in connections among scholars. Most often, this interest has taken the form of bibliometrics: the study of patterns and relationships in the formal record of scholarly communication. Occasionally, this approach has been supplemented with information about conference proceedings, funding streams, personal website links and other “altmetrics” – usually in attempts to give deeper meaning to citation patterns. Visualization techniques, particularly network graphs, have also helped to harness bibliometric...
Social Connections in the Humanities

As Weedman has noted, humanists are often portrayed as solitary and isolated figures [4]. Yet this perception has been based primarily on studies of formal communication. Studies of informal communication among humanists have stressed the similarities between humanists and non-humanists. For example, Weedman’s study of scholars of children’s literature found that the informal communication needs and behaviors of humanists were similar to those of researchers in other disciplines and that more than 50% of those surveyed said at least half of their ideas could be traced back to conversations with others [5].

The presence of informal intellectual exchange among humanists should come as little surprise. Historically, it has been common to discuss schools of thought, both in Western and Eastern intellectual history. Randall Collins distinguishes four senses of this term: (1) individuals with similar modes of thought (who need not be contemporaries), (2) intellectual influences among scholars, (3) chains of personal relationships and (4) organizations where authority and property are passed through succession [6]. The third category is most relevant in discussing social connections among humanists and, in fact, can be seen as mediating the other three. Personal connections serve as vehicles for aligning thought and doctrine, for transmitting influence through circulating publications and ideas and controlling limited attention space in the field and for establishing and maintaining actual social organizations.

The specific mechanisms through which personal connections exert their influence may be explained in terms of social psychology. Morrow and Sula hypothesize that uniformity pressure and confirmation bias work in tandem to disseminate ideas, reinforce some and relegate others [7]. Uniformity pressure is a form of social pressure that induces members of a group to seek uniformity of opinion within the group, while confirmation bias subsumes several more specific psychological tendencies that lead individuals to seek and believe information that is consistent with their existing beliefs and to ignore, to disbelieve or to be more critical of information that is inconsistent with their existing beliefs. The presence of uniformity pressure is well documented in enduring social groups, which may include academic
departments, and it is considered by many psychologists to be among the largest problems for human reasoning.

Given the likelihood of social influence in the humanities, it is worth considering which types of social connections are most prominent, as well as the documentation that may exist about them. The remainder of this section focuses on two types of relationships: student/teacher relationships and departmental colleagues. Other relationships, including conference participation, are briefly discussed.

**Student/Teacher Relationships.** Student/teacher relationships are among the oldest and most significant ties in humanities, especially relationships between advisors and doctoral students. Uniformity pressure and confirmation bias may explain the significance of this relationship in the following way: newer students in academic departments find themselves unable to match the intellectual abilities of higher-status faculty and more advanced students and either adopt the views of the group in which they find themselves or gravitate toward those who already share their views. In either case, confirmation bias may further entrench whatever views are adopted, perpetuating them through several generations of scholars. Of course, rational mechanisms may intervene and override these other mechanisms, but it is no understatement to say that many students have followed in the footsteps of their advisors.

Data on these relationships is documented in dissertation front matter, which lists advisors and committees and often includes acknowledgements that offer further insights into the contributions of particular individuals. Since the mid-19th century, dissertation procedures have been formalized in Anglo-American and Continental institutions (and earlier in some other cases), providing a large source of this data over roughly a dozen decades. Though less significant, other teacher/student relationships can be gathered from attendance records or roughly inferred by comparing students’ dates of attendance in a program with the lists of faculty teaching in the program at the time, narrowed according to the students’ and faculty members’ areas of interest.

**Departmental Colleagues.** Another important relationship is that of departmental colleagues. A case study of community college faculty found that the average faculty member has three to five close collegial relationships and regards less intimate collegial relationships as a standard part of the college environment [8]. Gender, age, parental status, workload and physical proximity influenced the development and maintenance of these relationships, and departmental colleagues serve as information sources, discussion partners and readers of unpublished manuscripts. Universities maintain annual or biannual listings of departmental faculty, providing clear documentation of appointments. Specific relationships, however, may need to be inferred based on acknowledgments and citations in formal scholarly communication.

**Other Relationships.** While perhaps less common, relationships of correspondence, conference attendance, membership in the same professional associations, editorial relationships and personal friends also contribute to the social structure of the humanities, and some (for example, correspondence) are particularly important during particular time periods like the early modern period. Some of these are clearly documented (see Table 1), while others may not be. A fuller study of the significance of these connections and possible sources of documentation is needed.

<table>
<thead>
<tr>
<th>Type of Relationship</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student/Teacher Relationships</td>
<td></td>
</tr>
<tr>
<td>Advisor/advisee</td>
<td>Dissertation front matter</td>
</tr>
<tr>
<td>Classroom student/teacher</td>
<td>Various sources</td>
</tr>
<tr>
<td>Peer/Peer Relationships</td>
<td></td>
</tr>
<tr>
<td>Faculty colleague</td>
<td>University catalogs</td>
</tr>
<tr>
<td>Student colleague from graduate school</td>
<td>Dissertation acknowledgments, degree dates*</td>
</tr>
<tr>
<td>Other Relationships</td>
<td></td>
</tr>
<tr>
<td>Conference participant</td>
<td>Conference programs, proceedings, CVs</td>
</tr>
<tr>
<td>Correspondent</td>
<td>Letters, references, acknowledgments</td>
</tr>
<tr>
<td>Editor/Contributor</td>
<td>Anthologies, journals</td>
</tr>
<tr>
<td>Member of an association/society</td>
<td>Organization rosters</td>
</tr>
</tbody>
</table>

The * indicates that relationships may be roughly inferred from these documents, though further study should be conducted to determine the confidence interval of these inferences.

---

**TABLE 1. Social relationships and documentation of them.**
Visualizing Social Connections

While additional data on social connections would provide a fuller picture of the humanities, it also presents challenges of representation, particularly with respect to longitudinal data. Where textual representations might be nearly impossible to comprehend, visualization may help to amplify cognition, extend working memory and allow for greater exploration of such data.

Network graphs have been used to aid social network analysis from its beginnings. However, large-scale networks with many nodes and overlapping connections have also been shown to hinder pattern recognition – the main reason for employing visualization in the first place. Several proposals have been studied for simplifying social network graphs, including algorithms for reducing overlapping connections, fisheye techniques that focus on particular areas at a time, clustering or omitting fine detail, limiting “degrees of interest” to provide details only on demand and building flexible systems for network exploration [9]. It would be premature to speculate which methods work best for visualizing data on social connections in the humanities, and alternatives to network graphs should also be explored. Brandeis and Nick, for example, present an intriguing “gestaltline” approach that combines sparklines with Gestalt-based glyphs to visualize asymmetric relations in longitudinal social networks [10], precisely the type of relations one encounters in the humanities.

Simply put, there is no shortage of techniques for experimentation, and digital humanists should test different visualization methods for potential insights, such as emerging areas of research and “invisible colleges” that drive research in scholarly fields, including the humanities [5], clustering that suggests hidden subfields or potentially emerging breaks and so forth. This use of social data also need not exclude the use of bibliometric data. Starting with bibliometric information, different weights may be assigned to citations, multiple authorship and social connections to yield a hybrid visualization that is more inclusive than either of the simple visualizations alone. Determining the nature and weight of these connections is an important area for further study. These studies will also provide fruitful ground for comparison with traditional bibliometric analyses of scientific literature.

Conclusion

 Whereas traditional bibliometric analyses have focused on purely quantitative measures of formal scholarly communication among scientists, this essay has advanced the role of social connections in the humanities and their potential to bring qualitative nuance to bibliometrics. Social data may bridge the bibliometric gaps that exist in the humanities and provide critical context for references and acknowledgments. In addition, a fuller picture of the humanities will help to clarify the ways in which the humanities and sciences differ, beyond citation patterns and authorship practices. And if – as some suspect – the social structures of the humanities and sciences are largely the same, the methods of analysis and visualization developed for the humanities may, in turn, be applied to the sciences, yielding a richer picture of scholarship across the academy as a whole. ■
Resources Mentioned in the Article


Interactive Visualization for Multilingual Search
by Stan Ruecker, Ali Shiri and Carlos Fiorentino

Over the past 10 years we have been carrying out a series of experiments in rich-prospect browsing, where some meaningful representation of every item in a collection is combined with tools for manipulating the display [1]. Example projects include the Mandala Browser for the visual construction of complex Boolean queries of XML documents [2], the Texttiles browser for visual grouping of items in RSS feeds and the structured surfaces interfaces for the Just in Time Research (JiTR) project [3]. What these interfaces each provide is an environment for people to explore digital materials where some combination of grouping and searching supports both information access and information exploration.

In addition, we have been exploring the use of interactive visualization in the design of user interfaces that leverage knowledge organization systems as multilingual thesauri. The goal is to support a variety of information seeking tasks, namely searching, browsing, navigation and exploration. More specifically, depending on the level of sophistication that can be offered in the interface, visual user interfaces have the potential to support such interactive tasks as query formulation, modification or expansion. In pursuit of this goal, we have developed two visual user interfaces, called Searchling and T-Saurus that use the Government of Canada Core Subject Thesaurus, a bilingual thesaurus in English and French, and the UNESCO multilingual thesaurus in English, French and Spanish [7].

Theoretical Framework and Design of Searchling and T-Saurus

The design of the Searchling and T-Saurus user interfaces was based on two key elements – the first was the idea of rich-prospect interfaces; the second was the set of principles for design ideas for thesaurus-based search interfaces suggested by Shiri, Revie and Chowdhury in 2002 [4], including the following:

- providing hierarchical and alphabetical lists to support different strategies

Stan Ruecker is an associate professor of design at the IIT Institute of Design in Chicago. He can be contacted at sruecker@id.iit.edu.

Ali Shiri is an associate professor in the School of Library and Information Studies at the University of Alberta, Edmonton, Alberta, Canada. He can be contacted by email at ashiric@ualberta.ca.

Carlos Fiorentino is a visual communication designer and educator in design studies at the University of Alberta and Grant MacEwan University, Edmonton, Alberta, Canada. He can be reached by email at carlosf@ualberta.ca or fiorentinoc@macewan.ca.
allowing flexible ways of choosing terms
facilitating moving between a descriptor and its hierarchical structure
catering for the selection of alternative Boolean operators
providing a term pool option for saving the descriptors
integrating thesaurus and retrieved documents displays
making thesaurus options available in all stages of the search process.

The two interfaces provide the user with the following three spaces within a single screen:
query space: for formulating search statements
thesaurus space: for browsing and navigating the thesaurus
document space: for viewing document representations.

We created two versions of the Searching interface: one for the
Government of Canada Core Subject Thesaurus [5] and the other for the
UNESCO multilingual thesaurus [6]. Our comparative evaluation of T-Saurus
and Searching made use of the UNESCO thesaurus. The functional
prototypes of these user interfaces are available at http://thesaurusbrowser.info.

Searching User Interface

The Government of Canada Core Subject Thesaurus is a well-structured
and well-established thesaurus, which is currently being used by a number
of government agencies and information centers in Canada for indexing and
information representation [7]. The thesaurus is bilingual, which allows for
multilingual features to be designed based on the terminology of the thesaurus.

The goal of the Searching interface is to make the thesaurus information readily accessible to the user during the process of query formulation or
reformulation (Figure 1).

A tabular view allows quick navigation through the five kinds of data,
namely broader terms, narrower terms, related terms, and preferred and non-
preferred (synonymous) terms, to help inform the user where a given term falls
within the language of the thesaurus. In addition, a side panel presents a list
of the highest-level facets of the thesaurus, which gives a user unfamiliar
with the system a possible list of starting points. A language switch provides
a means of checking for corresponding terms in another language. These
terms are always visible as microtext satellites of the query terms. Their
persistent presence in the thesaurus table both reminds the user that more
than one language is available and also provides a quick means of switching
back and forth between languages. This function is also
served by an explicit language selection choice, made with a
radio button in the
panel to the right of
the main table. The
Searching user interface is similar to faceted search
user interfaces, except that it provides various thesaurus-based browsing
and search functionalities in addition to high-level facets.

In using the system, people can add as many terms as they like to the
Selected Terms list and can delete them at any time or choose to keep them
in only one language rather than both. When users have finished formulating
their query by selecting terms, choosing languages and combining them
using the Boolean operators below the Selected Terms list, they click
"Retrieve Documents" to get a list of relevant documents in the collection.

The evaluation method for Searching drew upon information search
behavior and usability techniques. Fifteen researchers were recruited from
documents the collection has to offer them on each topic). Most users also appreciated that they could keep as many items as they liked in the Selected Terms list and that they could keep them there without adding them to the search for documents by unchecking the language boxes beside each term [8].

**T-Saurus User Interface**

The T-Saurus user interface design (Figure 2) is based on the idea of representing the information as a set of visual elements rather than a series of text lists [9]. This approach allows users to interact dynamically and intuitively with the information as objects, optimize the process of retrieving information and obtain results more quickly. The interface shows a core of visual elements consisting of a set of “buckets” organized in the center of the screen. The number of buckets represents the number of terms found by the query. The size of the buckets represents the number of matches for that particular term, while proximity and opacity represent scope and accuracy of the term in relation to pre-established hierarchies for the query: main terms, related terms, more specific, more general and synonyms.

Users can also browse all the terms in the thesaurus using the panel on the left, which can be sorted either alphabetically or hierarchically by category. Again, each term has a number beside it in parentheses indicating how many documents in the collection contain the term. When a term in the list is clicked, it will appear in the center of the screen. When a term is selected by either method it is represented by a square in the central thesaurus space. By utilizing the checkboxes in the bottom of the right-hand panel, users can choose to view the thesaurus terms that are related, narrower (more specific), broader (more general) and preferred or
non-preferred (synonyms) compared to the selected term. These associated terms are also represented in the Thesaurus space by squares, and their relationship to the selected term is represented by their relative proximity and opacity.

People can use the checkboxes in the right-hand panel to show the terms in more than one language at once and also to view scope notes for selected terms. When users decide to add a term to their query, they do so by clicking on its square in the centre of the screen, at which time it is added to the Summary of Terms list, or term pool, at the top of the right-hand panel. Users can add as many terms as they like, delete them at any time, choose to keep them in only one language rather than multiple languages, and combine them using the Boolean operators below the list. When they have finished formulating their query they click Retrieve Documents to view the results (Figure 3). The red dots in the middle around the green box represent the results retrieved for the chosen term. The green box in the middle shows the thesaurus term, its French and Spanish equivalents and the number of documents indexed using that term.

When a person mouses over a red dot, the title of the document is immediately shown. In addition to the visual representation of the retrieved documents, the interface shows titles of retrieved documents on the bottom right-hand side of the interface.

**T-Saurus User Evaluation**

A comparative evaluation [10] was carried out to examine users’ attitudes, impression and thoughts about both the Searchling and T-Saurus user interfaces. As part of this study participants were asked to categorize themselves as either visual learners or linear thinkers. This decision was made to examine which user interface would be evaluated positively by the two categories of users. Twenty-five participants were recruited for this study by purposive, maximum variation snowball sampling. Although the participant pool included students and faculty members across various departments, multilingual volunteers – particularly those from the Department of Modern Languages and Cultural Studies – were specifically targeted throughout the recruitment process. The resulting participant pool was diverse, comprising professors and graduate and undergraduate students. This study used a wide range of data gathering tools, including a pre-test, post-test and usability questionnaires; interviews; audio, video and screen capture; the think-aloud technique; and direct observation. Three search tasks were designed to allow participants to interact with the interfaces, thesauri and query formulation mechanisms.

The results show that the visualization in both interfaces was found to be comprehensible to users. The Searchling interface was found to be more favorable and easier to use in terms of multilingual features, thesaurus and search functions, as well as users’ motivation to use such an interface for research purposes. Though T-Saurus was preferred by fewer users than Searchling, the most promising finding for T-Saurus is that it has the potential not only to support browsing, searching and query formulation, but also to transform these processes. It was found that linear thinkers prefer Searchling, whereas visual learners like T-Saurus. Searchling is a linear, sequential and visual interface that uses faceted structure as its default interface, and the thesaurus terms such as more general, more specific and related terms are shown automatically as soon as the user selects a term. T-Saurus, on the other hand, provides users with a more interactive and dynamic visualization interface where users need to interact with and choose the individual thesaurus term relationships to be shown.
Conclusion

Our research into the design, development and evaluation of a wide range of experimental visual browsing interfaces, including the two detailed above, has demonstrated that users’ cognitive and interactive preferences and skills may influence how they evaluate visualization user interfaces and environments. In particular, visualizing such textual information as thesauri can provide alternative ways of interacting with text and formulating queries. Within these two studies we showed that domain-specific knowledge organization systems, such as thesauri in the humanities and social sciences, can be effectively reused and repurposed to support information access and retrieval in semantically rich user interfaces that provide seamless support for searching, browsing and results displays. The approach taken in these projects can be extended to other kinds of thesauri as well as to other controlled vocabularies. There are many different domain-specific thesauri available in the areas of humanities and social sciences that can serve as sources of term selection, collection visualization and understanding, and query formulation.

Acknowledgements. The authors wish to acknowledge the Social Sciences and Humanities Research Council of Canada (SSHRC) for providing the funding support and our research assistants, Matt Bouchard, Mark Bieber, Ximena Rossello, Amy Stafford, Karl Anvik, Lindsay Doll and Paras Mehta, for their contributions to the above research projects.

Resources Mentioned in the Article


From Records to Data with Viewshare: An Argument, An Interface, A Design
by Jefferson Bailey and Trevor Owens

Creating and manipulating an interface to a digital collection is fundamentally an interpretive act. Viewshare (http://viewshare.org/) is a free and open platform that intends to make it easier for archivists, librarians, curators and scholars working with discreet digital collections within cultural heritage institutions to engage in that interpretive act. Through work in digital preservation, the National Digital Information Infrastructure and Preservation Program (NDIIPP) team is continually reminded of the critical need our partners have to provide access to the collections they are working to preserve. Access remains an essential way to demonstrate the value of these collections. Viewshare grew out of a project designed to enable NDIIPP partners to provide access to their diverse digital collections, but has widened its scope to support the broader cultural heritage community’s efforts to empower data sharing, the creation of dynamic collection visualizations and access and discovery of their digital collections.

In this essay, we describe some of the ideas and ideals that informed the design of Viewshare. This effort is, in effect, an argument for the design of this free, open-source tool. We will suggest how Viewshare’s design helps empower cultural heritage stewards to create new ways to encounter, manipulate and explore digital collections. In the next section we briefly describe three goals that have emerged through our iterative software design process. They include setting a low barrier for entry and emphasizing usability, allowing users to work with heterogeneous data and helping to reveal emergent patterns in collection data. We then provide a walkthrough of how to use Viewshare and examine a sample collection in order to illustrate how these broader goals are enacted in the design and use of the application.

Jefferson Bailey is currently a fellow focusing on digital preservation projects in the Office of Strategic Initiatives at the Library of Congress. Previously, he managed digital projects at Brooklyn Public Library and at Frick Art Reference Library and has done archival work at NARA and NASA. He can be reached at jbailey<at>loc.gov.

Trevor Owens is a digital archivist with the National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Office of Strategic Initiatives at the Library of Congress. At the Library of Congress, he works on the open source Viewshare cultural heritage collection visualization tool, as a member of the communications team and as the co-chair for the National Digital Stewardship Alliance’s Infrastructure working group. Before joining the Library of Congress he was the community lead for the Zotero project at the Center for History and New Media and before that managed outreach for the Games, Learning and Society Conference. He can be reached at trow<at>loc.gov.

EDITOR’S SUMMARY

Viewshare is a platform designed to support the work of cultural heritage institutions in their efforts to share data, dynamically visualize their collections and access and discover items. It was developed as a free and open tool for partners of the National Digital Information Infrastructure and Preservation Program. Its central goal is to reveal patterns within collections to lead to new understanding. Viewshare features a simple and intuitive interface and limited requirements for technical skills, and it supports a wide variety of digital collections and metadata types. Its users establish custom metadata, import and enhance collection data and create their own interface to support dynamic and interactive views of their digital collections. The process is illustrated with the Fulton Street Trade Card Collection from the Brooklyn Public Library.

KEYWORDS

museum collection management systems, interfaces, electronic visualization, human computer interaction
Values Embedded in Designing Interfaces

In developing Viewshare our goal was to create an application that is open, beneficial and widely adopted. To this end Viewshare leverages the existing open source Exhibit software as a lightweight structured data publishing mechanism [1]. Viewshare aspires to provide a dynamic tool useful to collection stewards and users who need to share, build and interpret digital collection interfaces. It accomplishes this task by designing features and workflow that can be utilized by non-technical users, by supporting ingest and augmentation of heterogeneous, extant data and by empowering the visualization, discovery and understanding of collection-wide trends and patterns.

Low barrier to entry. The software needs to provide collection managers without deep technical skills an intuitive way to articulate and visualize their deep knowledge of their digital collections. It needs to balance among the complex demands of being “sophisticated, robust, transparent, and easy to use” in order to attract a broad user base [2]. To this end Viewshare is designed entirely around a drag-and-drop interface.

Work with heterogeneous extant data. The software needs to support the heterogeneity of metadata across many different types of institutions and collections. It also needs to offer behind-the-scenes tools to transform this metadata into the kinds of data required for visual interfaces. To demonstrate the value of a range of digital and digitized collections, the tool needs to work with many different kinds of collections and with a heterogeneous mixture of extant content and metadata.

Reveal emergent patterns in collections. The software needs to enable visualizing cultural digital collections as unified sets of data and not just as discreet, individual objects. Visualization is thought of as a process for revealing and illustrating knowledge. Recognizing work from humanities scholars on visualization, we see the value of working with visual interfaces as part of a process in which we create new knowledge and understanding. According to Drucker, the process of visualization can be “generative and iterative, capable of producing new knowledge through the aesthetic provocation” [3, p. 41]. In embracing this subjective and iterative notion of visualization, the Viewshare platform attempts to make it easy for librarians, archivists, curators and others working with cultural heritage collections to exercise and surface patterns in their collections.

The Viewshare Workflow

Viewshare’s workflow is designed to support these particular goals. Users import and augment existing collection data, iteratively build interfaces to their collection data and ultimately are able to share the interfaces and views that they have created. We will briefly explain the steps in this process and how they are connected to the design goals and then further illustrate those connections in an example.

Viewshare interfaces are built entirely on user-uploaded metadata. Recognizing the heterogeneity of collection data, Viewshare allows multiple methods of importing data. Users can build or work from existing simple spreadsheets or MODS (Metadata Object Description Schema) records or import Dublin Core metadata via OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). To make this data usable, Viewshare includes a set of data augmentation tools to work from this extant data. For example, Viewshare enables users to derive latitude-longitude coordinates from plain-text geographic place names and then use these coordinates to plot their items on a map. Similarly, plain-text expressions of date information can be used to derive ISO 8601 formatted dates for plotting items on a timeline. With its ease-of-ingest and data augmentation features Viewshare understands and facilitates the use of the unique and sometimes idiosyncratic nature of cultural heritage collection metadata. At the same time it also allows users to enhance this metadata in order to enable the creation of dynamic interfaces.

After importing and augmenting collection data, users can begin creating interfaces. The tool’s primary purpose is building dynamic, interactive views of digital collections. Through a drag-and-drop workspace, users can create multiple views including maps, timelines, charts and other dynamic visualizations. Users can then choose which facets they want to include, and these facets will be available for each view, creating unique ways of manipulating the data presented in each of them.
For instance, in a collection of postcards, a tag cloud facet set to display subject information will show the relative frequency of the subjects throughout the collection. If a user clicks on one of those subjects, Viewshare will limit the display of whatever view they are using to show only the objects associated with that term. As a user selects the data values they want to use in a given facet, and the particular views they want to display, they can use the “show preview” function to continually toggle back and forth between building their interface and a fully functional preview of what their resulting interface will look like. In this way, the tool supports an iterative and exploratory approach to creating these interfaces.

A Working Example: The Fulton Street Trade Card Collection

The Brooklyn Public Library’s Fulton Street Trade Card Collection, a collection of late 19th and early 20th century merchant trade cards from Brooklyn, New York, is an excellent example of a “hidden” special collection that generates significant patron interest and use; it is also a good demonstration of Viewshare’s ability to empower new ways of navigating, interpreting and understanding collections. This colorful, idiomatic group of hundreds of advertising cards offers many insights into the commercial and social world of turn-of-the-century Brooklyn [4]. Special collections such as this one are not just a grouping of individual items, but should also be thought of collectively as datasets documenting the trends, influences and styles of this time period and place. Using Viewshare, a collection manager or curator, with his or her deep knowledge of the contextual details of the collections, can create views that allow users to interact with this collection in ways not possible via a static web interface; and those views can also empower users to uncover collection-wide relations not evident or interpretable through traditional online gallery display or through item-by-item browsing. A working example of the view we are about to describe is available on the Viewshare website (http://viewshare.org/views/jefferson/fulton-street-trade-cards-collection/) and an image (Figure 1) at left.

By uploading a spreadsheet of collection data that includes links to the web-accessible image files, a collection manager can begin building new interactive views. After deriving points of latitude and longitude for the cards, the user can create a map view. The map view shows the locations of each card’s business. A clickable pin on the map allows users to see a thumbnail image of the item and select metadata elements. When users add a facet to the view, they can click on any facet element, such as the Augmented Subject element “flowers,” as demonstrated in Figure 1. The map will then update to show only the location of the trade cards with flower imagery. Adding other facets such as date or subject will allow a user to further manipulate the display. A pie chart view can also be constructed from this same data set. The pie chart view defaults to a creator-defined metadata element. Each pie slice, when clicked, shows the calculation of the total number of items with that particular element in the collection and its percentage of the collection as a whole. For example, in a chart built on the element describing business type, the pie slice for “confectioners” when clicked will show there are 15 trade cards from “confectioners” representing 4.4% of the collection. Faceting can be used on charts, too, so if viewing a pie chart of subject headings, a user can then apply the business type facet.
and see that the 15 confectioners used four different image subjects in their trade card advertising images. Switching to a gallery view, a user could use the same faceting and see thumbnails of those specific cards. As users create and iteratively interact with their collections they are simultaneously building an interface and deepening their own understanding of the collection and the collection’s metadata.

Interfaces for Visualization and Understanding

There are many ways to build interactive interfaces and visualizations of cultural heritage collections. However, the time, cost and expertise necessary for creating these kinds of interfaces leave them outside the reach of many cultural heritage collections and users. To this end Viewshare is intended to empower librarians, archivists, curators and other cultural heritage professionals to create interfaces that help illustrate the value of the content they are working to preserve and make accessible. By making it easy to rapidly and intuitively create these interfaces, we hope that Viewshare can enable cultural heritage professionals to experiment with and explore their collections. In this respect, Viewshare aspires to John Bradley’s characterization of another digital humanities tool: “It is meant to be a tool that blends so well into the task of the development of an interpretation... as to be almost invisible” [5, p. 263].

In sum, working with idiosyncratic extant collection data, Viewshare enables new ways to visualize and analyze cultural heritage collections. By providing these new levels of interactivity, it encourages users to see digital collections less as discreet items and more as broad, unified datasets that can be manipulated and interpreted to form new modes of understanding. In this sense, Viewshare supports the ongoing efforts to involve existing cultural heritage assets in the growing use of information visualization in digital humanities scholarship as it also provides cultural stewards a free, intuitive tool for the display and use of digital collections.

Resources Mentioned in the Article


Beyond the Score: Music Visualization and Digital Humanities
by Margaret Lam

EDITOR'S SUMMARY
Among the humanities, music is especially powerful for reflecting time and place and evoking personal experiences and memories. Yet music research in the digital humanities has attended little to socio-cultural context, tradition, history and performance knowledge, instead focusing on the musical score and technical elements. To have an impact on music scholarship, digital research in the music domain must be considered more broadly. An example is the work of Alexander Chan, whose cutting edge music research may be appreciated as interactive installations but gets little recognition within digital humanities. Music research has fallen short in its use of digital technology to understand theory, composition, musicology and the preservation and transmission of music knowledge. Despite music's evocative power, its conceptual scope in digital research is narrow.

KEYWORDS
music
humanities
electronic visualization
research and analytic methods
information technology

Margaret Lam investigates challenges related to the transmission, preservation and organization of music knowledge. She works independently as a researcher, writer and designer as well as a collaborator on various projects. You can learn more about her work at www.margism.com.

Digital Humanities and Information Visualization

Music – beyond the scores, recordings and signal analysis that occupies information scientists and music librarians – has the amazing ability to bring human history to life.

The ethos of a time and place is inevitably reflected by the humanities, and music is no exception. We know this effect through our own personal experiences: the way couples have their song, the way music takes you back to a time in a different place, the soundtrack of car radios, movies and television shows that accompany us through various life stages. For many music scholars the socio-cultural aspect of music is an important part of their research. However, current conceptualization of music research within digital humanities does not fully address the socio-cultural element in music research. Efforts within the digital humanities are focused primarily on music as either information or artifact. While the elements of music such as pitch and rhythm are common to all music researchers to various degrees, it is not fully representative of the diverse approaches that music research can take [1]. Conceptualization of what constitutes music within digital humanities is necessarily narrow as we experiment with different applications of technologies to support music research. Yet, as I will suggest in this short piece, a broader conceptualization of music is necessary for digital humanities research in the domain of music to have real impact on music scholarship.

Music as Information, Music as Knowledge
When we conceptualize music as information, we are referring to elements of music such as pitches, chords, tempo and dynamics on one level, and contextual information such as genres, performers, dates and instrumentation on another. The development of specialized markup languages for music (see www.recodare.com/musicxml) and the largest hub for music information
research and cutting edge innovations is not an insurmountable one. It may be worthwhile here to outline the spectrum of interests and specializations that may be the context for a music researcher, based on personal experience. Broadly speaking, these areas are the following:

- Theory and Composition: the creation of music; composition techniques, musical systems
- Performance: performance practices, historical performance
- Pedagogy: methodology for the transmission of different musical practices
- Musicology: analysis of Western music, historical musicology
- Ethnomusicology: anthropological and sociological approach to studying music

While music scholars may align themselves with one or more of these research areas, most are familiar with the type of research in all the areas. The fields that are most preoccupied by music information (that is, pitch, musical structure and so forth), are those engaged in theory and composition and musicology. For all the other areas of music research, while it is no doubt a necessary skill to be able to work with music information, it is not their primary research interest. Instead, it is music as a socio-cultural and essentially human phenomenon that they are interested in. There is recognition among music scholars of the way changing and emerging technologies are changing music practices around the world, and by proxy, the way they conduct and share their research [4]. Chan’s work, for example, would be recognized as the work of a composer who is working with technology at an advanced level. The use of technology is seen as a natural extension of what composers have always done.

Such specializations within music – which most would regard as already a specialized area of study and research – create subtle divisions not only between the broad areas of research listed above, but also within each of those research approaches. Anyone who self-identifies as a music researcher will no doubt have something to say about the way I have divided and organized the list. I do not intend it to be an authoritative way of organizing, but rather as a way of highlighting the nuances behind the concept of music research.

The Scope of Music Scholarship

While the examples of Chan’s work appear to belong more in the realm of digital arts rather than digital humanities, it is without question a product of music research, in the sense that his research output is an interactive installation, rather than a musical score or a research paper. Yet, despite the common reliance on elements of music, or music information, in music information visualization and in digital humanists’ current approach to music, it is noteworthy that research output such as Chan’s is an area that is not often addressed within digital humanities.

The difference between the traditional conceptualization of music retrieval (MIR) research at the University of Illinois at Urbana-Champaign (www.music-ir.org/) reflect a score-centric approach as the dominant research paradigm in music information research today [2]. In other words, music information research is striving to enrich existing musical scores with metadata for better indexing and search on the one hand, while also working to extract information from audio signals and translating it into decipherable musical notations on the other [3].

These problems are unique research challenges that only arise within in the domain of music, but there is little research on issues dealing with music as knowledge, such as how to perform music, how to discern or attribute human intention within a piece of music and how to situate a piece of music within a musical tradition. Such issues are important to music researchers and have much more overlap with the research interests in the humanities than the score-centric approaches.

Information visualization of music has also focused on music as information, by offering a visual way of highlighting the relationship between selected structural elements (see for example: http://blog.chenalexander.com/2011/baroque-bach-cello/). In other examples of Alexander Chan’s work (http://work.chenalexander.com/), the creation or composition of music is dependent on input from external environments (such as the movement of commuters in a subway station in Sonata for the Unaware) and virtual instruments with which we create music, such as a line drawn on a screen (as in his work Crayong).
Looking Forward

Music has the ability to bring not only history to life, but also to bring the lesser-known corners of our contemporary world to the fore. In conversation with ethnomusicologists, who are among the most actively boundary-crossing music researchers encountered, it is clear that collaboration between digital humanists and music researchers is not only a great possibility, but also extremely necessary. Ethnomusicologists and their peers deal with unique challenges related to the archiving of, and fair access to, their data from their field work. Traditional music archives do not allow them to balance the need to give the communities they work in ownership and access to their field recordings. There is currently no easy way for them to share their research in a meaningful way with a general public or to actively preserve musical cultures that are transient or dying. While there is no lack of ideas from music researchers as to what new technology can do for their work, there is a lack of time, resources and expertise available to them. In the context of the music-related research within the digital humanities, there is a tremendous opportunity to further the way we preserve and transmit musical cultures that are transient or dying. This opportunity is only possible if we broaden current conceptualization of what constitutes music research within the digital humanities and if we strive to better understand and appreciate the diversity of research approaches within music as a field itself.

Resources Mentioned in the Article

The recently published international standard ISO 25964-1:2011 – Thesauri for Information Retrieval presents a data model for thesaurus structure which is more extensive than any published previously. It is intended to provide a rigorous presentation of the entities and relationships that will not only clarify and standardize the varying and conflicting interpretations that exist, but which can also be implemented consistently in automated systems. The SKOS (simple knowledge organization system) format is designed to present KOS data in a format that is suitable for machine inferencing and particularly for use in the Semantic Web. This standard is largely compatible with the ISO model, but does not yet implement all its features. Discussions are continuing on possible extensions to SKOS to cover these other features.

Structure Based on Concepts, not Terms

The model is based on the understanding that thesauri show the relationships between concepts – units of thought – and distinguishes these from the terms that are used to label these concepts. These terms may be in one or more languages, and one term per language is chosen as a preferred term for each concept. One or more additional terms for the same concept may be recorded in the thesaurus as non-preferred terms. This linkage of multiple terms to the same concept is another way of expressing the traditional equivalence relationship between terms normally indicated by the tags USE/USE FOR, although the model does also show that relationship for compatibility with existing systems. It additionally provides a “role” attribute that allows the nature of the relationship to be specified if desired, for example, that the relationship between a preferred and non-preferred term may be abbreviation/full form, formal/informal, obsolete/current or...
scientific/popular. It was thought unnecessarily complicated to provide for such relationships between one non-preferred term and another.

**Compound Equivalence**

A more complex case is that of compound equivalence, where a compound concept, such as coal mining, does not exist in the thesaurus but has to be expressed as a combination of two or more simpler concepts which are there. This case is shown symbolically as

\[
\text{coal mining} \quad \text{USE+ coal} \quad \text{USE+ mining}
\]

with reciprocals such as “coal UF+ coal mining.” Because the complex concept is not in the thesaurus, there is no provision for recording its attributes or attaching a scope note to it – it has to be interpreted from the scopes of the component concepts. As a thesaurus is normally used for post-coordinate indexing, the indexer would assign the two terms coal and mining to a document without expressing any relationship between them. A searcher would be expected to construct a search statement combining these terms with a Boolean AND operator.

In the terminology of set theory, coal mining applies to the “intersection” of the set of documents that deal with coal and the set of documents that deal with mining. On the other hand a compound concept may apply to the “union” of two or more sets of documents rather than their intersection. Although ISO 25964 does not specifically deal with this case, it is generally better for the thesaurus builder to add such a compound to the thesaurus, showing its components as narrower concepts, rather than expressing it as a compound non-preferred term. For example, rather than

\[
fossil fuels \\
\text{USE+ coal} \\
\text{USE+ natural gas} \\
\text{USE+ petroleum}
\]

Fossil fuels is better to have

\[
fossil fuels \quad \text{NT coal} \\
\text{NT natural gas} \\
\text{NT petroleum}
\]

**Hierarchical Relationships and Transitivity**

Hierarchical relationships between concepts are modelled, and the traditional symbols such as BT/NT are retained for consistency with current practice, although these designations are to be interpreted as meaning “broader concept/narrower concept” rather than “broader term/narrower term.” There is provision for each relationship to be specified by an optional “role.” This role can be used to distinguish the three types of hierarchical relationship – generic (kind of), partitive (part of) and instantial (instance of) – and even to subdivide these types further if required, but in a way that allows the distinctions to be ignored by systems that do not use them.

The first level of distinction is important in automated systems and for compatibility with ontologies, where it is necessary to recognize whether a relationship is transitive or not, that is, whether the relationship holds between concepts which are related hierarchically but where one is not the direct child of the other. A hierarchical chain in which all the relationships are generic/specific will maintain transitivity, but if it is mixed with whole/part relationships it will not. For this reason, among others, the standard recommends that partitive relationships should normally be used only in a few specific cases: disciplines or fields of discourse, geographical locations, systems and organs of the body and hierarchical social structures. The first of these could be interpreted as generic in any case – is physics a “kind” or a “part” of science? Geographical locations are a special case because the concepts have proper names that label individual instances rather than classes, so that a generic relationship is not possible. This is different from the instantial relationship, which is used to show that an instance is a member of a class.

**Top Concepts**

Each concept can have a pointer linking it to the concept at the top of any hierarchy in which it occurs. These top concepts can be facet names, for
example, and this link can facilitate browsing by clearly indicating which facet a concept is in. It can also be used for validation, because hierarchical relationships are valid only if the two concepts are in the same facet. A concept can also have a Boolean (true/false) attribute to indicate whether it is a “top concept.” This feature can be useful in producing a list of top-level concepts from which to start browsing.

These links and attributes are, strictly speaking, redundant, because top concepts could be identified by navigating up the hierarchy until no more broader concepts could be found, but as this search would use substantial processing resources it will generally be more efficient to store the information rather than determining it every time it is needed.

Associative Relationships

Similarly, associative relationships can optionally specify the nature of the relationship, such as cause/effect, process/product or person/discipline, while allowing these all to be treated as the catch-all “related concept” (RT/RT) when necessary. This allows a thesaurus to come closer to the approach taken in ontologies, where the nature of all relationships is specified.

Arrays and Node Labels

Groups of sibling concepts, which have a common parent concept, may be organized into arrays introduced by node labels. These labels are an important and helpful feature for navigation, browsing and selection of terms when hierarchical displays of thesauri are presented in a human interface, and many existing systems do not handle these array displays well. The order in which concepts are displayed within an array may be different from the alphabetical order of preferred terms, perhaps following some inherent sequence such as number, size or age. Node labels, which normally contain a characteristic of division (such as “by age” in the node label “people by age”), do not represent concepts and do not have hierarchical or associative relationships with concepts. They are not preferred or non-preferred terms, although the limitations of some thesaurus software force them to be treated as such.

Concept Groups

Many thesauri group concepts into subsets, often discipline based, called “themes,” “microthesauri,”
“domains” or “groups.” The box in the model called “concept group” provides for such groups. The concepts within such a group may or may not have any hierarchical or associative relationship with each other and may be drawn from distinct hierarchies or facets of the thesaurus, such as activities, people, places or things. Concept groups may be nested and may have a scheme of notation distinct from that used for concepts or arrays, thus providing the possibility of a classified arrangement which complements the generic hierarchy of the thesaurus itself, as in a “Thesaurofacet” or “Classaurus.”

Notes and Attributes
The model provides for notes of various types to be associated with concepts and terms, as well as allowing the addition of custom notes to cater to the particular needs of special applications. In addition, many of the boxes in the model include several attributes, and where possible these have been drawn from other standard schemes; many of the attributes of the thesaurus as a whole, for example, are those of the Dublin Core.

Version History
There is provision for attaching a version history to a thesaurus, recording the various versions that have been made available and, for each, showing what distinguishes that version from others and whether it is still current. Dates of creation and modification can also be attached to each concept and each term.

Coming Soon – Part 2: Mapping
The model given in ISO 25964 is for a single thesaurus. It may be multilingual, but the structure of concepts does not differ among languages. Mapping, or the creation of relationships between two or more thesauri or other types of knowledge organization schemes, will be discussed in Part 2 of the standard, currently in draft. To extend the model to cover such mapping would require models for each scheme to be shown side-by-side with relationships between the concepts of one and the concepts of the other.

The data model in diagrammatic form is publicly available on the website for the ISO25964 project, at www.niso.org/schemas/iso25964/. An XML schema intended for use when exchanging thesauri in whole or in part has been derived from the data model and is on the same site together with related documentation and a test document illustrating how a typical thesaurus conforming to the ISO 25964 data model can be serialized in an XML format.

Obtaining the Standard
The full ISO 25964 standard may be purchased directly from ISO in Switzerland (www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53657) in pdf or paper format or from national standards organizations such as ANSI (http://webstore.ansi.org/RecordDetail.aspx?sku=ISO+25964-1:2011) in downloadable pdf format only.

Acknowledgement
I am grateful to Stella Dextre Clarke, leader of the ISO 25964 project, for helpful comments on a draft of this article.
What Happens When Architectural Questions Are Not Asked?
by Thom Haller, Associate Editor for IA

This past fall, I disappeared from Facebook. It wasn’t intentional; I like the connections with old friends, the content friends send my way and posting my own photos and observations. Last November, for example, I posted all sorts of interesting photos – vivid foliage, images from my 11.11.11 adventures and unusual signs (such as one with an arrow that said “dirt”). During this period, I dropped in random content about my urban life (such as my experience sharing a bus stop with a man who muttered while peeling hard-boiled eggs). Important Facebook content, right?

But, did these images or comments receive any comments or likes? Nope. Nothing.

How sad. I battled ennui and loss for a while but then wondered if the lack of content had to do with a recent Facebook upgrade. I decided to investigate. First I inquired via email. “How come I never hear from you any more on Facebook?” I asked a friend.

“How come you no longer post?” she responded.

Aha. Something was amiss. I went sleuthing into new privacy options to find out more. The privacy screen (Figure 1) gave several options. In one cluster I had the choices Public, Friends, Only Me and Custom. In another cluster of information I had these choices: All, Close Friends and See all lists.

What’s the difference between these labels?

In my head, I faced these questions: “What’s the difference between Public and All?” “Why am I offered two groupings?” “Are these classification schemes different or the same?” “Are there hierarchical or associative relationships between the first and second set of links?” “Has anyone at Facebook considered this architectural quandary?”

I selected All. I made this choice because I didn’t care to filter content and figured I would allow All to see my content. Wrong. In this Facebook interface, All means nothing.

What was going on with the architectural relationships in these screens? I posed my questions online, and one of my former students, Jeremy Cluchey, sent this response:
“I believe All in the second grouping refers to all your Lists – which you would need to have set up, assigning friends to different lists, for that to work effectively. Lists were FB’s answer to Google’s Circles, but I don’t think they have been widely understood or used.”

When I work with students to structure information, we often ask what clusters of labels might be called. Typically, we find mutual understanding about why information is clustered together if the content has an understandable label.

So I asked, if the second cluster of information shows Lists, what would the first cluster be called? Jeremy provided his perspective:

“I guess I’d call the first group Who can see this? That should be the default options list for users, with choices like Everyone, Friends of Friends, Friends, No one (just yourself) and a custom option that lets you exclude specific people by name… such as share with all my friends EXCEPT for these friends.”

What’s the relationship between these clusters of information?

What about the relationship between the two clusters, Who Can See This? and Lists? Obviously, that’s the architectural challenge that prompted my user error. Even with new labels, I have to ask, “Don’t the choices Who Can See This? and Lists mean the same thing? Or is there a hierarchy? These are questions that all developers of communication products need to ask, even if they do not go by the label “information architect.”

This Facebook example shows what happens when these questions are not asked. If they were, the information might reveal hierarchical relationships rather than equivalent groupings that confused me. My student Jeremy suggested a hierarchy that could be viewed by users via a secondary prompt, such as Use Friend Lists. Here’s his suggestion:

You could offer choices like Everyone, Friends of Friends, Friends and Custom. If you select either Friends or Custom, I’d then offer a secondary prompt, something like USE FRIENDS LISTS. Clicking that would reveal all friend lists set up by the user, and you could narrow among your friends this way (if you wanted to).

Update on My Facebook Status

I’m not surprised to learn that Facebook has changed the way people can adjust their privacy settings. As it has shifted more toward new Timeline views, it has also narrowed my choices for who can see what. Gone is the nefarious All option that caused my hardship. But the unnecessary complications remind us that information architects have the opportunity to make the complex clear. And there’s a lot of work to be accomplished.