**2016 ASIS&T AWARDS COVERAGE, PART 1**

<table>
<thead>
<tr>
<th>13</th>
<th>Reflections on Receiving the ASIS&amp;T 2016 Award of Merit</th>
<th>by Peter Ingwersen</th>
</tr>
</thead>
</table>

**SPECIAL SECTION: INFORMATION STANDARDS**

<table>
<thead>
<tr>
<th>15</th>
<th>Introduction</th>
<th>by Timothy J. Dickey, guest editor</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>From Records to Things: Managing the Transition from Legacy Library</td>
<td>by Carol Jean Godby and Karen Smith-Yoshimura</td>
</tr>
<tr>
<td>24</td>
<td>Standards for Archives</td>
<td>by Morag Boyd</td>
</tr>
<tr>
<td>29</td>
<td>Practical Application of the Dublin Core Standard for Enterprise Metadata Management</td>
<td>by Camille Mathieu</td>
</tr>
<tr>
<td>35</td>
<td>The Modern Records Management Program: An Overview of Electronic Records Management Standards</td>
<td>by Jennifer Seymour</td>
</tr>
<tr>
<td>40</td>
<td>Increasing Patient Findability of Medical Research: Annotating Clinical Trials Using Standard Vocabularies</td>
<td>by Michael Panzer</td>
</tr>
</tbody>
</table>

**COLUMNS**

<table>
<thead>
<tr>
<th>44</th>
<th>IA COLUMN: The IA You See When You Can’t Read the Language</th>
<th>by Laura Creekmore</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>RDAP REVIEW: The Center for Research Data and Digital Scholarship at the University of Colorado-Boulder</td>
<td>by Shelley L. Knuth, Andrew Johnson, Thea Lindquist, Debra Weiss, Deborah Hamrick, Thomas Hauser and Leslie Reynolds</td>
</tr>
</tbody>
</table>
This issue is dedicated to the many ASIS&T members who have devoted their time and energy to the development, promotion and implementation of information standards. Standards in our field have played a vital role in allowing the collaborative creation, automation and wider exploitation of information resources, particularly bibliographic data and knowledge organization tools. Now there is also a focus on linked data standards that may allow these resources to be used in other contexts.

At an official level, the ASIS&T standards effort is spearheaded by the ASIS&T Standards Committee and its dedicated co-chairs Timothy Dickey and Mark Needleman. The committee is our representative to the national and international standards organizations and serves as liaison to us. Timothy Dickey is the guest editor of our special section on standards, which is special not only in being a focused treatment, but also in its approach. Instead of being a dry recitation of what the standards are, it presents case studies of the implementation of several important standards, including those for description, information management, record-keeping and technical interoperability, with the issues and opportunities accompanying these innovative efforts.

While our major coverage of the Annual Meeting will appear in the next issue, we are pleased to present Peter Ingwersen’s acceptance speech for the ASIS&T Award of Merit, our highest honor, as well as greetings and an update from 2017 ASIS&T president, Lynn Silipigni Connaway. I would also like to call your attention to our interview with ASIS&T communications officer, Yolande Nanayakkara. Communications officer is a new position, and Yolande is the first person to fill it. Her energy and imagination are already making a difference at the Bulletin.

We also welcome a new RDAP associate editor to the Bulletin. Amelia Kallaher is scholarly resources librarian at the Jane Bancroft Cook Library, New College of Florida, USF Sarasota-Manatee. If you have ideas or suggestions about RDAP coverage, she can be reached at akallaher<at>ncf.edu. The RDAP Review in this issue reports on the recently created Center for Research Data and Digital Scholarship at the University of Colorado-Boulder.

Finally, IA associate editor Laura Creekmore reflects in the IA Column on how vital information architecture is in airports and other situations where we are trying to find our way when we don’t speak the language.
I am honored and excited to be serving as the president of ASIS&T for 2016-2017. Our first ASIS&T Annual Meeting outside of North America was a great success. There was great energy at the conference even though the weather in Copenhagen was a bit blustery. Thirty-seven countries were represented by the attendees, an increase of 17 represented countries over those at the ASIS&T 2015 Annual Meeting.

Diane Sonnenwald and Lauren Harrison, co-chairs of the Annual Meeting, did an excellent job of creating a stimulating program and providing many networking opportunities. The idea of hosting an Annual Meeting outside of North America was Diane’s, and she and Lauren worked very hard to make the event a success. Thank you, Diane and Lauren!

We had several new events at the Annual Meeting this year. We hosted the Women Leaders of the Information Field Happy Hour and had a great turnout. Marcia Bates, Toni Carbo,
The year 2017 marks the 80th anniversary of ASIS&T. We will celebrate this event throughout the year but will plan special celebrations during the 2017 Annual Meeting in Crystal City (Washington, DC area).

Candy Schwartz and many other leaders joined the fun. The miniBARcamp, sponsored by the European Regional and European Student Chapters and ASIS&T, was well attended and a different way to meet people and share research ideas. OCLC sponsored a Storify Contest with five $500 cash prizes for students attending the conference, who documented their experiences using Storify. The five winners are Sarah Polkinghorne, Charles Sturt University, Australia; Wade Kelly, Charles Sturt University, Australia; Agnes Mainka, Heinrich Heine University Dusseldorf, Germany; Frances Ryan, Edinburgh Napier University, UK; and Laura Ridenour, University of Wisconsin, Madison, USA. Check them out at www.asist.org/events/annual-meeting/annual-meeting-2016/student-activities/. SIG CON was dedicated to Dick Hill’s 27 years as executive director of our Association. The banter was lively and fun, and the cake and champagne were wonderful! And this was one of the few Annual Meetings where registration included lunch every day at the hotel restaurant. Thank you, ASIS&T staff!

I also want to thank my predecessor, Nadia Caidi, who initiated new ways of engagement for members, including the Meet the Author Series, the Doctoral Students’ Research Videos and Regional Meetings. I had the opportunity to work closely with her and past president Sandy Hirsh throughout the year. Working with them was not only a great experience, but it also provided me with the understanding of the issues important to our members.

My goals for the coming year build on Nadia’s work. I plan to work to continue to attract an international membership, students, early career professionals and researchers by facilitating new forms of member engagement, such as webinars, workshops and events through collaborative initiatives with similar professional associations and ASIS&T chapters, committees and SIGs. I will also work to promote and support the value of the chapters, committees and SIGs. I have developed a Working Group for Presidential Initiatives that includes Becky Wilson (ASIS&T New Leader who has agreed to work with me during the year), Kendra Albright, Ron Larsen (representing the iCaucus) and Dietmar Wolfram (representing ALISE).

The year 2017 marks the 80th anniversary of ASIS&T. We will celebrate this event throughout the year but will plan special celebrations during the 2017 Annual Meeting in Crystal City (Washington, DC area). I have created the ASIS&T 80th Anniversary Advisory Group whose members include Tatjana Aparac-Jelušič, Dania Bilal, Michael Buckland, Toni Carbo, Ken Herold, Nouf Khashman and Javed Mostafa. It is chaired by Kathryn La Barre. The Advisory Group will provide content on the history of IS and ASIS&T to celebrate ASIS&T’s 80th Anniversary at the 2017 Annual Meeting and also for webinars and other events throughout the anniversary year.

One of the best aspects of ASIS&T is the members, and I have been fortunate to have the people below agree to chair or co-chair our Association’s committees. If you are interested in joining or becoming involved in a committee, you can do so on the “Become a Volunteer” page on the ASIS&T website, or, if you would like to find out more about a committee’s work,
This is going to be a very busy year, trying to increase our international membership and engagement, searching for a new executive director and planning for our 80th anniversary celebration.

Please feel free to contact the committee’s chair(s) or any Board member.

- Award Nominations Committee: Lai Ma & Anita Komlodi (Board Liaison: Abebe Rorissa)
- Awards & Honors: Soo Young Rieh & Ken Fleischmann (Board Liaison: Abebe Rorissa)
- Budget and Finance: June Abbas
- Constitution and Bylaws: Linda C. Smith (Board Liaison: Lisa Given)
- Education and Professional Advancement Committee: Keren Dali & Bill Kules (Board Liaison: Heather O’Brien)
- Executive Committee: Lynn Silipigni Connaway
- International Relations Committee: Mega Subramaniam & Elke Greifeneder (Board Liaison: Dania Balil)
- Leadership Development: Ixchel Faniel & Laura Creekmore (Board Liaison: Kathryn La Barre)
- Membership: Naresh Agarwal (Board Liaison: Fidelia Ibekwe-SanJuan)
- Nominations: Nadia Caidi
- Publications Committee: Rong Tang & Yuelin Li (Board Liaison: Jamshid Beheshti)
- Standards: Mark Needleman & Timothy Dickey (Board Liaison: Fidelia Ibekwe-SanJuan)

We have been extremely busy since the Annual Meeting. The European Regional Chapter hosted two events in Italy and one in Barcelona. The Asia Pacific Chapter scheduled an event at Nankai University on November 29. The Taipei Chapter also has a busy month ahead:

- Events at the National Taiwan Normal University on December 5
- A Taipei Chapter Reception on December 7 at National Taiwan University
- Events at the National Cheng-Chi University on December 8 and at Tamkang University on December 9. I hope to see you at one of these meetings.

This is going to be a very busy year, trying to increase our international membership and engagement, searching for a new executive director and planning for our 80th anniversary celebration. However, I’m looking forward to the exciting road ahead, so fasten your seat belts!
As we now traditionally do each year, we will provide extensive coverage of the 2016 ASIS&T Annual Meeting in the February/March 2017 issue of the Bulletin. At that time, we will include both photographic and substantive coverage of many of the events, the sessions, the parties and the fellowship.

But we’ll whet your appetite in this issue with the complete list of 2016 ASIS&T Annual Award winners, as well as the acceptance speech given by Peter Ingwersen in accepting the prestigious ASIS&T Award of Merit. More details and photographs of the awards ceremony will be provided in the next issue.

When one Annual Meeting concludes, planning for the next is already well underway. In 2017, our Annual Meeting returns to the United States with plans to meet in suburban Washington, DC, and to celebrate the 80th anniversary of our Association. Committees are already hard at work seeking submissions for all aspects of the technical program, suggestions for pre- and post-conference undertakings and ideas for social events and new ways to engage attendees in networking and team-building activities. Mark your calendars for October 27-November 1, 2017, and plan to join us in Crystal City, Virginia.

The theme for the 2017 Annual Meeting is Diversity of Engagement: Connecting People and Information in the Physical and Virtual Worlds, focusing on the diverse ways in which people from different backgrounds, cultures and disciplines forge connections with each other, discover and use information and engage with technology. We will address the opportunities and challenges of navigating through physical and digital/virtual realm with computers, tablets or smartphones to interact in work and everyday activities. Within this information environment, there also are those who choose to disengage, and those who have no access to or knowledge of technology, widening the digital divide. ASIS&T 2017 will provide an interactive platform for exploration of these complex and rapidly changing technological and socio-cultural developments.

As always, the conference committees welcome contributions from all areas of information science and technology. For more information about submission ideas, formats and deadlines, please visit the ASIS&T website at www.asist.org.

Mark Your Calendars:

ASIS&T 2017 Annual Meeting
Diversity of Engagement: Connecting People and Information in the Physical and Virtual Worlds
October 27 – November 1
Crystal City, Virginia
New ASIS&T President Sets Ambitious Schedule

Lynn Silipigni Connaway, ASIS&T president installed at the Copenhagen Annual Meeting to serve for the 2017 administrative year, is off to a running start as the international face of ASIS&T in our 80th anniversary celebration year. Here’s a brief look at some of the presentations to which she is already committed on our behalf.

At the end of November, she attended the ASIS&T Regional Meeting at Nankai University in Tianjin, China, where she gave a brief overview of the Association, its activities and the benefits offered its members. It was then on to Edinburg and the ASIS&T Annual Lecture sponsored by the European Chapter.

In early December, Lynn heads to National Taiwan University and National Taiwan Normal University where faculty members at the two institutions will host a meet-and-greet for LIS faculty members and ASIS&T Taipei Chapter members to visit with the ASIS&T president. While in Taipei, Lynn will make presentations on both her work at OCLC Research and the benefits offered its members. It was then on to Edinburg and the ASIS&T Annual Lecture sponsored by the European Chapter.

In early December, Lynn heads to National Taiwan University and National Taiwan Normal University where faculty members at the two institutions will host a meet-and-greet for LIS faculty members and ASIS&T Taipei Chapter members to visit with the ASIS&T president. While in Taipei, Lynn will make presentations on both her work at OCLC Research and the benefits offered its members. It was then on to Edinburg and the ASIS&T Annual Lecture sponsored by the European Chapter.

In early December, Lynn heads to National Taiwan University and National Taiwan Normal University where faculty members at the two institutions will host a meet-and-greet for LIS faculty members and ASIS&T Taipei Chapter members to visit with the ASIS&T president. While in Taipei, Lynn will make presentations on both her work at OCLC Research and the benefits offered its members. It was then on to Edinburg and the ASIS&T Annual Lecture sponsored by the European Chapter.

In early December, Lynn heads to National Taiwan University and National Taiwan Normal University where faculty members at the two institutions will host a meet-and-greet for LIS faculty members and ASIS&T Taipei Chapter members to visit with the ASIS&T president. While in Taipei, Lynn will make presentations on both her work at OCLC Research and the benefits offered its members. It was then on to Edinburg and the ASIS&T Annual Lecture sponsored by the European Chapter.
ASIS&T Participates in International Day of Giving

For the first time, the Association for Information Science and Technology is a part of the #GivingTuesday (www.givingtuesday.org) phenomenon. The global day of giving brings together diverse organizations and communities around the world to give back to people and projects in need of various resources. For ASIS&T’s introduction, the Board selected TechSoup Global as the recipient of donations made in the ASIS&T name. TechSoup Global provides software products to nonprofits, charities and libraries around the world.

The international day of giving was Tuesday, November 29. Final participation numbers are not yet known, but ASIS&T’s first foray into the movement was seen as a chance to make a magical impact on the lives of people without access to technology. Yolande Nanayakkara, ASIS&T communications officer, explains it this way: “#GivingTuesday is a wonderful way for the ASIS&T community to come together and work for a good cause within their field. The Board felt TechSoup Global was a good match because they provide assistance to libraries as well as other types of nonprofits that need help with software on a global basis. We hope to have more programs like this, that allow the membership to participate in activities that do good and bring us together as a community.”

If you missed the magical day, TechSoup Global would still love to have your donation! Visit the TechSoup Global donation page at https://donatenow.networkforgood.org/1426356 to make your donation. Please make sure to indicate that this is for #GivingTuesday in the Donation Program Designation field.
Editor’s Summary
Yolande Nanayakkara has been selected as the first communications officer by ASIS&T, a role prompted by the Board’s view that the organization needs to be more visible and to engage more effectively with international members. In an interview with ASIS&T Bulletin editor Irene Travis, Nanayakkara noted that ASIS&T seeks to break down geographical barriers and enhance the free exchange of ideas and networking. She intends to focus on consistent branding throughout communications and marketing, emphasizing the connection between information science and other fields of study and industries, and expanding use of technology to promote interaction. Nanayakkara looks forward to applying her personal creativity, academic background in organizational psychology and broad career experience in the management of nonprofits to help strengthen and build ASIS&T through innovative marketing and communications.

Keywords
Association for Information Science and Technology
organizational communication
marketing
association membership
interviews

Editor’s Note: Recently ASIS&T selected its first communications officer, Yolande Nanayakkara. Me, myself and I (not to mention our readers) want to know all about it, though in lieu of getting out our Quick Quotes Quill, we interviewed Yolande via that favorite muggle medium, email. Welcome aboard, Yolande!

ILT: Communications officer is a new position at ASIS&T. What is the job, officially? How does it relate to other positions? Why did ASIS&T decide to create it?

YN: ASIS&T 2016 president Nadia Caidi explained that the board felt the new communications officer position was necessary in order to expand the organization’s visibility and also to increase engagement with our international members. Though my job description hasn’t solidified and direction from the board is still pending, I believe a relevant part of the role is to develop a strategic communication plan that is systemic in its coordination and innovative in the use of available technologies to enhance outreach to existing and prospective members.

ASIS&T wants to break the geographical barrier to allow easy access in the exchange of ideas, resources and global networking. It is imperative to the organization that all members feel connected to ASIS&T, as they are the very heart of the organization. We are an international organization, and technology can provide the opportunity for all members to interact for the benefit of expanding their research, enhancing their applications and accelerating their professional careers.
In addition, this position will lead marketing efforts to ensure consistent branding, design collateral that entices prospective members and partnerships and illustrates how information science connects to multiple disciplines and industries.

The communications and marketing process is a collaborative effort. It ensures the smooth flow of communication internally as well as externally. It promulgates the efforts of the organization, be that staff, committees or the board. Therefore, I’m looking forward to working with everyone in showcasing ASIS&T’s assets and value.

ILT: What inspired you to apply for the position?

YN: I love marketing and communication, the psychology behind the field, the creativity. I applied for this position because the idea of this position being newly created intrigued me. As with most positions however, you are applying blind, not knowing much about the organization. My calls with Dick and the board convinced me that this was the right position for me and to turn down some other offers I had received around the same time (when it rains it pours!).

ILT: What is your initial impression of ASIS&T now that you are on the ground?

YN: I truly enjoy working with everyone I’ve met, and Dick has created a pleasant office atmosphere that you don’t always find in associations. My only regret is seeing Dick retire. I would have liked to have worked and learned from him.

ILT: Tell us a bit about yourself – your professional background and other things it might be interesting for us to know.

YN: I have spent my whole career working within nonprofits, with the majority of the time within associations. I started in membership, where I comprehended the importance of marketing and communication skills to excel in reaching and connecting with members. Once I started down this road, I realized how much I love the field and that I had a natural ability. I love the creative aspect of marketing, and I started writing poetry in third grade, so I’ve always been a writer. I have directed numerous departments within association management at some point in my career – membership, marketing and communications, education, conference and event planning, community outreach, fundraising and grant writing, as well as some public policy. I was the executive director for a community nonprofit and a consultant where I honed my skills in board management, finances and HR. In addition, the concentration of my major was in organizational psychology.

When I’m not working, I can be found volunteering on committees and leading projects for the Taproot Foundation. I am currently writing a novel loosely based on my grandmother, and I love sailing and kayaking.

ILT: What is your vision for ASIS&T communications, and what are your immediate priorities?

YN: I believe ASIS&T’s potential is great. It’s exciting to be a part of this new chapter. My hope is to see ASIS&T grow and strengthen as it reflects a changing field through communications that connect with stakeholders and marketing that illustrates its value.
The Future of Our Informational Environment
by Eva Ortoll and Josep Cobarsi-Morales

The first-ever ASIS&T-organized event held in Spain was the workshop The Future of Information Environments, Thinking and Building with ASIS&T, held in October in Barcelona, hosted by the Universitat Oberta de Catalunya (UOC). In addition to ASIS&T, organizational support was provided by OCLC (Online Computer Library Center), the Librarianship and Documentation Department of the Carlos III University of Madrid (UC3M) and the UOC’s Knowledge and Information Management in Organizations (KIMO) research group.

The Future of Information Environments, Thinking and Building with ASIS&T began with introductions by Marta Aymerich, vice-president of research and strategic planning of the UOC; Jordi Sánchez-Navarro, director of the studies of information and communication sciences of the UOC; Lynn Silipigni Connaway, president of ASIS&T and senior research scientist and director of user research at OCLC Research (Online Computer Library Center, Inc.) and Agustí Canals, director of the UOC’s KIMO group.

Next, the introduction of Virginia Ortiz-Repiso, professor at the Carlos III University of Madrid and chair of the European chapter of ASIS&T, gave an overview of the trends in innovation in digital information environments, highlighting different facets of the open concept (open data, open learning, open innovation, etc.), in a context that is a challenge for the design of information products and services and for innovation in the training offered to this area of knowledge. A key source for this trend briefing is an extensive study of undergraduate and postgraduate education provided by the International Academic Network Information Schools, which includes the Carlos III University and the UOC.

Then, regarding a more specific initiative, Lynn Silipigni Connaway explained her current project – Digital visitors and residents: how people engage with technology. The project aims to identify how people relate to technologies to acquire information and to communicate with others. How do they access information? How do they communicate in the digital environment? Why do they make the decisions they make when using a technology or a social network?

The theoretical framework of visitors and digital residents (V&R) postulates that the way a person relates and uses technology does not depend so much on the age or generation of the individual, but more on the individual’s current situation and context. Thus, a digital visitor is characterized by making a functional use of technology, usually linked to a formal need and accustomed to having a fairly passive presence in social networks. In contrast, a resident has a fairly significant online presence and a high level of collaborative activity; his contributions in the digital environment usually leave a certain trace (uploading videos, photos, commenting on networks, etc.). The V&R typology is not a dichotomy, but rather has to be seen as the two
extremes of a continuum. In addition, each of us can behave in different ways in different digital spaces, whether it is the intranet of our organization or Facebook, Twitter, etc. The project has been conducted in the United States, United Kingdom, Italy and Spain, including UC3M and UOC.

The application of the Digital Visitors & Residents project in the UOC is especially relevant in order to compare the forms of interaction with the technology of students and teachers in a fully online learning environment with the forms of interaction in hybrid or face-to-face learning environments. In this sense, the following interventions were carried out by Agustí Canals, Eva Ortoll and Josep Cobarsí, professors of the UOC and researchers of the research group KIMO. Some results are curious and remarkable. On the one hand, it is evident one of the postulates of the theoretical framework is the fact that the age of the users does not mean a more intensive use of the technologies, as the comment of one of the interviewed teachers illustrates: “At home at lunchtime, on the weekend, we are together at the table (my wife and my college kids) and it’s the guys who say ‘Do you want to leave the cellphone, please?’” Also during the presentation of the results of the project at the UOC, the importance of different elements became clear. On the one hand, it is clear that there is a need to pay close attention to the human and communication factor in the processes of interaction with the technologies. Interviewees in the project, whether they have a resident profile or have a visitor profile, depending upon the content and situation, consider that the intervention of people in the access and use of digital information is essential to help give context and credibility to the content with which they access and use. On the other hand, users have a need to deepen the personalization of certain services and products.

A final aspect on which the presentation focused was related to digital infoxication or infoxication. Applying filters to the overabundance of information and communication or applying strategies of digital disconnection seem to be areas in which we must be very attentive.

Also related to the V&R project, some mapping tools were presented by Lynn Silipigni Connaway and Titia van der Werf, senior program officer, OCLC Research, in an interactive session.

The last talk was by Lynn Silipigni Connaway to explain the role of ASIS&T as an association that offers its members great potential for professional networking in an international environment, with some 2,000 partners worldwide. Virginia Ortiz-Repiso discussed the activities and opportunities to interact with hundreds of members in the ASIS&T Europe Chapter.

All in all, the workshop gave us some clues as to where we should continue to work with researchers and information professionals to offer the best experience of interaction and communication to the users of this so-digital world. In this regard, the potential of the Digital Visitors & Residents conceptual approach to typify user behavior and design (or redesign) services and products was revealed in both the results discussed and in an interactive workshop and subsequent discussion of information.
Reflections on Receiving the ASIS&T 2016 Award of Merit
by Peter Ingwersen

EDITOR'S SUMMARY
At the ASIS&T Annual Meeting in Copenhagen, Peter Ingwersen expressed surprise and gratitude for being honored with the Association’s 2016 Award of Merit. The professor emeritus from Denmark’s Royal School of Library and Information Science thanked peers for recognizing him as a mentor and research originator. Ingwersen observed that his own areas of research in interactive information retrieval (IR) and scientometrics/webometrics have come together as converging lines of interest, though research and analysis are often restricted by system and feature limitations. Experimentation, Ingwersen stated, is too little focused on the effective workings of IR models and contexts, repeatability and negative research results, and is instead constrained by practicalities and pragmatics. Ingwersen expressed pleasure at the acceptance of web redundancy, citing his theory of polyrepresentation and noting the value of weighting term and search results underlying ranking of web search results. He credited the ASIS&T meetings and publications for developments stemming from original IR experimental studies.

KEYWORDS
Association for Information Science and Technology
honors
research and development
information retrieval
webometrics
term weighting
redundancy
reproducibility
ranking

Dear colleagues and friends; ladies and gentlemen!

It was a nice and highly appreciated surprise to me to have the honor to be presented with the prestigious ASIS&T Award of Merit, given in recognition of my contributions to our field of information science. And then here in Copenhagen, my home town, at the first ASIS&T Annual Meeting outside the United States. We are very proud and thankful to Diane Sonnenwald and ASIS&T for this choice. Also, I wish to thank the jury that selected me and the nominators Peiling Wang, University of Tennessee, Knoxville and Virginia Ortiz-Repiso Jiménez, president of the European Chapter, who created the two nomination packages.

I have now had the opportunity to look into the two nomination texts and the many supporting letters from outstanding IS researchers worldwide. I am very grateful for their acknowledgements of my contributions to the field over four decades of research. As with citations received to your work, you obtain a range of new perspectives on your research and activities. It seems that I have had much stronger roles as mentor as well as initiator of research than I perceived. I really appreciate the recognition of that achievement.

Looking back, it is interesting to note two phenomena in association with my lines of research:
1) How scientometric and information retrieval (IR) research are technology driven and dependent on available features in IR and citation systems.

2) How redundancy and the ranking principle have become mainstream in research and commercial retrieval environments.

Speaking to the first point, my two lines of research, interactive IR and scientometrics/webometrics, have recently started to be seen in conjunction, including at this conference. In that endeavor I have followed in the steps of B.C. Brookes, Jean Tague-Sutcliffe and Don Swanson. In particular, in scientometric and webometric research it is interesting to observe how much we are dependent of the system structures and features available (or rather NOT available) in the citation databases and search engines on the web. With link-searching features having disappeared from Yahoo (originally in Altavista), webometric analyses have been seriously inhibited. We must rely on independent crawlers and repeatability becomes more difficult to perform. As long as Web of Science and Scopus keep their online analysis features available, research evaluation is doable and repeatable online and independent from the Leiden system.

It is also my opinion, that the IR experimentation, by and large, has been and is a technical venture. Experiments and tests are very often done because they can be done, based on technical progress and constrained by the availability of test collections of special kinds. There is less focus on why and how IR models actually do (not) work or work better than other models, in particular in IR interaction – and in which topical and media environments they are most effective. Repeated experiments as well as negative result papers are rarely submitted or published. But fortunately we see an increasing amount of user-driven IR experiments.

Turning to my second point, I want to comment on how quickly redundancy became accepted by the information profession. Up to the start of the web, redundancy was to be avoided; it was unacceptable except in faceted classification. My theory of polyrepresentation from 1994 and onwards makes deliberate use of redundancy. Polyrepresentation is a consequence of the cognitive perspective on IR, promoted by Nick Belkin, Pia Borlund, Birger Larsen and many others also present today in this room. Today redundancy is all around us, often in the form of chaos on the web. But without redundancy no weighting of terms, records and so forth can effectively take place. And from that follows logically the ranking principle in IR.

With respect to the ranking principle and other algorithmic solutions provided by IR research, we all remember how difficult it was to convince and transfer that ranking principle in experimental IR to major online vendors and applied search environments 1976-1996. They did not believe in it! With the commercial search engines (Altavista, Google, Yahoo) ranking first became a natural way of searching, and it caught on very fast to become the only available and accepted search mode on the web – and elsewhere. Remember that most of those features are grounded in experimental IR and information-seeking studies done since the 60s, in particular driven by research channels like TREC, ACM-SIGIR and, not least, JASIST, ARIST and the ASIS&T meetings.

Once again, allow me to thank you all for this distinguished honor.
Information Standards: Introduction
Timothy J. Dickey, guest editor

EDITORS SUMMARY
Standards and interoperable systems are crucial in the work of information professionals, applying to bibliographic descriptions, taxonomies, data exchange formats and markup and more recently for linking data. They enable access to archival information and associations to data in other contexts. The ASIS&T Standards Committee represents the Association with the International Standards Organization, the American National Standards Institute and related organizations and serves as an informative liaison with the membership. This special section explores standards as they relate to metadata for linking library data, recognizing MARC as an early and evolving standard and reviewing standards specifically for archival description. Discussions of records management at the Jet Propulsion Laboratory, Battelle Memorial Institute and the Mayo Clinic illustrate the necessity for standards-driven metadata and systems within large, complex organizations. The Program for Cooperative Cataloging and OCLC’s Virtual International Authority File exemplify projects working toward best practices for linking data.

KEYWORDS
metadata standards
linked data
interoperability
information access
interchange formats

Standards may be viewed properly by librarians and archivists as blueprints for survival, because the interoperability of systems is now essential to the continuation of high-quality library and archival systems [1]. Standards have indeed become essential to how information professionals, broadly construed, manage change. As we transition from an older role as primarily custodians of physical information objects to managers of interoperable information systems, the technical standards to assist the machine use of our resources and records become ever more central to our daily lives. We have long been accustomed to wielding standards for bibliographic description and standards for construction and use of identifiers such as ISBNs and ISSN and standards for working with taxonomies, to name a few. Data exchange protocols from Z39.50 for database searching to format and markup standards such as PDF and XML are no less important to what we all do to make information of all media searchable by – and available to – our various clients and patrons and customers. As machine apps take on more capacity for linking data and real-world objects, the technical standards remain necessary to our information infrastructures and thus remain of central importance to how we all craft and use our information tools.

The ASIS&T Standards Committee acts as the Association’s representative to the International Standards Organization (ISO), the American National Standards Institute (ANSI) and others to advise on the creation and maintenance of standards impacting our work as information professionals. We also exist to inform the membership about standards and to maintain internal dialogue. It is for this second goal that we offer the following special section of the Bulletin of the Association for Information and Technology.

Timothy J. Dickey is co-chair of the ASIS&T Standards Committee and chair of the Central Ohio chapter of ASIS&T. He teaches on the iSchool faculties of Kent State University and San José State University and serves the Columbus Metropolitan Libraries. He may be reached at tdickey1<at>kent.edu.
The contributors to this issue address how standards – standards for description, information management, record-keeping and technical interoperability; indeed, standards past, present and future – can impact our basic job processes every day. Our application of standards to our work today carries even more impact than in the past on the long-term viability of what we do in a world where machines and technical processes can and will make more decisions about our information resources. From a librarian working with print books in the catalog to an archivist with boxes of papers and other media, to a custodian of research materials or medical records to the designers of systems for use by any information professional, each of these areas has provided articles in this issue that can be treated as use case studies about the increasing importance of technical standards to us all.

This special section begins with two explorations of how changing standards impact more traditional library descriptive practices and the increasing potential of metadata in a connected world. Carol Jean Godby and Karen Smith-Yoshimura, research scientists at OCLC Research, explore some milestones on the path from MAchine-Readable Cataloging (MARC; ISO 2709) catalog data to the new potentials of linked data on the web. The authors recognize the good results that have come from librarians’ application of the decades-old standard for bibliographic description. However, we live in a new “culture of description” that empowers mechanical creation of data linkages such as Google Knowledge Cards. Projects at OCLC, such as the Virtual International Authority File and multilingual record clusters, and initiatives by the Program for Cooperative Cataloging (PCC) towards studying URI applications in linked data, are steps along the path of transition to best-practices for linked data standards within the library community.

Similar issues are present in other collections of legacy metadata and the systems built around them, which lead seamlessly to the review of archival standards by Morag Boyd of the Ohio State University Libraries’ Special Collections cataloging department. Boyd offers important context for the 1993 introduction of the General International Standard Archival Description (ISADG; ISO 2788) and the more complete 3rd edition of the implementation standard Encoded Archival Description (EAD3; conforming to ISO 8879 for markup standards) in 2015. Boyd highlights the impact of newer standards for making rich archival metadata available in semantic web contexts and in social and other linkages.

The special section then proceeds to two informative case studies in enterprise-level electronic records management (ERM). Camille Mathieu, of the California Institute of Technology, discusses how the standards of the Dublin Core Metadata Element Set (ISO 15836) may be helpfully applied at the Jet Propulsion Laboratory (JPL). The JPL Library, as with many other organizations involving knowledge creation as a central part of their businesses, has sought an enterprise-level solution to the description, sharing and reuse of knowledge within the enterprise. Mathieu takes her readers through an assessment of Dublin Core as a standard for description of resources, some issues of clarifying namespaces in application to JPL work products and lessons learned in the implementation.

The need for careful application of standards is even more critical in the case of the Battelle Memorial Institute, due to the vast array of individual scientific endeavors and business units, argues Battelle librarian and archivist Jennifer Seymour. Seymour’s contribution to the issue discusses the challenges of developing a comprehensive records management program for such a complex organization, while maintaining compliance with President Obama’s 2011 directive on government records maintenance and other relevant standards from the Department of Defense, among the needs of other distinct business markets. Amid these myriad challenges, Seymour describes the implementation of a flexible enterprise content management system at Battelle, as well as hopes and challenges for the next generation systems operating at the intersection of different standard requirements. Seymour concludes that this level of information management is “no longer … an added value but a default expectation,” and her finding resonates with today’s final contribution by Michael Panzer, former editor-in-chief for the Dewey Decimal System and chief ontologist for the Mayo Clinic.

Panzer concludes the special section with a study from within the burgeoning field of medical informatics. He makes a strong case for the application of standard vocabularies such as the Medical Subject Headings (MeSH from the United States National Library of Medicine) and SNOMED CT (for description of clinical trials data) to improve the performance of
technical operations in medical databases and on the website ClinicalTrials.gov. The hoped-for result is that users – cancer patients, in this case – are better able to locate information regarding treatments and to do so faster and with greater accuracy. This final case study takes us full circle to the positive impact that standards can have on our professional lives and, through us, on all of the lives we touch with better systems and better information.

Resources Mentioned in the Article

To make linked data work, the library community needs good data that is structured, unambiguous and published in a format that enables linking with data produced by other communities. Library data also needs to be more about the Things or the people, organizations, places and topics that users care about and that the library community has something to say about. These qualities are the keys to integrating libraries into the web, where users are now most likely to begin their quests for information.

Such conventions are already at work in Google’s production of Knowledge Cards, which integrate information mined from billions of web documents to produce simple and actionable displays about real-world Things or entities that underlie a search request issued in a particular language. For example, a search for “Chicago” returns the display shown in Figure 1.

![Figure 1. A Google Knowledge Card for the city of Chicago](image)

Chicago, on Lake Michigan in Illinois, is among the largest cities in the U.S. Panned for its top architecture, while a skyline punctuated by skyscrapers such as the iconic John Hancock Center, 1,151 ft. Willis Tower (formerly Sears Tower) and the map-Gothic Trinity Tower. The city is also renowned for its museums, including the Art Institute of Chicago, the Shedd Aquarium, and the Field Museum.

Planes
- 3-star hotel averaging $216, 5-star averaging $471
- 3.5-5.2 min flight, around $190

Elevation: 554
Weather: 71°F (22°C), Wind 5 mph (8 km/h), 43% Humidity
Local time: Sunday 2:35 PM
Population: 2,719,000 (2013)
Area code: 312
Sports venues: Chicago Bulls, Chicago Bears, More

Points of Interest
- Navy Pier
- Shedd Aquarium
- Art Institute of Chicago
- Willis Tower
- Field Museum

Colleges and Universities
- University of Chicago
- Loyola University Chicago
- DePaul University
- Chicago State University
- USP

Carol Jean Godby is a senior research scientist at OCLC Research. She can be reached at godby<at>oclc.org.

Karen Smith-Yoshimura is a senior program officer at OCLC Research. She can be reached at Smithyok<at>oclc.org.
this display, the user can get stuff done that would be much more difficult to accomplish if the search had returned a list of documents instead.

Unfortunately, the library perspective is not well represented in Google’s Knowledge Cards, perhaps because much of our data is still confined to silos and is not comprehensible to the web at large. But Knowledge Cards offer both a glimpse of linked data’s promise and a warning that the library community has its work cut out to realize it.

In the rest of this article, we illustrate some milestones on the path from MARC to linked data by showing how translated works can be represented in a more deliberately defined entity-relationship model that can be expressed using persistent identifiers and revealed in MARC if certain practices are followed. We then discuss the efforts of the Program for Cooperative Cataloging (PCC) to convert recommendations by individual researchers for attaining greater machine understanding into best-practices conventions for the wider library community.

Defining Identifiers for Things

Researchers at OCLC and elsewhere are working to unlock the digital knowledge about creative works and their creators that libraries have been accumulating since the dawn of the computer age and make it available as linked data. Although libraries have been creating descriptions for decades using the MARC (or MAchine-Readable Cataloging) format, the results are intended primarily for human readers and are machine-processable only within library systems, not by third-party data consumers such as Google. To move forward, data scientists need to make progress on three goals:

1. Define entities and relationships that are important to the library community. Many are commonplace and well understood, such as authors, subjects and publishers of creative works. But they are obscured in MARC and other library standards that rely too heavily on text and the tacit knowledge of human readers.

2. Use the best features of MARC while recognizing that the 51-year-old standard is on the path to obsolescence. Although a MARC description may consist primarily of text, it may also feature encodings that are machine-understandable and unambiguous. These encodings are linked-data gold.

3. Replace text with identifiers that may originate from librarianship but conform to linked data conventions. For example, a library authority file is a collection of records that define or establish names or subject headings and are typically associated with control numbers or record IDs that are usually understood only in local environments. Figure 2 shows a more web-friendly definition of identifier and shows that the Thing, or the real-world referent, is more important than the heading. For example, the internationally important entity known in English as the French National Library or the National Library of France is represented by many other text strings that only humans can read. But when associated with persistent identifiers instead, machine processes can assert that they all refer to the same Thing.

The need for these developments is also recognized by the Program for Cooperative Cataloging (PCC), which was founded in 1994 to promote greater uniformity and reduce cataloging costs by actively managing the evolution of MARC and other library standards. In the process, cataloging best practices are defined and promulgated throughout the library community. When necessary, the PCC also proposes incremental changes that respond to evolutionary pressures of librarianship. In the past two
For example, Wikidata creates machine-understandable descriptions for entities harvested from Wikipedia, assigns identifiers to them and associates other identifiers such as those defined in national authority files, VIAF, the International Standard Name Identifier (ISNI), Freebase and GeoNames. Wikipedia pages in multiple languages may embed one or more of these identifiers, thus enhancing library authority file descriptions with photographs, biographies or historical context and expanding the range of names or labels for an entity.

Figure 4 shows how the description of Kawabata is affected. The content of each Wikipedia page differs, but each refers to the same person and includes the same set of identifiers, including the VIAF identifier in Figure 3. The result increases the options for an international audience of readers to enter the linked-data cloud of trusted and authoritative information about a Nobel Prize-winning author.

For example, Wikidata creates machine-understandable descriptions for entities harvested from Wikipedia, assigns identifiers to them and associates other identifiers such as those defined in national authority files, VIAF, the International Standard Name Identifier (ISNI), Freebase and GeoNames. Wikipedia pages in multiple languages may embed one or more of these identifiers, thus enhancing library authority file descriptions with photographs, biographies or historical context and expanding the range of names or labels for an entity.

Figure 4 shows how the description of Kawabata is affected. The content of each Wikipedia page differs, but each refers to the same person and includes the same set of identifiers, including the VIAF identifier in Figure 3. The result increases the options for an international audience of readers to enter the linked-data cloud of trusted and authoritative information about a Nobel Prize-winning author.

For example, Wikidata creates machine-understandable descriptions for entities harvested from Wikipedia, assigns identifiers to them and associates other identifiers such as those defined in national authority files, VIAF, the International Standard Name Identifier (ISNI), Freebase and GeoNames. Wikipedia pages in multiple languages may embed one or more of these identifiers, thus enhancing library authority file descriptions with photographs, biographies or historical context and expanding the range of names or labels for an entity.

Figure 4 shows how the description of Kawabata is affected. The content of each Wikipedia page differs, but each refers to the same person and includes the same set of identifiers, including the VIAF identifier in Figure 3. The result increases the options for an international audience of readers to enter the linked-data cloud of trusted and authoritative information about a Nobel Prize-winning author.
A Use Case for Identifiers: Multilingual Description

The cream of the world's cultural and knowledge heritage is shared by being translated. Great works are often translated multiple times. The most important works will more likely have at least one MARC record in the WorldCat database, which is more complete or comprehensive than other catalogs, with more than 350 million records representing holdings of the world's libraries, of which more than half are in languages other than English. Aggregating all the translations of a given work allows us to leverage the value of the most complete records. For example, as long as one cataloger has supplied a uniform title, we can use it for the entire Work cluster, even if no other cataloger supplied one. We can take advantage of the Cyrillic-Russian records to display the Russian title in Cyrillic on the piece, rather than one of different romanization schemes few Russian speakers could anticipate. Machines can process the identifiers for a work, author and its translators and then present the labels in the preferred language and script of the user (for example, 川端康成 for a Japanese or Chinese reader and Kawabata Yasunari for an English reader).

The relationship of a work (with an author) and its associated translations (with their respective translators) is relatively straight-forward, with each translation linking back to the original work via the property isTranslationOf. A work can have any number of translations, and there can be multiple translations into the same language, which is why identifying the translator is so important. Since it's been library practice not to include a controlled form or added entry of the translator's name, we need to parse the statement of responsibility using a long table of “translated” or “translator” in different languages to extract the translator from the MARC record. But once extracted, this information (work, author, translator, language of translation) could be presented in the same kind of Google Knowledge Card presented for “Chicago” at the beginning of this article, with the option to borrow the book from a local library.

Although translations are often described in text, MARC also allows machine-understandable encodings of the most important relationships. The conversion to linked data could be automated if a MARC record for a translated work included the following descriptors:

- The language code of the original and the language code of the translation (and, if appropriate, the language of the intermediate translation)
- Uniform title, in the script of the original
- Added entry for translator
- Roles for each personal entity

A bibliographic description from a MARC record for an English translation of a Chinese work might look like this when mapped to Schema.org:

```
# Original Work (in Chinese)
<http://worldcat.org/entity/work/id/1215997>
  a schema:CreativeWork;
  schema:creator <http://viaf.org/viaf/102266649> ;
  # “Gao, Xingjian”
  schema:inLanguage “zh”;
  schema:name “靈山”@zh-hant.

# Translated Work (in English)
<http://worldcat.org/entity/work/id/145209748>
  a schema:CreativeWork;
  schema:creator <http://viaf.org/viaf/102266649> ;
  # “Gao, Xingjian”
  schema:translator <http://viaf.org/viaf/81663420> ;
  # “Lee, Mabel”
  schema:inLanguage “en”;
  schema:name “Soul Mountain”@en ;
  schema:translationOfWork <http://worldcat.org/entity/work/id/1215997>
```

If the information curated by libraries and museums is formulated as linked data according to our recommendations, a machine process can customize the display of search results according to the user's language preferences. But the underlying data is not only more accessible but even richer because library resource descriptions such as VIAF are already linked with third-party resources such as Wikidata. They can be discovered and mined more easily to support scholarly inquiry. For example, a researcher...
interested in gauging information sharing across cultures might ask these questions: Which authors are translated the most? Which works have been translated into the most languages? How many translations are from the original work and not from a translation of a translation? Which countries or regions are the focus of the greatest translation activity, and what are the most common source and target languages? Answers to such questions can be gleaned from searches on today’s library databases, but only with much manual intervention. Linked data implementations promise more precision and comprehensiveness with less effort. Experimental prototypes have demonstrated some impressive results.

From Innovation to Best Practice

The examples described so far illustrate what catalogers have probably always understood about MARC: the standard supports a range of options for description – from text strings intended for human readers to terse but precise encodings that are designed primarily for machine processes. Since linked data is about greater machine understandability, it makes sense to craft descriptions higher on the scale.

Adopting linked data as a replacement for MARC is a long-term goal for the library community, expressed in strategy documents published by the Library of Congress and many national libraries. Progress towards this goal is often achieved through the insights of individual researchers and catalogers. But innovation must be propelled into best practices and amplified into community-wide change through engagement with standards groups.

Since 2009, MARC standards committees have recommended the addition of URIs to bibliographic and authority records. According to a MARBI (Machine-Readable Bibliographic Information Committee) position paper published in 2009

The use of a URI instead of plain text is particularly applicable to situations where the value of the...element comes from a controlled vocabulary, which could be an authority list or formal thesaurus (e.g., a name from the LC Name Authority File or a topic for an LCSH heading) or any other list of controlled codes or terms (e.g. the MARC Code List for Languages). [3]

Sample outcomes are a MARC subject field such as ‘650 #0 $a Courtship $v Fiction $0 (uri) http://id.loc.gov/authorities/subjects/sh2008100298’ or a slightly more complex MARC author field such as ‘110 2# $a University of Texas $b Dept. of Anthropology $0 http://lcn.loc.gov/n86041077 $4 spn http://id.loc.gov/vocabulary/relators/spn.’ In these examples, URIs identify authoritative web locations for the topical heading “Courtship–Fiction,” the corporate name heading “University of Texas Department of Anthropology” and the relationship code “spn” or “sponsor.”

These uses of URIs promote greater consistency and machine understanding, but it is not obvious that the real-world Things behind the headings are identified. The intent is clearer in an authority record for a living person. The example below is an excerpt from the Library of Congress Name Authority record for Donald Trump, with details pulled while he was a 2016 U.S. presidential candidate. Of interest here are the 024 fields, some listed below, which list URIs associated with the entity named by the heading “Donald Trump.”

024 7_ |a http://dbpedia.org/resource/DonaldTrump |2 uri
024 7_ |a http://vocab.getty.edu/ulan/500082105 |2 uri
024 7_ |a http://id.worldcat.org/fast/174117 |2 uri
024 7_ |a https://viaf.org/viaf/49272447 |2 uri
024 7_ |a http://www.imdb.com/name/nm0874339 |2 uri
024 7_ |a http://id.ndl.go.jp/auth/ndlna/00476339 |2 uri
...
100 1_ |a Trump, Donald, |d 1946-

Nevertheless, the resources accessible from the URIs illustrate a variety of options for human and machine consumption. Some are only human-readable English-language text, such as the Internet Movie Database, or IMDB. Some are modeled as linked data and exported as a human-readable view, such as the Japanese-language texts published by the National Diet Library (NDL). And some are primarily about controlled headings for Donald Trump, while others are about Donald Trump the person. Perhaps such noise or ambiguity is inevitable, but all are labeled as URIs in the same MARC field. As a result, a machine process can mine this information and
construct a composite view of Donald Trump as an author, a performer, a media celebrity and a political figure, which leverages the knowledge of many stakeholders and spans many languages and domains of interest.

This example illustrates what might be interpreted as a full embrace of URIs as defined by the linked data conventions. URIs refer to a Thing, something real in the world. When URIs are dereferenced, a machine-understandable description links to other descriptions of the Thing, or to related Things, to echo Tim Berners-Lee, who first articulated the linked data vision over a decade ago. When the library community retires the MARC standard and adopts the linked-data paradigm, these examples are a hint of where they might end up – where it becomes possible to project the library’s knowledge onto the web and to assimilate knowledge contributed by others. After all, other communities may not understand library data, but all groups benefit when they come together around the task of merging what is known about famous people.

But the to-do list for realizing this vision from the starting point of existing library standards is still daunting. We have to manage the transition from the old to the new culture of description. We have to create MARC records now that can be easily converted to linked data later, even as the still-experimental research projects being conducted at OCLC and elsewhere are advancing to production. We need to identify the machine-understandable URIs more clearly so that the benefits of linked data can be realized even before the transition to linked data is complete. We have to recognize that data expressed in MARC is only part of the problem, because the same issues arise in all collections of legacy metadata and includes descriptions of all kinds of objects collected and managed by libraries, not just the published books we have highlighted. Translating the good ideas that emerge from this work into best-practices recommendations is the charge of the URI task force sponsored by the PCC, which has been working since late 2015 to identify and address policy issues surrounding the use of URIs in MARC records that should be resolved before implementation can proceed on a large scale. But once implemented, URIs empower every individual with expert knowledge to contribute a link, a fact or a simple association to a collective effort that has the potential to be transformative for libraries and the information-seeking public.

Resources Mentioned in the Article


Standards for Archives

by Morag Boyd

EDITOR’S SUMMARY

Formal standards and professional practices characterize modern archival administration, increasing consistent archival description and interoperable metadata as well as the authenticity and reliability of the archives themselves. The International Council of Archives’ General International Standard Archival Description identifies 26 data elements to describe archives, being extended for the semantic web. Archives in the United States follow three sets of archival description standards. Describing Archives: A Content Standard, 2nd Edition (DACS), used together with the other standards and with MARC, describes archival materials and authority records about material creators. It stresses principles for arrangement, order and hierarchical organization. The Encoded Archival Description (EAD) contains elements to describe archival materials and interrelationships. Like DACS, it stresses respect de fonds, keeping records together in original order. Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF) describes information about people and organizations reflected in an archive. It was adopted by the Society of American Archivists in 2011 and has been used to derive 6.6 million EAC-CPF records from EAD finding aids and authority records. Archival descriptions are complex and unique. Using standardized and required descriptive elements and special search interfaces would maximize the advantage of EAD encoding and extend opportunities for data sharing between institutions.

KEYWORDS

archives
archival science
standards
data curation
interoperability
Encoded Archival Description

Morag Boyd is associate professor in the Ohio State University Libraries. She can be reached at boyd.402<at>osu.edu.

Records, papers and manuscripts have been collected, organized and retained by institutions for their long-term value probably for as long as there have been records. As for other information institutions, the quantity of materials in archives along with their diverse user communities requires systematic handling of the collections. Modern archival administration has been marked by its consolidation around recognized professional practice and the formal adoption of standards. In particular, standards for the encoding and content of archival description have led to more consistent and interoperable archival metadata.

To be clear about the scope of archival standards, archives are defined by the Society of American Archivists (SAA) as follows:

... materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control; permanent records [1].

Archives vary widely in format and in the nature of their content, representing the range of human activities that generate documents in the broadest sense. They are collected in repositories (themselves often referred to as archives). An essential role of these institutions, as noted by the International Council of Archives (ICA), is to ensure that an archive is authentic and reliable and that the context of creation of has been documented [2]. Standards-based practice for the description of archives is one way that an archival repository can fulfill this duty.

International Context

Internationally, the theory and principles of archives have different histories in countries and regions. Despite these differences, the ICA reached
agreement on a core set of 26 data elements for the description of archives with the publication of the General International Standard Archival Description - ISAD(G) in 1993, now in its second edition [3]. ISAD(G) is both a schema – defining a set of elements for description – and a content standard, as it also offers guidance for how to provide data within the element set. There are six required elements:

1. Reference code
2. Title
3. Name of Creator
4. Dates of Creation
5. Extent of the Unit of Description
6. Level of description

For example, the required element title has the purpose “to name the unit of description” and the rule of completing the title begins “[p]rovide either a formal title or a concise supplied title in accordance with the rules of multilevel description and national conventions.”

Currently, the ICA’s Expert Group on Archival Description (EGAD) is working on extending these concepts into the semantic web through developing a “conceptual data model for archival description that identifies and defines the essential components of archival description and their interrelations in order to further shared international understanding, facilitate the development of the next generation of archival descriptive systems, further regional, national, and international collaboration, and promote collaboration with allied cultural heritage communities.” [4] This work is delving deeper into resolving some differences among archival communities of practice to achieve a shared model.

In the United States, there are three core archival description standards: Describing Archives: A Content Standard, 2nd Edition (DACS) [5], the Encoded Archival Description (EAD) [6] and Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF) [7]. These standards are the national application and extension of ISAD(G).

**Describing Archives: A Content Standard, 2nd Edition (DACS)**

The American archival community uses Describing Archives: A Content Standard, 2nd Edition (DACS) as the content standard and statement of principles for arrangement and description of archives. DACS is used in conjunction with the encoding standards EAD and EAC-CPF.

DACS covers both description of archival materials and archival authority records that represent the people and organizations that created the materials. There are 25 archival elements covered by DACS, which are a refinement of the ISAD(G) 26 elements. Many of these elements are encoded into more than one EAD tag, as these tags are often broken into sub-elements that collectively represent a single DACS element. DACS can also be used with other encoded standards, including MARC.

These are the required elements for a multilevel description:

- Reference Code Element (2.1)
- Name and Location of Repository Element (2.2)
- Title Element (2.3)
- Date Element (2.4)
- Extent Element (2.5)
- Name of Creator(s) Element (2.6) (if known)
- Scope and Content Element (3.1)

Note: In a minimum description, this element may simply provide a short abstract of the scope and content of the materials being described.

- Conditions Governing Access Element (4.1)
- Languages and Scripts of the Material Element (4.5)
- Identification of the whole-part relationship of the top level to at least the next subsequent level in the multilevel description. This may be done through internal tracking within a particular descriptive system; if so, the output must be able to explicitly identify this relationship.

Each subsequent level of a multilevel description should include all of the elements used at higher levels, unless the information is the same as that of a higher level or if it is desirable to provide more specific information [7, Chapter I].

In addition to providing rules for the data to supply in an archival description, DACS also lays out the principles for arrangement and description of archival collections. One of the key principles is respect de fonds, or keeping collections separate from those originating from other sources and keeping them in original order whenever possible. The
principles also emphasize the importance of arranging or identifying logical groupings of materials and that the description reflects these groupings. This principle emphasizes the importance of hierarchical arrangement. Hierarchy is a key organizational technique, with context and description inherited from higher levels of description, allowing the whole and the parts to be described and understood in context. This approach can be contrasted with item-level description in which the description for one document is able to stand alone, as one commonly sees, for example, in library catalogs.

The final principle of archival description in DACS is that “the creators of archival materials, as well as the materials themselves, must be described” [8]. DACS directs archivists to identity all entities significant in the creation of the materials, to provide biographical information about the entities (particularly as it relates to the materials) and to use a standardized form to represent the names. DACS does not itself provide guidance for formulating names; rather, it recommends using other standards and tools such as the Library of Congress Name Authority File [http://id.loc.gov/authorities/names.html] and Resources Description and Access (RDA) instructions for constructing authorized access points.

DACS can be, and often is, used to create MARC-encoded descriptions for library catalogs. Library catalog records are typically only the front matter of the finding aid, with archival repositories depending on links to the full EAD finding aid to provide access to the components of a collection as well as longer descriptions.

**Encoded Archival Description (EAD)**

The Encoded Archival Description (EAD) is a metadata transmission standard for archival materials. Introduced in 1998, EAD recently went through a major revision; EAD3 was released in 2015. The Library of Congress is the official maintenance agency of EAD, and the SAA Standards Committee steered the recent revision by the SAA Technical Subcommittee for Encoded Archival Description (SAA TS-EAD). Although developed and maintained in the United States, the standard has been implemented in many other areas of the world.

EAD is expressed in XML with a published schema and DTD (data type definition) and is the set of elements that can be used for the description of archival materials and the relationship between the elements. EAD is designed for the intrinsic nature of archives, particularly in the American archival tradition. A finding aid, previously produced as an unstructured document, provides description of an archive as an aggregation of materials. This approach to description is intended to represent the context of creation and use of the materials, following the principle of respect de fonds – that records should be kept together in the original order as created or organized by the original creator whenever possible. EAD is an expression of ISAD(G)-defined elements for the components of a finding aid as structured metadata.

It can be helpful to think about physical boxes, folders and items in understanding EAD, but keep in mind that archives can include digital items, objects, artwork and more. Moreover, physical and intellectual order do not need to be the same. The components of an EAD encoded finding aid include the front matter with elements such as a scope and contents note which summarize the entire collection. At the lower levels of the hierarchy, more specific description can be provided at that level, without the need to repeat information from higher levels. For example, a series named “Plays” means that each box within that series – and each folder within the box – does not need to include the word plays for the user to understand that is what the item is.

Example:

```xml
<co1 level="series">
  <did>
    <unittitle>...</unittitle>
  </did>
</co1>
<co2 level="file">
  <did>
    <container localtype="box">3</container>
    <container localtype="folder">18</container>
    <unittitle>Parent-Teacher Association of Fondsville</unittitle>
    <unitdate unitdatetype="inclusive" normal="1959/1972">1959-1972</unitdate>
  </did>
</co2>
```
Archival Context (SNAC) \[10\] project has demonstrated the potential of the standard for interoperability and the usefulness of extracting and using information about organization and people from EAD-encoded finding aids. SNAC is hosted at the University of Virginia and has received support from the U.S. National Endowment for the Humanities, the U.S. Institute for Museum and Library Services, and the Andrew W. Mellon Foundation.

The SNAC prototype derived 6.6 million EAC-CPF records from existing data in EAD-encoded finding aids and authority records. These entities are then linked to related EAD finding aids in over 4000 repositories that contributed their holdings. A researcher can search for an entity and then identify archival collections that include documents created by or about that entity (Figure 1).

**FIGURE 1. SNAC record for Vannevar Bush, 1890-1974; http://socialarchive.iath.virginia.edu/ark:/99166/w6cv4jx3**

---

**Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF)**

Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF) is the XML schema for expressing information about the people and organizations represented in archives. EAC-CPF was adopted by the SAA TS-EAD in 2011 and is jointly maintained by SAA and the Staatsbibliothek zu Berlin. Like EAD, EAC-CPF grew out of an IAC standard, the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families \[9\]. The 2014 tag library included many elements, including many focused on relationships and contexts, as these are attributes valued by the archival community. For example:

```xml
<functionRelation functionRelationType="controls">
  <relationEntry>Establishment and abolishment of schools</relationEntry>
  <descriptiveNote>
    The second responsibility of the Department is to control the establishment and abolishment of schools.
  </descriptiveNote>
</functionRelation>
```

EAC-CPF is in the early stages of adoption. The Social Networks and
Conclusion

Descriptions of archives reflect the nature of these types of information resources; the descriptions tend to be complex, rich and lengthy. It is not uncommon for a finding aid to be thousands of words long. Professional theory and practice led to more consistent content in finding aids, but until the arrival of XML in the mid-1990s, there was limited ability to encode this rich data fully. International standards focused attention on the required elements of archival description, and EAD was developed as an encoding standard designed for archives. Specialized search interfaces that can take fuller advantage of the EAD encoding are certainly a benefit of standard-based archival description. However, projects like the Social Networks and Archival Context are demonstrating the incredible opportunities for reusing and building connections between the data held in many institutions. Just as archives are complex, so are the relationships among them and their many potential uses by researchers. Through early and collaborative adoption of XML and well-documented specialized implementations, the archival community has created the capacity for engagement in the emerging linked-data environments.

Resources Mentioned in the Article

7. EAD-CPF: http://eac.sstaatsbibliothek-berlin.de/about.html
10. SNAC: Social networks and archival content [website]: http://socialarchive.iath.virginia.edu/
Practical Application of the Dublin Core Standard for Enterprise Metadata Management
by Camille Mathieu

EDITOR’S SUMMARY
Large organizations relying heavily on knowledge work require effective capture and reuse of information, enabled through consistent use of standardized enterprise content metadata. The Jet Propulsion Laboratory (JPL) has undertaken a standardization effort, building an internal content schema based on established metadata field standards that are content- and application-agnostic but locally customizable for application to a broad variety of repositories. The JPL adopted the Dublin Core standard, with its Simple and Qualified properties as well as further refined Custom sub-properties. The JPL Resource Schema serves as an enterprise-wide metadata standard, while specific application profiles state the available fields and field labels for each repository or content management system. The schema’s terms are drawn from two distinct but semantically related vocabularies and linked by an intermediary registry tying granular listings for specific applications to enterprise-level terms. The registry mappings permit the use of both local metadata and higher level or external systems. The effort has demonstrated the importance of consistent application of both granular and general metadata for information capture and revealed important lessons about adopting the Dublin Core standard in a large enterprise setting.

KEYWORDS
metadata
Dublin Core
document schemas
electronic document management systems
information reuse

When information is generated by knowledge workers in an organization, the organization has vested practical and financial interests in maintaining the usability of that information. Once created, such information can be reused many times in an enterprise setting, minimizing redundancies in knowledge work products and increasing the overall efficiency of an organization. However, despite these benefits, knowledge-producing organizations have long struggled to provide workers with a straightforward means of reusing business information and this gap in enterprise knowledge management that can exact significant cost on an organization [1].

Problems associated with information capture and reuse are amplified in large organizations, where greater numbers of employees, departments and content repositories allow for the nearly unlimited siloization of enterprise knowledge. The Jet Propulsion Laboratory (JPL), a NASA research institute managed by Caltech and a leader in robotic planetary exploration, is one such large enterprise with vested interest in effective knowledge management. Most of JPL’s 5,000 employees regularly perform knowledge-work tasks – and utilize hundreds of content repositories and other systems in managing information products. As JPL works to increase the sophistication of its knowledge capture and retrieval environment, one primary area of concern is in the effective management and practical standardization of enterprise content metadata.

The Value of Enterprise Metadata
Though the business value of well-managed enterprise metadata is evident in theory, in practice many organizations neglect to consistently apply metadata to their work content because they do not see any...
immediate benefit in doing so. Further, if the application of metadata to a document or dataset is a laborious process, many knowledge workers may feel that applying metadata to their work products is outside the scope of their current task. Concerns such as these may help explain why recent surveys of enterprise information management practices suggest that about 50% of organizations have no metadata standards in place [2, 3].

This resistance or hesitation in adopting standards for organizational metadata is ill-founded, however, since metadata is vital for a number of business and information technology operations. The digital preservation of documents, intranet search and retrieval, the aggregation of like content across systems and repositories, document rights management, information validation, and records management and disposition are all key tasks facilitated by controlled content metadata [4]. The value of quality metadata for enterprise content is not evident immediately, but cumulatively. Over time, consistently applied metadata will yield greater and greater returns, while a lack of such metadata will progressively compounding retrieval issues and further stress organizational efficacy. In order to better leverage the information generated by knowledge workers, organizations should seek to develop enterprise-level standards for metadata application and management.

Standardizing JPL Metadata

JPL presently has no formal metadata standard for internal content, although attempts to develop enterprise vocabularies or to define core metadata attributes have been made at intervals over the last decade. Recently, members of the JPL Library and other stakeholders have undertaken a new standardization effort. The aim is to create a standard schema that can be used to describe JPL’s internal content, regardless of where that content is housed. Stakeholders for this effort have included not only information specialists from the JPL Library, but also repository and application managers. The content they write or manage is in need of standardized metadata to adequately describe it.

These stakeholders defined a series of parameters based on JPL’s specific information management requirements to guide in the selection and adaptation of a metadata standard. The first parameter determined that standardization efforts would center on metadata field standards, rather than on metadata value or format standards. That is, stakeholders determined at the outset that the present standardization effort would focus primarily on the creation of an enterprise-level schema instead of on the specification of value standards (such as thesauri or controlled vocabularies) or format standards (relating to metadata encoding, such as XML, RDF, etc.). By prioritizing field standardization, the present effort finds a middle ground between the detailed value standards approach and the broad strokes format standards approach, since the field standardized metadata schema can be adopted at the enterprise-level and implemented at the application-level while neither oversimplifying the content metadata nor being stalled by too much repository-specific detail.

The second parameter required that the JPL metadata standard make significant use of some established external metadata standard while still providing a level of customization that makes the metadata useful in local organizational applications. Satisfying internal information sharing requirements is of foremost importance, since it is these enterprise-specific requirements which ensure the efficient local retrieval and management of content. Adherence to external standards is similarly important, however, for sharing information outside of the organization and for making content application-neutral so that it will not be locked in to any specific application or content management system. This second parameter seeks to ensure the long-term usability of JPL content, both within and outside of the organization.

Finally, the third parameter, taking into account all of the different content types, applications, repositories and departments that comprise the JPL information environment, specified that the JPL metadata standard be both content-neutral and application-neutral and be able to be consistently applied in each of JPL’s hundreds of active repositories. Such a parameter again emphasizes the importance of internal cohesion between enterprise contents in various repositories while also looking ahead to ensure that enterprise search and aggregation systems can benefit from the consistency and neutrality of the standard organizational metadata.
Assessing Dublin Core

The Dublin Core standard, as described in ISO 15836 [5] and more extensively on the Dublin Core Metadata Initiative (DCMI) website [6], was selected as the basis of the JPL Resource Schema after a review of several established metadata standards. The Dublin Core standard falls within all predefined parameters established by JPL stakeholders, as it is general enough to describe a variety of content types in a variety of contexts, but also refined enough and customizable enough to be useful in specific application instances. For clarity, metadata properties defined by or allowed by the Dublin Core standard can be broken out into three groups:

- **Simple Dublin Core** properties are those original 15 elements first defined by the Dublin Core Metadata Workshop Series in the mid-1990s. Though this categorization is conceptually useful, Simple Dublin Core is a somewhat deprecated notion now subsumed into the *dc/terms/* namespace as high-level properties.

- **Qualified Dublin Core** properties are refinements of the original 15 elements (with some elemental additions defined more recently by the DCMI), which are presently managed in the *dc/terms/* namespace.

- **Custom Dublin Core** properties are custom refinements of the controlled *dc/terms/* elements made by local schema developers. While the Dublin Core standard does not allow for custom elements to be asserted, it does allow for the custom refinement of existing elements through the Dumb-Down Principle. This principle states that local refinements on the Dublin Core elements are supported as long as external applications can “ignore any qualifier and use the description as if it were unqualified” [7]. Adherence to this principle ensures that all enterprise-specific metadata elements can be dumbed down and ingested by external systems, even with some loss of specificity, since all custom elements are sub-properties of controlled Dublin Core elements.

In early stages of JPL schema development, existing content metadata from several repositories was mapped to either a Simple, Qualified or Custom standard element field to determine how many JPL-specific metadata properties would be supported by the established Dublin Core and how many would require custom refinements within the standard schema to remain useful in enterprise operations.

Using Namespaces, Building Profiles

Knowledge workers at JPL utilize hundreds of content repositories and applications to create, manage and store their digital work products. Each of these repositories describes content with a set of metadata fields that are distinct and disambiguated within that individual repository. While certain of these metadata fields are understandable regardless of their repository context (generic fields like “author” or “title,” for example), more specialized repositories and applications may make use of highly specific fields that are usable only in the context of a single repository. This situation leads to an unavoidable ambiguity as to how metadata fields are understood at the enterprise level. For example, if a meeting-notes repository makes use of only one “document identifier” field to uniquely identify content, but a specialized engineering database makes use of not only a “document identifier” field, but “parts identifier” and “revision identifier” fields as well, then the concept of an “identifier” is not universally understandable at the enterprise level, but only at the individual repository level. In order to make an “identifier” field understandable at the enterprise level, the field has to be refined enough to allow for the different permutations of identification that each repository will necessarily need to employ.

Thus, because of the nuance required to adequately describe work products at JPL, the standard (Simple/Qualified) Dublin Core schema had to be customized and further refined to meet JPL content management needs for use as a standard within the organization. JPL schema developers created a series of custom refinements of the established Dublin Core terms and organized these refinements in a separate */jpldc/* namespace, meaning that, in an XML record, custom properties are prefixed by the “jpldc:” prefix instead of the standard “dcterms:” prefix. All properties described in the (Custom) */jpldc/* namespace are refinements of properties in the established (Simple/Qualified) *dc/terms/* namespace and are expressible in RDF as sub-properties of these *dc/terms/* terms. Because of this property/sub-property relationship between terms in the two namespaces,
all the terms in the /jpldc/ namespace inherit effectively the same definition and usability as that of their parent dc/terms/ property. Rather than being assertions of new properties, the custom terms in the /jpldc/ namespace serve to refine the standard Dublin Core terms so as to make them usable within the specific contexts of the enterprise – while, at the same time, also ensuring that these specific fields could be understood more generically by external systems because of their sub-property relationship with terms in the dc/terms/ namespace (Figure 1).

FIGURE 1. A sample of the terms in both the dc/terms/ namespace and the /jpldc/ namespace, with sub-property relationships identified. Terms which are not sub-properties of any other properties are akin to Simple Dublin Core aspects, which cannot themselves be altered but merely refined.

<table>
<thead>
<tr>
<th>dc/terms/</th>
<th>/jpldc/</th>
</tr>
</thead>
<tbody>
<tr>
<td>term</td>
<td>subproperty of</td>
</tr>
<tr>
<td>Date</td>
<td>none</td>
</tr>
<tr>
<td>Created</td>
<td>Date</td>
</tr>
<tr>
<td>Modified</td>
<td>Date</td>
</tr>
<tr>
<td>Valid</td>
<td>Date</td>
</tr>
<tr>
<td>dateSubmitted</td>
<td>Date</td>
</tr>
<tr>
<td>dateCopyrighted</td>
<td>Date</td>
</tr>
<tr>
<td>Subject</td>
<td>none</td>
</tr>
<tr>
<td>Contributor</td>
<td>none</td>
</tr>
<tr>
<td>Creator</td>
<td>Contributor</td>
</tr>
</tbody>
</table>

Although the creation of custom properties in the /jpldc/ namespace, used in combination with the controlled dc/terms/ properties, makes the Dublin Core standard useful and usable as an enterprise-level generic schema, its design is too capacious for implementation at the application level. The JPL Resource Schema defines a large series of controlled properties, and while most applications-specific fields can be mapped to this schema, it is not practical to encode the entire schema in broad strokes for documents at the individual application level. There will be many fields in the generic, application-neutral JPL Resource Schema that are not useful for certain applications, but that are vital for adequately describing the content in other repositories. Thus, while the JPL Resource Schema may represent the enterprise standard for metadata, the implementation of this schema at the application level requires the creation of application profiles for each repository or content management system where the standard will be adopted. These application profiles will specify which standard fields are utilized by individual applications, as well as the local repository-specific label for each controlled field. By mapping between application-specific fields and those controlled properties in the JPL Resource Schema, and by selecting for inclusion in each application profile only those fields which are core to describing an application’s content, the JPL Resource Schema can be implemented at the application level with limited impact on the existing workflows of the repository owners or users.

JPL Resource Schema

The JPL Resource Schema has been in development since May 2016. Presently, the schema comprises 54 properties, 34 of which are controlled dc/terms/ terms and 20 of which are custom (sub-)properties in the /jpldc/ namespace (Figure 1). Because it draws terms from two namespaces, the JPL Resource Schema can be considered a composite schema, built from two separate yet semantically interrelated vocabularies. This composite schema represents all of the resource metadata properties that JPL tracks at the enterprise-level and the relationship of these properties to one another.

In order to bridge the gap between this enterprise-level composite schema and the granular field listings of application profiles, an intermediary registry has been proposed which will relate generic, controlled schema terms to the existing fields used by individual applications. This registry, which has at its core the composite standard schema, will associate controlled properties from the composite schema with uncontrolled application-specific fields (Figure 2). This approach of mapping individual application metadata profiles to an enterprise-level schema interferes minimally with the application’s existing metadata usage, while also allowing the application-specific content to better interoperate between applications. Further, because the JPL Resource Schema is constructed in compliance with the Dublin Core metadata standard, the...
architecture of mapping an application-specific field to a JPL schema property means that the application-specific field is ultimately readable as a standard term in the dc/terms/ namespace. By applying the JPL Resource Schema in this way, applications are able to continue operating with specific, local metadata while, at the same time, registry mappings allow enterprise-level and even external systems to understand application-level metadata (though certainly with less specificity than would be captured at the application level). Thus, the application loses none of its required metadata granularity, and enterprise-level information search and aggregation systems are able to, through the metadata registry mapping, treat specific fields more generally and relate generic property fields across repositories.

FIGURE 2. A depiction of how the properties in the two namespaces interact to form the composite JPL Resource Schema and how this schema is applied at the repository-level through the definition of controlled application profiles for each repository/application.

Implementation and Lessons Learned

As the metadata registry is developed, implementation of the standard at the application-level will occur in stages, with individualized analysis and mapping of local fields to schema properties required for each of the hundreds of JPL content repositories. As repositories at JPL progressively begin to adhere to a common enterprise-level schema, finding relevant information will become easier for JPL employees, regardless of the repository in which the information is housed. Search connectors, which currently work to index content from a variety of repositories for JPL’s enterprise-wide search, can be updated to also index controlled metadata fields, improving the relevancy of free-text searches and the accuracy of cross-repository federated aggregations. Similarly, the standardization of the organization’s content metadata in this way will make it more usable in event tracking systems, records management modules and any other enterprise content management systems that work by aggregating content from a variety of sources. Additional future work in the development of controlled vocabularies and thesauri as metadata value standards will only further these benefits, since enterprise systems must be able to rely on not only a known set of fields, but a defined set of values for those fields as well. Future work in modelling entities in the JPL information environment, similar to the DCMI’s work on its own Abstract Model (DCAM) [8], will also factor heavily in the continuing effort to standardize JPL metadata.

The practical application of the Dublin Core standard in an enterprise environment has not been without lessons learned. Efforts to standardize JPL metadata in accordance with the ISO 15836 standard brought to light some difficulties in practically utilizing this standard, as it is very brief and does not provide many implementation guidelines. The 2016 revision ISO/NWIP 15836 is currently up for a vote by ISO, and if the revision is ratified its formalization of properties in the dc/terms/ namespace and of aspects of the DCAM model will make the standard more practical and applicable. However, even with this expansion, implementing the standard requires frequent use of content distributed throughout the Dublin Core website. Any institution wishing to standardize metadata according to the Dublin Core standard should familiarize itself with both the ISO 15836 standard and Dublin Core website materials, should work within the stated confines of the standard when developing customized aspects and should realize that the precise method of implementation will be determined more by the needs of the organization than by any external guidelines.

Conclusion

Much work remains for organizations which, like JPL, are seeking to tackle longstanding issues surrounding information storage and reuse in the
enterprise. As applications, repositories and other information technologies are used increasingly in the workplace, it is important to remember the vital role the consistent application of metadata plays in making content more accessible to users. Enterprise-level metadata schemas, especially those which are built on established standards, are integral to increasing the interoperability of information both within and outside of the organization. Practically speaking, an organization looking to gain the most value from its information and digital content will construct a metadata standard that is granular enough to be used locally in and between an organization’s repositories, yet is also general enough to effectively incorporate an established standard.

**Acknowledgements**

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. An additional special thanks to the JPL Ontology Working Group and to Robert Powers and Sara Tompson for their support in this work.

---

**Resources Mentioned in the Article**

The Modern Records Management Program: An Overview of Electronic Records Management Standards
by Jennifer Seymour

EDITOR’S SUMMARY
Standardization is fundamental for bringing a vast variety of electronic records under control. It enables capturing and preserving original records as well as evidence of any access or change to the records. Creating standards within and across organizations is an extreme challenge that must be met. The 2011 Presidential Memorandum on Managing Government Records and subsequent 2012 Directive established values and strategic direction for managing federal electronic records without creating strict standards. The Battelle Record Management Office relies on a Defense Department standard for enterprise content management systems to be secure, searchable and capable of preserving contextual relationships and on the Code of Federal Regulations regarding equivalence of electronic records and signatures to paper. The result demonstrates a record’s integrity and authenticity and enables it to be discovered and accessed. Defined access permissions and an audit trail add further assurances. Interoperability through application program interface layers is another requirement, being addressed through advanced platform development, which may provide the solution for authenticity and contextual preservation.

KEYWORDS
standardization
records management
electronic documents
electronic document management systems
digital object preservation
document access
authenticity
interoperability

Jennifer Seymour is document control/records management coordinator in the Battelle Records Management Office. She can be reached at Seymour<at>battelle.org.

The vast array of electronic media, file formats and record types produced by the conduct of scientific research and the work of its supporting business units would be unmanageable without some form of standardization. The intricacies of the connections between the creator organizations would be incomprehensible, and the breathtaking volume of records would be unnavigable. Devising and implementing global standards across the parent organization, however, is nearly inconceivable. Every semantic business unit within each business line within the market verticals conducts business in the manner best suited to its success. As a result, even standardizing the records retention schedule can be a challenge. How many international non-profit 501c(3) charitable trusts dedicated to scientific discovery and its government and commercial applications across multiple market verticals have public-facing retention schedules to consult for comparison? In developing a comprehensive records management program, we aim to identify and create executable standards in each arena, building them into processes and procedures – frameworks designed to satisfy staff, clients and regulatory agencies – based on internal organizational and departmental expertise, exercising determined and judicious conversation and collaboration. Passion for artefactual preservation aside, our mission is to provide solutions to the needs of the business via the capture and preservation of the evidence of their operations.

Effective records management programs balance efficiency, cost-effectiveness, transparency and risk. In November of 2011, President Obama signed the “Presidential Memorandum – Managing Government Records,” officially beginning the federal government’s design and implementation of
a modern records management program that forces the adaptation of government activities to the electronic environment. Self-assessment, the reduction of redundancy and knowledge management are all identified within the memorandum as the values of sound records management – not the preservation of the public record for its own sake or a democratic ideal. The directive puts efficiency first in the General Records Schedules as well, aiming to reduce schedules by aggregation of records series in a move that could result in more efficient dispositioning via reducing the burden on both users and records managers. Archivists will be familiar with Greene and Meissner’s minimalistic “More Product, Less Process” approach, a modern compromise between adequacy and efficiency when faced with overwhelming volume. In records management, similar principles guide big bucket scheduling or scheduling records by aggregation. They’re facets of the same attempt to preserve the mission of records and archives, carrying out retention activities efficiently without losing context and value in the face of skyrocketing volume.

The National Archives and Records Administration (NARA) and the Office of Management and Budget released the Managing Government Records Directive in 2012, demanding that all email be managed in an accessible electronic format by the close of 2016 – now an issue of national prominence under the focus of an election year’s media spotlight – and all permanent electronic records be managed in an accessible electronic format by the close of 2019. A significant step toward standardizing the misunderstood and mistrusted format, the directive did not provide guidance or goals for other day-to-day records of government activity but is rather representative of the realization of a shift still in flux. Through providing strong strategic direction for email and permanent records, endorsing the transition from paper to electronic records management across the federal government, the agencies began to cement this change to information culture. But digitizing permanent paper collections, as is suggested, has been the default solution to preservation issues since the heyday of microfiche – so nothing much changes there. Born-digital records are different: as a society we are still wary of what is viewed as a manipulable, falsifiable medium. The trappings of context – primarily demonstrable via metadata – have become the lynchpin of authenticity in the electronic age. Standards dictate authenticity, not any inherent quality of the record.

The most stringent standards that we apply to our role in the business are the Department of Defense standard 5015.2 (“DoD 5015.2”) Electronic Records Management Software Applications Design Criteria Standard and Title 21 Part 11 of the Code of Federal Regulations (“21CFR11”), establishing the United States Food and Drug Administration regulations on electronic records and signatures. While 21CFR11 is not a generalized electronic records standard, it can be discussed here as one of the most detailed standards in common practice that directly addresses and provides a standardized solution for the active management, in records management and its sister disciplines, of electronic records and their integrity, authenticity and reliability.

21CFR11

21CFR11 addresses the conditions under which “electronic records, electronic signatures, and handwritten signatures executed to electronic records” are to be considered “trustworthy, reliable, and generally equivalent to paper records and handwritten signatures executed on paper,” applying to “records in electronic form that are created, modified, maintained, archived, retrieved, or transmitted, under any records requirements set forth in agency regulations... [or] submitted to the agency under requirements of the Federal Food, Drug, and Cosmetic Act and the Public Health Service Act” (21CFR11 Sec 11.1 Scope). The standard specifically notes that these are rules by which an electronic record can be considered acceptable in lieu of paper, implying that electronic records inherently have no authenticity or integrity. On the basis of that assumption, the standard goes on to describe the regulations and requirements for achieving those acceptable conditions. A facility must be able to demonstrate designed procedures and controls that ensure authenticity, integrity and confidentiality, requiring the validation of closed systems, the ability to produce exact copies for inspection, the administration of access controls, computer-generated and time-stamped
audit trails and uniquely identifiable authority checks, among other controls. Records requiring signature validation must be verifiably unique, cannot be reused and where not based on biometrics must meet several further qualifications.

Some of these requirements have been decried as obstructions to the conduct of business, too burdensome to implement, but they serve as a lens through which we might analyze the significance of the commonplace assumption that electronic records have one sure quality in common: manipulability. It is easier to question and disprove a record’s authenticity than it is to prove it, and if context is key to authenticity, then the record’s physical integrity, its chain of custody and the security of the repository in which it is stored must be unquestionable. Determining how to prove that a SAS file, for example, remains unaltered during the archival process involves developing and validating an automated capture application that demonstrates the integrity of the data, represented by an MD5 hash check. Every time that an electronic file is opened or moved, there is a risk that the object could be altered in some way that challenges or endangers the integrity and validity of the data. Once generated from the original file, generating a new hash and checking it against the original proves that the file remains untampered; the validation of the process serves as documented proof that the archival mechanism does not permit alteration and therefore invalidation. The mechanism also maps structural and descriptive metadata, including the MD5 hash, from the electronic record to the enterprise content management system. The system captures provenance and chain of custody, then at ingest initiates an audit trail, a key feature of the DoD 5015.2 compliant application, to provide proof of security.

**DoD 5015.2 and the Enterprise Content Management System**

The cornerstone of a modern records management program is the enterprise content management system (ECM). The current software Battelle has implemented is DoD 5015.2 compliant – not always a required standard but endorsed by NARA and expected as a best practice for federal contractors. Most of the standard is below the notice of everyday activity, but key features are tied to security, searchability and the preservation of contextual relationships. The records must be made visible via the development and application of standardized and rich descriptive metadata; the application must accommodate dates and date logic – for example, the ability to search for a range of dates; it must support meta-tagging and organization-defined metadata; it must be capable of meeting particular security compliances, featuring strict access controls and audit capabilities. It is due to the flexibility of both description and searching within the application that we are able to accommodate so many varied needs in terms of access and description, and due to access controls and audit trails that we are able to depend upon the application to provide the requisite security. The combination of the application design standards of DoD 5015.2 and the technical demands of 21CFR11 ensure an electronic object’s integrity and authenticity can be proven, and it can also be made accessible and discoverable by the guidance of principled metadata standards such as Dublin Core.

Though security begins with network authentication controls and other security measures, long before ever reaching the controls featured in application design, the ECM’s ability to define access per records collection and, if necessary, per record, is what enables the records manager to guarantee authenticity and integrity. The system administrator identifies and assigns the appropriate permissions, defining who can take action with a record: accesses can be set such that although one can view the descriptive metadata record, they cannot view the electronic object attached to it, for example. The audit trail feature, which tracks those actions, enables the administrator to view a complete history of all actions performed on a record from the point of ingest, demonstrating the authenticity and integrity of the files in storage. The context-based metadata that create the audit trails are both structural and descriptive, with the potential to either manually enter that data or, ideally, automatically capture it at ingest. Some elements are captured by the ECM by default, such as Title (File Name), Date Created, and Date Registered, but not all, and so we develop what we need to maintain not only the integrity and authenticity of the electronic object but also the accessibility of its valuable content.

Developing standardized metadata for the entire organization might be
out of the question, but developing standards based on provenance is not. Records collections are identified first by the creator organization, and collection management activities are tailored to the specific needs of each functional group, usually delineated by business unit. The Battelle ECM provides the ability to create bibliographic record templates called Record Types that use organization-defined and out-of-the-box metadata fields to describe the records to which they are attached. Keeping in mind the principles of Dublin Core, NISO, PREMIS and other descriptive metadata element standards, we have full autonomy to create our own data dictionaries and metadata elements. The fields can be string fields, text fields, dates or numbers, among other data types. Data types can be standardized by feature, such as a maximum number of characters, or restricted to organization-defined controlled vocabularies, which can be used like linked data. Any field created can also be queried, enabling multifaceted searching. A search using the Date Registered field, for example, can be filtered by Author and even Record Type itself. In concert with metadata searching, document content searching enables a text-based search of text-based file types, or files with an OCR layer, that are indexed for it by the ECM. Information can be identified as useful or relevant to an information request based on the actual content of the record, and not just an interpretation of the context.

On ESI and E-Discovery

Consider that Rule 34 of the Federal Rules of Civil Procedure inform us that one party may serve another with a request for “any designated documents or electronically stored information – including writings, drawings, graphs, charts, photographs, sound recordings, images and other data or data compilations – stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form” (Rule 34 (a)(1)). Consider next that the “responding party” to which the rule refers is not the records management office but the organization as a whole, meaning that anything in the organization’s possession, whether it has met retention but has not yet been destroyed, whether it is a copy or an original, whether it is the archived version or not, can be produced for review for litigation. The functions of an ECM that provide for multifaceted Boolean searching, multi-field filtering and document content queries to the benefit of enterprise-wide knowledge management can take the volume of e-discovery requests into the terabytes.

Volume is rapidly becoming a more difficult defense to an objectionable request. An ECM can reasonably perform a content-based query through tens of thousands of records in moments, identifying thousands of potentially responsive records based on a single search term or phrase. The operational risks are escalating dramatically. Redgrave, Peay and Bulander of Redgrave LLP presented an excellent narrative in the Richmond Journal of Law and Technology on the rapid transformation of e-discovery rules and precedents [1], reminding us that accessibility and discovery are not the end of identification, as attorney expertise will ultimately serve as the final filter before production. I will call particular attention to Case Assumption #3 and the myth that “Preservation of Electronically Stored Information Is Getting Easier with the Passage of Time.” Courts are not yet setting official precedent for sanctioning parties that fail to adequately preserve and identify potentially responsive information under spoliation rules, but the risk is quite alarming, creating the potential for a save-everything failsafe. Does this conflict with the precedent being set by the federal government to reduce detail and simplify retention schedules? How minimalistic can schedules become before they can no longer be defended? Finding the balance between transparency and risk is avoidable by the establishment and execution of not only legal hold procedures, but well-designed, standards-conscious electronic records retention policies and procedures.

Conclusion: On Marist, Rockefeller, and the API layer

Multifaceted, interoperable, complex content management systems have moved information culture out of the sphere of document management and into that of information management, no longer as an added value but as a default expectation. The ECM’s API layer is crucial to its success as an interoperable repository software. The ability to write code bridging the myriad business and collaboration applications in operation to the ECM, automating the intelligent capture of electronic records, lessens the burden...
on both the end user and the records management team. Often this relies on careful management of the records by their owners, or administrators of the systems in which the records reside. Users can work from SharePoint and their records can be captured with little to no overhead, capturing both provenance and original order; the code written for this task maps metadata from document libraries to the ECM, capturing key fields like Site Title, URL, Date Created, and Date Modified, as well as the organizational hierarchy of the folders themselves. The same tool can be tweaked to capture anything from a Windows Explorer style hierarchy. The interoperability provided by application program interface (API) layers enables the fluid exchange of records and information from origin to archive, and we’re already seeing significant progress in this arena.

This year it was announced that Marist College and the Rockefeller Archive Center were taking this technology to the next phase [2], working together to develop a platform, with an API layer, that can support the complexity of managing electronic records originating from near-limitless varieties of creators and account for rapid changes to technological contexts. If they are successful, information professionals could be witnessing the technical solution to the problem of electronic records authenticity and contextual preservation.

Resources Mentioned in the Article


Increasing Patient Findability of Medical Research: Annotating Clinical Trials Using Standard Vocabularies

by Michael Panzer

Among industries that rely heavily on the use of terminology management in a broad sense, health care, along with libraries and financial institutions, might be facing the toughest challenges in developing appropriate strategies for the use of vocabularies or ontologies to organize its knowledge. One reason is the variety of knowledge assets encountered in a clinical environment, which goes far beyond the classical paradigm of the text-based document or bibliographic resource. Rather, medical institutions like Mayo Clinic are confronted with the need to transform these assets (from symptom lists over care process models to clinical decision rules) into actionable artifacts that transport a specific standard of patient care in a way that supports physicians and patients alike in their shared decision making processes.

Which role can metadata play in operationalizing this knowledge? At Mayo, several groups are involved at the same time in different forms of knowledge representation, for example, clinical knowledge management with its focus on physician support or medical informatics with its focus on automated data and natural language processing (in addition to data governance and other data standardization efforts).

In contrast, the ontology group (by taking a more LIS-centric approach) collaborates closely with content generators inside Mayo, such as editors of consumer-oriented health information for the mayoclinic.org website. A specific strength of the group is the application and curation of a variety of controlled vocabularies (from value lists to ontologies) in an advanced environment using semantic technology. Through directly working with editorial content, the annotation workflow informs the ways the underlying knowledge standards evolve. While other groups are working with unmediated clinical data, the expertise in metadata design and implementation of standard vocabularies unique to the ontology group oftentimes serves as
connective tissue between research, physicians, consumers and clinical knowledge assets.

One specific example for the role of standards and metadata framework design in clinical information management is the reworking of the way clinical trials are published on the Mayo Clinic websites. The overhaul of clinical trial management was part of a larger initiative to improve comprehensive research management, but in this article I focus specifically on issues of interoperability and findability addressed by implementing standard value vocabularies, leveraging a metadata element set derived from domain modeling.

**Publication of Clinical Trials – Status Quo Ante**

What is a clinical trial (or, more precisely, a clinical study; I am using the two terms interchangeably from this point on)? According to the definition of the National Institutes of Health, “clinical trials are research studies that test how well new medical approaches work in people. Each study answers scientific questions and tries to find better ways to prevent, screen for, diagnose or treat a disease.” Recruitment of participants becomes crucial for the success of clinical studies. A smaller sample size than optimal based on the study design usually results in less reliable scientific outcomes. Recruiting participants, of course, requires that people are able to find appropriate studies in the first place, preferably based on their conditions or health interests.

Indeed, one of the key challenges for clinical research institutions is recruiting and retaining participants in clinical trials and other research studies. One of the main goals of redesigning the publication process and website for clinical trials at Mayo Clinic by leveraging standard vocabularies is to increase trial participation.

Several key weaknesses were identified in the legacy process for recruitment, which lead to a situation where not only prospective participants had trouble finding trials, but also study coordinators. The system was mainly browse-focused, based on tags assigned in an ad hoc manner by IT teams. The lack of consistency in breadth and depth of indexing was aggravated by the trial summaries being written for specialists, not including participants as one of the audiences. At some point, MeSH (Medical Subject Headings, edited by the National Library of Medicine) was adopted as a standard source for medical terms, yet it was only applied in a fairly limited, lexical way (using the preferred headings only).

**Excursion: Standard Metadata Needs Robust and Reliable Data**

A prerequisite to annotating clinical trials with standard metadata was shifting the workflow of capturing core study data to include the ontology group’s TopBraid environment (called semantic services environment, or SSE). Together with Sitecore as the content management system, these two components form an integrated KCMS (knowledge content management system). Without going into more detail here, core study data from various sources such as clinicaltrials.gov is integrated by epiCenter (a study protocol information system), which sends an appropriate subset of such studies to KCMS, that is, Sitecore and SSE. An ontologist in the workflow annotates the clinical trial before the annotated catalog item gets published to the web in various ways (see below). Also, the annotations are then shared back to epiCenter and also, for Mayo-sponsored studies, all the way back to the original registration at clinicaltrials.gov (see Figure 1).

**Designing a Metadata Framework for Clinical Trials**

The metadata framework relies on the development of a clinical study domain model/ontology, as well as on the selection of candidate clinical
vocabulary. Clinical trials are conceptualized as a subset and extension of a larger domain model encapsulating a broad view of the entities and relationships involved in Mayo research as a whole, as represented by the research web. The research web domain model tries to account for all entities that play a role in the domain such as Person, Organization (with a subclass for Department, etc.), Location and Information Resource (with a subclass for Publication).

The annotation design, on the other hand, attempts to be much more specific, as you can see in Figure 2. The class of clinical trials is at the center of the design. The relationships can be grouped into three broad categories: Which disorder(s) and medical procedure(s) is the study investigating? Which drug will be investigated as an intervention? Which clinical departments are conducting the research study?

In order to convey the first two categories in a standardized way, SNOMED CT and RxNorm, respectively, were selected as value vocabularies. SNOMED CT is a comprehensive clinical health terminology with more than 300,000 concepts, governed by an international body from 28 member countries. SNOMED CT allows us to capture the investigated condition and intervention at a very granular level and also provides for post-coordination of concepts and inference of properties. For drug intervention, RxNorm as a vocabulary provides normalized names for clinical drugs, but also links to many other drug vocabularies.

Obviously, much more needs to be specified to ensure consistent application of vocabulary terms. Such rules include that the primary condition captures the primary topic of the study, while the secondary condition may capture a condition of the population being studied if different from the primary condition, for example, diabetic neuropathy in patients with untreated diabetes. Both are constrained to concepts from the UMLS semantic group of Disorder.

Both SNOMED CT and RxNorm alone, while providing the appropriate features to code clinical concepts in an interoperable manner, do not address the problem of translating the clinical idiom into terms used by consumers of health information and, by extension, increasing findability of studies by prospective participants. RxNorm provides generic and brand names of drugs, which helps in that regard. SNOMED CT concepts contain a rich set of synonyms, which also helps closing the jargon gap.

As a third vocabulary, the Mayo-curated Consumer Health Vocabulary (CHV) was included in the design to address some of these findability issues. CHV is a fairly compact SKOS-based scheme of consumer-oriented concepts of conditions, procedures, symptoms, devices and human anatomy (~5000 concepts with a rich set of relationships connecting symptoms, diagnoses and treatments). Augmenting SNOMED CT with CHV also allows for a tighter integration with health information and patient education content on mayoclinic.org, most of which is already annotated with CHV concepts, whereas SNOMED CT allows for interoperability with clinical content.

The ontology group already curates a mapping of SNOMED CT to CHV, which is leveraged in the design to derive the relationships “associatedProcedure” and “associatedCondition.” The connection to CHV also allows acquiring related body systems to further enrich the annotation with relevant terminology.
The value for the third category of properties is selected from a list of organizations curated as part of the larger research web ontology.

In summary, the design increases findability by allowing search for conditions and procedures (based on SNOMED CT and CHV, including synonyms), for drug names (brand and generic, based on RxNorm), for associated organizations (for example, clinical departments) and body systems. Figure 3 shows an example of a clinical trial instance (i.e., a catalog item) with most properties present.

Interoperability, Findability

As we have seen, SNOMED CT as a vocabulary provides the access points from a clinical standpoint and is thus closely aligned with the language of medical research. But to be able to increase visibility and get studies in the path of potential participants visiting the websites, CHV serves as the main access point. Figure 4 gives an overview of the various locations on Mayo websites at which clinical studies show up automatically based only on shared annotations alone. Clinical trials are accessible on mayoclinic.org from diseases/conditions topics (for example, breast cancer), treatment/procedure topics (for example, liver transplant) and clinical departments (for example, the Breast Clinic). On the research site mayo.edu/research, they are directly integrated with research centers conducting trials (e.g., Cardiovascular Research Center).

Findability improvements are mainly reached through enhanced site search, incorporating the additional access points explicated above. A second main driver of search is a new clinical trials landing page, leveraging the core descriptive metadata from epiCenter (study phase, open/closed status, location and so forth) in the form of facets, and the model-based annotations as an autocomplete feature in the search field.

Evaluating Outcomes

No formal evaluation has been attempted since the launch of the search and (complementary) annotation enhancements went live, but some data is still available for a quick reality check. The Mayo Clinic Cancer Center saw an increase in inquiries from 3500 to 5000 a year after the improvements went live, with the volume of inquiries continuing to increase. As we have seen, successfully implementing a metadata framework that promotes interoperability and findability of assets is a multi-stage process involving strategic and operational choices, from reliably acquiring source data, designing a model and crafting an annotation workflow to enabling the use of annotation in search and information architecture.
I traveled recently to Mexico for a meeting. I speak just a tiny bit of Spanish – I took the language for three years in school, and I always enjoyed it and had a decent facility for it. But since then, I’ve only had a few occasions to use it. I knew that my trip would be a fun time to refresh my Spanish a little bit, at least on some routine conversational interactions.

But I didn’t expect to spend my time thinking about the information architecture of travel. Yet that’s exactly what was glaringly obvious to me the moment I stepped off the plane in Mexico City.

When you strip away the information provided by understandable language, the architecture sticks right out at you. I don’t mean the design of the building, of course. I mean the design of the information. This feature begins to become obvious at an international airport, where you may find signs in multiple languages. My own home airport in the United States has some signs in Spanish, but I pay little attention to them (or to the signs that aren’t translated) since I’m looking at the English words.

At the Mexico City airport, with the situation reversed, I realized immediately how much the architecture of the information could give me hints – when it was analogous to the way I’d seen it handled in U.S. airports – or confuse me, when it was different.

I knew we had arrived at Terminal 2, and for a while, we were trying to figure out how to get to Terminal 1 to meet some friends. My companions and I stood in a large hall with at least five exits. Three went outside, in different directions. Two went to long interior hallways. I saw no icons for trains, yet we knew there was some kind of train or tram that led to the other terminals. I didn’t see any numbers that appeared to represent the other terminals either. Fortunately, we solved that problem by contacting our pre-arranged driver and asking him just to come over for us, instead of vice versa. I wouldn’t be at all surprised to learn that there were signs that explained exactly what we needed to know, but we didn’t see the cues we needed. If we hadn’t been able to contact our driver, we’d have had to find an English-speaking employee (of which there were certainly many – but part of the fun of international travel is seeing what you can come up with on your own).

I also had trouble getting my phone to work, despite having contacted my carrier prior to travel and ensuring I had the appropriate plan. I decided the thing to do would be to get on wifi once we arrived at the hotel and talk to my carrier to resolve the situation.

While we were waiting at the airport, I noticed a sign referring to free wifi. I could tell enough to figure out the network name and the password. But I couldn’t read Spanish well enough to tell who provided the service – so I
felt unsure about whether I should trust the security. I decided to give it a try but do nothing other than contact my carrier via their app. A dumb move or a smart one? I’ll tell you when I get my next bill. But again, I realized that being familiar with brand names and knowing how official government programs are named are very important to establishing trust. We’ve all seen scams that use a near-miss, official-sounding brand or government type name. Frankly, I’ve judged people who fall for them. After my trip, I realized how sophisticated your understanding has to be to know whether or not you can trust information.

When we returned to the airport to leave a few days later, I found even more ways that the Mexican airport architected its information differently. We started at the airline ticket counter to check in a bag, and there were no arrivals or departures listed there. In fact, we had to walk all the way down to the security line at the other end of the hall to find flight information. We weren’t sure that was the right way to go, in fact, because we still didn’t know our gate number, but those were the screens we saw. Once we arrived, we learned we were in the right spot.

The other really useful reminder to me: I realized I made a lot of my decisions in the airport based on whether or not I thought I would look foolish. I’m not a person who spends a lot of time thinking about what others think about me. But I realized that when I was very uncertain of the right choice, and I believed I was surrounded by people who I felt must obviously see the right choice themselves, my desire to appear confident in the face of my ignorance was especially strong. Now surely no one was wasting their time to notice whether I had to double-back or was otherwise confused – but add a little stress to an already increased cognitive load and see how rational your decisions are!

When we’re managing IA problems, understanding the context and the emotions our audience bring to the table is so important. Sometimes it may help to put yourself in a situation where things are confusing to you to see how much your audience depends on the non-textual cues you create with your information architecture.

What information are you conveying (or hiding) with your architectural decisions?
The Center for Research Data and Digital Scholarship at the University of Colorado-Boulder

by Shelley L. Knuth, Andrew Johnson, Thea Lindquist, Debra Weiss, Deborah Hamrick, Thomas Hauser and Leslie Reynolds

In October 2016 the Center for Research Data and Digital Scholarship (CRDDS) was created with the mission of supporting the University of Colorado Boulder (CU Boulder) and the general community in developing an advanced data infrastructure and of assisting CU Boulder researchers with developing data skills and advancing their digital scholarship. CRDDS is an active collaboration between the CU Boulder libraries and research computing, the group that provides large-scale computational and data infrastructure on campus. The center’s mission includes data education, data quality consulting and assistance with data analytics, visualization, management, storage and preservation. CRDDS will build on existing services and skills across the CU Boulder campus to build a research and scholarly data ecosystem that is cross-disciplinary and connected to many local, regional and national collaborations. The main purpose of the center is to assist researchers with the following tasks:

- How to determine what data exists and how to retrieve it
- How to prepare data for use
- How to analyze and visualize data using the latest tools
- How to preserve and manage data for proper future use and reuse.

It is anticipated that, by providing these services to researchers, the center will advance the state of the art in digital data and scholarship, by specifically contributing to:

- discovery and reuse
- access and publication
- management, curation, and preservation
- analysis and visualization
- training and education.

CRDDS will be a partner in research and scholarship projects primarily for campus groups, but also for the community by providing support services, new training and education approaches, and consulting. Initially the center is organized around four initiatives, each with an initiative director: research data management, cyberinfrastructure, education and training, and digital scholarship. Each initiative is described below.

The Four CRDDS Initiatives

Research Data Management. The Research data management initiative of CRDDS builds off the existing research data services (RDS) group at CU Boulder [1]. This group, the first formal collaboration between the libraries...
and research computing at CU Boulder, provides support for research data management on campus (https://data.colorado.edu). RDS provides support for compliance with data management planning and data sharing requirements for federal funding agencies and journal publishers. RDS offers one-on-one consultations, seminars and workshops to assist researchers, with about 40 consultations and 10 workshops per year [1]. RDS also administers the DMPTool (http://dmptool.org)/ [2], creating 240 data management plans as of September 2016.

Cyberinfrastructure. The cyberinfrastructure initiative of CRDDS leads development and deployment of the central CU Boulder data infrastructure. This infrastructure includes storage, big-data analytics approaches and data management and curation software and applications. The PetaLibrary, the largest storage infrastructure at CU Boulder, is managed by research computing and is, in conjunction with the CU Boulder ScienceDMZ, the core of the current research data infrastructure of CRDDS [1]. CRDDS-affiliated research staff will work with researchers to provide consultation surrounding this and other storage services across campus. Support will also be offered for data management, analytics, visualization, cleaning and processing. The cyberinfrastructure initiative will also assist users in discovering knowledge that exists in big data and in leveraging cloud or other computing services for their data analysis needs.

Education and Training. This initiative will work closely with the other initiatives of CRDDS, as well as other campus groups, to ensure the CU Boulder campus training needs are met. This initiative will develop new or incorporate existing courses and informal seminars centered around data science and management skills. This training program will also develop short online modules or videos about data science subjects designed to reach an even larger audience. In addition to the informal training approaches, several formal for-credit classes related to data are being developed to create a data-related certificate program. The tuition revenue from the certificate is anticipated to help fund the center.

Digital Scholarship. The digital scholarship initiative promotes exploration and integration of digital scholarship tools and methods – such as social network analysis, geospatial analysis, text and data mining, and digital exhibits – into research and teaching. This initiative will facilitate support networks for all stages of the digital project life cycle, from project planning to dissemination of scholarly outputs. It also will promote researcher engagement with cultural heritage data, especially digital primary-source collections, and offer guidance in scholarly communication, including authors’ rights, open access and publishing in the institutional or other repositories.

Structure and Inclusion of Other Groups on Campus

The governance structure of CRDDS includes four initiative directors, two executive directors, an executive board, an advisory board and a set of affiliates. The initiative directors include representatives from both the libraries and research computing, while the executive directors are the senior associate dean of the libraries and the director of research computing. The executive board consists of the associate vice chancellor for IT, the dean of libraries, two deans from CU Boulder colleges or schools, and the vice chancellor for research.
CRDDS will strive to create an inclusive and cross-disciplinary culture to encourage participation from other groups across campus who are participating in data-related center activities. This effort will be particularly relevant through the advisory board and the affiliates program. The advisory board will consist of up to 12 members including a broad representation of stakeholders from faculty, students, staff and community advocates. The board will include at least five tenure-track CU faculty, one post-doctorate, one graduate student and one undergraduate student. Affiliates will include personnel across campus and within the community who are interested in formal and informal partnerships with CRDDS.

Future Work
As of this writing, the CRDDS has been an official entity for less than one month. The initial plans are to develop a consulting and seminar place within the main library on campus, with the intention of having this space available in early 2017. The seminar space will have a direct network connection into the ScienceDMZ to enable visualization of large datasets. Seminars will be broadcast to the larger community via streaming. Other plans are to develop the advisory board and set of affiliates. Plans to incorporate other campus groups are well underway. The center will play an integral role in the development of the PetaLibrary 2.0 and other data related campus infrastructure.

Resources Mentioned in the Article
