

Spatial Discovery

Expert Meeting

June 17–18, 2015
Santa Barbara, California

FINAL REPORT

Savannah Cooley, Sara Lafia,
Antonio Medrano, Denise Stephens, Werner Kuhn

Center for Spatial Studies
UCSB Library
University of California, Santa Barbara



Supported by
UCSB Library

UCSB Center for Spatial Studies
ESRI Inc.

Website: <http://spatial.ucsb.edu/2015/spatial-discovery>

Table of Contents

Introduction	1
Ten Key Insights from the Meeting	2
Goals and Agenda	4
Keynote Addresses	
User Perspective: Anabel Ford "Discovering the Complexity of Spatial Data: The Evolution and Organization of the Maya Forest GIS"	6
User Perspective: Alan Liu "Digital Humanities and the Reorientation of the Humanities Knowledge Space"	7
Spatial Perspective: James Boxall "Re-positioning Spatial Discovery"	7
Library Perspective: Christopher Gist "History of Geodata and Metageodata at the University of Virginia Library"	8
Library Perspective: Marcel Fortin "Spatial Data Discovery and Support in a Library Setting"	9
Breakout Sessions	
Library Needs and Requirements	11
User Needs and Requirements	13
Spatial Needs and Requirements	15
ArcGIS Online Demo	17
Plenary Sessions	
Unified Vision: Werner Kuhn	18
Best Practices: Marylynn Francisco	20
Identifying the Gaps	23
Moving Forward	26

Introduction

This report summarizes a two-day expert meeting on “Spatial Discovery,” organized jointly by the Library and the Center for Spatial Studies of the University of California, Santa Barbara (UCSB), and held on June 16–17, 2015 at the Upham Hotel, in Santa Barbara. The 24 participants contributed expertise in Library Science, as well as knowledge pertaining to spatial information and relevant research on data-seeking behavior. Five keynote addresses as well as several plenary and break-out discussions explored the challenges, best practices, and potential strategies associated with the cross-platform discovery of spatial data in the context of modern libraries. The meeting was made possible by a generous grant from ESRI (www.esri.com) and collaboratively organized by the UCSB Library and the UCSB Center for Spatial Studies.

The UCSB Library provides integrated services and research support for scholars pursuing data-driven inquiry. Its history includes the ground-breaking collaborative development of the Alexandria Digital Library project (ADL), as well as ongoing development of data-centric programs designed to cross disciplinary boundaries. The Library’s recent Spatial Discovery grant enables new collaborative investigation of technologies to facilitate the identification of spatial data content in the burgeoning and complex Internet. Bringing meta-data expertise and a history of data services to the effort, the UCSB Library seeks to build a data services architecture that expands the accessibility to and persistent identification of spatially referenced data assets far beyond the University.

The UCSB Center for Spatial Studies is dedicated to promoting campus-wide events, research, and teaching for all disciplines that recognize the importance of spatial thinking in science and in artistic endeavors, the development of spatial analytic tools, and the importance of place in society. The Spatial Discovery meeting was held in the 25 year tradition of specialist meetings organized by the Center, promoting intensive discussion on themes related to theoretical issues, technological developments, and applications of geographical information science and spatial thinking in science and society. These meetings are intended as catalysts for new research and teaching programs, new software developments, and funded research initiatives.

The organizing committee for the meeting consisted of:

- Denise Stephens (University Librarian, UCSB Library)
- Werner Kuhn (Director, Center for Spatial Studies)
- Antonio Medrano (Research Coordinator, Center for Spatial Studies)
- Karen Doehner (Administrative Coordinator, Center for Spatial Studies)
- Isabella Madarang (Executive Assistant, UCSB Library)
- Savannah Cooley (Center for Spatial Studies Research Intern, Clark University)
- Sara Lafia (Geography Graduate Student, Center for Spatial Studies)

Prior to the meeting, all participants were asked to submit a short biography and a perspective on the topic of the meeting. A list of all participants along with their submissions is available online at <http://spatial.ucsb.edu/2015/spatial-discovery/participants>.

Ten Key Insights from the Meeting

While the topic of spatial information discovery is difficult to delimit and the meeting discussions were intentionally scoped broadly, it seems worthwhile to distill key themes and related insights that resulted from the meeting. These ideas may not necessarily be shared by all participants; they are stated here only as a point of departure for further investigations, as well as for experiments with library discovery tools, rather than as conclusions.

1. The hub of any future “spatial university” is the library. The *role of spatial information and information technologies* in libraries and for interdisciplinary research is fundamental to their re-invention. At a time when libraries reorganize contents, curation, and discovery, a spatial view of these three aspects has the potential to transform the library user experience.
2. A spatial view of contents is a key enabler of *transcending the traditional “information carrier” view* of library contents. Spatial discovery is not (only) about the discovery of maps and images. It is about the discovery and linking of contents related to some location, no matter on what information carrier or in what physical location it is being held.
3. The cataloguing expertise in the library community and the spatial data and reasoning expertise in the GIS community ideally *complement each other*. The two approaches have traditionally come together in the organization of map and imagery collections, but are now expanding to support the discovery of and access to all sorts of contents, in particular to research data.
4. Pursuing this vision comes down to supporting library users in two basic tasks: to *“do the map”* (i.e., to view their data geographically) and to *link contents to a map*. One of the keynote speakers described “linking a map to a book” as a formative childhood experience of hers. With today’s mapping and semantic technologies, such experiences can now extend to any student and researcher in the sciences, social sciences, humanities, arts, and engineering.
5. *Discovering and accessing contents by their relation to places* is a key enabler of interdisciplinary research. Neither libraries nor GIS are currently supporting place-based inquiries well. The “where” questions that library users can currently ask are limited. However, the revival of gazetteers and the broad take-up of semantic technologies have laid the ground for new capabilities.
6. Traditional libraries, with physical bookshelves and study areas, enable *serendipity in discovery* and for browsing for related items. With digital contents, we have lost this support for intellectual neighborhoods and chance encounters and must recreate it. Once we figure out how to do this (for neighborhoods in geographic, historic, and thematic spaces), the impact on serendipity, learning, and transdisciplinarity will likely surpass any traditional search and discovery mechanisms.
7. Library staff is under enormous pressure to catalogue and curate traditional and rapidly growing new content. It is unrealistic to add spatial indexing of contents to their tasks, particularly not at any broader scale. Thus, *automated generation of metadata* from

contents, particularly from research data, is imperative. It must be combined with human intervention for quality assurance and in some cases with crowdsourcing. One approach to enable this is to define something like “spatial aboutness”: library data are about objects and events, many of which are situated in geographic spaces. One needs to determine suitable spatial referents for the data (what spatial entities can they be about?) to discern how library contents should be spatially organized.

8. As is the case for spatial data, discovery of library contents in general may undergo a transition *from catalogues to contents with more intelligence* in them, rather than held separately.
9. Tackling *granularity and history* are the two quintessential requirements for spatial discovery.
10. This effort can fundamentally be seen against the background of open access policies and measures taking hold in libraries and academia at large: *Once information is accessible, how will users find it?*

Goals and Agenda

The aim of the Spatial Discovery meeting was to develop an interdisciplinary research agenda to address the challenges that libraries and researchers face in trying to discover data via spatial metadata on diverse platforms and in a variety of contexts. This topic was approached from three distinct perspectives—user, library, and spatial—each with its own set of requirements and research questions on “spatially referenced content”:

User requirements:

- What are researcher needs in terms of using and sharing spatially referenced content?
- What discovery capabilities would users like so they can connect with other research and researchers through spatially referenced content?
- What should be the relationship between library catalogs and GIS?
- What support do users need for their data management plans?

Library requirements:

- Do libraries want a single point of access to spatially referenced content?
 - How should spatial search interact with general library search tools?
 - Should spatial metadata be included in the general catalog?
 - What genres or classes of material in the general library collections should be retrievable via spatial search?
- What metadata standards should be used by a catalog?
- Should library users be allowed to deposit their own spatial data into the collection? If so:
 - What sorts of vetting and controls should be placed on self-depositors?
 - What sorts of incentives can be offered to self-depositors to encourage well-formed and described data?
 - What version controls need to be available?
 - What metadata standards and procedures should vendors follow?
- How should offsite content interact with locally controlled content? Should it be discoverable in the general library catalog?
- How should map- and text-based searches interact?
 - How can text documents be indexed spatially?
 - Should the library expend effort on extracting spatial metadata from text documents?

Spatial requirements:

- How should we define and delimit “spatially referenced content”
 - When is it useful (and when is it not) for a spatial reference to be defined?
- What is required to make content spatially referenced?
 - Gazetteers
 - Semantics through linked data
 - Automated metadata generation
- How can spatial discovery cover all aspects of “spatial”?
 - Spatial extents

- Place names
- Feature types
- What spatial metadata provide effective search?
 - place names, descriptions, valid time / transaction time, extents
 - provenance
 - coordinate systems
 - what standards offer these elements?
- How can spatial discovery mechanisms become robust and adaptable to future technologies?

The meeting agenda was customized to maximize interdisciplinary interaction between participants and, at the same time, to target research questions as an outcome. Each interactive breakout session was led by a moderator and recorded by a designated recorder; a reporter then summarized the group's discussion for the other groups. The complete agenda is available online,¹ and the summarized agenda was as follows:

Day 1 (Wednesday, June 17)

- Welcome and introduction
- Keynote addresses:
 - **User Perspective:** Anabel Ford (MesoAmerican Research Center, UCSB) "Discovering the Complexity of Spatial Data: The Evolution and Organization of the Maya Forest GIS"
 - **Spatial Perspective:** James Boxall (GISciences Centre Libraries, Dalhousie University) "Re-positioning Spatial Discovery"
 - **Library Perspective:** Christopher Gist (GIS Specialist, Library, University of Virginia) "History of Geodata and Metageodata at the University of Virginia Library"; and Marcel Fortin (Map & Data Library, University of Toronto) "Spatial Data Discovery and Support in a Library Setting"
- Breakout Session: *Needs and Requirements*
- Plenary Session: *Unified Vision*
- ArcGIS Online Demo

Day 2 (Thursday, June 18)

- Plenary Session: *Best Practices*
- Breakout Session: *Identifying the Gaps*
- Keynote address and Discussion: Alan Liu (Department of English, UCSB) "Digital Humanities and the Reorientation of the Humanities Knowledge Space"
- Library Collaboratory Tour
- Plenary Session: *Moving Forward*
- Concluding remarks and discussion

¹ Spatial Discovery Meeting Agenda: <http://spatial.ucsb.edu/wp-content/uploads/sd-Agenda-6-15-15.pdf>

Keynote Addresses

The contents of the five keynote addresses are summarized below. The presentation slides and other relevant documents are available online on the Spatial Discovery Meeting website.²

User Perspective: Anabel Ford

MesoAmerican Research Center, UCSB

“Discovering the Complexity of Spatial Data: The Evolution and Organization of the Maya Forest GIS”

Anabel Ford is recognized for her discovery of the ancient Maya city center of El Pilar, on the contemporary border of Belize and Guatemala. At El Pilar, Ford is advancing programs that simulate “Maya Forest Gardens” as an alternative to conventional monocrop farming. Her interdisciplinary research involves gathering regional and site-specific spatial data showing the environment and settlements of El Pilar, including information about soil type, monument locations, forest cover, elevation, drainage, place names, roads, and boundaries. The UCSB Maya Forest GIS comprises this information, which takes a range of GIS data formats—including LiDAR, aerial images, and shapefiles.

Ford expressed a desire to make the UCSB Maya Forest GIS data accessible and displayable online. Initially, this was accomplished through the ground-breaking ADL³ spatial search portal, which was in operation between 1999 and 2008. ADL was the first distributed digital library with collections of georeferenced materials hosted online and with search capabilities that included spatial criteria. This project was the inspiration for many future spatial search technologies, including the open-source GeoBlacklight system (<http://geoblacklight.org/>) and the internal system used by the National Geospatial-Intelligence Agency (NGA). However, when ADL funding ended the Maya Forest GIS data was no longer available through that portal (although the footprints of the data and summaries can still be found on the ADL website). Today the UCSB Maya Forest GIS data circulates within a small group of people directly relevant to the project, mainly local government members, archaeologists, and the non-profit El Pilar Maya Forest Garden Network (EPFGN). Putting the UCSB Maya Forest GIS data back online would facilitate collaboration with other people interested in this project beyond the circle of colleagues currently able to access the data, including researchers, policy-makers, and local advocates in the surrounding area of El Pilar.

Another point of interest that Ford discussed involved finding support for data management, particularly because most NSF grant applications now require a data management plan as part of the proposal. Organizing poorly labeled data contributions to the Maya Forest GIS has proved to be a time consuming task. Ford expressed a desire for the library to fulfill the role of creating a centralized space with resources that inform GIS users on best practices for data management. The discoverability and usability of data, once it ends up in a library repository, depends on the quality of the data and metadata and on whether the metadata conforms to the standards used by the library.

² Keynote presentation slides: <http://spatial.ucsb.edu/2015/spatial-discovery/agenda>

³ Alexandria Digital Library legacy page: <http://legacy.alexandria.ucsb.edu/>

The library could offer additional resources that support GIS users in learning how to meet those standards.

User Perspective: Alan Liu

Department of English, UCSB

"Digital Humanities and the Reorientation of the Humanities Knowledge Space"

In his capstone talk at the meeting, Alan Liu discussed many different approaches to thinking about "Spatial Discovery." Among those mentioned were the use of knowledge-finding methods to discover spatial resources, the use of spatial wayfinding aids and analogies to discover knowledge resources, and the "knowledge space" of scholarly methods, which in the case of the Humanities knowledge space, is changing drastically in the digital age.

A core part of ancient, pre-digital scholarly methods that upheld the humanities knowledge spaces involved ingesting and archiving content. Liu drew from "Archives as a Place" by Luciana Duranti, describing the place of deposit of the archival documents in the most remote part of the archival building, isolated from any possible source of contamination or corruption, and untouchable by social and political forces. Duranti also discusses the "archival threshold," the space where the officer of the public authority takes charge of the documents, identifies, and associates them intellectually. Liu emphasized that this space does not necessarily have immaculate order because the organization of resources is based on *respect des fonds*, as the physical space implies order. In the digital age, although the physical archival structures that relied on the spatial organization of content are lost, there is the opportunity to reconstruct the process of ingesting, archiving, and discovering digital content with spatial logic instead of purely relational logic. This transformation could take the form of enabling any combination of text-based, temporal, and spatial search through federated catalogs. The new paradigm of thinking about information spatiotemporally mitigates the loss of physical discovery spaces and provides users with a new sense of curiosity and wonder as they ask questions and search for the content that can reveal answers.

Spatial Perspective: James Boxall

GISciences Centre Libraries, Dalhousie University

"Re-positioning Spatial Discovery"

James Boxall established historical context for the meeting, illustrating the past vision of the ADL project panel on which he served, and painting a picture of what the future could hold for spatial discovery. Looking toward the future, he felt it is important to determine if spatial data libraries like ADL will be at the heart of research and the promotion of geo-literacy, or if they will simply become irrelevant. His argument was for the former, but with the caveat that as progress is made, it is important not to lose sight of the foundation. That is, progress should not be made for the sake of progress; it should always be mindful of the original objectives of the problem and the lessons learned from past work.

Although systems such as ADL and ideas like the Digital Earth offer a means to begin connecting everything and everyone spatially, complexity is a common barrier to engaging users. A recent Canadian Geomatics Study exemplified a core community of experts that do not communicate with the growing group of spatial users on the periphery. There is a need to align user visions and leverage expertise to create an “invisible infrastructure” that supports spatial data discovery.

Organizing big metadata is central to Boxall’s research and teaching, as he has grappled with diverse challenges ranging from finding use for more than 100,000 photos taken from the International Space Station to making sense of 1TB of data/day generated from Dalhousie University’s Ocean Tracking Network.

Boxall cautioned the audience not to treat spatial data infrastructures like a game of Jenga. As foundational pieces of the puzzle—such as geo-literacy and continuity—are removed, systems become unstable. With the expanding availability of spatial data and computational complexity come increasing opportunities for spatial conceptualization; however, progress must not come at the cost of abandoning ideas that gave inspiration and meaning to the technology in the first place.

Library Perspective: Christopher Gist

GIS Specialist, Library at Scholars’ Lab, University of Virginia

“History of Geodata and Metageodata at the University of Virginia Library”

Christopher Gist began his talk by reviewing some of the free online spatial data discovery and display tools used at the University of Virginia Library. Gist first introduced the Historical Census Browser⁴ tool and the Social Explorer⁵, both of which have access to the same historic demographic data. However, Social Explorer’s spatial search and display functionality is more advanced and enables users to customize maps with side-by-side map comparisons, satellite views, storyboard tools, and annotations. Although the Social Explorer mostly renders the Historical Census Browser tool obsolete, it remains available because some K-12 textbooks still point to this resource and use the Historical Census Browser because of its differences in spatial data search capabilities, such as the longitudinal data search.

Gist also mentioned: the Virginia Gazetteer (with text-based search and spatial delivery); Map Scholar⁶ (a platform for geospatial visualization featuring high-resolution images of historic maps); the David Rumsey Map Collection⁷ (featuring a spatial-temporal search using MapRank⁸ that searches and displays a historical map collection of more than 60,000 historical and current maps and images); and Neatline⁹ (a geotemporal exhibit-builder developed by the Scholar’s Lab in the University of Virginia Library that allows the user to build complex maps, image annotations, and

⁴ University of Virginia Historical Census Browser: mapserver.lib.virginia.edu

⁵ Social Explorer homepage: www.socialexplorer.com

⁶ Map Scholar: mapscholar.org

⁷ David Rumsey Map Collection: www.davidrumsey.com

⁸ Map Rank: www.mapranksearch.com

⁹ Neatline: neatline.org

narrative sequences from Omeka¹⁰ collections of archives and artifacts, and to connect maps and narratives with timelines).

From 2006–2007 the Scholar’s lab collaborated with the Open Source Geospatial Foundation¹¹ and Stanford University to create standards for metadata and tools. In 2009 the Scholar’s Lab participated in the Scholarly Communication Institute’s 7th annual meeting, titled, “Spatial Technologies and the Humanities,” which focused on the nature and implications of spatial tools and methods for scholarly communication. All of these projects advanced Gist’s dream “for users to seamlessly discover and use spatial data through traditional and spatial searches.”

Finally, Gist discussed open source platforms that are considered to be the present-day state-of-the-art for data discovery, including OpenGeoportal¹², GeoBlacklight¹³ for spatial search, as well as the OpenGeoMetadata¹⁴ initiative working toward an open source geospatial metadata standard.

Library Perspective: Marcel Fortin

Map & Data Library, University of Toronto

“Spatial Data Discovery and Support in a Library Setting”

Marcel Fortin’s presentation illustrated the increasing number of roles librarians have at the University of Toronto with respect to geospatial data support. Librarians are responsible for data acquisitions, data management, and data dissemination. The geospatial collection size increased from 3 Gigabytes to 4.2 Terabytes between 1999 and 2012. Along with this increase in data comes a demand for geodata literacy skills, software support, and research consultations, including geostatistics and analysis (see Figure 1). The University of Toronto Library uses a geoportal called Scholars GeoPortal¹⁵, built by the Ontario Council of University Libraries. It has a text-based search showing the spatial extent of the data with download options for a subset of data with a defined bounding box as well as a preferred coordinate system and data format.

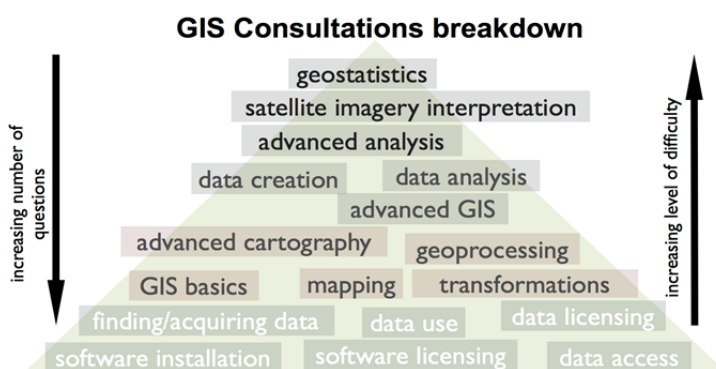


Figure 1. Breakdown of GIS Consultations at Toronto Library based on frequency and difficulty

¹⁰ Omeka: omeka.org/about

¹¹ Open Source Geospatial Foundation: www.osgeo.org

¹² OpenGeoportal: opengeoportal.org

¹³ GeoBlacklight: geoblacklight.org

¹⁴ OpenGeoMetadata: github.com/OpenGeoMetadata

¹⁵ Scholars GeoPortal: geo2.scholarsportal.info

One limitation of the Scholars GeoPortal identified by Fortin is its inability to search at the layer level of datasets. Further, he noted the gap between how many shapefiles the Library stores (23,000) and the number of shapefiles available in the geoportal (only 700). This gap results from the level of detail required in the metadata standards for quality control and searchability, coupled with a lack of implementation of automated metadata harvesting technologies. This discrepancy was repeatedly referenced throughout the meeting, underscoring its importance. Yet this is not a new problem; as early as 1952 it was brought up in Arthur Robinson's book, *The Look of Maps: An Examination of Cartographic Design*, in which he states "The ability to gather and reproduce data has far outstripped our ability to present it." This remains one of the seminal challenges in information science.

Other issues Fortin noted—particularly pertinent in Canada—revolve around copyright and licensing that occur when GIS users do not want to agree to ambiguous terms of data use. These cases primarily arise in federal data license agreements. For instance, many researchers do not want to consent to avoiding projects that could risk bringing, in the "opinion of Canada," any "disrepute to . . . the reputation of Canada." This commitment inherently bans a fundamental component of scientific inquiry: the ability to explore in the pursuit of truth, unencumbered by social, economic, or political interests. The consequences of breaking the license agreement cause researchers another disincentive to use the spatial data; for instance, if the city is sued, the consequence is "reimbursing the City for everything which you cause the City to suffer." Fortin articulated a need for the research community to negotiate with the Canadian government to change the licensing agreements of the spatial data it provides.

Fortin concluded his talk by bringing up new challenges experienced at the University of Toronto Library, such as the demand for web mapping support, adjusting to the change from 2D to 3D data, and the need for geo-literacy, especially with the growing internationalization of the University of Toronto's student population.

Breakout Sessions

Library Needs and Requirements

Moderator: Susan Powell

Recorder: Savannah Cooley

Reporter: Jon Jablonski

Participants: John Ajao, MaryLynn Francisco, Matt McKinley, Stephanie Simms, Denise Stephens

A central issue discussed in this breakout session related to the question of whether libraries should offer a single point of access to spatially referenced content; and more specifically, how closely integrated spatial data should be with the non-digital collection. One difference between spatial data and published content is the lack of prior external editing and reviewing of datasets when compared to the publishing standards of books. Traditionally it has not been necessary for libraries to monitor or take responsibility for the validity of the content in books, whereas libraries are expected to fill this role with datasets submitted for archiving and distribution. Data has analysis potential in which results will be significantly distorted if the dataset is incomplete or was updated without the metadata reflecting that.

The NGA Library combined digital (e.g., digital text documents, maps, images, and various formats of spatial data) and non-digital content into one catalog. The catalog system that supports this is an enterprise content-management solution that combines geospatial data management with a browser-based search interface. The NGA Library's search offers a sort-by format search that provides a useful option for users who know exactly what data format they need while still allowing those users who don't have a primary resource to discover content.

The discussion then assessed the needs and requirements associated with integrating spatial search into the general catalog search system. One requirement for enabling spatial search across the entire corpus would necessitate system compatibility across metadata standards. An alternative to creating a metadata "agnostic" system that recognizes all metadata standards involves using automated metadata generation and transformation technologies to ensure that all content ultimately migrates to the same standard. For instance, all repository content could map to the MARC metadata standard with different extensions based on the individual resource type. In either scenario, archival and curation workflows that facilitate discovery represent important components of library ingest methods.

Many libraries struggle with addressing the disconnect between the large quantity of spatial data they have access to or that is generated on-site compared with the relatively small subspace of those items available and searchable through the main catalog or their online geoportal. While self-deposit data systems could alleviate the workflow bottleneck of data ingestion, they pose a challenge to enforce the quality of both the content of the data and the metadata submitted by researchers. If discovery depends on the quality of the metadata, when a library chooses to support self-deposited data it plays a crucial role in data quality control and quality assurance.

The discussion led to how the NGA Library created a self-deposit data system. Effective quality control and quality assurance methods for self-deposited data require that the library offer the user

ample resources for guiding the process that go beyond passively listing data requirements. The NGA provides resources for NGA staff so that they can navigate the deep web safely, discover the required data, and, once they generate the desired content, adhere to all of the standards the Library specifies for ingestion.

For any library, basic guidelines to follow in a self-deposit scenario include: (1) connecting the data generator with resources informing best practices and standards for finding and generating spatial data; (2) creating incentives to adhere to those standards (such as an online user-rating system for a public library, criteria for tenure evaluation or promotions for a university library or corporate setting, etc.); (3) providing access to automated metadata generation and transform tools; and (4) training archivists to review the data before integrating it into the repository.

The discussion then moved to data licensing and ownership. Once a library commits to holding content for a certain number of years, it can use metrics to evaluate data value by how often it is accessed by users. If a library doesn't own or house data, the library can license it and make it discoverable for users. In the attempt to provide users with more resources through cross-institutional sharing, libraries relinquish "ownership" of content. This phenomenon challenges the assumption that collection size represents a useful metric for overall success of a library because it doesn't account for the resources available through data sharing consortiums. Therefore, a more holistic metric for quality would assess the library's access to outside resources as well as the post-discovery support it offers the user for the research the information is collected for.

If institutions combine data, they must clearly label what is available and how to access the content, as well as offer a search filter option that allows the user to only see results he can immediately use or access within a short time period. To ensure that users can access data once discovered, libraries can implement a system—used in GeoBlacklight and termed the *geomonitor*¹⁶—in which the server pings services to verify there are no broken links.

Summary:

- Ideally, the entire corpus of digital and non-digital content is spatially referenced and spatiotemporally searchable in the main catalog.
- Self-deposited data systems need quality assurance and should require minimal review from cataloguers. This could help close the gap between what is available and what is discoverable.
- Life cycle management can be informed by metrics recording uses of data.
- Collection size is an outdated indicator of overall success of the library. Focus on the library's access to content as well as the post-discovery support it offers instead.
- Searches should account for licensed use restrictions (e.g., offer an option to only display results that are "downloadable now").

¹⁶ GeoBlacklight geomonitor code on GitHub: github.com/geoblacklight/geomonitor

User Needs and Requirements

Moderator: Tom Brittnacher

Recorder: Sara Lafia

Reporter: Karl Grossner

Participants: Anabel Ford, Don Janelle, Alan Liu, Mary Whelan, Christopher Gist

The first point of discussion assessed researchers' needs in terms of using and sharing spatially referenced content. Knowing where research takes place can aid the discovery of new information. One of the library's important roles is to help users identify connections between themes and places. However, the established practices of identifying related data tend to be defined within disciplines. Researchers read studies and request data from authors. If the affiliated researcher is no longer available to provide data, the fate of the data is unclear. Libraries can help the researcher sustain the usefulness of his information.

Researchers from areas such as the digital humanities would also benefit from spatial data discovery even when they do not think of space as a construct useful for research. For example, discovery of spatially referenced physical artifacts could enhance a literary scholar's non-spatial research. Providing conduits between specialized collections known from particular campus libraries would allow for self-discovery of relevant information.

Existing catalog platforms would benefit from the incorporation of location information. The WorldCat¹⁷ identities page constructs a biography on the fly devoid of locational information, which is arguably a key element of discovery and event significance. It is not always enough to just have access to a map, since the temporal dimension of discovery is often equally vital. In the past, subject specialists could provide this type of expertise to point out such connections. Incorporating a linked data context in building data, discovery systems might expose these undiscovered connections and supplement existing spatial knowledge. The Social Networks and Archival Contexts (SNAC)¹⁸ project, which uses machine-learning methods to establish social relationships between authors, would also benefit from incorporating place information.

A broader discussion followed, focusing on the discovery capabilities that could connect users with other researchers via spatially referenced content. It is conventional for the author generating metadata to offer spatial descriptors in the form of bounding boxes or lists of place names, but it would be beneficial to consider information the user would look for or expect when seeking this data. This type of scoping exercise might include development of an ontology of kinds of places to build a hierarchy of place priorities (i.e., place of publication, place of birth, significant places for different cultures). A prosopographical approach to the organization of location with regard to biographical information could inform the role of place in a catalog's geospatial and temporal aspects.

¹⁷ OCLC WorldCat: www.worldcat.org

¹⁸ SNAC: <http://socialarchive.iath.virginia.edu/>

Another important distinction between asking research questions and asking teaching questions was raised. Research and question-asking as a skill might be fading with the decline of the baby boomer generation and mixed media in catalogs. Many students expect Google-like search capability but this approach is a “cul-de-sac” of information that is designed to omit outliers because responses build on themselves. In the research process, the discovery of outliers is very important. Conversely, teaching emphasizes the convergence of results into general trends. Both opposing needs should be supported within a single system. Randomness and serendipity also remain important elements to search.

With regard to specific user needs, better guided searches that walk users through semantic relations in metadata would be useful. Some current limitations of search methods include literal interpretations of data. Moving beyond Boolean to a system that supports weighted queries would resolve some of the issues that arise from treating all tagged content equally. An ideal search interface would retrieve geographic information on the fly. Currently, spatial queries allow users to generate metadata on the fly but are fixed and do not enable discovery of imaginary places or places that existed in the past. In terms of database restrictions, general trends point to lightweight noSQL databases like MongoDB. However, current trends are pushing in the direction of top heavy metadata. These two divergent trends must be reconciled.

Educating researchers about ways to describe their resources for discovery is a cultural imperative. If all researchers would start spatially and temporally coding observations, the need for better descriptive information about resources would be partially resolved. This first step would bring library catalogs closer to the functionality of a web GIS. Traditionally, libraries have been in the business of preservation, not offering user interfaces. A library catalog should allow users to preview layers, attributes, view metadata, and query or manipulate it much like previewing a book on a shelf. Functionality, such as geospatial analytics in the cloud, as well as platforms that would allow for mashups with a researcher’s own data would also be useful. There is hope that libraries will play a critical role in making materials discoverable and dynamic in the digital age.

Summary:

- Libraries should serve as facilitators between researchers publishing their data and users seeking access to data resources.
- Guided search would provide a solution to incomplete content discovery today.
- Existing platforms should consider inclusion of place as a data-curation dimension.
- Spatial and temporal referencing of all research could unify the discovery process.

Spatial Needs and Requirements

Moderator: Angela Lee

Recorder: Antonio Medrano

Reporter: Erin Mutch

Participants: James Boxall, Bernard Szukalski, Krzysztof Janowicz, Werner Kuhn

The discussion began with the question “When is spatial referencing useful with regards to library content?” The answer is that there is a broad array of content that could be spatially discoverable, where spatial referencing may have varying level of utility. Clearly, the lowest hanging fruit exists where the spatial utility of the content is high is (in maps and remote imagery). But even with this content there are challenges with the indexing system the library implements. Once the focus turns to the general catalog, where the content is not so explicitly spatial, the utility of spatial referencing becomes less clear. One insightful suggestion was that “perhaps the spatial referencing is useful when one can automatically determine it.” If it is easy for an automated metadata scheme to determine spatial referencing of an item, then the spatial association of that item is strong enough to be a useful resource in a spatial discovery search. But to classify something as spatially referenced requires specification of such characteristics as location, scale, granularity, temporal information, and authoring.

Next, the question, “What is the taxonomy of spatial content in the library?” was asked. Again, some answers were obvious such as paper maps, data in GIS formats (GDB, shapefiles, etc.), statistical databases (lat/long, address); and this extends into book holdings (e.g., the history of Goleta or a fictional novel set in Chicago). However, the taxonomy becomes less clear if the object is not physical. The digital content revolution has ushered in a period of transition where people are departing from traditional roles. Libraries are used less for traditional use, Google search has replaced library card catalogs, and librarians must update their competencies so as to remain capable of providing research assistance using these new digital contents.

On the topic of Google, people have grown very accustomed to using Google search for all types of search queries in their everyday lives. Does it suffice for library catalog search as well? To this, the answer was an emphatic “no!” Google provides results to a very simple search, which only uses keywords as search terms rather than a more guided search mechanism. The success of prior spatial search projects such as ADL and GeoBlacklight can be attributed to the inclusion of specialized guided search features that allow for refined spatial queries.

Another question considered was whether libraries should be collecting all data available to them. Is there value in collecting raw data in addition to finished products? On one hand, it requires vast resources to store all data indiscriminately; but on the other hand, sometimes data can have unexpected re-use applications. An example cited for this was the work by some computer scientists on determining the effects of climate change by detecting the presence of green pixels in webcam feeds from around the world.

The volume of available data continues to grow exponentially, but human resources remain approximately constant. How can libraries deal with cataloging data resources in such a scenario? It

is generally accepted that automation must take the place of human resources for the process of cataloging data, but even so, smart applications are extremely difficult to write, and they don't make data any smarter. On the other hand, smart, self-describing data (linked data) will make all future applications more robust, since this makes the data itself independent of having to conform to any of the numerous metadata standards. Linked-data eliminates the data/metadata distinction; this would be a promising approach for future work on the spatial discovery project.

At this point, the discussion turned to the process of spatial discovery. Part of the discovery process is the serendipity aspect of finding unintended items that spark new ideas and have great value to the research process. Current spatial search schemes use square extent, place boundaries, and polygon search for spatial queries, but the serendipitous discovery process might be assisted by incorporating an optional fuzzy extent to the search. In terms of the keyword aspects of the search, some semantics and natural language processing may be necessary to clarify search terms. People are not always able to describe exactly what they want, and it may require an ad-hoc iterative process to get to an accurate specification of what they are looking for. But librarians can play a role in improving people's ability to search content. If automation is able to take over the massive burden of cataloging library content, the librarian's role would transition from just curation to teaching information literacy.

Summary:

- Spatial referencing is obviously useful for maps and remote imagery, but perhaps when it can be automatically determined, it could be considered useful for general content as well.
- Spatially referenced requires such characteristics as location, scale, granularity, temporal information, and authoring.
- Google search uses a simplistic brute-force paradigm; spatial discovery requires a smarter approach.
- The question remains as to whether libraries should collect all spatial data or curate only "good" spatial data.
- In either case, the volume of data growth will require automation for cataloging.
- Discovery requires an element of serendipity to provide unexpected results that researchers will find of high value.
- In addition to curation, librarians should also emphasize teaching information literacy, particularly with regard to the nuances of spatial search.

ArcGIS Online Demo

Bernard Szukalski

Chief Technology Advocate and Product Strategist, Esri

"A Very Spatial Update"

Bernard Szukalski began the demonstration by pointing out how rapidly users are adding data to ArcGIS Online (AGO)¹⁹. Currently, AGO stores 3.5 million items, and the collection increases on average by 8–12 thousand additional items per day. As the amount of hosted data grows, so do AGO's capabilities and tools that support the "living Atlas" of information. For example, in the July 2015 release of AGO, organizations in AGO will be able to make formal metadata available for their content, allowing content authors to provide more information about an item than what is available on the details page. AGO will also apply a style to support the ISO 19139 and the Federal Geographic Data Committee (FGDC) metadata standards. For organizations that enable metadata, a "Metadata" button will appear on details pages for all of their item types. The button opens the built-in metadata editor and allows anyone who has access to the item—including those items accessed from open data sites—to view and download the metadata. This new capability offers technical specification that supports users in producing spatial data that a library could more easily ingest.

Additional group capabilities enable administrators to create groups whose members can update items shared within that group, while only the owner of the item can delete or share it. Updates to the items can include changes to the item details, tags, metadata, and content. Another new feature Esri is developing for AGO is a "hub," which will enable a specific community of organizations to share data. For instance, a planning department and a utility department that both fall into the Los Angeles County hub could discover and use the spatial data corresponding to the other department.

Szukalski navigated to Esri's Open Data website to show how organizations make data available and where it becomes federated into a common portal for anyone to use. He then illustrated several examples of creative uses of AGO's geoprocessing, analysis, and visualization tools such as the Story Map. AGO is designed as an out-of-the-box framework with developer-oriented tools and APIs. At the same time, it provides an accessible way to use spatial data for geoprocessing, analysis, and sharing for users who have no programming or in-depth GIS experience.

Summary:

- The content and capabilities of AGO are growing quickly.
- Given a self-deposit system, libraries can direct GIS users to AGO for geoprocessing, analysis, and metadata support.

¹⁹ ArcGIS Online Website: www.esri.com/software/arcgis/arcgisonline

Plenary Sessions

Unified Vision

Moderator: Werner Kuhn

Recorder: Sara Lafia

Participants: Everyone

The plenary session began by making a “case for space” in the discovery process. Location and proximity are integrators of contents. The discovery, sharing, and linking of library content can be facilitated using GIS. Discovery architecture offering a single point of access for spatially referenced content is the future vision for the geolibrary.

Inspired by the book, *This Idea Must Die*, Kuhn proposed the retirement of several ideas: that libraries collect; that libraries own contents; that content providers will produce metadata; that metadata are different from data; that users want to discover containers of information first; and that polygons are good for search. Polygon generation should be the result of some representational need rather than a means of storing data. When conceptualizing data, a hierarchical grouping of entities ensues, so clustering rather than bounding data within polygons may prove a more effective means of organization.

The current baseline shows an interdisciplinary pull for data-publishing mandates. Library staff is already overburdened by the management of library catalogs and infrastructure updates, as well as by the increasing demand for user support. Hybrid solutions of commercial and open-source platforms address a number of challenges ranging from automated metadata generation to integrating gazetteers with search. New standards for metadata that utilize geospatial and semantic frameworks are leading the charge. In moving forward, lessons from other types of content, such as music, are important to consider.

Using core concepts of spatial information supports transdisciplinary research by providing a set of lenses for organizing content. Thinking about data in terms of content concepts (location, field, object, network, and event) and quality concepts (granularity and accuracy) is an alternative approach to organizing content for users. Unpacking the logic behind these entities is a more logical means of organizing content than people, animals, places . . .

The development of workflows to create, share, discover, and link spatially referenced library contents holds great promise. Cycles of automated metadata workflows for spatial reference and provenance can enable library staff to meet the demands of spatial data users. Automation of metadata contains an element of certainty by guaranteeing that the metadata meets minimum standard requirements. Allowing for mining of the digital corpus and enabling automated semantic annotations of metadata as well brings up the need for humans to look over the machine output a posteriori. If a library implements an automated metadata generation system, catalogers remain essential for verifying the quality of the automation system. Their job would transition from the task of cataloging to reviewing and correcting the output mechanisms of the metadata generation system and they could devote more time to helping users transcend the gathering of static resources and engage in spatial analysis to “do the map.”

The interested crowd can be a boon. Human-infused tags can complement computer-produced metadata to pick up on cultural nuances and local trends. The New York Public Library (NYPL)²⁰ represents an institution that has successfully implemented a crowd-sourced metadata generation platform. One of NYPL's crowd-sourced information games, called "Map Rectifier," allows users to "rectify" historical maps by overlaying them on modern ones. Another NYPL game called "What's on the Menu?" enlists the public in the transcription of historical menus. NYPL initially released 8,700 digitized menus, which were provisionally transcribed in just four months. NYPL recently ramped up digitization efforts to meet demand. Another example of an effective crowdsourcing metadata platform is Metadata Games,²¹ which is an open source crowdsourcing game platform that represents more than forty-four collections at ten institutions. The games feature tens of thousands of media items (images, audio, video) that have generated 150,000 tags, with even outlier images garnering nearly 100 tags each. At this point, implementations have been developed by the British Library, Boston Public Library, The Open Parks Network, Digital Public Library of America, and the American Antiquarian Society, among others.

Both NYPL and Metadata Games metadata generation projects stand out amid many other game competitors as helpful contributions to educational institutions as well as providing subliminally educational uses of online time. If a library intends to create a crowdsourcing metadata platform, an important design consideration involves targeting audiences and anticipating incentives for people to participate. Specialists in a certain field would be most interested in working with a specific collection that relates to their area of expertise. If this process is coupled with automated metadata generation and final review by archivists and catalogers, then the library would not risk sacrificing data quality standards. Integrating these ingestion systems into the library's workflow addresses an issue articulated by many librarians at the Needs and Requirements Breakout Session: the need to increase capacity for making the spatial data it already houses, or has access to, more discoverable.

Hybrid solutions of commercial and open source software often complement each other and can support library catalogues and search infrastructure, including gazetteers and spatial and semantic technologies. Hybrid solutions become increasingly relevant as the roles of a library expand to include not only ingesting, curating, and making spatial content discoverable, but also to supporting library users in the skills to "do the map."²²

Summary:

- The overarching goal is that innovations are needed to enable dynamic, reliable identification of spatial objects within the catalog.
- Data innovations in libraries, while necessary, must account for compelling quality control and persistence goals prevalent in these organizations. The question is how might dramatic innovation and the control of descriptive architectures work in tandem?

²⁰ New York Public Library: www.nypl.org

²¹ Metadata Games: www.metadatagames.org

²² Mapping the Dollars and Sense of Land Use Patterns: video.esri.com/watch/4185/

- Since human resources for cataloging are finite, metadata automation technologies save time and offer a systematic approach to ensuring standards.
- Additionally, crowd-sourced metadata can effectively enhance content discoverability by leveraging outside knowledge and resources while at the same time stimulating involvement from interested users.
- A combination of automated core metadata generation and crowd-sourced information, with oversight and review by archivists and catalogers, could solve the problem of cataloging the vast amount of spatial library content, while enhancing the items' discoverability.

Best Practices

Moderator: MaryLynn Francisco

MaryLynn Francisco serves as the Chief Librarian at the GEOINT Research Center (GRC) of the National Geospatial-Intelligence Agency (NGA). The GRC provides geospatial intelligence materials, research services, and access to open source information resources to support internal NGA and external government customers. The GRC collection is comprised of 5 million items including maps, charts, hand-held imagery, commodity data, and books.

In the Plenary Session, Francisco first discussed her initiative of re-engineering the NGA Library. The project took five years to convince management to fund it and an additional two years to define the contract, which ended up specifying 591 requirements. The requirements maintained that the geospatial data management system support geospatial search of data with the MARC metadata format using a browser-based search interface with an unclassified internet system (i.e., not on the World Wide Web). In 2014 the NGA Library purchased and implemented an enterprise content management solution that met all of the requirements.

In a geographic and geospatially visual world there is the need to develop a federated information-retrieval technology that can access and probe multiple, disparate databases or repositories. In a federated search, a user makes a single query request that is distributed to the search engines participating in the federation. Implementation of a successful search engine requires an agreed-upon set of standards and data structure coupled with content to which metadata tagging have been liberally applied at the data level.

A challenge related to federated repository systems is the loss of information that institutions have not archived well. Initiatives like the California Digital Library e-Scholarship²³ project strive to bridge this gap, documenting conference proceedings in a consistent way. Describing these contents in a smarter way is an important next step toward building a distributed catalog that supports cross-federated searching. Without the initial ingestion of legacy content, this vision for collections cannot be fully realized.

²³ University of California e-Scholarship: <http://escholarship.org/>

Francisco shared that the NGA Library has been working on a standard that encompasses the Dublin Core Metadata Standards²⁴, international tagging standards, and federal systems of metadata standards to probe the multitude of federal geospatially enabled databases. They discovered that providing coordinates (latitude and longitude) to the tagging requirements assisted item retrieval in the same search along with subject probes. Many of the digital objects they use have embedded metadata that begins the long content-tagging process.

When the discussion turned to gazetteer development, a central challenge arose regarding how to define place names geographically. Most use cases involve searching with place names; however, place names constantly evolve semantically and geographically and are thus difficult to define. The NGA Library has specialists working on gazetteers for current-day place names that look for new place names and define or update the coordinates so that the gazetteers are accurate and up-to-date. Many gazetteers (used by NGA or Google, for instance) support modern names. Historical gazetteers require a different kind of diligence to create. Although they may not require constant coordinate updates, only explicit and clear definitions about the spatial and temporal extents of historical, art-historical, and archaeological periods (e.g., defining and agreeing exactly when the Age of Enlightenment in Europe took place) will enable discovery of historical collections across disparate databases. The process of coming to a general consensus will help scholars and students see where period definitions overlap or diverge.

The aim of the Pelagios²⁵ gazetteer project, which stands for "Pelagios: Enable Linked Ancient Geodata in Open Systems," is to help introduce Linked Open Data into online resources that refer to places in the historic past through linking gazetteers. Pelagios is a collaboration between archaeologists, historians, and geographers to define a publicly available Core Ontology for Place References (COPR) that permits modular extensions for different kinds of documents so that details specific to each type will not add unnecessary complexity for users wishing to publish data in conformance with the core ontology. The ontology is designed to help users answer two kinds of query: (1) "Within this document (text, map, database), give me a list/map of all the ancient places in it"; and (2) "Given this place, give me a list of documents (texts, maps, databases) that reference it."

The Pelagios project aligns place references in their own document types to Uniform Resource Identifiers (URIs) in the Pleiades gazetteer then generates Resource Description Framework (RDF) based on the ontology. Although Pelagios uses URIs based on Pleiades²⁶, the ontology will be able to work with other gazetteers (including those based on modern place names). All software code will be made publicly available through an open repository on GitHub.²⁷

Rather than aiming to create a common, standardized data model, Pelagios standardizes the creation of links between objects through defining a syntax for "descriptive records." The records contain the minimum information necessary to identify and disambiguate places, and build a

²⁴ Dublin Core Metadata Initiative: dublincore.org

²⁵ Pelagios: pelagios-project.blogspot.com

²⁶ Pleiades: pleiades.stoa.org/home

²⁷ GitHub: Pelagios Gazetteer Interconnection Format: github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format

searchable index external to the gazetteer, so that we can relate search queries in a third-party application (such as the Pelagios API) to the original entry in the source gazetteer. The descriptive records should include links to entries in other gazetteers to indicate similarity. Each gazetteer exposes such records about all of its primary entities.

Summary:

- The NGA Library catalog system successfully implements spatial-temporal search of their entire holdings, including digital and non-digital content.
- Libraries have open source resources to support creating gazetteers and linking them.
- To link gazetteers and create a federated spatial search, libraries do not need to agree on data models or metadata standards, however, they do need to agree on ontologies or “a syntax for descriptive records” that define the nature of the relationships between objects stored in the repositories.

Identifying the Gaps

Moderator: Karl Grossner

Note: This session divided attendees into three breakout groups, but since much of the discussion was overlapping, this section combines the comments from all three groups.

The gaps identified in the discussions spanned the entire breadth of the discovery process. Foremost, the need and potential of expanding quantity and accessibility of data can occur through creating greater cross-institution collaboration. Sharing data across libraries and other information-gathering organizations (such as local governments) would enable libraries to offer more resources to their users. Furthermore, developing federated search capabilities across multiple repositories with Linked Data approaches would create better systems to symbolize and represent indeterminate place areas over time. This involves negotiating license agreements as well as leveraging open source enabling technologies.

Another major gap involves the expert culture of traditional library roles. Libraries can learn to relax control and entrust the crowd with important tasks, such as generating metadata. Engaging expertise at the end of the metadata ingestion cycle rather than at the beginning is a possible solution. Many communities of experts, such as surveyors, are also interested in filling these gaps. With the rise of configurable apps such as Esri's GeoForm²⁸, closing the metadata gap through crowdsourcing is realizable.

Within an individual library, increasing the amount of available data could involve formalizing a self-deposit system for ingesting locally generated spatial data. In addition, libraries could integrate automated metadata-generation technologies and crowdsourcing to increase ingestion capacity. These approaches would help libraries make data "smart" and user-friendly before being made discoverable.

Another gap to fill involves establishing analogy-based search of the entire library catalog with topic and spatial-temporal similarities. Libraries could enable this kind of search on an online discovery portal that emulates the kind of physical, space-oriented hunting-and-gathering discovery that occurs in library buildings. This would include topic similarity organization, but also would intermingle a balance of serendipitous discovery with the intention of enhancing the user experience of digital spatial-temporal search—which often loses a dimension of exploratory discovery if not prioritized. Once a library user finds a spatial dataset that meets his needs, a library should offer post-discovery guidance to support spatial data users by providing resources for geoprocessing and analysis of the spatial data. By doing so, the library addresses existing gaps in user geoliteracy, exemplified by the underutilization of advanced search options. The Google paradigm of typing in a phrase and receiving an answer in the first result has oversimplified user research expectations. The serendipity of discovery as well as opportunities for deep thought are being left behind as technology advances. Content is increasingly delivered as snippets and soundbites. At the same time, information is becoming available at exponential rates, resulting in the properties that define

²⁸ Esri's GeoForm: www.arcgis.com/home/item.html?id=931653256fd24301a84fc77955914a82

big data: velocity, volume, and variety²⁹. Making sense of big data through visualization and other means promises to slow users down and consider data meaning in lieu of instant gratification.

A set of technological gaps included resolution for spatial data retrieval. Many platforms use bounding boxes, place names, or a combination of these for search. Issues result from the inclusion of extraneous content due to geometry. If a user selects a square extent of California, parts of Nevada are also included due to the shape of the states. MARC standards also attempt to capture spatial extent using a bounding box based on the sheet area of the map. Starting search with a bounding box and then filtering results in a tenable solution, yet coupling systems in a smarter way could be even better. For instance, GeoBlacklight first prompts users for a keyword (what) and then derives a box (where).

Balancing search system design to minimize expertise and maximize general knowledge is a challenge. The gap between expert developers and non-expert users is widening. Performing usability studies to learn which services are most popular is important for designing effective systems. For example, index maps point to series of maps based on topics; they have been heavily used in cartographic libraries. A digital counterpart to this search paradigm could offer a metaphor as a solution for search in an abstract space.

Data should be responsive to the three questions a guest is typically asked at a dinner party: "What is your name, where are you from, and what do you do?" Most systems only ask one of these questions and tend to not do it well. These questions should also allow for flexible order, since specialists from different domains may think in contrasting order. For instance, a remote sensing specialist might seek LiDAR data first, then narrow results by place, while a historian might search by place, then narrow results by data type.

Summary:

- **Cultural:** To increase ingestion capacity, libraries can use self-deposit systems, automated metadata generation technologies, and crowdsourcing.
- **Archival:** Smart documentation of loosely organized content is a first step to providing a distributed catalog with cross-federated search capability.
- **Technological:** Provide search methods that combine bounding extents with filtered thematic search.
- **Geoliteracy:** Reviving traditional learning perspectives is a solution to modern expectations of instant gratification and handling the three V's of big data.
- **Analytical:** Investigating tangible services that are most used as a proxy for digital usability, answer questions about the design of better non-expert systems.
- Spatial search can be facilitated with consideration of three main questions for data: "What is your name? Where are you from? What do you do?"

²⁹ Data Science Central: *The 3Vs that define Big Data*: www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data

- Libraries can improve quantity and availability of data by participating in data-sharing consortiums and working within that consortium to link gazetteers to create a federated spatial search.
- By making data “smart” and user friendly before it is discovered and by offering post-discovery GIS analysis resources, a library’s role in providing spatial data can go beyond discovery to support users in the next step of research.

Moving Forward

Moderators: Werner Kuhn and Denise Stephens

The beginning of this discussion included a timeline that forecasted the related future activity of the Center for Spatial Studies and the UCSB Library. The first outcome is a final report of the meeting (this document) followed by a more formal white paper that conveys the spirit and discussions to broader communities. Participants modified and added to the outline of the white paper, producing the following overall structure:

1. The Problem (Motivation, Narrative on Library in 2025, Problem Statement, Relevance)
2. Background (History, state of the art)
3. Use cases and scenarios
4. Gaps identified
5. Research Agenda and Project Contexts
6. Project Goals (what part of the problem do we plan to solve)
7. "Experimental" design
8. Evaluation criteria
9. Data management and broader impact
10. Implications for Standards and Knowledge Organization

The next item in the timeline after the white paper is a prototype of a GIS-enabled library search tool for the UCSB Library. This project serves as the first project created from the Interdisciplinary Research Collaboratory (IRC), an outcome of the partnership between UCSB's Library and the Center for Spatial Studies. Among the longer-term goals of the IRC is the creation of a social space for "guidance" in producing, sharing, finding, and linking "smart data" with a spatial reference.