John Ajao is the Director for Systems & Repository Operations at the University of California, Santa Barbara Library. He is primarily responsible for managing the Library IT operations including Alexandria Digital Research Library Project. The UCSB Library maintains a unique and complex Information Technology (IT) environment, and offers a number of services both internally and to the campus as a whole. The organization has a strong online presence, with a large, complex, mission-critical website offering dynamic content and 24-hour access to online catalogs. Continual scanning efforts have yielded a growing collection of Terabytes of digitized aerial photography and maps and depend heavily on in-house workflow software. The Library is an active participant in the field of digital library development, with ongoing research projects emphasizing geospatial data search, digital archiving, and logistical networking. As the IT team, responsible for managing UCSB library’s geospatial data and our past involvement with Library of Congress NDIIPP/NGDA project, the challenges identified by the project team with regards to developing a comprehensive spatial discovery solution locally at the time are still real today.

Position Statement

Organizations have complex and fragmented architecture that make the cohesive collation and dissemination of spatial data sets challenging. Spatial Data are the bulk of UCSB Library’s digital collections. These are unique sets of data that present their own management and storage challenges. Emerging analytic solutions have played an important role in enabling some intelligent and manageable steps towards achieving this goal. As organization and Institutions are now requiring a smart-data preservation platforms. Data-intelligence are now the norm for most institutions and the information industry, and spatial data is at the core. With UCSB Library local implementation of ADRL platform, we are looking at various technologies for providing cheap and scalable storage solutions, leveraging technologies such as Ceph File System (CephFS) and Inter-Planetary File System (IPFS), while exploring metadata & rules framework for large-scale file management.
Chris Barnett has been building geoportals and tools for handling geospatial data and metadata for about a decade: first at MIT, where he was also an undergraduate, and currently at Tufts University. He is also a committed member of the free and open source software community. His open projects have included the Open Geoportal and associated tools, as well as being a contributor to OpenGeoMetadata. Currently, Barnett serves on the local organizing committee for FOSS4G 2017.

**Perspective Statement**

*How can spatially mediated discovery provide single-point access to research data, across distributed repositories and catalogs?*

Good metadata is essential for discovery and in determining the usability and reliability of a data set, which is a part of accessibility. When federating catalogs, this is doubly true. Yet, having seen and worked with a great deal of metadata from top-tier institutions, I can say with confidence that poor metadata is the norm.

Our approach with metadata in Open Geoportal and OpenGeoMetadata has been, foremost, to share. Access to good metadata records provides good examples. Metadata authors are commonly data stewards in small groups (often of 1!) or the data producers themselves. It is somewhat rare that either of these groups have both the geospatial and library knowledge needed to create good XML geospatial metadata records in isolation. At a broader level, best practices emerge from the collective and are shared with newcomers and observers. Specific questions can be posed to the community.

Because records are publicly available, metadata issues that arise during the process of ingest to a catalog or that are surfaced in discovery tools receive greater scrutiny. In the case of OpenGeoMetadata, which is hosted as a GitHub repository, comments and issues regarding specific records or at the collection level can be tracked and resolved.

Our metadata will continue to improve, but realistically we are unlikely to ever solve all issues. Geospatial metadata creation is simply too time consuming and underfunded. So how do we build robust discovery systems around less than ideal records?

Spatial search as it commonly exists relies on bounding boxes of data sets, so, in theory could depend only on machine generated metadata. Open Geoportal’s implementation uses a combination of spatial queries that may be taken as somewhat representative:

1. An axis-aligned bounding box algorithm filters out data sets that do not intersect the query bounds.
2. Relevance is weighted by a comparison of the area of intersection to the area of the result record. (bounding boxes)

3. Relevance is weighted by the inverse distance from the center point of the query bounds to that of the result record.

This type of search works reasonably well for many spatial data sets, but there are some problems and some paths forward.

It can be argued that bounds derived directly from the data are not always correct. A search for the “area of study” may be more appropriate than the strict data bounds. A linked data approach that associates the data set with something like a GeoNames ID could help here, but we are then back to the problems of metadata curation.

Rectangles are very approximate. Data sets where the area of the bounding box doesn’t match well with the area covered by the actual geometries can be problematic. Native spatial indexing in tools like Lucene should allow performant queries of polygons more complex than rectangular bounds. Using a generalized convex hull instead might work better for many of these cases.

Relevant scale is very hard or impossible to discern from most geospatial metadata records. For example, there may be a data set with state-wide extent that has a higher level of detail than another with municipal extent. The data set with state-wide extent is likely to be buried in the search results without factoring scale into a relevance score. Some useful measures that could help with this problem and can be derived easily from the data are feature density and minimum segment length.

**How can the discovery of research objects in general be spatially supported?**

Any spatial components would have to be derived from extant metadata (library catalog) records for the research objects. A reasonable starting point would be to process place keywords, which a human cataloger has likely decided is relevant to the research object. These could be geocoded and matched with a linked data aware gazetteer like GeoNames. The linked data relations can be used to attach bounds, aggregate spatially, etc. If a controlled vocabulary has been already used, perhaps a linked data service already exists (LOC), and it might make more sense to bridge the two relations.

Questions arise once we encounter an object without any explicit geographic keywords. One could easily imagine use cases where it would be useful to relate other metadata values, such as publisher or author with one or more geographic entities. But, clearly, these indirect relationships must be represented differently. How?

**How can spatial discovery be applied to topic spaces, not just geographic ones?**

Spatial navigation through non-geographic topics might be useful, particularly in searching for related items, but would need a specialized user interface. Geographic maps act as a common frame of reference for spatial search UI’s. What is “north” in a knowledge graph?
Wade Bishop is an assistant professor in University of Tennessee School of Information Sciences whose research explores how humans organize, access, and use geographic information (GI). He received an Institute of Museum and Library Services (IMLS) grant titled Geographic Information Librarianship (GIL), with co-PI, Professor Tony H. Grubesic, Ph.D. (Arizona State University), to develop and deliver coursework covering the knowledge, skills, and abilities to locate, retrieve, analyze, and make spatial data discoverable. Their new Springer book, Geographic Information: Organization, Access, and Use, outlines findings and future directions for research in the area of metadata creation and data discovery. Bishop’s dissertation and other projects related to the study of answering location-based questions (i.e., spatial discovery) within the context of traditional information agencies.

Challenges to Spatial Discovery

S
ingle-point access to research data, across distributed repositories and catalogs, as well as other spatial discovery challenges require an understanding of organization, access, and use of geographic information. The user roles, communities, and types of spatial data impact aspects of discovery and require further study of the human information-seeking behavior used to locate and the digital curation processes implemented to manage. It is important to consider the implications for spatial data discovery related to usability, functionality, accessibility, information needs, information seeking behaviors, and curation throughout the data lifecycle.

Data discovery refers to the process of identifying and obtaining observational data for use and analyses. There are a variety of challenges worth noting. First, much of the observational and/or environmental, cultural, or socio-economic data published on the Internet is not easily discoverable. Although there are many open data efforts, many of these data are buried within databases that are not indexed by commercial search engines or crawlers designed to index the deep web. Second, both syntactic and semantic heterogeneity in these data largely prevent their discovery, integration, and synthesis. For example, syntactic heterogeneity refers to the way data are encoded or organized. GI is available in hundreds of different encodings, including the relatively common shapefile (.shp) format, but also layer files (.lyr), Google Earth documents (.kml or .kmz), compact data format (.cdf) or many other encoding schemes. This can make it difficult for end-users to both determine the best format for acquisition and use, as well as the software platform to be used for analysis. Semantic heterogeneity refers to the identification of semantically related objects in different databases and the resolution of schematic differences among them. For example, if the source of information and the potential receiver of that information operate within different contexts, the ontological terms and associated metadata for a database may differ dramatically. Further, the repositories of these data
and their underlying structure may also diverge. In short, the process of data discovery is not always straightforward.

Another key challenge for discovery relates to fitness for use, which is a multifaceted concept that refers to data suitability for a particular application or purpose. Facets include data quality, scale, interoperability, cost, metadata, syntactic and semantic heterogeneity, among others. Data quality may be difficult to determine given the persistent and pervasive lack of data documentation related to positional accuracy, temporal accuracy, attribute accuracy, logical consistency, and completeness. In most cases, this information is transmitted from the producer to the user via metadata. The problem with metadata is that many users choose to ignore it (even when metadata are present). Without detail upon creation, subsequent GI, and geo-enabled information lacks the data provenance necessary to ensure future use and inform reuse.

Finally, in other topic spaces other than geographic ones, spatial discovery could be relevant. The relative importance of location in the broader context of geo-enabled information, maybe of secondary interest, or in some cases inconsequential, to some users. Still, equity in access to create spatial data is growing and certainly the need to understand spatial discovery for other geo-enabled information is needed.
TOM BRITTNACHER  
Library  
University of California, Santa Barbara  
Email: tombritt@ucsb.edu

Tom Brittnacher is the University of California, Santa Barbara Library’s Geospatial Data Curator. He is currently participating in the development of the library’s online repository, Alexandria Digital Research Library (ADRL), specifically with respect to ingesting and displaying scanned maps and geospatial data. He is envisioning, documenting and implementing geospatial metadata creation workflows, and designing new techniques for using geographic information systems (GIS) to generate metadata for individual sheets within map sets. He is also assisting ADRL developers in creating new data models and display frameworks for scanned map sets and index maps, enabling easier user navigation within large sets.

Brittnacher also works with UC Santa Barbara researchers to incorporate data management best practices into workflows and to improve GIS data storage and distribution, and is the current chair of the University of California’s Data Curation Common Knowledge Group (CKG).

Brittnacher’s background is in cartography, GIS analysis, and transportation planning in addition to GIS librarianship. He holds a Bachelor of Science in Geography, a Master of Science in Urban Planning and a Master of Library and Information Science, along with GIS Professional (GISP) certification.

Perspective

In my view, the key to any effective discovery across distributed repositories and catalogs is the use of consistent metadata standards and schemas that are flexible and easy to use. Those standards and schemas should be shared across platforms, or at least be sharable via appropriate crosswalk transformations. (Not all transformations can accurately match a metadata element with an equivalent counterpart in another schema.) The spatial elements within metadata schemas are complex, and there are a variety of ways to express location. This adds more difficulty to any synchronization of spatial information. Whether location is graphically expressed by a bounding box, polygon, or point, or is textually expressed by a keyword place name (chosen from a variety of gazetteers), the transformation of a place described by one method into another method introduces inaccuracies, error, or even incompatibilities.

In the bibliographic cataloging world, Resource Description and Access (RDA) content standards and the Machine-Readable Cataloging (MARC) encoding schema have become ubiquitous. Librarians all over the world can borrow and contribute records through the Online Computer Library Center (OCLC). As a result, single-point access to discovering materials across distributed catalogs is happening already, such as through the WorldCat union catalog. Can this philosophy be applied to research data using spatial discovery tools? Of course, you can’t access the materials themselves through WorldCat!
The Open Geoportal (OGP) community began as a collaborative effort to develop open source, federated spatial web applications. The community has expanded to include the OpenGeoMetadata project where libraries share metadata for their geospatial repository holdings using the JSON file format. The metadata files incorporate bounding boxes using the georss encoding specification, which is in turn used to drive spatial search in a map interface. Several universities contribute their metadata to the project’s GitHub repository, which other repositories can then harvest. Again, a shared metadata schema has enabled discovery across distributed repositories. In this instance, the materials can be accessed via any repository that ingests the metadata through web links. But this is far from ubiquitous.

The key is figuring out how to enable the quick and easy creation of metadata that incorporates a compatible representation of location (even when the data aren’t inherently geospatial), and then figuring out how best to share that metadata in such a way that repositories can discover, ingest and share the information, as well as hopefully providing direct access to the research data itself.
Julien Brun
National Center for Ecological Synthesis and Analysis (NCEAS)
University of California, Santa Barbara
Email: brun@nceas.ucsb.edu

Julien Brun is a Scientific Programmer at NCEAS, where he supports and advises scientists on data modeling, data analysis and data preservation. He also teaches programming and open science best practices to the synthesis working group participants of the Science for Nature and People Partnership (http://snappartnership.net/) and the Long Term Ecological Research Network (LTER; https://nco.lternet.edu/). Brun’s scientific background is in ecohydrology and remote sensing. Prior doing his PhD in the Civil and Environmental Engineering Department at Duke University, he was a GIS and Remote Sensing consultant for international institutions, focusing on vegetation monitoring and natural disaster impacts assessment.

Perspective

Space provides a universal reference for any act of observation or measurement, such that an observation’s location and spatial extent comprise a “key” to relate it with other spatially referenced objects, helping identify shared attributes or infer causal connections by providing context. Spatial characteristics of an object are therefore a powerful way to discover and link heterogeneous data. Although the mathematical expression of a geolocation on Earth is defined in various well-established ways, the storage of this information in standardized and interoperable metadata formats is still a challenge for data discovery across various data archives. Web technologies and standards should provide inspiration and maybe solutions to tackle these interoperability challenges.

When thinking about spatial discovery, there are several different types of spatial search that come to mind. For example, the geolocation based discovery mentioned above: it can be based on latitude and longitude of a coordinate system, but it can also be based on names of spatially well-defined locations or areas (e.g., administrative boundaries), as well as fuzzier (e.g., neighborhood) or ambiguous (e.g., different locations with same name) spatial concepts. Another example of spatial discovery could be a user wanting to search for data using a spatial relationship in regards to another spatial entity, such as “close by” or “within.” I personally think an important challenge of spatial discovery is to enable all the different types of spatial data search one could want to do, while integrating the variable accuracy of the concepts used to conduct the research, the potential hierarchical relationships among these spatial concepts, as well as their variability in time (e.g., changes in species distribution).
James Frew is an Associate Professor of Environmental Informatics in UCSB’s Bren School of Environmental Science & Management. He has worked in remote sensing, image processing, software architecture, massive distributed data systems, digital libraries, computational provenance, digital archives, and crowdsourced geographic information. His current research addresses the automatic generation of data citations, and the applicability of array database technology to environmental data management.

Perspective

In an era of fine-grained web services, it is useful to recall the comprehensive functionality of the traditional academic citation. With a single (well-formed) citation—for example[1]—we (ideally) obtain:

- **credit**: We acknowledge the contribution of other scholars, specifically their influence on the work in which the citation is embedded.
- **provenance**: The cited work’s age and venue help situate it in the broader scholarly landscape, and inform our judgement about its relevance and credibility.
- **access**: The information provided in the citation is sufficient for us to locate (and hopefully access) the cited work (or a copy thereof.)

All of these aspects of a citation are important for discovery. Although we tend to think of traditional citations as driving manual searches through the published literature, in fact services like Google Scholar use them quite effectively to automatically locate cited works, even when the citations are incomplete or ambiguous. Moreover, human-readable citations, as opposed to opaque persistent identifiers, allow the works containing them to retain their meaning independent of the presence (or proper functioning) of the PID infrastructure.

While the scholarly community has hundreds of years of experience with creating and using citations to documents, citations to data (including spatial data) are still in their infancy. Document citations exploit the relative simplicity of document organization hierarchies (monograph, book→chapter, journal→article, etc.) Data citation, on the other hand, wrestles with issues like granularity (e.g., spatiotemporal subsetting) and allocation of credit (e.g., for collaboratively-maintained datasets). Content standards for data citations do not yet address these complexities. My colleagues and I propose[1] that the most appropriate solution to this problem (and one that also exploits some key properties of the database systems in which much citable data currently resides) is for data providers to generate correct and complete citations for each data request. In my lab at UCSB, we have demonstrated the feasibility of this approach by extending the OPeNDAP spatial data
access web service to support automatic citation generation for each unique OPeNDAP request. So, as a discovery tool, what will “Google Data Scholar” look like?

Reference

In thinking about the topic of this meeting, I like to begin at the beginning, by asking why someone might want to search for spatial data. Otherwise we risk our thinking being constrained by the legacy of previous approaches to spatial discovery, without the liberating effects of newer technology. The old media for disseminating spatial data included the map, the globe, the remotely sensed image, and the atlas. They provided a compact, filtered synthesis of a vast amount of original data that could not reasonably be produced, stored, and disseminated in any other way. These arrangements were copied into the digital world, by structuring our spatial-information technologies around layers, which simply replicated the contents of the old analog media in digital form. Even today the layer concept persists, often in the form of a list of layers occupying a panel on the left-hand side of the screen.

I find it helpful to think of layers as integrating spatial information horizontally, by storing one type of attribute, feature class, or variable over a spatial extent. But was this driven by a careful analysis of what users needed spatial data for, or was it a legacy of the earlier media, driven in part by...
the economics of spatial data production? It makes it easy to compare places with respect to the same attribute, feature type, or variable. But it makes it comparatively difficult to assemble multiple attributes or variables for one location. Despite this we persist in claiming that GIS is the answer to spatial data integration, but in reality we have made it remarkably difficult to integrate across layers, or what we might think of as vertical integration. Moreover integration based on location only works perfectly if location is measured perfectly, which it never is.

So let’s begin with the use cases, or the reasons why users might want spatial data. How many queries concern “What is at $x$?” Retrieving separate spatial data sets and integrating them vertically provides a very expensive, tedious, and ultimately inexact answer. The query “Where is $z$?” on the other hand is best answered by accessing a single layer that can be searched for instances of $z$. A query “How high is Mt Everest?” has a spatial answer, but it is likely not best obtained by searching a DEM, which will not readily identify features or their names. The query “How do we know that Everest is 8848m high?” is similarly not readily suited to traditional spatial search and discovery, because it asks for information about provenance that is typically not available through the traditional process of spatial data production.

This line of inquiry is especially important to users of a facility for spatial discovery. We risk artificially limiting the potential of such a facility if we model it on the processes of the past. However successfully we deal with metadata and semantics, we will still be locked into the old model of delivering an entire, horizontally organized data set.

The Center for Spatial Studies is already the hub of a campus network, and has tapped into many if not most of the applications of spatial thinking in the various disciplines. I suggest that the center institute a systematic study of the use cases, the tasks for which researchers and instructors need spatial data. Two to three months of a dedicated person should be sufficient to tease out some interesting perspectives on the problem, and to produce a publishable analysis.

I published an article in *D-Lib Magazine* (Goodchild, 2004) that reflected on the Alexandria Project, which is now more than a decade in the past, discussed its limitations, and raised some of these issues.

**Reference**

Karl Grossner is an independent GIScience researcher working to develop novel models, standard formats, and semantically-enabled software and systems supporting the emerging genre of digital historical atlases. A founding co-chair of the GeoHumanities SIG within the Alliance for Digital Humanities Organizations (2013), Grossner is an active member of that global and trans-disciplinary community. He is currently renewing the activities of World Heritage Web, a non-profit corporation he founded in 2003 to support digital research and education in world-historical geography. After earning a Ph.D. in Geography at the University of California, Santa Barbara in 2010, Grossner worked for five years as a digital humanities research developer at Stanford University, building several significant interactive scholarly web applications in partnership with faculty members. He is currently working on a few related projects: a World-Historical Gazetteer just under way at the University of Pittsburgh, GeoJSON-T, a temporal extension to the GeoJSON data standard, and Linked Places, experimental web software for representing, sharing, and analyzing data about historical geographic movement, including journeys, flows, and named routes.

Blog; Twitter: @kgeographer

Linked Places, Linked Pasts

It is interesting to note the three fundamental questions put before this meeting’s participants refer not to discovery of spatial data per se—as was the explicit goal of the 2014 and 2015 meetings—but of research data and objects generally. Spatial discovery in many institutions is (so far) taken to mean discovery of paper maps, satellite imagery, and other spatial datasets having coverage for a given place and time, relevant to a particular topic. Now we are talking about discovery of any research data and objects meeting those search criteria.

From my perspective this expansion is most welcome and suggests a couple of things: that an increasing volume of research data across many disciplines have either intrinsic spatial attributes or important spatial metadata; and that libraries are increasingly interested in providing data repository services. The third question is an exciting area of inquiry, and aligns with a position I expressed at the 2014 meeting: that “spatial search” often is—and will increasingly become—“geographic search.” That is, search criteria incorporating all three dimensions of geographic information per Berry’s geographic matrix, space, time and theme. This is also reminiscent of the next-stage geolibrary Mike Goodchild wrote of in 2004 as the ADL project wound down.

Spatial discovery for historical research data and objects presents some distinctive challenges, as compared to that for current or recent periods. The two most commonplace identifiers, name and location, are problematic in most historical data. Names change and locations are hard to pin down.
Much of the data required to create true historical gazetteers remains locked in texts and on historical maps, is being extracted piecemeal by individual research projects, and has been, until recently, not readily shared. The temporal attributes of historical data become more uncertain (imprecise, absent) the further back in time one goes. Many historical objects can be dated only to a named period likely to have contested extents. The Pelagios Commons project is a focal point for the growing global initiative by historical researchers to build a distributed network of linked gazetteers for both places and named periods. In my view, efforts at improving spatial discovery capabilities should account for historical data.

I am currently serving as a co-coordinator of the Linked Pasts Working Group within Pelagios Commons. Linked Pasts will facilitate the collaborative development of a technical infrastructure and best practices for linking any and all historical data, moving beyond the search dimension of place, to include periods and data about people, artifacts, and events. A large distributed system along those lines would enable the thematic spatial searches suggested by Question #3. Many of these will be of the form “where has this been so?” as opposed to (or in combination with) “what exists or happened here?” This is beginning to occur at limited scale within individual projects. For example, one might search for data about artifacts and events of particular types associated with places along the path of a particular journey; an event in the life of the “who” of the journey provides a “where” and “when,” which are the basis of a spatial-temporal “what” query. And so on!
As we learned in the first spatial search meeting, the phrase has many different meanings—from the cognitive behavior associated with navigating our surroundings, to the engineering and design tasks associated with describing documents using their spatial semantics—and many points in between. And around. And there it is: spatiality is even embedded in language.

As I write this position paper, I am keenly aware of how much humanity depends on all these definitions. I’m traveling in China, a country where I speak very little of the language and am functionally illiterate. Yet every city is navigable thanks to how human beings have built the environment. The design of locations like train stations and airports is optimized to help me navigate easily. My past experience, along with thousands of years of human practice, allows me to move freely through the cityscape. With a smartphone, for the first time ever, I am not even bothering to write down my destinations before leaving my hotel. My paper notebook has been relegated to a storage place for my American SIM card. Instead, I have four global maps of various levels of detail,
accuracy, language, and usability. Each has a blue dot that shows me my almost exact location—regardless of whether or not I know where I want to go. The only piece of analog technology I carry is a magnetic compass, because the arrow on the blue dot tends to get confused.

Spatial search is something I do continuously while traveling. One extended example. With a couple hours to kill this morning, I glanced at the map and noticed a park nearby my hotel. Chinese parks are generally walled off, with only one or two entrances. On past visits, if I saw a park in a guide book or on a paper map, I would set off with a carefully planned itinerary. If I arrived at the park, as I did today, but could not find the entrance, or even figure out if what I was looking at was indeed the park (signs are often in calligraphic characters that are unrecognizable to my eye), I would often abort the trip. Today however, I could zoom and see that the entrance was a narrow choke point between two buildings.

Thanks to the blue dot, beyond the park I found a farmers market and a lunch spot—purposely place between the park and a housing development. In the library, classification systems arrange materials by topic in order to facilitate exactly this sort of serendipity. After lunch, instead of retracing my steps back to my hotel, I was able to make a wider circle and find an open air gathering spot for dog and cat breeders.

We continue to search for the equivalent of the blue dot in the electronic library space. We have built a huge complex of buildings, but struggle to help our users navigate it. We have access to tens-of-millions of electronic documents, but lack an overarching navigation structure. For all of our sophistication, we still force our users to navigate our information spaces with multiple organization schemes, conflicting ontologies, and clunky interfaces.
Greg Janée is director of the newly-launched Data Curation Program at UC Santa Barbara, which offers consultation services to campus researchers and he is working now to establish a curated research data storage facility at UCSB. He also works at the California Digital Library, where he is principal developer of the EZID persistent identifier service. His previous work was on the Alexandria Digital Library (ADL) and related efforts. In addition to serving as principal developer of the ADL system, his research areas included query languages and frameworks, query rewriting, metadata mapping, spatial ranking, search result visualization, and provenance analysis. He is co-author of the ADL Gazetteer Protocol.

Statement

In my view, spatial search (specifically, geographic search) is hampered by mismatches and bifurcations. The problems are well known:

- Spatial extent can be described precisely: think of the complex, projection—and datum-aware geometries that can be expressed in GML. But many spatial search user interfaces and search engines operate only on bounding boxes, which suffer from precision problems.
- Reconciling fundamentally different ways of addressing space—coordinate-based descriptions and placenames—remains difficult.
- There are multiple ways any given spatial extent can be described due to differences of scale and simply differences of choice. Polygon, box, or simply a point? Which placename(s)? And in practice, there is little uniformity in how spatial resources get cataloged.
- Spatial and non-spatial search are poorly integrated, if at all. A resource lacking a spatial extent typically won’t show up at all in a spatial search engine, and conversely spatial extents typically provide no input to a non-spatial search engine.

All these problems make for a spatial search experience that is fragile in the face of the limitations of real-world metadata, which is often incomplete, incorrect, misplaced, inappropriate, or simply missing. Contrast this with web search, which is much more resilient and works quite effectively despite having to contend with malformed HTML, spelling errors, broken links, and the like. Might the advances that enabled the web search we enjoy today provide a path to better spatial search? Web search has improved over the last decade, not by getting better at finding words, but by collapsing file formats and data types into a single search space, by recognizing relationships between documents, and by employing sophisticated ranking algorithms. Are there analogies for spatial search?
TOMI KAUPPINEN
Aalto Online Learning
Aalto University
Email: tomi.kauppinen@aalto.fi

Tomi Kauppinen is a project leader and docent at the Aalto University School of Science in Finland. He holds a habilitation (2014) in geoinformatics from the University of Muenster (WWU) in Germany, and a title of docent (2014) and a Ph.D. (2010) in media technology from the Aalto University. From April 2014 to September 2014 he was appointed as the Cognitive Systems Substitute Professor at the University of Bremen in Germany, and since 2015 he is a Privatdozent at WWU. He has been active in opening and sharing data, and created semantic recommendation and exploration engines. A central theme in his work and teaching is data science and information visualization applied to spatiotemporal phenomena, and supporting understanding of related cognitive processes. Kauppinen has actively created online tutorials on these themes and run related courses and tutorials in international conferences and universities. He has organized several international workshops on information visualization, linked data, spatial thinking, educational resources and linked science. He is also the founder and community leader of LinkedScience.org. He is currently coordinating a new Aalto University wide strategic development project, Aalto Online Learning A!OLE.

Perspective on the Challenges

Why is it essential to support discovery of information via spatial metaphors or in situ in locations? What kind of links can we evidence and essentially represent to support connecting of the variety of information together via space, time and theme? What are essential spatial relations that best facilitate spatial discovery and underlying reasoning procedures? How do we support communicating of information by visual means and storytelling to facilitate discoveries and for improving of understanding of phenomena around us? How can augmented reality and virtual reality be employed to create blended topic and geographic spaces where we navigate through information and learn new paths? What can we learn from human spatial thinking and spatial cognition for spatially distributed information to be discoverable? What kind novel online learning settings can we build by arranging information spatially, and then serving it timely via location-based when learners are ready to acquire new knowledge. What happens when we weave together spatial and social, and which interesting participation characteristics can we then evidence? What kind models can we build to explain those spatial-social phenomena? Why places and events in them work so well in engaging people to increasingly participate? How can we support community building via places and spatial configurations in them? How do we truly support learning, remembering or decision-making by communicating research data
by spatial and visual means? Which spatial metaphors to employ for creating engaging information spaces, and how to employ virtual reality to allow for exploring those novel spaces.

As of now (2017) we are lucky to have both massive amounts of observations—essentially data—and tooling for making use of them. Visual analytics, contextualized semantic recommendation engines or virtual communication tools are all employed by numerous platforms, and often without users truly realizing it. Further on, we have emerging, novel ways for virtualizing or augmenting spaces. What is missing is to understand how people in fact can best be supported to discover new insight, learn new skills, or make educated decisions with tooling. Spatial configurations around us, be they about urban spaces, indoor spaces or nature itself, have educated our ancestors and us to cope with the world, often via trial and error. For instance, we are teaching our children to be careful in traffic (urban, modern spaces), not to run stairs (indoor spaces) or avoid jumping on icy cliffs (nature). When we create virtual and augmented spaces, the hypothesis is that we can similarly learn to find our paths through information spaces or blended information/world spaces.

This calls for ensuring quality of information objects via identity resolution and link discovery, and employing suitable spatial metaphors to represent (conceptually) and present (visually) both objects and links in between them. Indeed, virtual reality and augmented reality are calling for engaging visualizations of data. Haptic interfaces can bring these opportunities to yet new levels where we can spatially organize and arrange information as we would do organize and arrange physical objects. These research perspectives are getting timely as we now evidence that people (like teachers or market analysts) and organizations (like university libraries or companies) are looking for new ways to index, organize, share data and make data and its connections understandable. Trying of different ways to empower people to discover new insight by spatial means can lead to models explaining human behaviour in terms of how learning can best be supported. For this I give as a reference a new cross-cutting research topic covering cultural psychology, interdisciplinary physics and big data (in ICT and digital humanities) I am co-editing for *Frontiers*.

**Reference**

Claudia Knudson
MesoAmerican Research Center
University of California, Santa Barbara
Email: caknud@gmail.com

With a degree in geography in 2014, Claudia Knudson has emphasized her interest in GIS and its spatial applications. Knudson’s interests lie at the intersection of spatial analyses and data management that are at the crux of research that engages with the vast data resources that can be incorporated into the GIS platform. Her current work is on a geographic data compilation with the UCSB Maya Forest GIS. The Maya Forest GIS includes more than twenty years of data collection from Belize, Guatemala, and Mexico, based on digital data acquired from agencies, NGOs, as well as direct field and lab sources. Working on a new vetting strategy for GIS data, Knudson’s Digital Data Acquisition and Testing for Archiving (DDATA) project is designing a GIS data protocol with the intent to make data accessible, usable, and prepared for library archival storage and use. Her work includes creating a management information system and quality control to define standard protocols and parameters to deal with data and the problematic metadata. The challenge is in the implementation of the protocols while the data are incomplete, in use, proliferating, and annually updated.

Perspective Statement

Over the past 20 years with Anabel Ford at the MesoAmerican Research Center, the UCSB Maya Forest GIS is a compilation of disparate yet comparative data on the tropical lowlands of Mesoamerica. Overwhelmed by the diversity of data sets, the value of the data in the study of conservation and development in the Maya forest, as with all GIS data, depends on their complete and consistent metadata. Establishing best practices is the ideal, but in reality, at this time it seems to be a last step, if not forgotten. This needs to change. Working with the UCSB Library’s Geospatial Data Curator Tom Brittnacher, we have identified key areas of weakness and target areas for improvement in the development of the Maya Forest GIS. Our Digital Data Acquisition and Testing for Archiving, DDATA, is developing a standard protocol designed to build and integrate a useable, consistent, and well-documented file naming and storage system for our GIS. Implementing our DDATA protocol will ensure that the UCSB Maya Forest GIS has consistency and validity for use on the El Pilar project. More importantly, the success of the DDATA protocol will serve as a model tool for research professionals who are ready to share their stored data.
Werner Kuhn holds the Jack and Laura Dangermond Endowed Chair in Geography at the University of California, Santa Barbara, where he is professor of Geographic Information Science. He is also the director of the Center for Spatial Studies at UCSB. His main research and teaching goal is to make spatial information and computing accessible across domains and disciplines. Before joining UCSB in late 2013, Kuhn was a professor of Geoinformatics at the University of Munster, Germany, where he led MUSIL, an interdisciplinary semantic interoperability research lab. Kuhn is described as a leading expert in the area of geospatial semantics and especially known for his work on Semantic Reference Systems as well as his work on interaction metaphors for Geographic Information Systems. Recent research projects include the Linked Open Data University of Muenster (together with the university library), and a series of EU projects on geospatial services in the semantic web.

Kuhn holds a doctorate from ETH Zurich (1989) and was a post-doctoral researcher with the National Center for Geographic Information and Analysis (1989–1991) as well as with the Vienna University of Technology (1991–1996). He is a co-founder of the COSIT Conference Series (since 1993) and of the Vespucci Initiative for Advancing Science through Geographic Information. His publications range from GIScience and usability engineering through cognitive science to formal ontology. He is a member of several editorial boards of peer-reviewed international journals, such as the International Journal of Geographical Information Science (IJGIS), the Semantic Web Journal (SWJ), Applied Ontology (AO), Spatial Cognition and Computation (SCC), the International Journal of Spatial Data Infrastructures Research (IJSDIR) and the Journal of Spatial Information Science (JoSIS).
SARA LAFIA
Department of Geography
University of California, Santa Barbara
Email: sara.lafia@gmail.com

Sara Lafia is a graduate student in the Geography Department at UCSB. She works in partnership with the UCSB library and the Center for Spatial Studies on improving the spatial discovery of research data and documents. Her research addresses the question of how to spatially enable discovery of connected data and publications in a setting that allows for mapping and analysis using a Geographic Information System. She is also interested in the application of spatialization frameworks to non-spatial data, such as text, to gain new insights about themes of contents across formats. Her background is in Urban and Regional Planning. She has experience working on projects in collaboration with the Region 10 Environmental Planning Agency, the South Coast Air Quality Management District, and the Jet Propulsion Laboratory, applying GIS to assess social issues while also improving the accessibility of spatial tools.

Challenges

Location is an integrator of disciplinary perspectives and enables data producers to increase the discoverability of their products by offering a framework upon which data from multiple sources can be integrated. Platforms, for example Esri Open Data, have enabled format-agnostic approaches for data hosting and description at UCSB, which allows for a single-point of access across already hosted services, types, and domains. There are great opportunities for coordinating data production, dissemination, and use across university campuses specifically. However, while abundant in principle, data portals that span multiple repositories and integrate contents based on location are still far from common in practice. Current efforts, like Geoblacklight, tend to focus specifically on the discovery of more traditional cartographic products, such as geospatial data, while other types of data, such as interviews or historic texts, fall by the wayside. How can we facilitate the comparable spatial discovery of qualitative data, such as ethnographic surveys and text, which are not described in spatially explicit ways?

My research has produced a simple and extensible workflow for describing research objects as a library service and for spatially integrating datasets across repositories. This addresses current common practices for data publishing in academia, which often results in datasets that are not easily discovered, are hard to integrate across domains, and typically are not linked to publications about them. For example, archaeological observations and specimen data collections that share a spatial extent in Mesoamerica, should be concurrently discoverable. Relevant publications or other documents that reference the datasets should also be linked. Additionally, the notion of integrating across scales poses challenges. How should the location of data observations be described across multiple spatial scales?
Lastly, expanding the notion of spatial discovery to topic spaces offers another view of data that is non-geographic. For example, recent efforts to produce a self-organizing map of campus research, based on the subject descriptions of dissertations, yielded interesting insights into clusters of researcher themes. Coupling topic modeling with place-based views of research, obtained from geoparsing the same research descriptions, allows for new insights, offering two entry points for discovery through both geographic and topic spaces. How can the application of techniques, such as topic modeling, to unstructured text communicate collaboration gaps and motivate further research across disciplines?
Erin Mutch is the director of the Spatial Analysis and Research Center (SpARC) at University of California, Merced (UC Merced). She supports project development, grant applications, and research support campus wide for projects needing spatial analysis support and implementation. Mutch has also developed and conducted workshops and training for UC Merced students, faculty, staff and off-campus partners and provides consultation to faculty and student research projects. Mutch completed her MGIS at Pennsylvania State University in 2007, with a capstone focus on Geographic Information System (GIS) enterprise implementation for new and growing organizations.

**Perspective**

The mission and goals of SpARC is to be the hub for spatial science, research, analyses, education, visualization, spatial data archiving and access to spatial science software. Our focus is to enhance our ability to support spatial data archiving with the collaboration with the UC Merced Library and are currently developing projects archiving historical maps, and integrating historical data into an accessible GIS. My challenge is to manage spatially pertinent research data, maintain data quality and integrity through documentation and data archiving through coordination with our library staff. In addition, we are working with faculty and students in implementing geospatial software applications that minimize the need for software training while enhancing research and visual communications. As the nation’s first doctoral research university of the 21st century, our goal is to utilize spatial analysis applications to integrate and support data archiving and provide spatially oriented portals to access library information.
NATHAN PIEKIELEK
Geospatial Services Librarian
The Pennsylvania State University Library
Email: nbp104@psu.edu

Nathan Piekielek promotes spatial research and teaching in all departments and units at Penn State. Recent activities include establishing a maps and geospatial information support center from within the library. His research interests are varied and have included scientific workflow software and targeting non-geography academic disciplines for GIS outreach. He is currently working to develop fully- and semi-automated methods to convert print geographic resources (e.g., maps and historic aerial photographs), to research-quality digital geospatial datasets to be posted to centralized data portals. He is a taskforce member of the Big Ten Academic Alliance (BTAA) geoportal project and serves on their collection development subcommittee (https://geo.btaa.org/). Piekielek holds degrees in ecology, conservation ecology and sustainable development, and geography as well as postdoctoral experience working on climate change.

Perspective Statement

- How can spatially mediated discovery provide single-point access to research data, across distributed repositories and catalogs?
- How can the discovery of research objects in general be spatially supported? and □
- How can spatial discovery be applied to topic spaces, not just geographic ones?

Spatial discovery across repositories/catalogs is often mediated by metadata, specifically metadata elements that are populated and structured in common (i.e. compatible schemas). The metadata records and/or the data themselves are then aggregated (and cross-walked if necessary), and discovery tools built on top of the aggregated metadata records. This has been our approach in the BTAA geoportal project with the caveat being that we have only aggregated the metadata records—the data themselves remain in each individual institution’s repository. Issues with this approach are that it requires a substantial amount of effort (i.e., mediated deposit), to get an item into the collection and the collection itself requires constant update and maintenance to remain relevant. When considering applying this approach to research data one might realize that not every research project interacts with geographic space or frame of reference in the same way—for example in lab-based experiments, does the geographic location of the lab where the experiment took place have meaning? Was there a specialized instrument used in the experiment with a unique frame of reference that is meaningful to the results and for which “spatial” search should be enabled? Inconsistencies in representation of space and frame of reference make it difficult to define and enforce a common metadata element on which spatial discovery can be performed. For example, searching Penn State’s mediated deposit research data repository (DataCommons), by geographic search criteria will not produce every dataset in the result because a few (~30) escape geographic
description. Contrast this by searching Penn State’s self-deposit repository (Scholarsphere), where you cannot search by geography and where many datasets are not documented well enough for reuse.

In general, my perspective is that supporting interdisciplinary research data discovery is the far more tractable problem when compared to trying to get researchers to document their data well enough to support reuse. Rather than “metadata,” I might call this barebones characterization of data “documentation” including some kind of assessment of quality so that others may know for what other research questions the data are appropriate? What good is discovery if the data are not reusable by anyone except for (or including), the original researcher?
MARK SCHILDAUER
National Center for Ecological Analysis and Synthesis (NCEAS)
University of California, Santa Barbara
Email: schild@nceas.ucsb.edu

Mark Schildhauer is Director of Computing at the National Center for Ecological Analysis and Synthesis, NCEAS (http://nceas.ucsb.edu) at UCSB, where he oversees NCEAS’ cyberinfrastructure and technical staff, focusing on meeting the scientific computing needs of NCEAS’ associated projects and researchers.

Schildhauer’s technology research interests are in conceptual data modeling, the semantic web, scientific workflows, and Open Science--all in the context of facilitating integrative environmental and conservation science. Mark is currently involved in the NSF’s DataONE (http://dataone.org) Provenance and Measurement Semantics Working Group, and the NSF’s EarthCube GeoLink project (https://earthcube.org/group/geolink), which are advancing the use of techniques in Linked Open Data and the Semantic Web to describe and confederate disparate earth science data resources.

Schildhauer has a Ph.D. in Ecology, Evolutionary, and Marine Biology from the University of California, Santa Barbara, and an A.B. in Biology from Harvard.

**Perspective**

- **How can spatially mediated discovery provide single-point access to research data, across distributed repositories and catalogs?**

Access to spatial data is increasingly important to researchers in ecology and the environmental sciences, as the focus shifts beyond solely the experimentally-controlled “plot” scale to the landscape if not global scale of investigations. While the volume and variety of relevant spatial data for these investigations continues to rapidly grow, finding and harmonizing these data for synthetic analyses has been and will increasingly become more problematic. We believe this is due to several things. First, spatial data are broadly distributed—often housed in incompatible “siloes” due to idiosyncratic modeling of the “same” concepts, and these siloes not intercommunicating due to mismatches in terminology and schema. Knowing which of potentially numerous spatial data portals to visit, and then learning how to find and access the data one needs, becomes a time-consuming chore. Then, one finds that the data are documented with a variety of (often) sparse metadata, adhering to multiple potential standards or non-standards. There follows a conceptually and often computationally challenging and tedious process of transforming and integrating the data for one’s particular analyses. The difficulty of data integration is compounded by variation in the grain and precision of the data, as well as errors or omissions in their creation (i.e. variation in data quality).

How can we resolve some of these issues? We believe the development and more effective promotion of open-source, community-vetted, standardized models and vocabularies for describing the essential aspects of spatial phenomena and data—along with free, accessible tools and frameworks to link these to existing spatial data repositories—can alleviate some of these problems,
and represent a significant opportunity for improving this situation. The OGC is a good example of how these types of efforts can be productive, but there is danger of over-engineering solutions such that they become impractical for implementation, except for the most well-resourced participants. That is, highly complex metadata or other schemas and services can hinder uptake and participation. This is not ideal—both production and consumption of spatial data should ideally be democratized to the broadest possible population of consumers AND producers. In addition, any solutions should be fully interoperable with Web technologies—that is, involving globally unique, web-resolvable and dereferenceable identifiers for objects at grains of interest relevant for researchers, that are using semantically-defined terms curated by community-authorized consortia, and represented using W3-sanctioned languages, such as RDF/OWL.

- **How can the discovery of research objects in general be spatially supported?**

  The advent of increasingly miniaturized and affordable GPS sensors, along with other sensor technologies, is making it feasible to provide highly accurate geospatial information along with any field or remote-sensed observations and measurements. In the case of ecology, biodiversity, conservation, and associated environmental sciences, however, while “raw” geospatial information can enable understanding as to the specific location, “distance-between,” trajectory, areal extent, etc. of the observations—it is the opportunity to correlate a multitude of other measurements/observations through spatiotemporal coincidence that will also be extremely useful. For example, a query such as: “Where will drought most severely impact agricultural productivity” recommends spatial resolution to advance both general understanding and planning purposes—but fuzzy notions such as “drought” and “agricultural productivity” will need to be defined, as well as linked to locations in order for such queries to be resolvable. The spatial boundaries on the locations to which these thematic terms are linked may also be fuzzy or probabilistic: representing this is a research challenge.

  *Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.* (Mandelbrot)

  The fractal geometry of nature, coupled with “fuzziness” can be particularly problematic when referring to named spatial phenomena. Is “Santa Barbara” part of “Southern California”? While spatial ontologies MUST address issues of membership, containment, adjacency, disjointedness, etc. (and medical anatomy ontologies provide some good precedents for how to do this), and do so in a consistent, unambiguous, web-accessible way—natural language allows for many variations on how terms can link to the concepts they represent. Words that are polysemes, or have homonyms, synonyms, or plesionyms create challenges relative to appropriate or useful query resolution. Named places in particular often are polysemous, especially if not bounded by fiat. Temporal instability of spatial structure and configurations add another element of complexity to recognize and deal with.

  These complications further recommend resolving the first point above, to enable spatial information to more effectively become an important “Key” linking thematically described data to other data of potential interest.
• How can spatial discovery be applied to topic spaces, not just geographic ones?

I am not clear here if “topic space” is being used in a narrow “Publishers” sense of being able to group thematically related sets of publications. But whether the connotations of “topic spaces” here are narrowly circumscribed, or intended to be interpreted more broadly, we’d recommend that the approach for developing “topic spaces” and assigning digital objects as members to one or more of these “topic spaces,” conform to emerging standards for the W3C semantic web approaches, specifically by being represented in formal RDF/OWL ontologies. This logic-based language enables us to define Classes that can have Instances, and Instances are members of Classes. Classes can also contain SubClasses, and Instances can be richly related to one another through Object Properties assigned to the Classes. The logical expressivity of OWL allows one to build a rich and W3C standards-interoperable specification of how objects and their digital representations might “belong” to various “topic spaces” (although again, I’m not familiar with the topic space approach).

Finally, at the meta-level, I still find the conceptual model of accumulating “thematic-defined, spatially-registered layers” to be both intuitive and useful.
Katja Seltmann is the Director of the Cheadle Center for Biodiversity and Ecological Restoration (CCBER) at University of California Santa Barbara. CCBER houses a large, regionally focused natural history collection of plants, algae, insects and vertebrates. Her research focus is the field of biodiversity informatics, or data science based research of digitized natural history collection records, arthropod diversity in restoration habitats, and insect evolution.

**Perspective**

Scientific data can often be difficult to find using keyword searching, and spatial searching provides an opportunity to discover otherwise unlikely, but useful, datasets and descriptive information. I am interested in thinking creatively about how to expose natural history collection data and biological data publications, in order to better match natural processes of human inquiry. Many biological questions involve spatial aspects, although those are not always place-based. Rather, they may involve time, habitat type, ecoregion, or other spatial connections. Also, many valuable observations about organisms, or the places they are known to occur, are found outside the scientific literature. This suggests that by crossing intellectual boundaries, and enabling discovery via historical documents, biography, poetry, photographs and other media, it may be possible to enable collaborative and cross-disciplinary research that would otherwise be difficult to realize.
**STEPHANIE SIMMS**
California Digital Library
University of California Curation Center
Email: stephanie.simms@gmail.com

Stephanie Simms is a Research Data Specialist at the California Digital Library where she provides technical, educational, and policy-related support for the DMPTool (https://dmptool.org/). She works with the UC campuses and other research institutions, organizations, and funders in the U.S. and abroad to deliver training for research data management and promote open scholarship. Her current focus is coordinating global efforts to transform data management plans from an annoying administrative exercise into a useful tool for researchers, funders, service providers, et al. by making them active and machine actionable. Prior to joining the CDL she was a CLIR Postdoctoral Fellow at UCLA where she worked on geospatial data initiatives to connect the library with other campus stakeholders. Her own experiences with research data management during archaeological field projects in Mexico and Guatemala (Ph.D. 2014, Boston University) involved plenty of spatial data: creating maps, integrating historical maps and remotely sensed data in a GIS, modeling and spatial statistics, handling sensitive spatial data, etc.

Twitter: @stephrsims@TheDMPTool

**Perspective**

At the 2015 meeting I enjoyed grappling with the technical and theoretical complexities of spatial data—e.g., time and fuzzy boundaries—and how to represent these through spatially enabled search. I look forward to balancing previous discussions with the new emphasis on discovery. Discovery, or discoverability, hinges on both the technical infrastructure and the quality of the data and associated documentation. These interdependent considerations are central to my work in research data management outreach and education across all disciplines. At present most researchers lack the basic skills to manage their own data effectively. And the few who take the steps to preserve their data in a repository tend to think of it as a dumping ground rather than a place for discovery, which results in data that are not reusable. As we talk about designing systems, I hope to contribute and learn from others’ experiences with advancing the cause of basic data literacy (and consider how to do this in the context of spatial data discovery).
ANDRÉ SKUPIN
Center for Information Convergence and Strategy
San Diego State University
Email: skupin@mail.sdsu.edu

André Skupin is a Professor of Geography and the Founder and Co-Director of the Center for Information Convergence and Strategy (CICS) at San Diego State University. He is also vice chair and co-host of CICS’ annual Left of Boom conference on proactive threat mitigation strategies. He combines a classic cartographic education with long-standing interests in geovisualization, visual data mining, and spatio-temporal modeling. He has developed novel methods for representing human mobility, demographic change, and environmental indicators in attribute space. Skupin been engaged in the visualization of abstract knowledge spaces for more than two decades, combining traditionally disparate approaches from natural language processing, machine learning, and cartography. These efforts have been funded by diverse sponsors, including the White House Office of National Drug Control Policy, National Science Foundation, and National Institutes of Health. He is co-inventor of several patent-pending technologies involving ontologies and text mining for knowledge engineering. As co-founder of the commercial spin-off BigKnowledge® and Associate Director of the Center for Entrepreneurship and Innovation at the University of Dubai, he has a strong interest in accelerated transition of technological innovation into diverse application areas, from biomedical knowledge management to financial analytics, demography, crime analysis, and environmental monitoring.

Metaphors—Methods—Products

Some have long advocated for a broader perspective on the “spatial” adjective, beyond the narrow confines of geographically referenced locations in physical space. Such a wider view seems appropriate even—and perhaps especially—among geographers, GIS experts and the like, a community to which I happily count myself. The various disciplines that profess an intimate concern with space as a central concept and a source of foundational principles, from geography to architecture, physics, etc, have a lot to contribute and to gain by applying their thinking, concepts, and methods to the core themes of this expert meeting, from spatial mediation to spatial support and spatial discovery.

By adopting spatial concepts (proximity, region, scale . . .) as a set of unifying principles, is becomes not only possible to “provide single-point access to research data, across distributed repositories and catalogs,” i.e., access to research artifacts, but also to the actors that produce and consume such artifacts in the course of various activities (Fig. 1). While the notion of “One Space” in the diagram is meant to reflect a unifying power, harkening back to John Snow and the spatial analysis tradition, there typically are multiple spaces (perhaps reminiscent of Gärdenfors’ separable domains) of distinct dimensions, via which and in which a given set of items could be meaningfully
represented. Examples include geographic location, certainly, but also locations in knowledge domains, like pediatrics, finance, or information science.

It turns out then that spatial approaches cannot merely “be applied to topic spaces”—note the tautology in question 3—but that discovery and “single-point access” based on topic spaces simply is just another form of spatially mediated discovery.

The spatiality of topic spaces, term spaces, concept networks, etc, is inherent in the inner workings of many search mechanisms, such as when ranked proximities between a text query and corpus content are computed. However, that spatiality is rarely made explicit to users. Notable exceptions include efforts at spatialization of abstract spaces, with the goal of making such spaces more accessible to human cognitive abilities. In that context, in 2000 I called for moving “from metaphor to method” in the engagement of cartographic principles for non-geographic information visualization. Much progress has since been made, in that (1) methods have been developed that explicitly link cartography, information science, and computing, and (2) a robust empirical body of evidence regarding corresponding cognitive issues has been built. There have also been sustained efforts at popularizing the visual mapping of knowledge spaces, notably the “Places & Spaces” exhibit (http://scimaps.org/) spearheaded by Katy Börner, now in entering its 13th annual iteration.

Still, I wonder whether the three questions posed going into the expert meeting are reflective of these being rather academic pursuits, with impact on broader society being implied rather than being realized through an overt translation of novel methods into products. In other words, are we doing enough to enable stakeholders in business, government, and academia to adopt spatially mediated discovery in the pursuit of their mission?
Denise Stephens joined University of California, Santa Barbara as the University Librarian in 2011. Her particular areas of experience and research are in organizational and change leadership, program assessment, and digital services. Stephens came to UCSB from the University of Kansas, where she served for many years as Vice Provost for Information Services and Chief Information Officer. Prior to that, she was Acting University Librarian at Syracuse University. She holds a master’s degree in library science from the University of Oklahoma and is an alumna of the Association of Research Libraries Leadership and Career Development Program.

As the University Librarian, she is leading revolutionary advances in information technology as well as the expansion and redesign of the University Library. The expansion will provide critical physical spaces to support professional expertise and state of the art information resources to the new generation of scholars, technology, and the community.

Stephens also served as UCSB’s Interim Chief Information Officer from October, 2013 to September, 2015. In that capacity, she led the transformation of the Office of Information Systems and Technology (OIST) and Administrative Services enterprise IT programs into the new Enterprise Technology Services (ETS), a coordinated, broad-based organization delivering enterprise-wide solutions to campus.
Position Statement

Geographic content permeates library collections, whether through places described in books, depicted on maps, newspaper articles or photographs, as well as the location of authors and publishers. Spatially mediated discovery, through a single-point access is quite complex, and unfortunately may supply haphazard results due to variations of content standards across different schemas. For example, library catalog data is quite structured, and due to historic content standards, is often a simplified, aggregate representation of intellectual (geographic) content. This is a very different approach than what is found in metadata found in library digital repositories (i.e. Dublin Core), which often contains unstructured metadata with varying levels of detail. Different still is the robust and precise geospatial metadata found within a GIS. How can we integrate these various worlds into a single search if the underlying content standards are based on different assumptions? Additionally, consideration must be given to known geographic issues of scale, spatial footprints, and temporality. Integration in a single search must take into account the variation of input standards and assume messy results.

One approach to expand the catalog data and therefore, the discoverability of research objects with place references is with linked open data. Through LOD, scholarship is facilitated both within and more importantly, outside the library environment. Library catalog place name data is evolving to LOD. The long-term desired outcome of the LD4L project is to facilitate this flow of information to the broader scholarly environment. There are numerous scholarly gazetteers with detailed curated place name references, among them, DINAA (Digital Index of North American Archaeology),
Pleiades, and Pelagios among many others. Further, there are broad based LOD resources, including the Getty Thesaurus of Geographic Names, GeoNames and Wikipedia/DBPedia, which provide robust gazetteer data. Challenges, here include the complication that that each discipline will approach the definition of a place, a region, and place-related time periods from their particular disciplinary lenses. Ultimately, however, the integration of library (spatial) search to a wide variety of linked resources will only serve to expand the scholarship of place.

Another approach to the challenges of single search may be by using a combination of spatial search (using geographic coordinates) and natural language processing of the textual elements inside gazetteers. NLP could be employed to evaluate and process gazetteer data (including library catalog ‘Name Authority’ data and other digital gazetteers). From a search request, a NLP/spatial tool could evaluate dates, alternate names, spatial extent, etc. from the gazetteers, and then perform a more precise spatial search, therefore improving search results.

Scholarship will greatly benefit as we continue to pursue these spatial discovery questions. The ultimate goal of a modern and relevant library discovery system should be a seamlessly integration with the wider semantic web, therefore, intellectual access to library collections alongside other GLAM and scholarly resources. It is part of human nature to want to understand the why of “where.” Surfacing human knowledge through spatial search will allow for more robust ‘where’ questions, and ultimately more meaningful answers.
DAWN WRIGHT
Chief Scientist
Esri
Email: dwright@esri.com

Dawn Wright was appointed Chief Scientist of Esri in October 2011 after 17 years as a professor of geography and oceanography at Oregon State University. In this role, she reports directly to the Esri CEO and works with him and other Esri directors on strengthening the scientific foundation for Esri software and services, while representing Esri to the national and international scientific community. Wright also maintains an affiliated faculty appointment in the College of Earth, Ocean, and Atmospheric Sciences at Oregon State. Her current research interests include data modeling, benthic terrain and habitat characterization, coastal/ocean informatics, and cyberinfrastructure.

Wright’s recent advisory board service includes the U.S. National Academy of Sciences Ocean Studies Board, the NOAA Science Advisory Board, the Science Advisory Board of Conservation International, the Board of COMPASS Science Communication Inc., and many journal editorial boards. She is a Fellow of the American Association for the Advancement of Science, the Geological Society of America, as well as a fellow of Stanford University’s Leopold Leadership Program.

She holds an Individual Interdisciplinary Ph.D. in Physical Geography and Marine Geology from UCSB, an M.S. in Oceanography from Texas A&M, and a B.S. cum laude in Geology from Wheaton College (Illinois). Other interests include road cycling, 18th-century pirates, apricot green tea gummy bears, her dog Sally, and SpongeBob Squarepants. Follow her on Twitter at @deepseadawn.

Web: http://dusk.geo.orst.edu/

Perspective Statement

1. How can spatially mediated discovery provide single-point access to research data, across distributed repositories and catalogs?
2. How can the discovery of research objects in general be spatially supported? and
3. How can spatial discovery be applied to topic spaces, in addition to geographic ones?

In addition to an individual research project, many researchers are involved in at least one major partnership, perhaps one ocean observatory, or one collaborative. The accompanying framework for spatially-mediated discovery of resources may be focused on a single discipline or subdiscipline, or a particular study region. The spatially mediated discovery obviously exists to support research, but also collaboration in data collection, spatial analysis, visualization, and communication of the science to multiple audience. These interactions likely take place at multiple scales as they traverse distributed repositories and catalogs: the scale that serves the individual research project, or scales of small workgroups within a lab, or of inter-organizational collaboration. There are also frameworks that cut horizontally across discipline and region, connecting to broader national or global initiatives such as NSF EarthCube, other NSF-funded Research Coordination Networks, GEOSS, or the like.
In order to make data most readily accessible, a “geospatial aggregator” or web services approach is one consideration, using common service interface specifications that build on international standards from the World Wide Web Consortium (W3C), the Open Geospatial Consortium (OGC) and others. Any “aggregator” should ideally consider the full data value chain that includes connecting to Earth observation including in situ sensor networks, providing mechanisms for storing and hosting content (especially when hosting is not possible at the data source), making content discoverable, and enabling use of content in different media, for both online and offline use. To further increase the visibility and use of content and information products, they may be disseminated to other global or national networks also serving as aggregators of sorts, such as the Group on Earth Observation, the United States Geospatial Platform, and others. And of course all data and metadata must be OGC- or ISO-compliant. Another important consideration is the enforcement of minimum requirements for inclusion, such as:

- well-maintained data and documented with Open Geospatial Consortium (OGC) or International Organization for Standardization (ISO)-compliant metadata;
- data services that are reliable and well performing (i.e., services run on servers experiencing minimum disruption or downtime, with holdings exposed via an OGC-compliant catalog services such as CSW (Catalog Services for the Web) and as an OpenSearch endpoint, accessible through REST API, and providing GeoRSS, KML, HTML, or JSON responses; and with registered resources monitored and synchronized according to any changes in the catalog service.
- web maps with a well-defined legend, on a well-focused topic, and with well-configured pop-ups; and
- completion of all required elements of a “home page” for the contributed item in ArcGIS Online containing an attractive thumbnail, informative item details, descriptive user profile, data or map access use constraints, credits or attribution, and search tags.

It is hoped that spatial mediated discovery means not only the provisioning *data services*, presumably in the cloud, but also the crosswalking and sharing of workflows and use cases, additional apps for mobile, web, and desktop, community-building events where people gather face-to-face, and close interlinkages to other platforms such as NSF’s EarthCube. And finally, as we contend with human impacts on the biosphere, we see that recent innovations in computational and data science are now facilitating community resilience to climate change (e.g., helping communities to monitoring air quality or drought, find available drinking water, determine habitat vulnerability, etc.). But not often discussed is a path toward digital resilience. If digital tools, found via spatially-mediated discovery, are to continue helping communities, it stands to reason that the tools and the discovery frameworks must engender some resilience themselves. The capacity to deal effectively with change and threats, to recover quickly from challenges or difficulties, even to withstand stress and catastrophe, can apply to them too.
Marten Hogeweg
Senior Project Manager
Esri
Email: mhogeweg@esri.com

Marten Hogeweg joined Esri in 2001 as a Consultant and Project Manager focused on products and solutions for spatial data infrastructures (SDI), open government, and enterprise information sharing programs. He has extensive experience in geospatial industry standards from ISO, the Open Geospatial Consortium, United States Geospatial Platform, and European INSPIRE programs. Since 2007, Hogeweg has supported the Group on Earth Observations initiative to increase discovery and access to Earth Observation data. Hogeweg managed the design and implementation of the National Spatial Data Infrastructure networking system in Indonesia between 2010 and 2015. The NSDI connects various government organization who can share geospatial information between themselves, as well as provides public users access through a public information portal. Hogeweg has served as project manager for the Geospatial One-Stop program from 2003 to 2011. He was responsible for the operation and maintenance of the portal as well as for the development of enhancements and outreach to the user community. He managed the successful retirement of the portal and the transition to its successor site as part of Data.gov. He was instrumental in building the relationship between the Geospatial One-Stop and Data.gov programs. He is currently working with the US GeoPlatform, AmeriGEOSS, as well as various organizations who are looking to create data discovery solutions. Since 2003, Hogeweg has developed the Esri Geoportal Server open source metadata catalog, that focuses on facilitating data discovery in heterogeneous SDI. The Geoportal Server implements various metadata and discovery specifications from both the geospatial domain and the World Wide Web consortium. Hogeweg has a Master of Science in Geographic Information Systems from the University of Salford, United Kingdom (2000) and a Master of Science in Mathematics, the Free University of Amsterdam, the Netherlands (1989).