

Jun 22nd, 10:30 AM - 11:30 AM

# Library applications, collaborations, and courses for geodata and geoinformatics

C. C. Miller

*Purdue University*, [ccmiller@purdue.edu](mailto:ccmiller@purdue.edu)

Follow this and additional works at: <http://docs.lib.purdue.edu/iatul2010>

---

C. C. Miller, "Library applications, collaborations, and courses for geodata and geoinformatics" (June 22, 2010). *International Association of Scientific and Technological University Libraries, 31st Annual Conference*. Paper 6.  
<http://docs.lib.purdue.edu/iatul2010/conf/day2/6>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# LIBRARY APPLICATIONS, COLLABORATIONS, AND COURSES FOR GEODATA AND GEOINFORMATICS

C.C. Miller  
Purdue University Libraries, U.S.A.  
ccmiller@purdue.edu

## ABSTRACT

This paper will overview several library GIS and geoinformatics projects, collaborations, and initiatives at Purdue that, taken together, formulate a reasonable plan of attack for geolibrarians interested in tackling some (not all) of the issues rolled up into e-geoscience and geoinformatics. Included as examples are a metadata harvesting/catalog project, collaborations with domain faculty, courses developed and taught, and e-data efforts. From this handful of examples an argument will be made that aggressive and efficient individual geolibrarians can impact and indeed steer geoinformatics and interdisciplinary research efforts project by project, then campus by campus, and simultaneously develop and evolve their library data services and strategies. But with a price.

keywords: geoinformatics, GIS, geographic information systems, geolibrarianship, geospatial librarianship

## INTRODUCTION

Where it is happening, geolibrarians pushing into data curation and/or management, application-building, and collaborative research — geoinformatics, let's say — are traversing craggy, rugged terrain. Among our challenges are A) how to quickly building agile, interoperable library infrastructure that can respond to the shifting demands and requirements of, well, agile, interoperable cyberinfrastructure; B) the technological acumen one must possess in order to at least speak the language of those building this geocyberinfrastructure (and let alone to contribute to it); C) where to find entrée at all into geoinformatics initiatives, and D) how to manage librarian contributions to collaborative geoinformatics initiatives against still-important, but more traditional GIS services and education roles on campus.

This paper will overview several library GIS and geoinformatics projects, collaborations, and initiatives at Purdue that, taken together, formulate a reasonable plan of attack for geolibrarians interested in tackling some (not all) of the issues rolled up into e-geoscience and geoinformatics. Included as examples are a metadata harvesting/catalog project, collaborations with domain faculty, courses developed and taught, and e-data efforts. From this handful of examples an argument will be made that aggressive and efficient individual geolibrarians can impact and indeed steer geoinformatics and interdisciplinary research efforts project by project, then campus by campus, and simultaneously develop and evolve their library data services and strategies. But with a price.

## GEOINFORMATICS & LIBRARY SCIENCE

Geoinformatics is happening in many places and at many levels of complexity and funding. These projects are built to sustain, interoperate, and laterally support scientific research and discovery across disciplines and therefore tend to require more strategy, more labor, and more robust technologies than projects with a less extroverted attitude. Further, their need for data archiving, dissemination and sharing (typically via web services or other portal structures and protocols) tends to draw together domain and information scientists from around the world who share a devotion to open, standards-adherent, data-rich, web-facing applications that make geodata more accessible and usable by, in turn, the scientific and scholarly community, other systems, and the general public.

Now imagine the mission of some geospatial librarian collective. Would it not promise to deliver and support, both in collaboration with and support of domain researchers or other academic clientele, “open, standards-adherent, data-rich, web-facing applications which make geodata more accessible and usable by, in turn, the scientific and scholarly community, other systems, and the general public”? It is no coincidence, as there is plenty of library science in geoinformatics and vice versa. So although it is not the primary concern of this paper, a silent premise behind the remainder of this discussion is that there is a dearth of geolibrarians currently collaborating within or contributing to geoinformatics efforts around the world. No study has been conducted, but anecdotal evidence suggests that scientists and other personnel currently doing the work of building geocyberinfrastructure — which includes robust data archiving; resource discovery, access, and delivery mechanisms; and education components — are doing so largely without assistance from the library community. On more than one occasion, this author has been the only librarian at a geoinformatics-heavy conference or workshop and has in fact been asked if and how library expertise, research, and infrastructure could be folded into future or ongoing work in the field.

Exactly why there appears to be a dearth of librarians in geoinformatics is beyond the scope of this paper and may or may not warrant study. Legitimate, easy guesses would be that they're too busy answering Census questions; they're squirreled away preparing for the next round of ArcGIS training workshops; or simply spread too thin with the rest of their library duties. A more ominous hypothesis is that geoinformatics and CI-building are too intense, too intimidating, too resource-heavy, or too far outside the position description (or worse still the spectrum of awareness) of most geolibrarians to warrant the intense commitment required of its builders. Whatever the reason, to have few librarians, and therefore little library science, in on the efforts happening in geoinformatics right now is both a loss for librarians and a loss for those projects.

## IT'S WAR!

Although it is ultimately laughably inappropriate to posit that librarians should be waging war against domain and interdisciplinary researchers as a means to gain entrée into geoinformatics efforts (we should not and will not — we're on the same team in every way), there is a case to be made for it being a rhetorically accurate approach to libraries research and data support. Note the language of force and struggle that has slipped into this paper already: “librarians *pushing...*”, this is a “plan of *attack*”, we face “*craggy, rugged terrain.*” If we can accept for a moment the trope that research and scholarship is a battleground, librarians would have to be portrayed as historically postbellum forces, swooping in to manage and distribute the spoils of the war — the articles, monographs, and other miscellaneous materials that were forged in the hard fighting of academia — but having little impact on the outcome itself.

But as scholarly publishing boils and roils and evolves and as data management and reuse becomes increasingly important, it is less and less acceptable for librarians to simply wait for the call. Managing the materials that come into the building (often by auto-populating selection profiles) is still important, of course, but for data, applications, and

other such rough-hewn and cumbersome, OPAC-unfriendly beasts, simply waiting for a deposit of clearly-defined objects will not do. Anyone who has been handed data that were generated without an assumption of future library deposit and reuse can attest to the value of having someone looking out – from the start – for metadata, cleanliness, interoperability, and format issues. All of the examples presented in this paper (successes each, naturally) are direct results of librarian involvement in the content generation cycle well upstream of where it has historically occurred. To have librarians advising data generating projects *before* the harvest of some final product (if indeed there is a final product) is a relatively new addition to the toolkit with serious benefits and implications to the profession. This “up-streaminess” requires that librarians take some ground, however, and in fact infiltrate and occupy the space held historically only by researchers and traditional educators (who, we can probably all agree, aren’t paid or tenured by worrying about what librarians worry about). It means librarians are team members, co-principal investigators, advisors, or at least consultants. It means they sit in on project planning and progress meetings asking questions about how the project data will live when funding ends. It means they develop research ideas and strategies by asking questions about how other scholars will find and use the resources to be created by this or that workflow. It means they write sections of the project proposals themselves, injecting data management techniques and resources or even just geospatial technologies and services. And it means they don’t wait for invitations to do 50-minute one-off class visits, but seize other or additional opportunities for instilling geodata literacy concepts and skills into future (or current!) faculty.

So it isn’t war, really, but it might behoove librarians to approach it as such if that’s what it will take for us to claim our rightful place near the heart of what is being called (for some time, now) a “revolution...” [Atkins, 2003]. Even a modicum of aggression, coupled with a healthy commitment to progressive librarianship, will help push the librarian presence in geoinformatics (or perhaps more accurately push geoinformatics closer to libraries).

What is perhaps surprising is how easy this can be. Given the freedom to seek and develop collaborations with faculty researchers on campus or to commandeer departmental curricula, it took very little time for Purdue Libraries to become partners on or participants in a number of projects across campus that put librarians in new territory outside the walls of traditional library service – behind enemy lines, let’s say. Without overwhelming detail, several of these projects will be described below as exemplars of techniques and perspectives that have so far been winning battles for The Libraries.

IsoMAP is a 3-year project sponsored by the National Science Foundation’s Biological Databases and Informatics (BD&I) program. Its primary objective is to deploy spatially-explicit isotope models and data online, but by using a user-friendly, modular, open architecture as a broker to the power of running complicated geostatistical processes on TeraGrid. The project ultimately began when the future principal investigator needed help firing off a sliver of python code from within Google Earth and inquired in Purdue Libraries: basic GIS reference, in other words. This transaction led quickly to a collaboration on the much larger IsoMAP effort and within one year here are librarian hands up in the gears and wires of a grid-enabled, web-facing geospatial machine, contributing web mapping and geoprocessing strategy and code, making links between existing metadata services and pieces of the workflow that need to consult them in automation, making sure workflow results (spatial datasets) are well-described by metadata, and making sure that the metadata are indexed and searchable, both internally and via various harvesting protocols in case other catalog systems want to include it.

Isee (Integrating Spatial Educational Experiences into Crop, Soil, and Environmental Science Curricula) launched in Summer 2008 when a Purdue Agronomy professor assembled a team of librarians, computer graphics technologists, and soil scientists to build mechanisms that could translate to the classroom his experience with in-the-field data exploration. The project is funded by a USDA Higher Education Challenge grant and is currently very close to an initial release of a web mapping application through which students (and anybody, really) can explore their landscape in an intuitive, fast, data-rich, but browser-based spatial environment. Libraries contributions to the project include identifying, converting, preparing, organizing and storing spatial data resources; metadata services, storage, and indexing; and preparing data for fast delivery not only through the chosen web platform (Google Earth API), but also through alternative data access protocols (WMS, WCS, WFS, WTMS, et al).

The generally-relevant lesson to be spun from these projects is that the technological barrier between libraries and geospatial research is surprisingly low. Most geolibrarians should already be versed in the software/hardware stacks in widespread use in the world of geoinformatics. Geodatabases like PostGIS, web mapping frameworks like OpenLayers, metadata servers like GeoNetwork, and geodata servers like GeoServer and MapServer are all standards-adhering and largely ubiquitous in the GIS world (well, perhaps with the sad, telling absence of almost *any* metadata tools). They are powerful, open source (all of them) utilities that can be strung together to support many, if not most geospatial data operations and any geolibrarian would be remiss to not be at least familiar with them. They are powerful and adopted more and more frequently because they are open, connectable, sustainable, and interoperable (like libraries!). For these reasons alone they should be of interest already to librarians (who seem to just now be emerging from a dark age of vendor-driven pricing and proprietary lock-out). But beyond these innate qualities there

is an easily-overlooked but logical aspect of geospatial technology that should quell any fears a librarian might have about being able to fit into the admittedly heavier technologies of geoinformatics: the same reasons that geoinformatics needs librarian expertise are the same reasons they need our technologies, too.

A prime example is how a geometadata catalog project being developed unfunded and in-house at Purdue has already been able to support the operations and planned operations of both the IsoMAP and Isee projects, playing an important role in iteratively building spatial model workflows in the case of IsoMAP and offering a more traditional metadata recording and storing service in the case of Isee data. The catalog itself doesn't warrant much real estate here, except to say that it is built entirely on open source technologies and was initiated and developed to be standards-adherent and therefore pluggable at any time into the Libraries' greater data initiatives (not to mention, if necessary, wider metadata collaborations such as the U.S. Geosciences Information Network or geodata.gov). In other words, it is an open, sustainable, efficient piece of library technology, meant to not only serve current library clientele but also be a robust and modular component of current and future cyberinfrastructure. In other words, here lies a geodata technology stack that not only acts as a fairly standard library service (allowing users to find, evaluate, and preview geodata stores), but by virtue of having been built with geocyberinfrastructure in mind it easily fills the same empty places in geoinformatics technology that a librarian can fill in geoinformatics strategy. The technologies needed by researchers who are building applications and generating data are already in libraries with GIS programs (or should be). That is, our guns are already loaded. And the librarian who can wield these weapons (or employs someone who can) will be very welcome on teams led by researchers who are down with the mission of libraries (open data stewardship, sharing and reuse), but who don't already possess the expertise or cannot spare the time and resources required to make it happen.

### **GUERRILLA EDUCATION**

The pomp and bombast about cyberinfrastructure has been intensifying in the last decade or so and has perhaps been in full throat since the Atkins-led Blue Ribbon report [2003]. Whether the hype ultimately proves true perhaps remains to be seen, but nonetheless it is encouraging to see that even heavy-hitters such as the NSF did not overlook the importance of education. Entirely. "Undergraduate curricula must be reinvented," writes the NSF in their "Cyberinfrastructure Vision...", "to fully exploit the capabilities made possible by cyberinfrastructure; and the education of the professionals that are being relied upon to support, develop and deploy future generations of cyberinfrastructure must be addressed." [National Science Foundation, 2007, p. 3]

So it's good to see, but a reader trained to be defensive and skeptical about stentorian claims of support for education from major national agencies could easily hear some of these statements as weak post scripts, as though the NSF remembered something the morning their report was due: "Oh, right. We — by the way, kinda sorta, if we have time — need to overhaul education to make sure people understand how interdisciplinarity benefits from robust, interoperable, open tools and data." So, it's easy to casually agree that young geoscientists need to grow more savvy about what it means to be a "member" of this infrastructure — that is, to use, build, and fund components of it — but it's still the least sexy side of the revolution by far. Library/faculty collaboration is the air war — high-profile, dramatic, liable to shock and awe those not accustomed to seeing librarians outside of the stacks — and education is the ground war. Rugged, skulking, slow, and costly.

In other words, perfect for librarians, who have been waging guerilla warfare on the information literacy front for years — taking ground inch by inch and losing it in the same way, but always fighting the fight while we wait for campus-wide information literacy mandates or even just a modicum of info literacy content injected into existing core curricula. So indeed one way to push libraries further into geodata and geoinformatics efforts is for librarians to wage little battles here and there, tiny little dust-ups where they quickly throw out a couple of ideas about the data lifecycle and where librarians fit into it. Besides, surely there is value in even just raising awareness in students and faculty that data are important, have lives beyond projects, need a place to go, and that many libraries are open for business in just those areas. This would include, yes, single-session drop-ins to existing courses, but also the kind of education being done during reference and consultation sessions — very traditional library services but with a new twist wherein the librarian lays a heavy rap about data access and reuse on the unsuspecting student that has stopped by for some help with this or that. It sounds like sabotage, almost, and requires a librarian who can make a fast case to their patrons about these issues. These librarians must be rhetorically blessed, actually, and able to steal any brief instance of time they can get in order to tell department heads that their students are falling behind by not having exposure to geoinformatics ideas and trends; to tell researchers they're essentially killing their data by not making them more available and usable in proper repository systems; or to make the especially obnoxious argument for spending extra time or effort on metadata for a project's data output.

Still, this is an obvious second choice and insufficient on the whole. It amounts to so much sniping at the enemy. Successful usually, but even then not enough to truly imbue interdisciplinary research and CI-building with the library services and science they need. Even a guerilla approach needs more formal campaigns, something like librarians help-

ing to develop courses or even curricula that address data literacy, the data lifecycle, and the expectation of future researchers to contribute to data solutions by maintaining a commitment to sharing, standards, and open systems.

And so the war wages on, with a course taught by librarians at Purdue University that addresses “data literacy, the data lifecycle, and the expectation of future researchers to contribute to data solutions by maintaining a commitment to sharing, standards, and open systems.”

“Geoinformatics” is an upper 500-level, 3 semester-hour course nominally offered through Purdue’s Department of Earth and Atmospheric Sciences but taught by a GIS librarian. The course takes as its premise that geoscientific technologies, data, and workflows should be examined together as irrevocably interrelated pieces of a growing cyberinfrastructure. Laid out on day one is the argument that it is important for current and future geoscientists to be aware of the arc of geoscience – not just modeling and analysis, say, but the fuller picture of data acquisition, manipulation, documentation, visualization, and sharing. Then through modules that visit and address online data sources, web services and data delivery protocols, desktop GIS packages and data formats, spatial databases, metadata, and general trends in interoperability and the mechanisms of data sharing the course endeavors to provide a fuller picture of how one finds and evaluates information, prepares and manipulates data, uses the tools of analysis and visualization (including GIS), and conducts their work so that data and information can be made reusable by the next scientist down the line. It is an holistic approach presented in a hands-on, project-driven course with opportunities to develop skills in visualization and analysis, using workflow management software to automate tasks, connecting technologies and data, and preparing geospatial data for reuse by the scientific community.

Throughout the course, GIS and geoinformatics technologies are applied to various disciplines according to individual students’ home disciplines and research interests. GIS and geoinformatics are largely discipline-agnostic tools that can be applied to many different problems in many different fields of study, so the course emphasizes how said tools can be made to enable the data and information products of a single discipline (the student’s own) to be made available for integration with the tools and data of other disciplines as a means to engender and encourage interdisciplinary exploration and scholarship. And while its wide scope means students complete the semester without a terribly rich experience with any given technology or procedure, the important issues are emphasized and students are reminded repeatedly that they will soon be out in the world, planning, directing, and completing research that generates a number of data and information products, and that if they remember then what they’re learning now that availability and interoperability of geoscience data and applications can be improved at their hand.

The course is, in some ways, guerilla gone legit. Developed by Purdue Librarians in 2007 well outside of any coordinated GIS, geospatial, or cyberinfrastructure (or library science) curriculum, it was taught for the first time in Spring 2008, again in Spring 2010, and is already on track to becoming a permanent Multidisciplinary Science credit as part of Purdue’s College of Science core curriculum.

## **CASUALTIES OF WAR**

Simultaneous ground and air wars don’t come cheap. Even though one of the most valuable lessons to be derived from the projects and perspectives presented here is the relatively low expense of the technological and education contributions made by librarians, the much steeper cost is a librarian’s time. Devoting percentages of research time to even a handful of sponsored projects isn’t a cakewalk. Developing innovative library solutions for geoinformatics work (or merely sufficient solutions, for that matter), is still time consuming work. As is developing and teaching courses, donating additional time by essentially volunteering on other projects (in a more traditional service role), developing and teaching 3-hour courses, doing classic GIS reference and consultation, sprinkling in still more of the standard librarian work (including web content authorship and development; professional development; literature reading; campus, local and regional service; faculty or administrative obligations). Throw in the dreaded “other duties as necessary” catch-all on many position descriptions and you have a recipe for busy semesters and slow, plodding progress.

Even though there is a built in efficiency to some of this work (e.g. a single metadata catalog, if built properly, can serve a number of different purposes in several contexts; libraries-trained graduate students can be dispatched to inject our technologies into various scenarios; code and applications for data viewing and delivery can often be recycled, etc.), the strain on individual librarians’ time is considerable. In other words, bolting collaborations and inter- or multi-disciplinary projects onto classic GIS librarian positions means something must give. There will be casualties. Perhaps there is less time for reference, fewer one-shot class drop-ins, fewer workshops, less activity in professional organizations, etc. Something will slide and some consideration must be given to the overhead of collaboration-developing, collaborative proposal-writing, and of course the rich, robust work of actually collaborating. For these same reasons, it remains to be seen just how far this approach can scale. One librarian can only commit to so many projects, and Purdue at least (but surely other institutions) evidently has no dearth of research on campus that needs geospatial support of some kind (and these range from fairly routine geoprocessing to quite experimental spatially-explicit information and data platforms).

## POSTBELLUM

If we buy into the hype surrounding cyberinfrastructure and various \*informatics approaches to solving the data and information challenges posed by the very volumes and scales it enables, our excitement is tempered by those same agencies, who almost always simultaneously caution against factors that could impede this progress. The NSF's "Cyberinfrastructure Vision...", for example, states in no uncertain terms that the country's success in science and engineering "will increasingly depend upon our ability to leverage this reservoir of scientific data captured in digital form, and to transform these data into information and knowledge aided by sophisticated data mining, integration, analysis and visualization tools." [p. 16] Clear-eyed librarians will quickly read into this our position, deep down near the heart of the matter, no less: for us to efficiently and successfully capture, mine, integrate, analyze and visualize these data, they must be cleverly, safely organized and stewarded and made available – the tenets of library science from any approach.

Solving the information problems facing researchers who are generating massive amounts of geodata – or even those generating modest amounts but who still need to intelligently and semantically integrate their data with systems or data implemented in other contexts – is library science. Datasets are heavier, more feral, and require more resources than, say, monograph shipments or e-journal subscriptions, but managing and improving the organization of and access to them is still the obligation of the library and information scientist. Librarian participation in geoinformatics means there will be a library presence in the vanguard of the new science, the new humanities, the new *whatever*. Given the trajectories of both geoinformatics and geospatial librarianship amidst a more general trend in librarianship toward faculty collaboration and e-data stewardship it behooves us to rally around geoinformatics efforts for the benefit of both the movement itself and our own profession. We are finding at Purdue Libraries that while these efforts do, indeed, require aggressive approaches to librarianship and a hard commitment to solving difficult challenges, this war is in fact not hell.

Atkins, D. E. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Arlington, VA: National Science Foundation.

National Science Foundation (U.S.), & Cyberinfrastructure Council (National Science Foundation). (2007). Cyberinfrastructure vision for 21st century discovery. Arlington, VA: National Science Foundation, Cyberinfrastructure Council. <http://purl.access.gpo.gov/GPO/LPS80410>.