Designing the Cyberinfrastructure for Spatial Data Curation, Visualization, and Sharing

Yue Li[1], Nicole Kong[2], Stanislav Pejša[3]

Abstract

Widely used across disciplines such as natural resources, social sciences, public health, humanities, and economics, spatial data is an important component in many studies and has promoted interdisciplinary research development. Though an institutional data repository provides a great solution for data curation, preservation, and sharing, it usually lacks the spatial visualization capability, which limits the use of spatial data to professionals. To increase the impact of research-generated spatial data and truly turn them into digital maps for a broader user base, we have designed and developed the workflow and cyberinfrastructure to extend the current capability of our institutional data repository by visualizing the spatial data on the web. In this project, we added a GIS server to the original institutional data repository cyberinfrastructure, which enables web map services. Then, through a web mapping API, we visualized the spatial data as an interactive web map and embedded in the data repository web page. From the user's perspective, researchers can still identify, cite and reuse the dataset by downloading the data and metadata and the DOI offered by the data repository. General information users can also browse the web maps to find location-based information. In addition, these data was ingested into the spatial data portal to increase the discoverability for spatial information users. Initial usage statistics suggest that this cyberinfrastructure has greatly improved the spatial data usage and extended the institutional data repository to facilitate spatial data sharing.

Keywords

Spatial information, data repository, visualization, cyberinfrastructure, GIS

Introduction

Although spatial information is widely used in many disciplines, it is usually saved in very specific formats that are not familiar to most of the researchers and general information users. To view the spatial data files, users need to have some background knowledge and skills about GIS software, which might not even be free. This software requirement hindered the wide adoption of spatial information. To overcome this barrier, the Association of Research Libraries (ARL) has set out the GIS literacy project to educate and equip librarians with the GIS skills necessary to provide access to spatial data in all formats since 1992 (Association of Research Libraries 1999). As a result, the value-added spatial reference services have benefited a range of users from GIS experts in special libraries to casual users in public libraries (Gluck & Yu 1999; Weimer & Reehling 2006). In recent years, the emerging technology in web GIS has shown promise for general information users to access spatial data via web maps (Kong et al. 2014; Batty et al. 2010). Web maps provides an easy and direct way for any web users to browse spatial information without any software license restriction or learning curve. Many spatial data providers, such as US Census Bureau and the U.S. Geological Survey (USGS), started to take advantage of the web mapping technology and provide online spatial data portals for users to view and download information.

In academic settings, research-generated spatial data is experiencing an exponential growth due to the availability of new sensor technologies and the broader adoption of spatial thinking skills across disciplines (Gregory et al. 2015; Matei et al. 2007; Kong 2015). Although institutional data repositories offer a great platform for data management, curation and sharing, they usually lack the web-based visualization capabilities for spatial data sharing. Adding the spatial data visualization function to institutional data repository can benefit any information users by providing an online interactive map, so that users can view the research generated spatial data in their browser in the same way as they can in any other online map tools, such as Google Maps. For GIS professionals, the visualization function can also help them to judge if the data is suitable for their research before downloading.

In this article, we introduce the project that we have developed to streamline the spatial data curation, visualization and sharing by connecting our institutional research data repository with the library's GIS server set and spatial data portal. Based on our campus users' needs, we have prototyped the theoretical workflow to curate, visualize and share the research-generated spatial data. With the prototyped workflow, we designed, optimized, and adjusted our existing cyberinfrastructures to fulfill the requirements. In our design, the research data is curated using the institutional repository, visualized using the library's GIS server, and then the generated web map is embedded in the data publication webpage via web mapping API (Application Programming Interface). To promote the data usage, we also share the data via our spatial data portal in addition to the institutional data by ingesting the metadata and link the published dataset from the portal. Additionally, we developed strategies to track the data usage statistics for both web map view and data download. We would demonstrate our first successful test case of spatial data publication, using this improved cyberinfrastructure and workflow. The initial statistics show that our cyberinfrastructure design has greatly improved spatial data usage and extended the institutional data repository to facilitate spatial data sharing.

Background

Spatial data and GIS technology development

Spatial data is data that is associated with a place on the earth's surface, implicitly or explicitly (ISO/TC 211 2009). It includes not only digital maps in various themes, but also any tabular data with location information such as census datasets, housing, marketing, and social media data. It provides a spatial dimension to simplify complex and dense information, and allows for the review of patterns, relationships and trends which may not be obvious in statistical and non-spatial datasets (Knowles & Hillier 2008; Ridge et al. 2012; Galbraith & Coonin 2001; Goodchild 2000). Spatial data serves as an important component in various disciplines as well as promotes interdisciplinary research development (Kong et al. 2016). In science, technology, engineering, and mathematics (STEM) disciplines such as environmental science, agriculture and civil engineering, spatial data has been widely used since the 1980's with the emergence of the Geographic Information Systems (GIS) software (Diamond & Wright 1988; Goodchild et al. 1993; Corwin & Wagenet 1995). In the humanities and social sciences, it is only in recent years with the "spatial turn" that researchers began to realize the power of spatial data and started to learn and use this type of data for their research (Knowles

2000; Warf & Arias 2008; Bodenhamer et al. 2010). As spatial data have such a broad user population across disciplines and technology levels, it is essential to design effective and efficient infrastructure and systems to curate, preserve, and share research-generated spatial data so that different user groups can benefit from it.

Due to its nature of recording the relationships between geographic features and their attribute information, spatial data is often saved in a more complex format comparing to other regular numeric or text files. The typical spatial data formats include shapefile, coverage, geodatabase or raster dataset, which require specific GIS software, such as ArcGIS developed by the Environmental Scientific Research Institute (ESRI). Although some open source GIS software exist, they were designed for professional GIS users and developers who already developed the understanding about geoinformatics.

With the development of web technology in the early 21st century, a service-oriented architecture and web-based system was introduced to GIS technology, extending the task oriented desktop-based system. Unlike desktop GIS, which is intended for professional users with training and experience in GIS, Web GIS is intended for a broader audience, including people without any knowledge about GIS (Kidd 2010; Batty et al. 2010). It is designed to be simple, intuitive, and convenient to use. Visualizing data with web GIS technology is a streamlined process which can be completed within a few clicks. Google and Microsoft launched online global basemap imagery, enabling people to access free geospatial information products from home and mobile devices. This exposure also lead to consumers' increasing familiarity with information in map form. In 2004, ESRI released ArcGIS server, a software that makes geographic information available to anyone with an internet connection. Government agencies, universities, and organizations began to build GIS infrastructure with web technology to share their data, maps and services online.

Institutional repository

Over the last decade there has been a push in the research communities across the disciplines to adjust and modify the foundation of scholarly communication, which often manifests itself as calls for a more open and transparent science. Reproducibility, replicability, and verifiability are increasingly being seen as a fundamental attributes of good science and responsible research (Casadevall & Fang 2010; Doorn et al. 2013). Access to the data that underlie the scholarly journal publications is an unavoidable prerequisite of transparent science. Some research areas have access to dedicated data repositories around which research communities gather and develop data archival practices, data publishing standards, and tools dedicated to data visualization and analysis. These repositories, however, are relatively rare. Many researchers work in disciplines that are lacking such an infrastructure and are facing difficult choices when project funders require a mandatory data management plan (DMP) that contains a proviso for access and preservation of data. This provides an opportunity for establishing institutional data repositories, such as The Purdue University Research Repository (PURR, https://purr.purdue.edu/).

PURR allows for data deposit for all Purdue faculty, staff, students and their collaborators. Purdue researchers can use PURR, to share and manage their research data with their collaborators in a secured environment, to publish and disseminate datasets with DOIs, to track and measure the impact of sharing their data, and to archive their data for long-term preservation and reuse. The PURR system has been developed based on HUBzero, an open source software platform for building powerful web sites that support scientific discovery, learning, and collaboration (McLennan & Kennell 2010). PURR is a combination of a virtual research environment (VRE), content management system (built on Joomla), and a data publication platform. PURR provides secure storage for research data and a safe space for online collaboration. The results of such collaboration can be published as datasets and shared with the research community at large. Researchers can publish through PURR not only their primary data in a variety of types and formats, such as observational and experimental data, software code, audio and video files, or GIS files, but also the necessary documentation, such as the research procedures, code books, technical drawings.

The process of data publication and archiving is straightforward in PURR (Figure 1). It is a team effort between the researchers, the PURR administrators, and the library subject liaisons. It starts with the researchers, who select and upload the data files and provide key metadata, including the dataset authors, description, and the license under which dataset can be used and shared. Then the dataset is submitted for review by the subject liaisons or the PURR administrator. The liaisons assess the technical aspects of the dataset, primarily whether all the necessary domain-specific scholarly attributes are present. Once the dataset is approved, it is published and the PURR repository assumes custody of the dataset. All published datasets are provided with unique digital object identifiers (DOI) so that they can be easily identified, cited and reused. After a 30-day grace period, an archival information package (AIP) is created and the archival package is deposited in a dark archive provided through the MataArchive consortium for a long-term preservation. The AIP contains not only the descriptive metadata, but also technical and preservation metadata that are essential for a successful preservation and long-term access to the archived data.

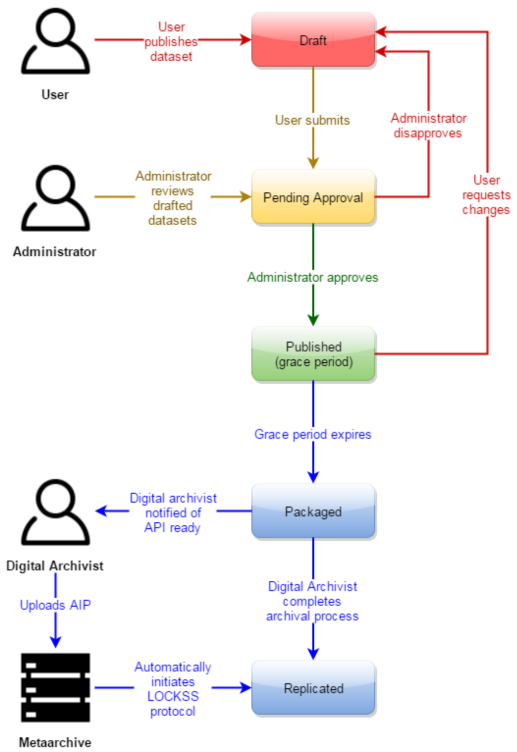

Figure 1. PURR data publication and data archiving workflow

The usage statistics of the published dataset in PURR, i.e. the number of views and downloads, are tracked and visible to the users of PURR. Figure 2 shows an example of a published dataset in PURR. PURR allows for version tracking. The individual versions are linked and all versions of a dataset can be accessed in the 'Versions' section of the published dataset webpage. Researchers are also able to post questions to the authors of the datasets or leave comments in the 'Reviews' section of the PURR publication.

Figure 2. Example of a published dataset on PURR. It allows users to read the metadata information and supporting document about the dataset, download the data, cite the data, as well as view the versions and usage statistics, leave review comments, and ask questions.

PURR was initially designed to provide generic data publication and archival functions for data in all disciplines. Different from many other discipline-specific data repositories which provide data visualization and analysis functions, PURR lacked customizations for special datasets, including the spatial data discussed in this article. PURR's development team was working with Hubzero team over the last few years to add more specialized data functions into PURR, such as visualizing tabular datasets and images. However, spatial data visualization was not available in PURR because it requires additional hardware and software support, which was relatively more challenging compared to some other common file formats.

GIS cyberinfrastructure review and our existing status

As geospatial information has increased exponentially in recent years, many academic libraries have started to build their own GIS cyberinfrastructure to support the user communities (Lage 2007; Kollen et al. 2013). These efforts include both setting up their own GIS servers to support web-based research data publications, and developing or deploying spatial data portals to facilitate spatial data discovery (Stowell Bracke et al. 2008; Carlson et al. 2011; Florance et al. 2015). The GIS servers can be used to create and manage web map services and web mapping applications. In an academic library setting, this server environment facilitates mapping of research-generated spatial data online.

The spatial data portals are web applications that facilitate discovery, preview and retrieval of spatial data. At Purdue University Libraries, the GIS team has made efforts in both directions in setting up its GIS cyberinfrastructure. This in-house infrastructure provides powerful, reliable, and customizable services for storing, organizing and sharing GIS assets (Figure3).

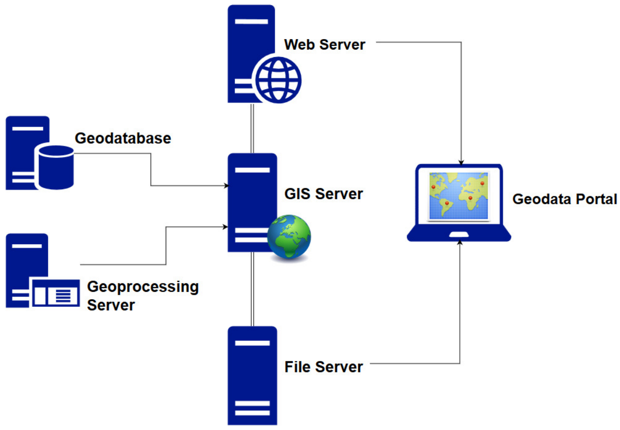

Figure 3. Purdue University Libraries' GIS cyberinfrastructure includes GIS server set for geodatabase, geoporcessing, web mapping, and file storage. This server set is also connected to Purdue geodata portal for spatial data discovery.

The GIS server cluster, which includes an ArcGIS Server, an enterprise geodatabase and two geoprocessing servers, is essential to creating and managing map services, applications and data. The ArcGIS server is used for publishing web map services and hosting web applications, which helps research data gain a broader impact by sharing the interactive maps online. It also acts as a file server to accommodate file download requests for offline maps and other spatial data sharing purposes. Using ArcGIS and Microsoft SQL technology, the geodatabase supports multi-user editing, security management and version controls in addition to other functions. The geodatabase and ArcGIS Server are used to manage data from Purdue Library map collections and Purdue research generated spatial datasets. In addition, two state-of-art high-performance compute nodes are dedicated to running computationally intensive geoprocessing jobs, which significantly decreases the processing time for users who have high computing demands.

Spatial data portals usually provide a web map interface which allows users to search for spatial data by both map location and keyword. Connecting the curated spatial data from an institutional repository to the spatial data portal can help information users to develop a better understanding about the background of the dataset as well as correctly cite the data source. It can also increase the discoverability of the research dataset from the institutional data repository.

To develop our institutional spatial data portal, we deployed the Open Geoportal project framework. In addition to connecting the metadata contributed from collaborative universities in the Open Geoportal project, we also connected our Open Geoportal instance to metadata shared by our state agencies and our library's own spatial data, including the map collection and research data. Although this geoportal server is a separate server from the GIS server cluster, they are connected for through the metadata when the geoportal server needs map preview and download functions from Purdue spatial data. This infrastructure is upgraded as needed and has been kept up-to-date since created over four years ago. It has supported various spatial data research projects across disciplines.

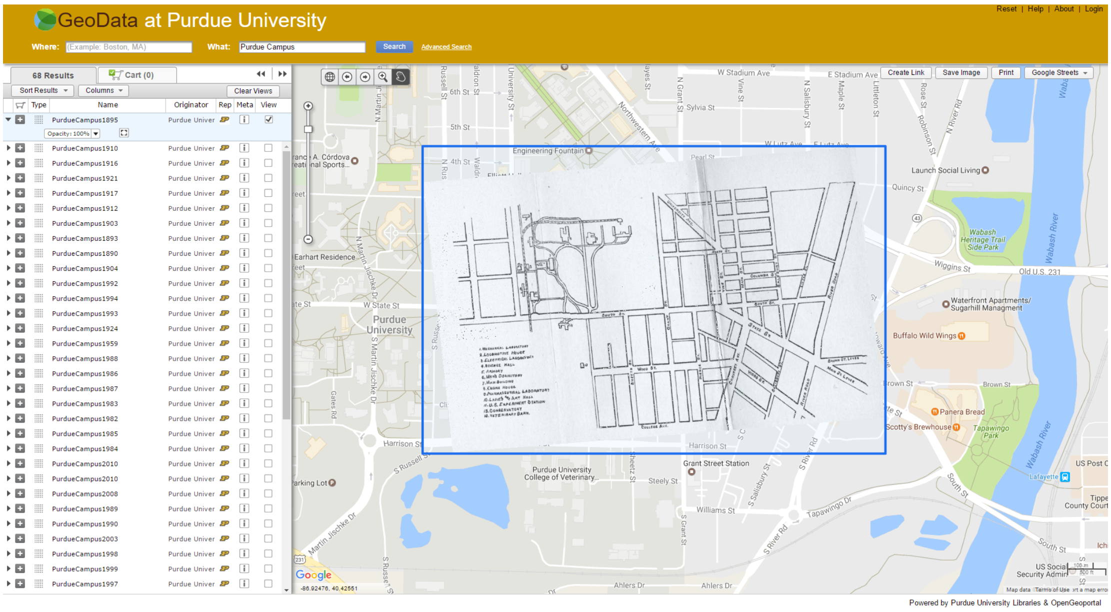

Figure 4. Geodata portal at Purdue supports spatial data discovery, preview and download function for data contributed by collaborative universities in Open Geoportal project, state agencies, and Purdue curated spatial data.

Additional spatial data service needs

Though the PURR and GIS cyberinfrastructure provided excellent tools for researchers to curate their research data and share online maps, we still have received additional service requests to meet researchers' expectations. One particular service area was to create customized spatial data sharing pages, so that the particular data user group could easily preview, download, cite and read more background information about the spatial dataset. Although PURR provided a convenient way for researchers to curate and share their data, it did not offer a map preview function for potential users to view the map without downloading it. Thus, requiring users to download and view GIS data with professional software, it limited the data sharing to professional GIS users. The geodata portal provided data preview and download functions, but it did not have a specific web link for a particular dataset. The researchers had to describe how to find their dataset in the portal to their potential data users. The GIS server cluster offered the opportunity possibility to meet the researchers' expectations, but we had to make an extra effort to create a customized webpage for the particular dataset.

One example of this kind of request is from a researcher whose project was about floodplain mapping using soil information (Sangwan & Merwade 2015). The research lab had developed an innovative approach to generating floodplain maps for the state. Although most of the project is conducted using GIS, their findings (i.e. floodplain map) can benefit a larger group of the population beyond GIS professionals, including homebuyers, property insurance providers, emergency management agencies, urban planners, etc. It is the research lab expects to widely share their research findings as online interactive maps so that users can zoom in to their area of interest to explore more detailed information about the possibility of floods. In addition, sharing the data as a set of GIS files can help GIS professionals reuse their findings in other related studies.

To accommodate this researcher's needs, we first investigated the existing servers and created a customized webpage to help disseminate the dataset (Figure 5). The webpage was developed using the ArcGIS server web map service and JavaScript API. This page has greatly helped the researcher to communicate his research findings with his potential user groups and funding agencies. The usage statistics have shown that the webpage has received almost 5,000 visits since its launch to the time that we developed this new project to enhance its functionality in 2016 (Figure 6). Although the initial dataset shared via our webpage is just for Indiana, it attracts visitors from many other states in U.S. and countries all over the world.

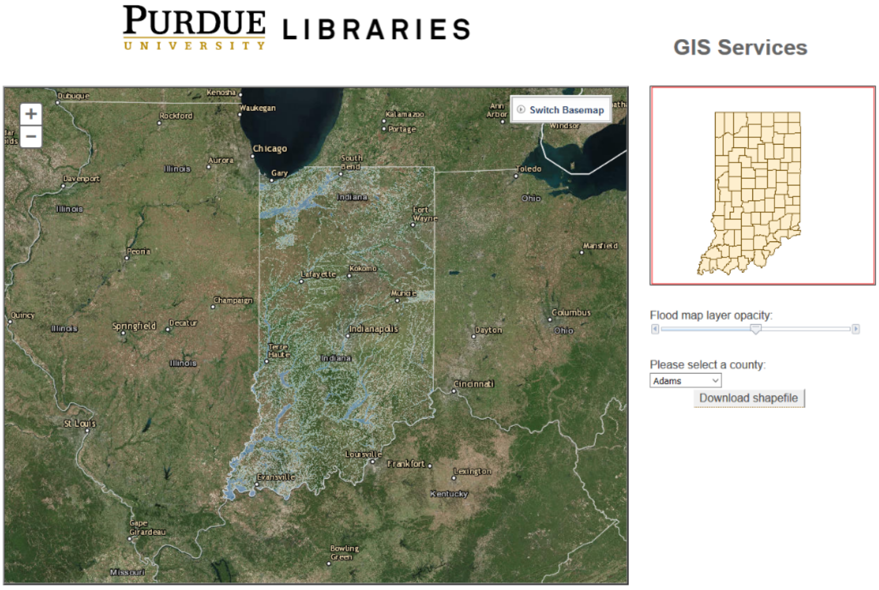

Figure 5. Customized webpage to disseminate the flood map information as both online interactive maps and downloadable GIS files.

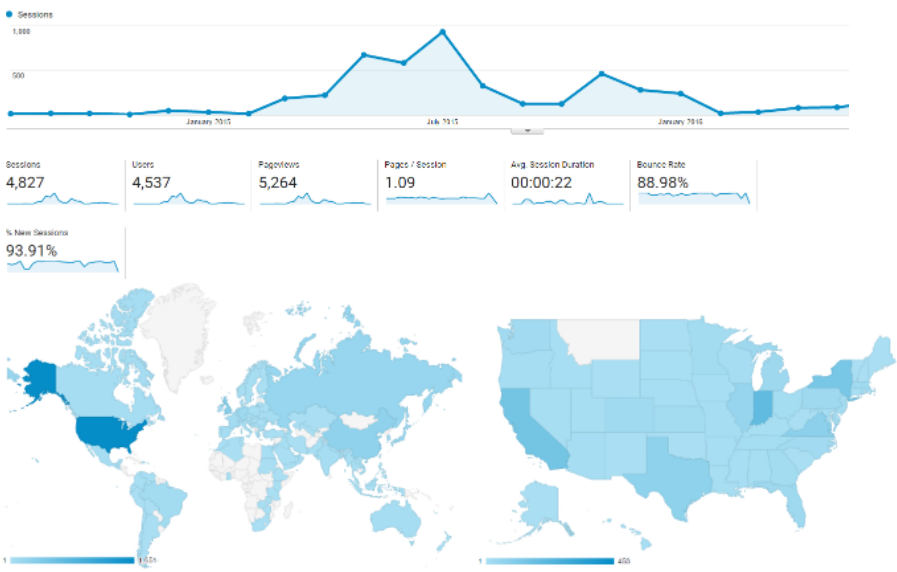

Figure 6. Usage statistics of the customized flood map data sharing webpage.

With this initial success, we looked for solutions to expand this data sharing solution to our institutional data repository (PURR) so that the dataset could be easily cited. Another advantage of connecting PURR with this data sharing solution is that PURR offers data downloading option. This will free up GIS server space for file hosting. In addition, PURR provides an archival solution for the dataset, so that the information can be safely preserved for a long period without the hassle of computer software/hardware upgrade.

Results and Conclusion

Project design - theoretical model

With an initial assessment of our existing cyberinfrastructure for PURR and the GIS servers, we have drawn the conclusion that the most feasible solution to connect these two systems is to extend PURR with spatial visualization capability supported from the GIS servers. Two factors helped us to make the decision. First, PURR has already been used by many researchers at our institution. By the time this article was written, there were more than 3,500 registered researchers on PURR. Second, PURR provides a great platform to manage users, and a user-friendly interface to curate and publish datasets. In our theoretical model, spatial data users can start the data publication process by uploading their datasets in PURR, and publish it with a DOI. Once they enable the spatial data visualization functionality, map services can be created based on their datasets using our GIS server set. The map service makes it possible to visualize the spatial data on the web using the JavaScript API. Then, the visualization will be integrated into PURR's data publication webpage to offer a visually-appealing and interactive preview. The spatial visualization enables users to explore the features on various scales, and zoom into an area of interest. The map visualization allows for user interactions beyond a textual and static description, and provides users with a convenient way to access more in-depth information embedded within the dataset.

Furthermore, to increase discoverability beyond the institutional data repository, we expect to ingest the spatial data into our geodata portal. Thus, general spatial information users can find the data and preview the map and its metadata from the portal. If they are interested in downloading the data, users will be redirected to the dataset webpage in PURR. By combining the data sharing and preservation platform of the institutional data repository with a spatial visualization from the GIS cyberinfrastructure, this efficient and comprehensive model solves the challenges in spatial data management, increases the discoverability and meets the needs of researchers and collaborators (Figure 7).

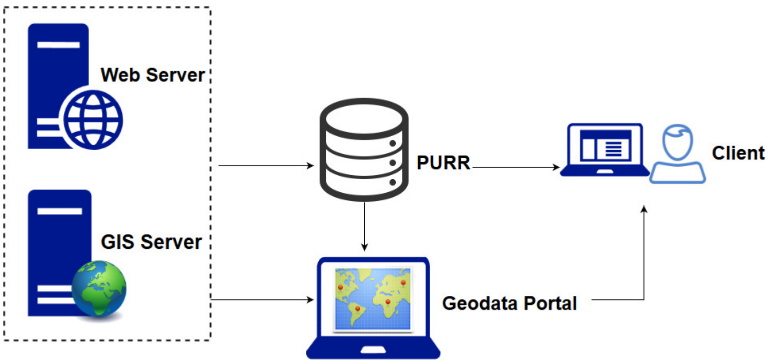

Figure 7. Theoretical model of connecting institutional data repository (PURR) with the GIS cyberinfrastructure.

Workflow

With the theoretical model, we further list the detailed spatial data curation workflows shown in Figure 8. First, the spatial dataset is submitted and published in PURR and given a DOI. To initiate this step, a project is created and shared with all collaborators by the researcher. During this process, project description, funding source, and other information need to be collected. After the storage space is allocated, the dataset can be uploaded and submitted for review by the data management specialist and subject librarian using the PURR publication wizard. Upon approval, this dataset is published and a DOI is assigned, so that the dataset is openly accessible for use and discovery. The dataset is now preserved and archived in PURR, and can be cited.

While the dataset is curated in PURR, it is also copied into our geodatabase and published as a map service using the GIS server. A web map application is created with the map service overlain on a basemap. Users can pan and zoom in or out to their area of interest. An overview map is included for users to have a broader picture, and users can choose a preferred basemap using the basemap switch. After that, the web map application is embedded in the project description in PURR to make the data previewable and interactive, along with a brief introduction on how to use the map and download the data. With this comprehensive insight into the dataset, users can download the data from the repository and cite it with DOI.

Finally, to increase usability spatial data is ingested into our geodata portal. Metadata is created using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM), including the link to the preview layer from the GIS server, and the link to the PURR page for the dataset. Upon users' download request, they will be directed to the dataset's webpage in PURR.

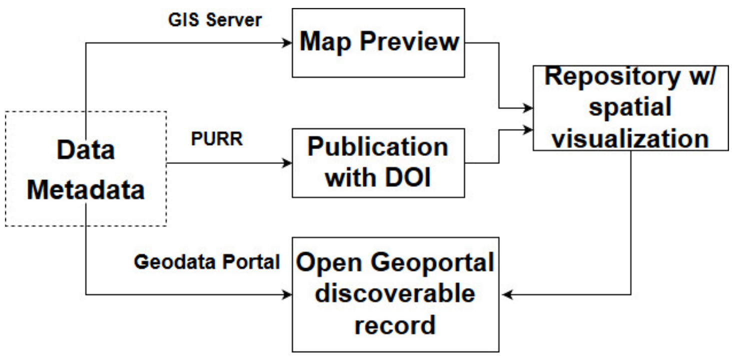

Figure 8. Workflow for spatial data publication.

Deployment

In the deployment process, we tried to utilize the existing functions in both PURR and the GIS server set and minimize our development efforts. We took the floodplain map for United States as the first sample dataset for this project. For the first step in the workflow, PURR already provides excellent functionality for data submission and publication. Therefore, there was no further development requirement in this step. However, since many spatial datasets are large, researchers usually organize them into smaller files by region, which created challenges in PURR to organize these individual files into one data collection. For example, when we worked with the floodplain map data for the United States, the researcher divided the map into 50 smaller state-level datasets, because the smaller files make the data uploading and downloading process much easier. In this case, we recommended they publish each small file separately with an independent DOI. Then we used PURR's data series function to list all these datasets as one data series (Figure 9).

To connect the dataset from PURR with our GIS servers, we have to manually check the quality of the dataset to verify if it fits the map service publication requirement, then initiate the map service publication process. To connect both systems, we developed a JavaScript-based web application template to visualize the published map service. In this template, the map service information has to be updated based on the new dataset that is published. Then, this web map application (or the JavaScript code) was updated in PURR so that the interactive map could be embedded into the dataset's webpage in PURR. To be compatible with the PURR system, the map service and web map application were both secured with HTTPS connections on our GIS server.

At the final step, we generated FGDC standard metadata based on the generic metadata information obtained from PURR, and updated the data preview link according to the map services created on our GIS server. This standard metadata was ingested into our geodata portal's indexed searching database. The metadata ingestion is an existing function for our geodata portal. In order to make our geodata portal recognize the dataset contributed from PURR, we modified the data download link in our portal to redirect the data users to PURR's data publication webpage.

With minimum development efforts, including a JavaScript web mapping application template and the modification or geodata portal download link, we were able to connect PURR with the GIS cyberinfrastructure and extended our institutional repository by adding the spatial data visualization functionality.

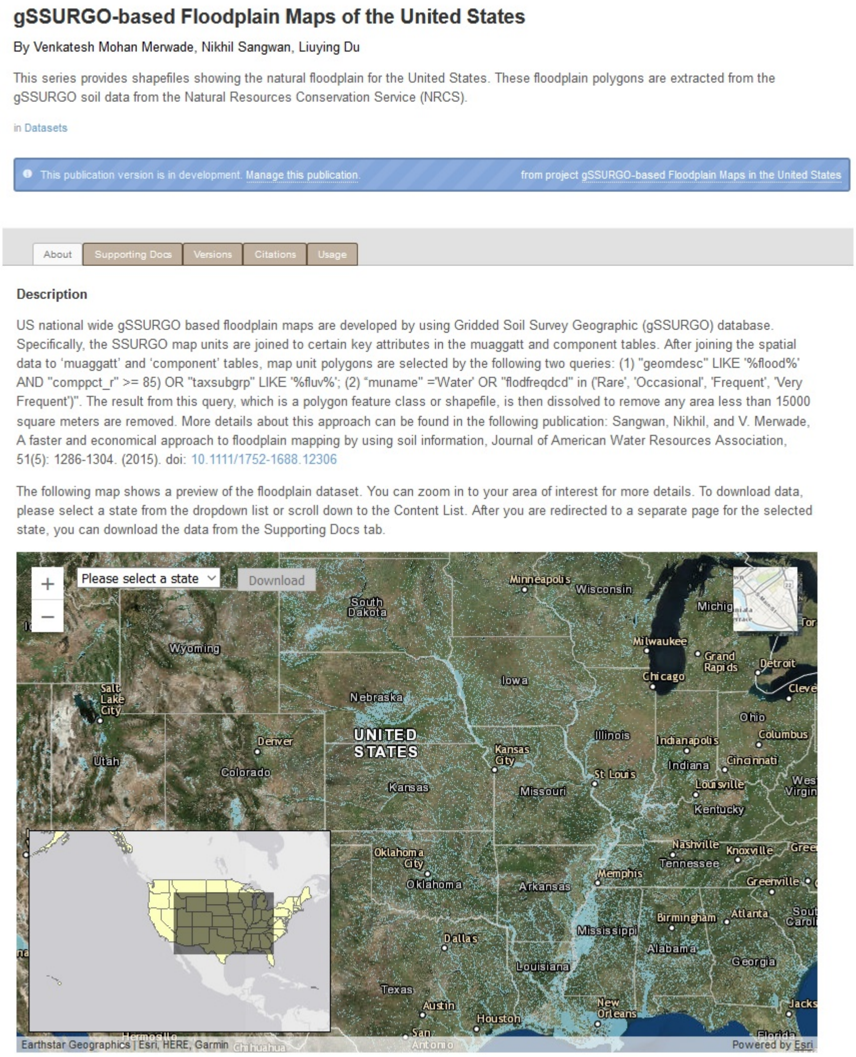

Figure 9. The spatial data series publication webpage with an interactive map.

Initial assessment

The datasets have been published in PURR for six months. Based on the visit statistics during this limited period, there were 5,683 total visits on the published dataset's webpages and 842 total downloads. The spatial visualization integration has been tested for about one month at the time this article was written. The online interactive map has received more than 500 visits even though it has been available for a very short time period. The statistics suggest that the datasets and the spatial visualization have received great attention by information users. There are no statistics at this time to track users' activities on this webpage, such as the interactions they had with the online map before they decide to either download the data or leave the webpage. We believe that the spatial visualization can help users to better understand the information contained in the dataset before they decide to either download the data or be satisfied with the information provided on the map. We will develop more logging scripts to capture user activities on the webpage to help us further understand users' needs.

Discussion

By connecting the institutional data repository to the GIS server, we were able to expand the spatial data sharing capability to serve a broader information user group. The spatial data visualization feature can help any web users to view spatial information related to their interested location without GIS software or skills. Thus, it helps to communicate the research findings to a broader audience. From a technical perspective, our efforts have suggested that such an expansion is feasible with limited development needs. Many academic libraries already have institutional data repositories and GIS server environments, or are in the process of creating that cyberinfrastructure. Our project can serve as an example to demonstrate the benefit of connecting these two sets of servers to create a better option for sharing spatial data.

There are still three manual steps in our workflow to share the spatial data with an interactive map visualization. These three manual steps include creating map services on the GIS server, updating the PURR data sharing webpage with the web map application, and updating the FGDC metadata with PURR publication information. We expect to automate or semi-automate these steps in the future if more resources are available. Once this happens, researchers will experience a more streamlined process to upload and share their spatial data using PURR interface.

A good data sharing practice not only needs an excellent cyberinfrastructure as the platform to support the data sharing mechanisms, but also requires researchers to fully recognize the value of data sharing, and to build up good practices around the data sharing process. In our practice, we have learned that there are several key points that the researchers and their graduate students need to pay special attention before publishing and sharing their spatial data. First, they need to develop good data management practices in order to prepare the data for publication. This includes critical evaluation of their datasets, and creation of human readable documentation in order to enable others understand and use their datasets, etc. In our project, we have worked closely with the researcher's team to learn about the dataset and created appropriate metadata for the data repository and spatial data portal. Second, they need to understand the data sharing license options, data ownership, and data publication options in order to share the data as they desire or as required by their funding agencies. Although we have prepared a good environment in this project to facilitate spatial data sharing, more education programs need to be developed in order to improve the data sharing practices.

References

Association of Research Libraries, 1999. *The ARL Geographic Information Systems Literacy Project*, Washington, DC. Available at: http://www.istl.org/06-fall/refereed.html [Accessed May 1, 2015].

Batty, M. et al., 2010. Map mashups, Web 2.0 and the GIS revolution. *Annals of GIS*, 16(1), pp.1–13. Available at: http://www.tandfonline.com/doi/abs/10.1080/19475681003700831 [Accessed September 3, 2014].

Bodenhamer, D.J., Corrigan, J. & Harris, T.M. eds., 2010. *Spatial Humanities: GIS and the Future of Humanities Scholarship*, Indiana University Press. Available at: http://site.ebrary.com/lib/purdue/reader.action?docID=10767195&ppg=20 [Accessed May 4, 2015].

Carlson, J. et al., 2011. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *portal: Libraries and the Academy*, 11(2), pp.629–657. Available at: http://muse.jhu.edu/content/crossref/journals/portal_libraries_and_the_academy/v011/11.2.

carlson.html [Accessed November 6, 2014].
Casadevall, A. & Fang, F.C., 2010. Reproducible science. *Infection and immunity*, 78(12), pp.4972–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20876290 [Accessed March 22, 2017].
Corwin, D.L. & Wagenet, R.J., 1995. Applications of GIS to the Modeling of NonPoint Source Pollutants in the Vadose Zone: A Conference Overview. *Journal of Environmental Quality*, 25(3), pp.403–411.
Diamond, J.T. & Wright, J.R., 1988. Design of an Integrated Spatial Information System for Multiobjective Land-Use Planning. *Environment and Planning B*, 15(2), pp.205–214.
Doorn, P., Dillo, I. & van Horik, R., 2013. Lies, Damned Lies and Research Data: Can Data Sharing Prevent Data Fraud? *International Journal of Digital Curation*, 8(1), pp.229–243. Available at: http://www.ijdc.net/index.php/ijdc/article/view/8.1.229%5Cnhttp://www.ijdc.net/index.php/ijdc/article/view/8.1.229/308.
Florance, P. et al., 2015. The Open Geoportal Federation. *Journal of Map & Geography Libraries*, 11(3), pp.376–394.
Galbraith, J. & Coonin, B., 2001. GIS in business: Building a core collection for business geographics. *Reference & User Services Quarterly*, 41(1), pp.9–17.
Gluck, M. & Yu, L., 1999. geographic information systems: Background, Frameworks and Uses in Library. In E. A. Chapman, ed. *Advances in Librarianship*. Emerald Group Publishing Limited, pp. 1–38.
Goodchild, M. et al., 1993. *Environmental modeling with GIS,* null, ed.,
Goodchild, M.F., 2000. GIS and Transportation: Status and Challenges. *GeoInformatica*, 4(2), pp.127–139.
Gregory, I. et al., 2015. Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research. *International Journal of Humanities and Arts Computing*, 9(1), pp.1–14. Available at: http://www.euppublishing.com/doi/abs/10.3366/ijhac.2015.0135?journalCode=ijhac [Accessed March 19, 2015].
ISO/TC 211, 2009. *Report from stage 0 Project 19150 Geographic information - Ontology*, Available at: http://metadata-standards.org/Document-library/Documents-by-number/WG2-N1251-N1300/WG2_N1292_TC211n2705_Report_19150.pdf.
Kidd, J.C., 2010. *Web-based mapping: an evaluation of free mapping applications and web GIS for library reference services*. University of North Carolina at Chapel Hill. Available at: https://cdr.lib.unc.edu/indexablecontent?id=uuid:f64751da-cb95-47aa-8d4f-3ec5338420af&ds=DATA_FILE.
Knowles, A.K., 2000. Special Issue: Historical GIS: The spatial turn in social science history. *Social Science History*, 24(3), pp.451–470.
Knowles, A.K. & Hillier, A., 2008. *Placing History: How Maps, Spatial Data, and GIS are Changing Historical Scholarship*, ESRI, Inc. Available at: https://books.google.com/books?hl=en&lr=&id=VN1v7rzhSQEC&pgis=1 [Accessed April 12, 2016].
Kollen, C. et al., 2013. Spatial data catalogs.pdf. *Journal of Map & Geography Libraries*, 9(3), pp.276–295.
Kong, N. et al., 2016. An Interdisciplinary Approach for a Water Sustainability Study. *Papers in Applied Geography*, 2(2), pp.189–200. Available at: http://www.tandfonline.com/doi/full/10.1080/23754931.2015.1116106.
Kong, N., 2015. Exploring Best Management Practices for Geospatial Data in Academic Libraries. *Journal of Map & Geography Libraries*, 11(September), pp.207–225.
Kong, N., Zhang, T. & Stonebraker, I., 2014. Common metrics for web-based mapping applications in academic libraries. *Online Information Review*, 38(7), pp.918–935. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/OIR-06-2014-0140 [Accessed January 18, 2015].

Lage, K., 2007. Cataloging Digital Geospatial Data. *Journal of Map & Geography Libraries*, 3(1), pp.39–55.

Matei, S.A. et al., 2007. Visible Past: Learning and discovering in real and virtual space and time. *First Monday*, 12(5). Available at: http://ojphi.org/ojs/index.php/fm/article/view/1836/1720 [Accessed March 19, 2015].

McLennan, M. & Kennell, R., 2010. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. *Computing in Science & Engineering*, 12(2), pp.48–53.

Ridge, O., Stefanidis, A. & Ber, D.O.E., 2012. Spatiotemporal Data Mining in the Era of Big Spatial Data : , pp.1–10.

Sangwan, N. & Merwade, V., 2015. A faster and economical approach to floodplain mapping by using soil information. *Journal of American Water Resources Association*, 51(5), pp.1286–1304.

Stowell Bracke, M., Miller, C.C. & Kim, J., 2008. Adding value to digitizing with GIS. *Library Hi Tech*, 26(2), pp.201–212. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/07378830810880315 [Accessed December 28, 2014].

Warf, B. & Arias, S., 2008. *The Spatial Turn: Interdisciplinary Perspectives* B. Warf & S. Arias, eds., New York, NY: Routledge. Available at: http://reader.eblib.com.ezproxy.lib.purdue.edu/(S(atxzqvrkm1hx0jd2yxokucki))/Reader.aspx?p=356413&o=58&u=JWvGAIMz8bM%3D&t=1430763856&h=1D90639E400C6AB96D982F7DF7281EF697C95DF2&s=36065559&ut=130&pg=1&r=img&c=-1&pat=n&cms=-1&sd=2# [Accessed May 4, 2015].

Weimer, H. & Reehling, P., 2006. A New Model of Geographic Information Librarianship : Description, Curriculum and Program Proposal. *Journal of Education for Library and Information Science*, 47(4), pp.291–302.

End-notes

[1] Yue Li is a GIS analyst at Purdue University Libraries.

[2] Nicole Kong is an assistant professor and GIS specialist at Purdue University Libraries. She can be reached by email: kongn@purdue.edu

[3] Stanislav Pejša is a data curator at Purdue University Libraries.