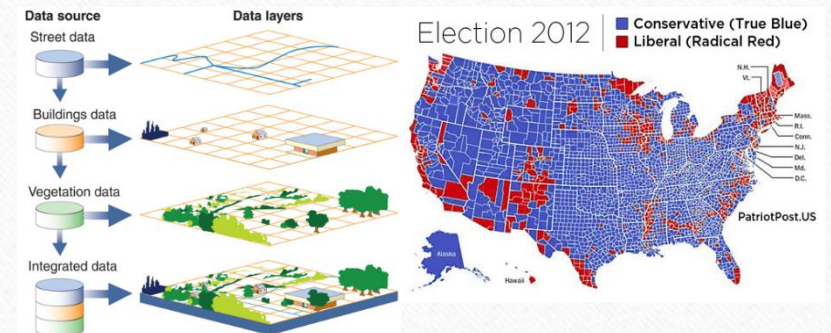


Designing the Cyberinfrastructure for Spatial Data Curation, Visualization and Sharing

Yue “Shirley” Li, Nicole Kong, Stanislav Pejso
Purdue University Libraries

Introduction

- Spatial data is the data that associates with a place on the earth's surface implicitly or explicitly
- Spatial data's specific formats limit its use to GIS professionals
- In academic settings, research generated spatial data is experiencing an exponential growth
- Many universities are developing their spatial data portals to facilitate spatial information discovery and access



Institutional Repository

- Purdue University Research Repository (PURR)
 - Create data management plan
 - Share and manage research data
 - Publish and disseminate
 - Archive

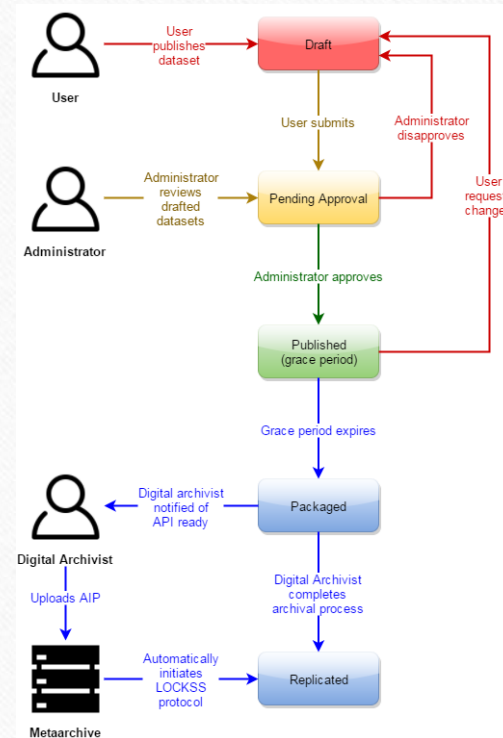


Figure 3. PURR data publication and data archiving workflow

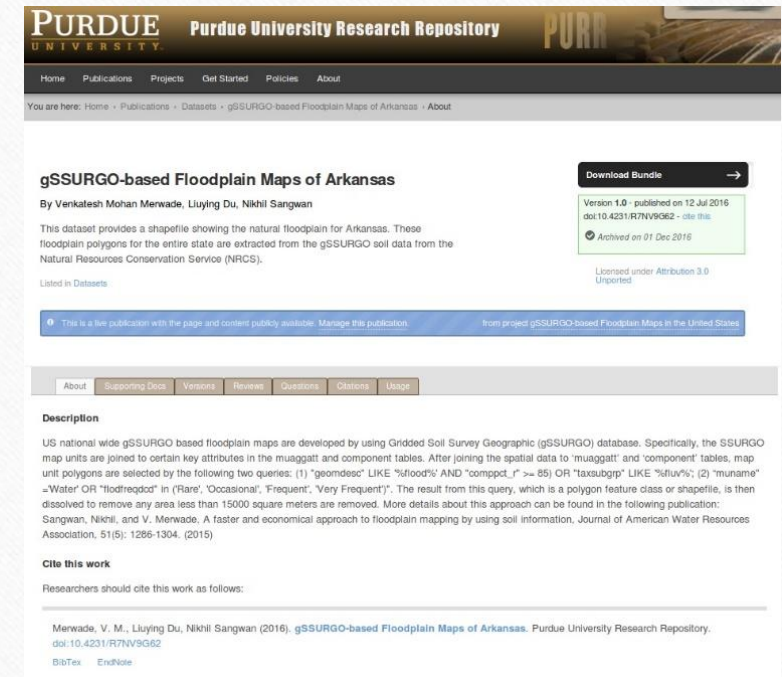


Figure 4. Example of a published dataset on PURR. It allows users to read the metadata information and supporting document about the dataset, download the data, cite the data, as well as view the versions and usage statistics, leave review comments, and ask questions.

GIS Infrastructure

- Provide powerful, reliable, and customizable services for storing, organizing and sharing GIS assets
- Support web-based research data publications
- [Geodata Portal](#) to facilitate spatial data discovery

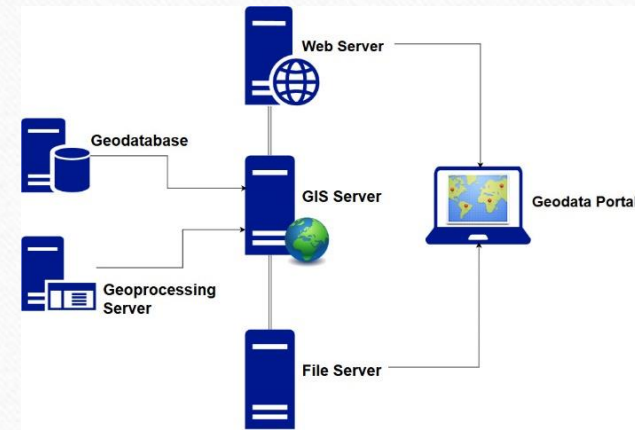


Figure 5. Purdue University Libraries' GIS cyberinfrastructure includes GIS server set for geodatabase, geoprocessing, web mapping, and file storage. This server set is also connected to Purdue geodata portal for spatial data discovery

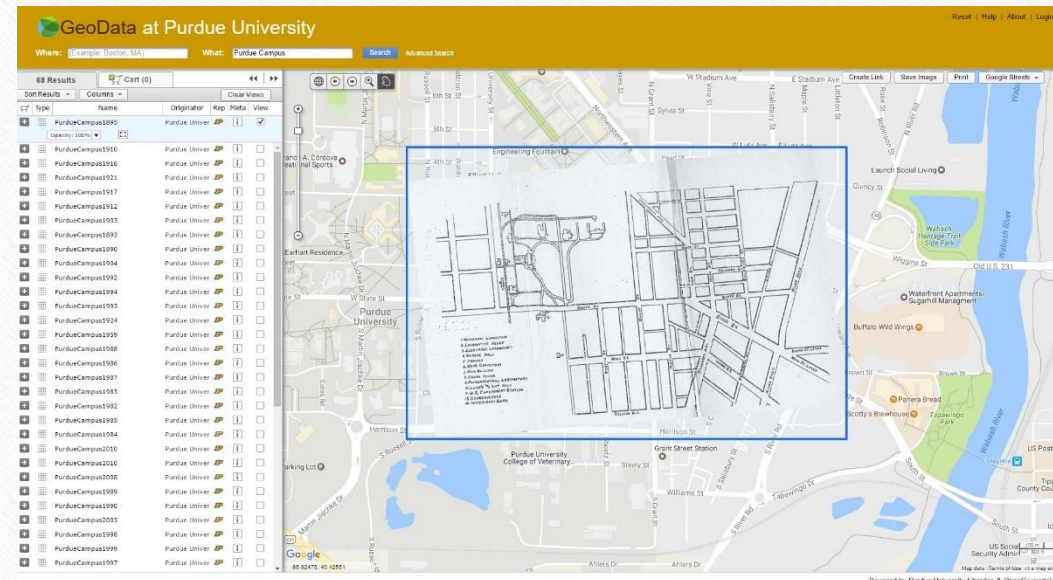
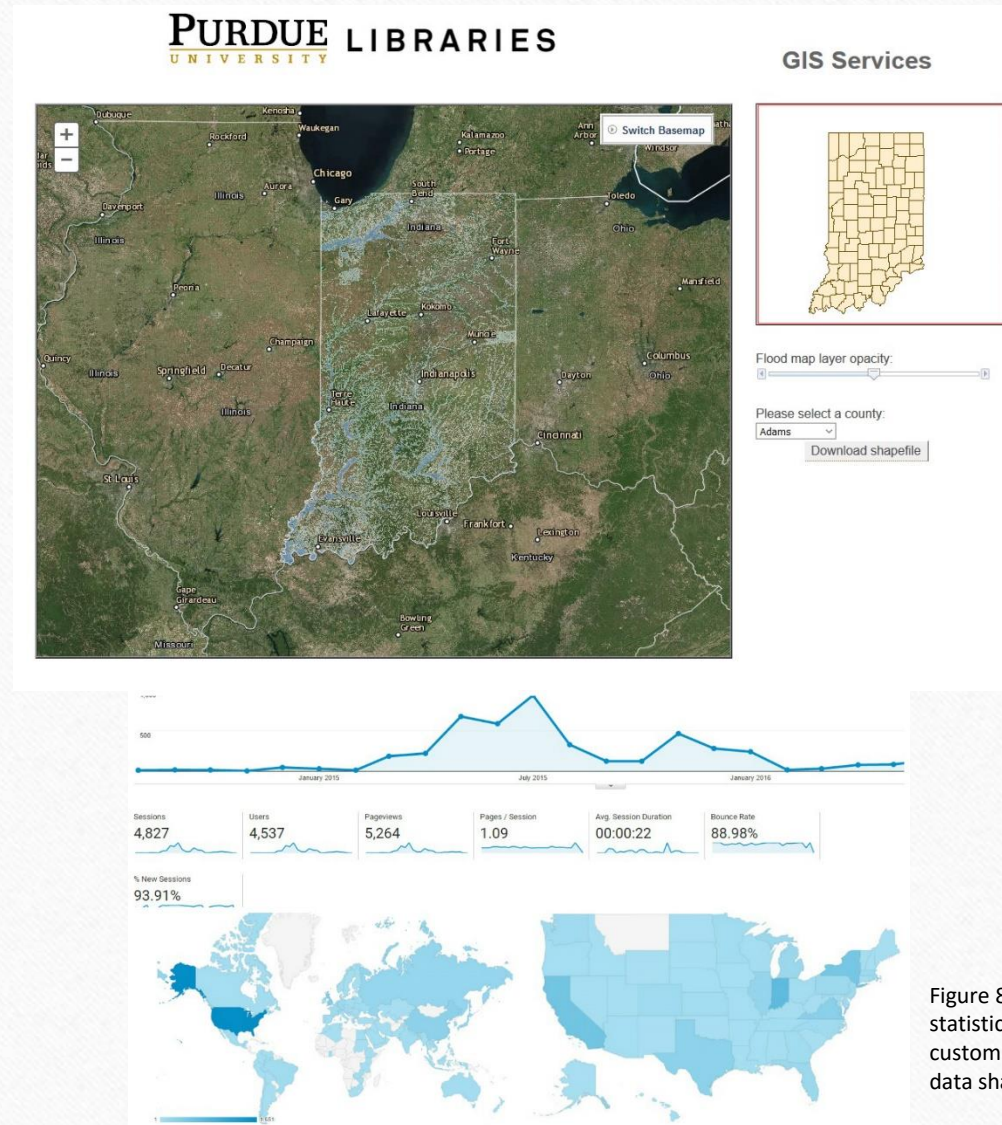


Figure 6. Geodata portal at Purdue supports spatial data discovery, preview and download function for data contributed by collaborative universities in Open Geoportal project, state agencies, and Purdue curated spatial data

Additional spatial data service needs

- To create customized spatial data sharing pages
- Case: floodplain mapping
 - To share research findings with online interactive maps
 - Utilized current GIS cyberinfrastructure
 - To improve this solution, connect with PURR for data download, citation and archive



Project Design – Theoretical Model

- Extend PURR with spatial visualization capability supported from the GIS servers
 - more than 3,500 registered researchers on PURR
 - PURR is a great platform to manage users, and a user-friendly interface to curate and publish datasets
- To increase discoverability beyond the institutional data repository, we expect to ingest the spatial data into our geodata portal

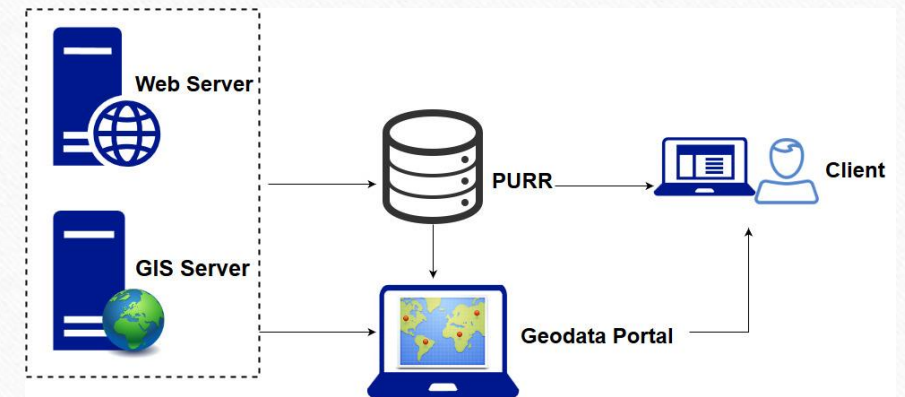


Figure 9. Theoretical model of connecting institutional data repository (PURR) with the GIS cyberinfrastructure

Workflow

- Spatial dataset is submitted and published in PURR with a DOI
- Dataset copied into our geodatabase and published as a map service using the GIS server. An interactive map with the map service is created
- Dataset ingested into our geodata portal

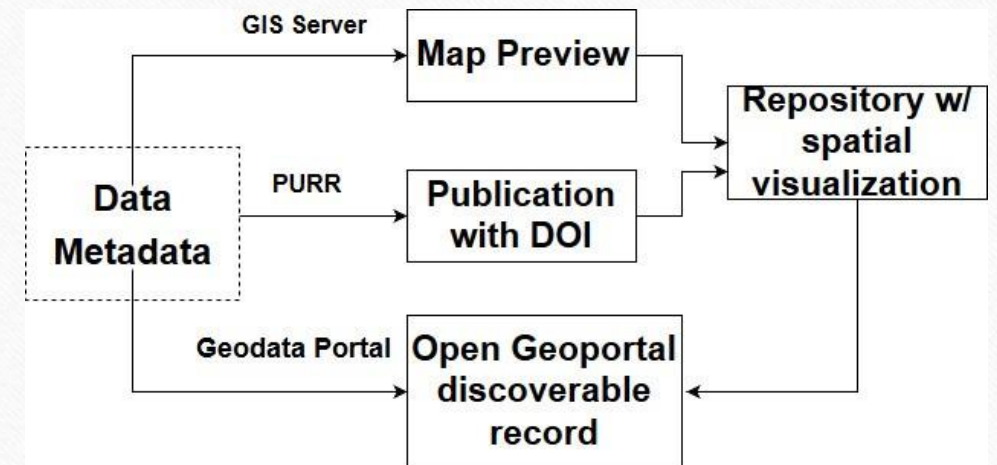


Figure 10. Workflow for spatial data publication

Deployment

- Embed interactive map into the dataset's webpage in PURR
 - Enable HTTPS connections on GIS server
 - Generate FGDC metadata with preview link and ingest into geodata portal

gSSURGO-based Floodplain Maps of the United States

By Venkatesh Mohan Merwade, Nikhil Sangwan, Liuying Du

This series provides shapefiles showing the natural floodplain for the United States. These floodplain polygons are extracted from the gSSURGO soil data from the Natural Resources Conservation Service (NRCS).

in Datasets

This publication version is in development. Manage this publication.

from project gSSURGO-based Floodplain Maps in the United States

About Supporting Docs Versions Citations Usage

Description

US national wide gSSURGO based floodplain maps are developed by using Gridded Soil Survey Geographic (gSSURGO) database. Specifically, the SSURGO map units are joined to certain key attributes in the muagatt and component tables. After joining the spatial data to 'muagatt' and 'component' tables, map unit polygons are selected by the following two queries: (1) 'geomdesc' LIKE '%flood%' AND 'compct_r' >= 85 OR 'taxsubgrp' LIKE '%fluv%'; (2) 'muname' = 'Water' OR 'floodfreqdcd' in ('Rare', 'Occasional', 'Frequent', 'Very Frequent'). The result from this query, which is a polygon feature class or shapefile, is then dissolved to remove any area less than 15000 square meters are removed. More details about this approach can be found in the following publication: Sangwan, Nikhil, and V. Merwade, A faster and economical approach to floodplain mapping by using soil information, Journal of American Water Resources Association, 51(5): 1286-1304. (2015). doi: 10.1111/1752-1688.12306

The following map shows a preview of the floodplain dataset. You can zoom in to your area of interest for more details. To download data, please select a state from the dropdown list or scroll down to the Content List. After you are redirected to a separate page for the selected state, you can download the data from the Supporting Docs tab.



Figure 11. The spatial data series publication webpage with an interactive map

Initial Assessment

- As of March 2017, 5,683 total visits on the published dataset's webpages and 842 total downloads
- The online interactive map has received more than 500 visits (online for just one month)
- The spatial visualization can help users to understand better about the information contained in the dataset before they decide to either download the data or be satisfied with the information provided on the map

Discussion

- Expand the spatial data sharing capability to serve a broader information user group
- An example to demonstrate the benefit of connecting institutional data repository and GIS infrastructure to achieve a better spatial data sharing option
- Researchers need to develop good data management practices
 - Metadata
 - Data sharing license, data ownership...

Questions?
