Spencer Naar
January 26th, 2023
LIS 545B | Term Project Part I
Word Count: 812

Data and Metadata Profile

Data

This profile looks at long-term cumulative groundwater depletion data in the United States between the years of 1900 and 2008. The data is published by the U.S. Geological Survey under the Department of the Interior and the dataset is publicly accessible on the USGS Water Mission Area NSDI (National Spatial Data Infrastructure) Node page. Data was collected from 40 separate aquifer systems and 1 land use area to estimate overall rates and the magnitude of change in the volume of groundwater stored in the earth's subsurface. The most reliable method for calculating these estimates was direct measurements of water-level changes in the aquifer systems.

The dataset is organized into *database* and *spreadsheet* folders. The *database* folder is organized by *aquifer (365 files)* and *basecamp (27 files)* folders and includes .xml, .dbf, .prj, .sbn, .sbx, .shp and .shx files. XML data can be viewed using a web browser however Oxygen XML developer or another xml editor would be helpful. The *spreadsheet* folder has six .xls formatted spreadsheets with groundwater depletion average rates from 1961-1970 and 2001-2008, depletion volume data from 1900-2000 and 1900-2008, Map Data and High Plains data and can be accessed using Microsoft Excel.The dataset does not come with any usage restrictions. It is intended for public access and no license information is provided.

The key stakeholders for the data are i) its *creators/managers*: employees of the U.S. Geological Survey and the larger Federal Department of the Interior ii) its *users*: environmental scientists and researchers, data analysts and estimators of groundwater depletion iii) *individuals*

& organizations: academics and scholars who monitor and study water resources (UNM Water Resource program for example) and nonprofit environmental organizations (NMWCA for example) iv) *policy makers* who may enact legislation around water rights or access based on the data and projections and v) the *larger public* interested in water and environmental issues.

Metadata

The data comes with relatively comprehensive metadata focused on *identification, data quality, spatial data organization, spatial reference, entity and attribute, distribution and metadata reference* information. The metadata is provided on the USGS Water Mission Area NSDI Node page. I was not able to access the original metadata URL from the data.gov page. The *identification information* provides details on citation, description of the dataset article including abstract, purpose and supplemental information, time period of content, status, spatial domain, keywords by theme and location, access constraints, use constraints, contact information, graphic and data set credit, security information, and details on the native data set environment. The *data quality information* centers around attribute accuracy, logical consistency, completeness, positionality and lineage. The *spatial data organization information* focuses on spatial reference methods and descriptions. The *spatial reference information* looks at horizontal and vertical coordinate system definitions. *Entity and Attribute Information* provides descriptions; *Distribution Information* focuses on contact information, distribution liability and ordering; and the *Metadata Reference Information* provides information on standards. The metadata is structured according to the FGDC Content Standards for Digital Geospatial Metadata. This standard was previously widely-used but is no longer a current standard for defining digital geospatial data. In 2010 the FGDC started encouraging federal agencies to

transition to ISO 19115 which is an internationally-adopted schema for describing geographic metadata.

Publications

The primary publication written based on the dataset is *Groundwater depletion in the United States (1900-2008)*, a Scientific Investigations Report by Leonard F. Konikow and published in 2013 by the U.S. Geological Survey. At least 82 articles cite this article and dataset in their scholarship. These resources were identified by locating the article in UW Libraries Search, going to references and clicking on sources which have cited this article.

Enrichment

To improve users' ability to discover the data set in a repository environment, I would assign new descriptive file names which better reflect the contents of each file as well as add tags or descriptors to the file metadata to provide additional context. I would also include information on the overall organization of the dataset. While there are designated folders, there is not a lot of information on how the files are arranged and what they specifically capture. This would be useful metadata while searching and when trying to quickly review data contents. To assist someone unfamiliar with the data and to make use of the dataset for new purposes I would add additional information on the process for creating the data. While there is some lineage or provenance metadata included I would add additional assumptions about the primary input data and limitations around calculations estimating groundwater depletions. I would also include additional context around data compatibility given that data was collected from over 40 different aquifer systems using different methods over a period of a century. Differences in samples as a result of equipment and procedure should be addressed, especially if it is anticipated that the dataset will be used for new research and scholarship purposes.

References

DataOne Best Practices Primer, accessed 26 January 2023, https://dataoneorg.github.io/Education/bestpractices/

Groundwater depletion in the United States (1900-2008): Metadata: USGS Water Mission Area NSDI Node, accessed 26 January 2023, https://water.usgs.gov/GIS/metadata/usgswrd/XML/sir2013-5079_Groundwater_Depletion.xml#stdorder

Konikow, L.F., 2013, Groundwater depletion in the United States (1900−2008): U.S. Geological Survey Scientific Investigations Report 2013−5079, 63 p., http://pubs.usgs.gov/sir/2013/5079.

Research Data Alliance Metadata Directory, accessed 26 January 2023, http://rd-alliance.github.io/metadata-directory/