

Spencer Naar
February 10th, 2023
LIS 545B | Term Project Part II
Word Count: 770

Repository Profile

The repository selected for the dataset detailed in my Term Project Part I is the USGS National Water Information System. This repository stores data on the quantity, quality, use, distribution, and movement of groundwater and surface water for all fifty US states. In addition to storage and preservation the repository serves to disseminate data to the public, State and local governments, public and private utilities, and other Federal agencies involved with managing water resources which is in alignment with the nature of the dataset. This data repository is closed to general data submissions but access to the research data repository is open to the public.

The repository accepts many different types of data including comprehensive information on data collection “site characteristics, well-construction details, time-series data for gage height, streamflow, groundwater level, precipitation, physical and chemical properties of water and water use data, peak flows, chemical analyses for discrete samples of water, sediment, and biological media” (Additional system background, 2011). The collected data fits into broad categories of surface water and groundwater but also water usage and water quality which includes figures on temperature, pH, nutrients and pesticides (About the USGS Water Data, 2021). All data is contributed by researchers and scientists at the United States Geological Survey (USGS) and all information and guidance on what the repository provides to the data submitters specific to the “submission information package” is not publicly viewable. All consulting details and metadata structure and standards information for the submitter appear to

be internal. All data submitted is received at the National Oceanic and Atmospheric (NOAA) Wallops Command and Data Acquisition Station (WCDAS) in Wallops Island, Virginia. Data is then sent to the USGS Water Science Centers for data processing. Data is backed up on receivers located at the Earth Resources Observation and Science (EROS) Data Center in South Dakota (About the USGS Water Data, 2021).

Data is publicly viewable in the repository by category of data and by geographic area. There does not appear to be a login required to download data and there are multiple mechanisms in place for accessing data within the repository. The repository has six “production services” to download and retrieve current condition data, daily values, hydrologic sites, groundwater levels, water quality and statistics. Users can also view graphs of current conditions, water levels, and water quality; tabular output in HTML and ASCII tab-delimited (rdb) data files and summary lists for selected sites. While XML and JSON are common files for sharing data, rdb data files are the primary output format for this repository. That being said, the new Water Services site does allow for instantaneous, daily values and groundwater levels data to be downloaded in the XML format and data-friendly formats, such as Microsoft Excel spreadsheets and KML, are being worked on to enable integration with Google Maps, Google Earth and GIS formats (About Output Formats, 2021). It is not possible for the entire repository to be downloaded in a single action due to the volume of data, however data can be acquired slowly over time across geographic areas. The only access restriction is that a single data request can not exceed 100,000 site records, “a limitation intended to prevent any one data consumer from unduly affecting other users of the system” (Automated Retrieval, 2021) The repository has an entire page dedicated to automated retrieval of its data which is significant because the designated community relies on active, current and up-to-date water data for research and policy decisions.

The repository does display metadata using the same content standard as the dataset detailed in Term Project Part 1 which is the FGDC Standards for Digital Geospatial Metadata. USGS provides a number of resources on metadata creation and following specific standards as part of the data description and input process. In terms of the “dissemination information package” the repository provides a number of custom options for output information and formats depending on the needs of the designated community.

One thing that was interesting to me was that the repository has a notification service to inform users of the automated data retrieval community about planned outages, unexpected system problems and changes to the system that might affect third party sites and research. Since water data is collected at millions of location sites around the country and maintained by different USGS Water Science centers, data can suddenly become unavailable or data output formats can quickly change. This communication feature which proactively engages its stakeholders highlights the credibility and integrity of this repository and its recognition of the importance of its designated community which relies on its repository for timely, accurate and quality data.

References

- About Output Formats* (November, 2011). USGS National Water Information System: Help System. <https://help.waterdata.usgs.gov/faq/output-formats>
- About the USGS Water Data for the Nation site.* (November, 2021). USGS National Water Information System: Help System. <https://help.waterdata.usgs.gov/faq/about-the-usgs-water-data-for-the-nation-site>
- Additional System Background.* (November, 2011). USGS National Water Information System: Help System. <https://help.waterdata.usgs.gov/faq/additional-background>
- Automated Retrievals.* November, 2021). USGS National Water Information System: Help System. <https://help.waterdata.usgs.gov/faq/automated-retrievals>
- Content Standard for Digital Geospatial Metadata. (CSDGM).* (1998). Federal Geographic Data Committee. <https://www.fgdc.gov/metadata/csdgm-standard>
- FAQ.* (November 2011). USGS National Water Information System: Help System. <https://help.waterdata.usgs.gov/faq>
- USGS National Water Information System.* (March, 2022) Registry of Research Data Repositories. <http://doi.org/10.17616/R3S333> last accessed: 2023-02-11
- USGS Water Data for the Nation.* (February, 2023). USGS National Water Information System <https://waterdata.usgs.gov/nwis>