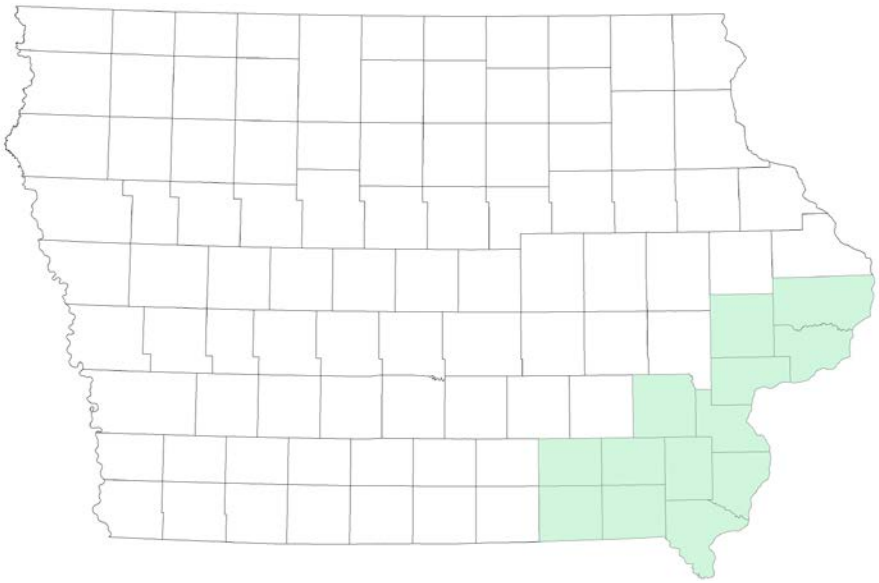
Digital Infrastructure and Metadata for Geologic Collections, Southeast Iowa Focus



Iowa Geological Survey Technical Report
National Geological and Geophysical Data Preservation Program
United States Geological Survey Award Number G14AP00147.
Term Dates: 09/15/2014 – 09/14/2015
Submitted: 09/15/2015

Jason A. Vogelgesang
Iowa Geological Survey
IIHR – Hydroscience & Engineering
The University of Iowa
100 Oakdale IGS
Iowa City, Iowa 52242-1304
O: (319) 335-4231 F: (319) 335-5238
Jason-Vogelgesang@uiowa.edu

Abstract

This report documents work performed by the Iowa Geological Survey for the National Geological and Geophysical Data Preservation Program (NGGDPP) under award number G14AP00147 for 2014. The project focused on preserving data in a thirteen county southeast region of Iowa. Grant objectives of creating and/or updating digital infrastructure and creating metadata were met. The majority of paper documents preserved to electronic format included well logs and paper documents. Metadata for these and other documents were created and uploaded to the National Digital Catalog.

Background

The IGS was established as the Iowa Geological Survey (IGS) in 1892 but existed discontinuously since 1855. Systematic collection of samples and records for wells began in the 1930s. While much of the data have been converted to digital records, large portions of the data are still held as paper documents and files in printed reports, maps, field notes, etc. Many of the paper documents are deteriorating, often poorly maintained, and difficult to access efficiently for research, resource or environmental assessment, among other uses. Conversion to digital formats and addition to GeoSam and NGGDPP document storage locations alleviates these shortcomings.

Figure 1. Well records and other document storage located at the Oakdale Iowa Geological Survey Building.

The IGS occupies two buildings on UI campuses. The main office is located in Trowbridge Hall on the central UI campus. The Oakdale Iowa Geological Survey (OIGS) building is on the west edge of the University of Iowa Research Park campus. The building houses the IGS' Rock Library, the repository for drilling logs, well cores, well chips, and rock specimens from exposures throughout Iowa. Additionally, the building provides storage for the bulk of the paper files that document physical collections (Figure 1). Included are a wide variety of older paper files and archived maps.

The majority of the most critical and widely used geologic information for the state is electronically held in the IGS Geologic Sample Database (GeoSam). Access is free and available to the public through a web-based interface (https://geosam.iihr.uiowa.edu), which also provides tools to query the database and view electronic documents. Geologic information accessible via GeoSam includes: site identifiers and location, well construction information, geologic logs (lithologic and/or stratigraphic), and water production information. GeoSam's web-based interface was recently redesigned to allow users to locate sites geographically as well

as through multiple text search fields. GeoSam forms the backbone of the current IGS data preservation system and will remain an integral component of its preservation plan.

Established 2014 Grant Activities

IGS included the following in its 2014 NGGDPP grant proposal:

> 1) Paper records from approximately 6,400 unique sites in southeast Iowa will be gathered, evaluated, and converted to digital format. These files contain drilling logs, well construction information, lithologic logs, pump tests, and other relevant site information. The records will be scanned in Portable Document Format (PDF) and will be added to GeoSam for public access. We will focus on thirteen counties in southeast Iowa as all other regions of the state were assessed during previous NGGDPP grants. Metadata for the southeast Iowa records will be developed and submitted to the National Digital Catalog.
>
> 2) Well information from approximately 400 drilling logs will be entered into the GeoSam database. These records contain valuable geologic data from across the state including: well locations, drill dates, lithologic information, well construction, and capacity information. We will begin preserving this newly discovered collection by focusing on records in the southeast portion of the state. Metadata will be developed and submitted for these records.
>
> 3) We propose to prioritize and scan approximately 4,000 pages of the remaining 10,000 pages of paper-only publications and upload them to our publically-accessible publications website. Metadata will be developed and submitted for these publications.
>
> 4) In addition to metadata created for tasks 1 – 3, metadata will be created for collections received as part of regular IGS activities. While the acquisition and preservation of these data are not completed as part of the NGGDPP, metadata will be submitted to the National Digital Catalog as part of the NGGDPP work.

This project focused efforts on capturing data from a thirteen county region of southeast Iowa, when applicable, to conclude our geographic data preservation approach. Prior grants have focused on other regions of the state, working systematically from the northwest to the southeast.

Results of 2014 Grant Activities

<u>NGGDPP Metadata Development</u>

The IGS manages a total of 10 collections defined for NGGDPP. Metadata records were created and submitted for all items associated with a geographic location, including rock cores, rock cuttings, well logs, and paper reports. Items contained in GeoSam, including rock cores, rock cuttings, and well logs, have been assigned a point location and horizontal accuracy as part of GeoSam entry. In the case of paper reports, centroid coordinates were assigned.

Metadata were prepared in CSV format for all of the collections submitted as part of this project. The metadata were generated using views and queries from MySQL Server databases. The CSV format was chosen as the metadata creators have the most experience with that format.

Metadata preparation was carried out as follows:

1) MySQL server tables, views, and custom functions needed to produce the necessary text strings were developed.
2) Queries were developed to select the data from the views prepared in step one to a temporary file.
3) The temporary file was imported into Excel and validated to ensure National Catalog compatibility.
4) The Excel file was saved to CSV format with pipe characters as delimiters and the file was uploaded to the National Digital Catalog website.
5) The final step was to request NGGDPP staff to load the data into the catalog, replacing previously submitted data sets.

The NGGDPP metadata elements, along with a description of the data items provided for each element, are listed in Table 1. In instances where required data elements were not available, metadata records were not prepared, e.g. if latitude and longitude had not been stored for a data item, the item was omitted from the metadata. If changes/updates were not made to a collection, metadata was not submitted. In this case previously uploaded metadata are still current.

The IGS metadata supplies data for the optional items in some cases. For most collections, optional items were either populated for all records or no records. In other cases, the optional items were populated where the data to be supplied was available.

Table 1. Description of metadata elements developed from IGS data.

Tag/Data Element	Description of data supplied
collectionID	Collection ID from previously defined Iowa NGGDPP collections.
title	Name of collection and IGS identifier for the item. Untitled documents were given a descriptive title.
abstract	General descriptive information about the site and/or data item including identifier and site type; other information, depending on the collection being documented, includes total depth, sample depth, completion date, collection date, aquifer, and author.
dataType	NGGDPP catalog supplied values.
supplementalInformation	Statement concerning archive location of samples/logs and/or a URL where additional information may be accessed, if available.
coordinates	Longitude and latitude in decimal degrees (NAD83) derived from coordinates stored in GeoSam. Centroid coordinates are provided for paper reports.
onlineResource (optional element)	Hyperlinks are provided where available.
datasetReferenceDate	The date the metadata for the collection was prepared

Additions to IGS Collections for Southeast Iowa

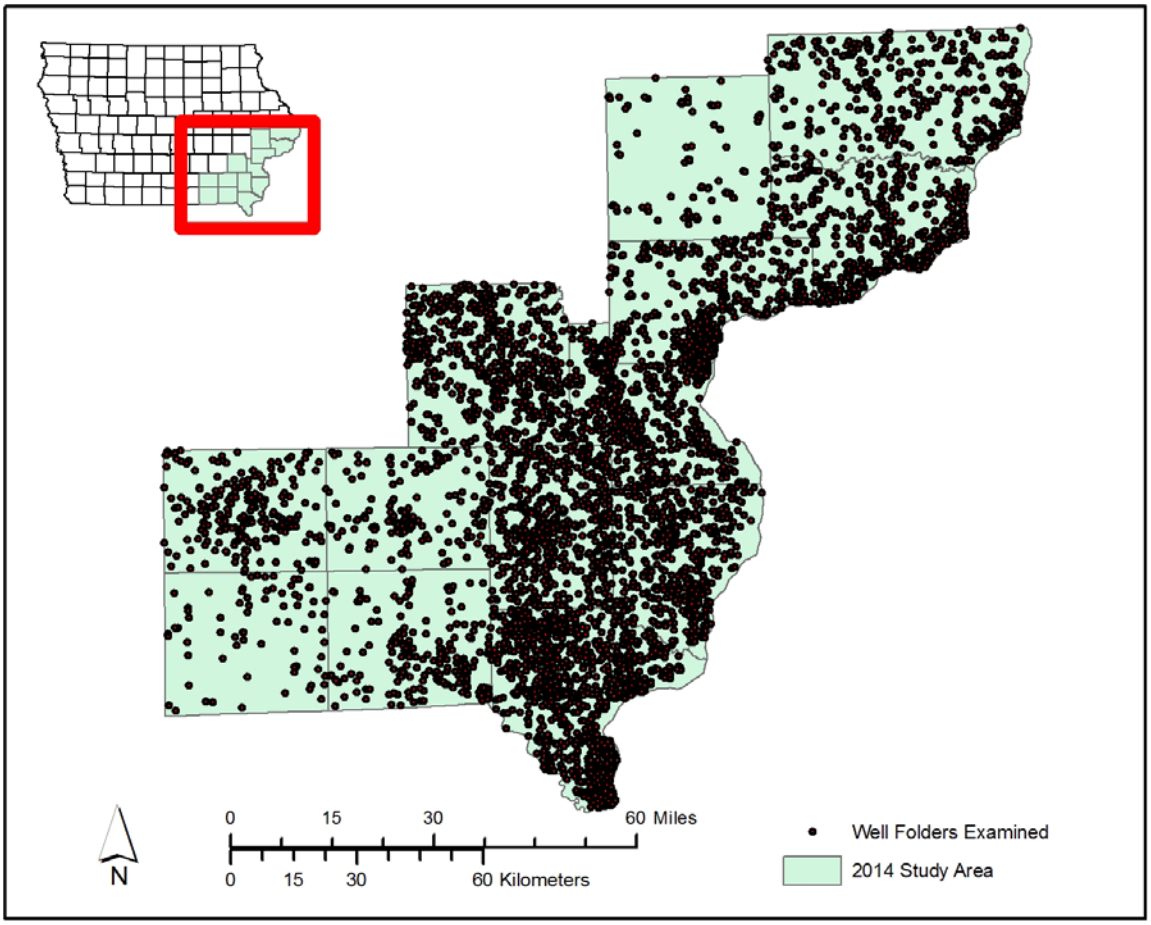


Figure 2. Southeast Iowa study area showing the locations of 6,384 well sites examined with relevant information scanned and added to GeoSam.

Previous NGGDPP grants have allowed the IGS to focus geologic data preservation efforts in the northwest, southwest, south-central, north-central, northeast, and east-central areas of the state. This year's work follows that regional approach and focuses on thirteen counties in southeast Iowa (Figure 2).

A total of 6,384 well file folders from east-central Iowa were examined (not including archived well logs described below). While GeoSam contained the basic information on each of these records, additional well log information existed only in paper format. Relevant information from each folder (Figure 2) was scanned and attached as PDFs to the existing well data in GeoSam.

Archived Well Logs Scanned and Entered in GeoSam

The transition of the IGS to the University of Iowa initiated a major inventory and organization of archived material. As such, a number of boxes of driller's logs were discovered. An investigation into these records disclosed that these logs had never been entered into the GeoSam database. As a part of this NGGDPP grant, 412 drilling logs were scanned and entered into GeoSam. This time consuming process involved looking through historic and current

landowner data, locating the well, entering all applicable information to GeoSam, and scanning and uploading the log(s) to GeoSam. The logs contain valuable geologic, hydrogeologic, well construction, and location data that is now accessible electronically.

Preservation and Cataloging of Paper Reports

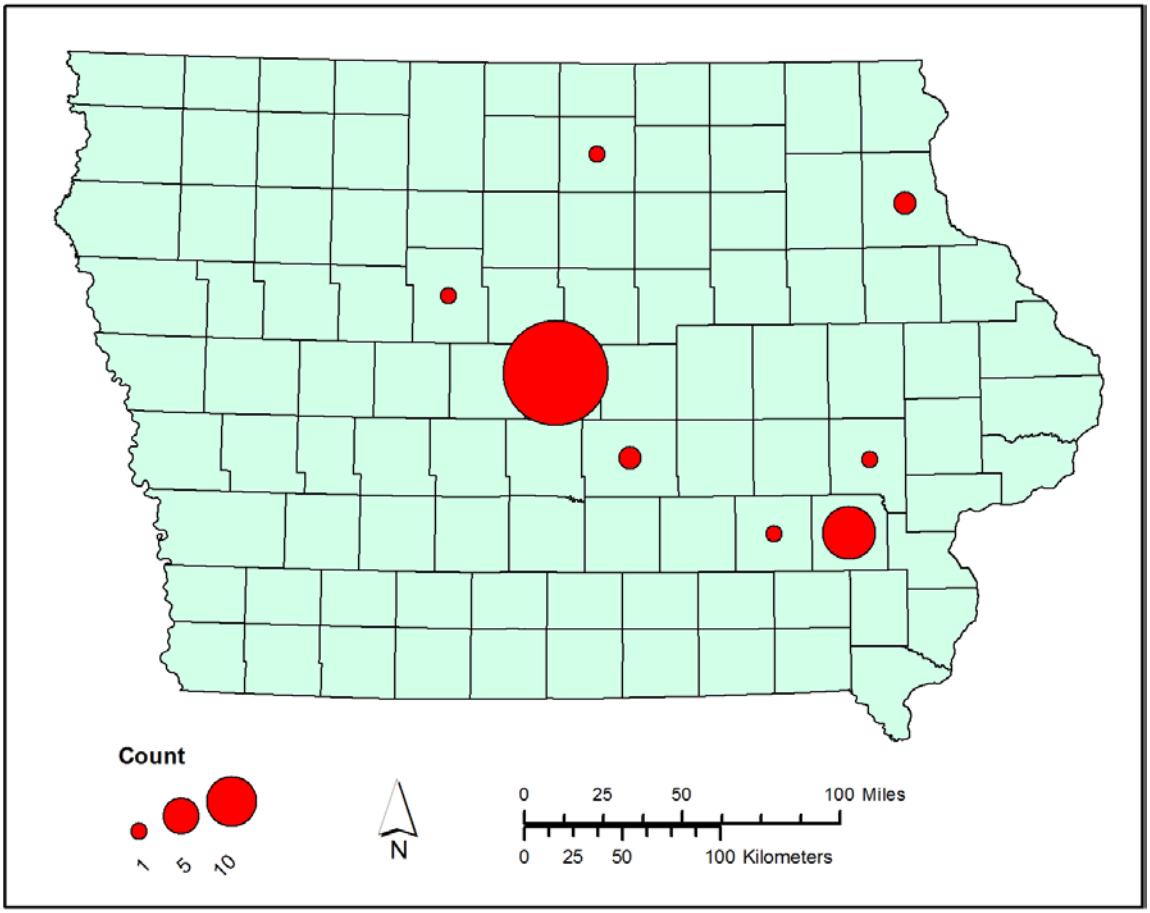


Figure 3. Map of Iowa showing centroid locations assigned to 62 paper reports that were scanned as part of year's work. The majority of the publications scanned this year reference the entire state, with centroids plotting near Polk County.

Throughout its history, the IGS has generated an extensive collection of paper reports documenting geologic studies. Reports include projects ranging in scope from statewide to narrowly-focused in geographic and subject area. In recent years, newly published IGS reports are created as electronic documents or as electronic and paper documents. The electronic versions are publicly available as PDF downloads from the IGS website (http://s-iihr34.iihr.uiowa.edu/publications/search). However, older reports exist as paper documents only. In some cases the reports have been superseded by newer work, but older reports represent foundation work and are important from a historical perspective. Metadata were created and submitted for paper reports scanned this year as well as those scanned in previous years.

Remaining paper-only publications were prioritized for digital conversion. Prioritized and preserved paper report categories included: 13 Water Supply Bulletins, 11 Report of Investigations, 8 Open File Maps, 6 Open File Reports, 5 Miscellaneous Publication Series reports, 4 Technical Papers, 4 Water Resources Information reports, 2 Miscellaneous Map Series maps, and others. A total of 5,310 pages were scanned during this project year. Scanned documents were added to the IGS publication website (link above). Metadata were developed and submitted for these paper documents.

Metadata Records Added to the National Catalog

Metadata describing the information held in GeoSam, as well as other geologic information from Iowa, were loaded into the National Digital Catalog in September 2015. The number of data items included in the metadata is summarized by collection as follows:

Table 2. Summary of new metadata provided to NGGDPP catalog.

GeoSam Metadata Records Submitted **Southeast Iowa Focus*	**New Records 2014**	**Total Records Uploaded**	**Approximate Collection Total**
Rock Cores	45	1,336	1,385
Rock Cuttings	132	36,747	39,500
Hand Samples	0	258	425
Thin Sections	0	1,865	5,200
Well Logs (drilling and strip)	3,282	84,751	85,000
Paper Reports	126	445	600
Maps	0	1,549	2,000
Geochemical Data	0	20,022	20,022
Photographs	0	0	20,000
Field Notes	27	4,199	8,000
Total	**3,612**	**151,172**	**182,132**

GeoSam Updates

2014 project goals included modifying GeoSam by developing tables and/or views to manage additional types of geologic data. With GeoSam serving as the IGS' main geologic data storage database, ongoing modifications for efficient data retrieval are necessary. During the 2014 year, revisions were implemented to better manage and retrieve geologic data. One such example is a new pump test category in GeoSam that allows scanned pump test data to be viewed separate from drilling records. Groundwater quantity continues to be an important concern in Iowa and efficient access to pump test data ensures aquifer parameters can be used in groundwater sustainability studies.

2014 Award Summary and 2015 Award Plans

Grant objectives of creating and/or updating digital infrastructure and creating metadata were met for the 2014 NGGDPP award. This project year focused on preserving data within a thirteen county region in southeast Iowa. The majority of paper documents preserved to electronic format included well logs and paper documents. Metadata for these and other documents were created and uploaded to the National Digital Catalog.

Previous NGGDPP grants allowed the IGS to convert valuable geologic data to digital format in the northwest, southwest, south-central, north-central, northeast, east-central, and now southeast regions of the state. This regional approach to preserving drilling records has successfully concluded with the 2014 NGGDPP project as data in all geographic regions of the state will have experienced focused preservation efforts. Proposed NGGDPP work for 2015 includes scanning and uploading the IGS' entire collection of down-hole geophysical logs, scanning and inputting information into GeoSam from a collection of archived drilling logs, and completing the scanning and uploading of the IGS' entire collection of paper reports and publications. In addition, an inventory will be created from a collection of working and final maps/documents at risk of disposal. Metadata from these tasks and work completed as part of regular IGS operations will be generated and submitted to the National Digital Catalog.