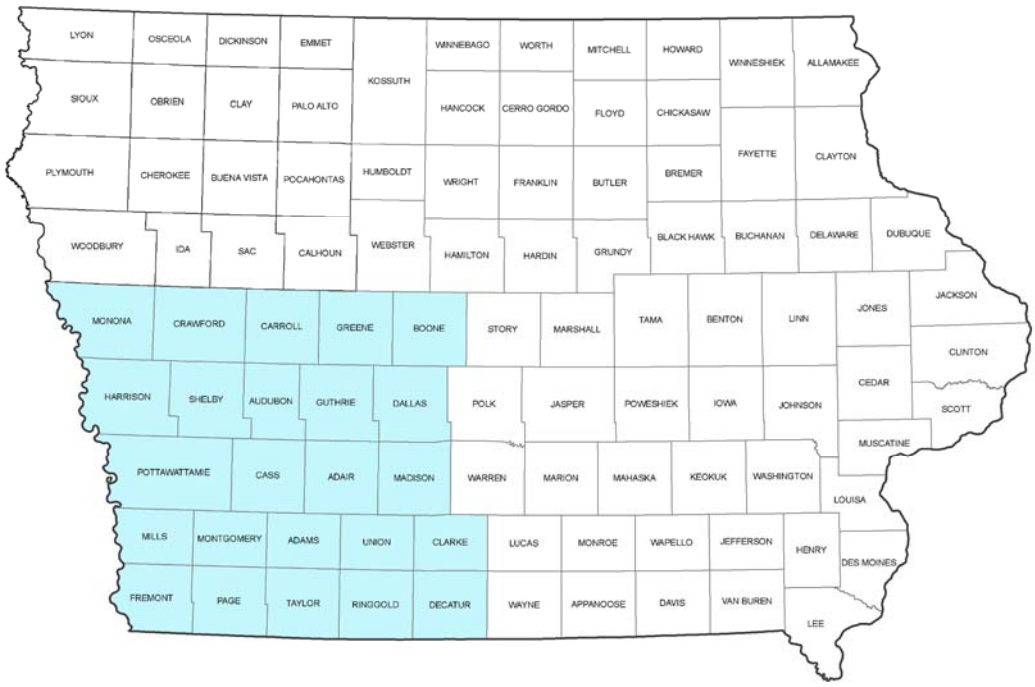
Metadata for Iowa Geologic Data Collections and Southwest Iowa Additions



IGWS Contract Report
National Geological and Geophysical Data Preservation Program
United States Geological Survey Award Number G09AP00127

Iowa Geological & Water Survey
Iowa Department of Natural Resources

The project described in this publication was supported by award number G09AP00127 from the United States Geological Survey. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS.

Introduction

This report documents the development of National Geological and Geophysical Data Preservation Program-standard metadata for the physical and electronic data collections maintained by the Iowa Geological and Water Survey (IGWS), and the conversion of data held as paper records to electronic formats for counties in southwest Iowa. This work was conducted as part of the U.S. Geological Survey - National Geological and Geophysical Data Preservation Program (NGGDPP), under award number G09AP00127 for 2009.

IGWS and Geologic Data Storage

The IGWS was established as the Iowa Geological Survey (IGS) in 1892. After periods of being part of the University of Iowa and later an independent state agency, IGS became part of the newly formed Iowa Department of Natural Resources in 1986. In 2008, the survey changed its name to Iowa Geological and Water Survey to reflect the agency's long-term role as the state's lead earth and water science agency.

Figure 1. IGWS Oakdale laboratory and sample repository building.

IGWS occupies two buildings on the University of Iowa Campus. Most staff offices are located on the main campus of the University in Trowbridge Hall, which is shared with the Department of Geosciences. The laboratory and sample storage building is located on the University's satellite Oakdale Campus. The 11,000-square foot Oakdale building (fig. 1), built in 1983, functions as the repository for most physical geologic material storage, such as cores, cuttings, and hand samples. Sample preparation, cataloging, and descriptions as well as a variety of laboratory work are carried out at the Oakdale building. The Oakdale facility provides storage space for the bulk of the paper files that document the physical collections. In addition a wide variety of older paper files and unpublished and archived maps are stored at the Oakdale facility.

The majority of the most critical and widely used information for the state, mainly from cores and well cuttings, logs, and construction data, are electronically held in the IGWS Geologic Sample Database (GEOSAM) as a combination of a relational database developed to reflect the site-based nature of the data and electronic documents. Access is provided through a web-based interface (http://www.igsb.uiowa.edu/webapps/geosam), which provides tools to query the database and view the electronic documents. GEOSAM and GEOSAM-compatible databases form the backbone of the current IGWS data preservation system and its ongoing preservation plan.

Summary of 2009 Grant Activities

IGWS included the following in its 2009 NGGDPP grant proposal:

- Develop metadata, in accordance with National Catalog standards, for all current GEOSAM sites
- Establish locations as needed for physical and paper data collections from 24 counties in southwest Iowa (fig. 2) that were not held in GEOSAM
- Make any needed changes to GEOSAM to accommodate different site/data types
- Move these locations and site types to GEOSAM, develop metadata for these additions
- Submit all developed metadata to the National Catalog

The data collections planned for addition to GEOSAM from southwest Iowa included:

- Outcrop records/descriptions from publications, field notes, and other sources
- Physically archived hand specimens and bagged Quaternary specimens
- Mechanical and clay mineralogical data for Quaternary materials
- Thin sections
- Geochemical analyses consisting primarily of general and contaminant groundwater analyses
- Municipal water supply well information

Of these collections, the need for three minor changes to the original plans became apparent during the work. First, the concept of "municipal water supply well information" as a separate "collection" became illogical, as data from these wells was

already included in other collections—cuttings, cores, construction data—that were already present in GEOSAM or contained in new collections such as groundwater quality analyses. So while a variety of municipal well data from northwest Iowa was added to GEOSAM, "municipal water supply wells" are not reported as a separate collection. GEOSAM is searchable by a variety of criteria such as "well type," which includes a municipal well category. Second, little in the way of "mechanical or clay-mineralogical data for Quaternary materials" was found to be archived for this part of the state. Most of IGWS Quaternary studies have focused on the central and eastern parts of the state, and we will likely populate this collection when those parts of the state are being done. Third, we added a "Field Notes" collection to the NGGDPP online inventory. This collection had been overlooked and/or included in other collections during our initial inventory, and it became apparent it should be handled separately.

Summary of Project Results

NGGDPP Metadata Development

The IGWS collections defined for NGGDPP are broadly defined and may include physical samples, derived data, and documents. Iowa's data was prepared in XML format as the bulk of the data from which metadata was extracted is stored in SQL Server databases. XML was selected over a delimited text format as the simpler of the formats to develop from a relational database. The necessary relationships between tables and views are modeled as views/queries for output to the required format from this source as described below.

Metadata preparation was carried out as follows:

1) SQL server tables, views, and custom functions needed to produce the necessary text strings were developed.
2) Developed queries to select the data from the views prepared in step one to a temporary XML-type data item then exported this data item to a text-based XML file
3) When the text-based XML was successfully exported it was validated against a simple "noNamespace" schema with a simple Visual Basic program.
4) When the local validation was successful, the file was uploaded to the NGGDPP site and the validity of the XML tested using the tool available there.
5) Following successful validation with the NGGDPP-provided tool, a request was made to the DBA to load the data into the catalog.

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.

The IGWS metadata supplies data for the optional items in some cases. For most collections, optional items were either populated for all records or no records. For example, the optional item, verticalExtent, is populated for most collections but was

omitted for the Geochemical Samples collection (water quality) as not relevant to the data being presented. In other cases, the optional items were populated where the data to be supplied was available.

Table 1. Description of metadata elements developed from IGWS data. Bolded elements indicate required data.

Tag/Data Element	Description of data supplied
collectionID	Collection ID from previously defined Iowa NGGDPP collections
title	Name of collection and IGWS identifier for the sample/log
alternateTitle	
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 a URL where additional information may be accessed
coordinates	Longitude and latitude in decimal degrees (NAD83) derived from coordinates stored in Geologic Sample Database (GEOSAM)
alternateGeometry	County, tier, range, section, and quarter sections for sites where this information is available
onlineResource/resourceURL	
browseGraphic/resourceURL	not used
dates/date	Last update of the record for some collections; sampling date for some collections
datasetReferenceDate	The date the metadata for the collection was prepared
verticalExtent	Total depth supplied for wells, length of section measured/described for field notes

Additions to IGWS Collections for Southwest Iowa

Twenty-four counties in southwest Iowa (fig. 2) were selected for more focused development of digital data from samples not yet completely cataloged in IGWS databases or from archived documents. Samples that were cataloged for the area and paper records converted to digital formats included the following:

- Field notes/measured section descriptions from 3,253 locations, notes scanned and in GEOSAM (fig. 1, table 2).
- Well file information scanned and pdfs attached to 2,298 records (fig. 2).
- Thin sections from 1,929 locations (some multiples per locations), descriptive information in GEOSAM-compatible format, data will be integrated into GEOSAM.
- A total of 232 large format maps were classified into six map types (table 3) and scanned. Of these, 219 were georeferenced and 36 cataloged.

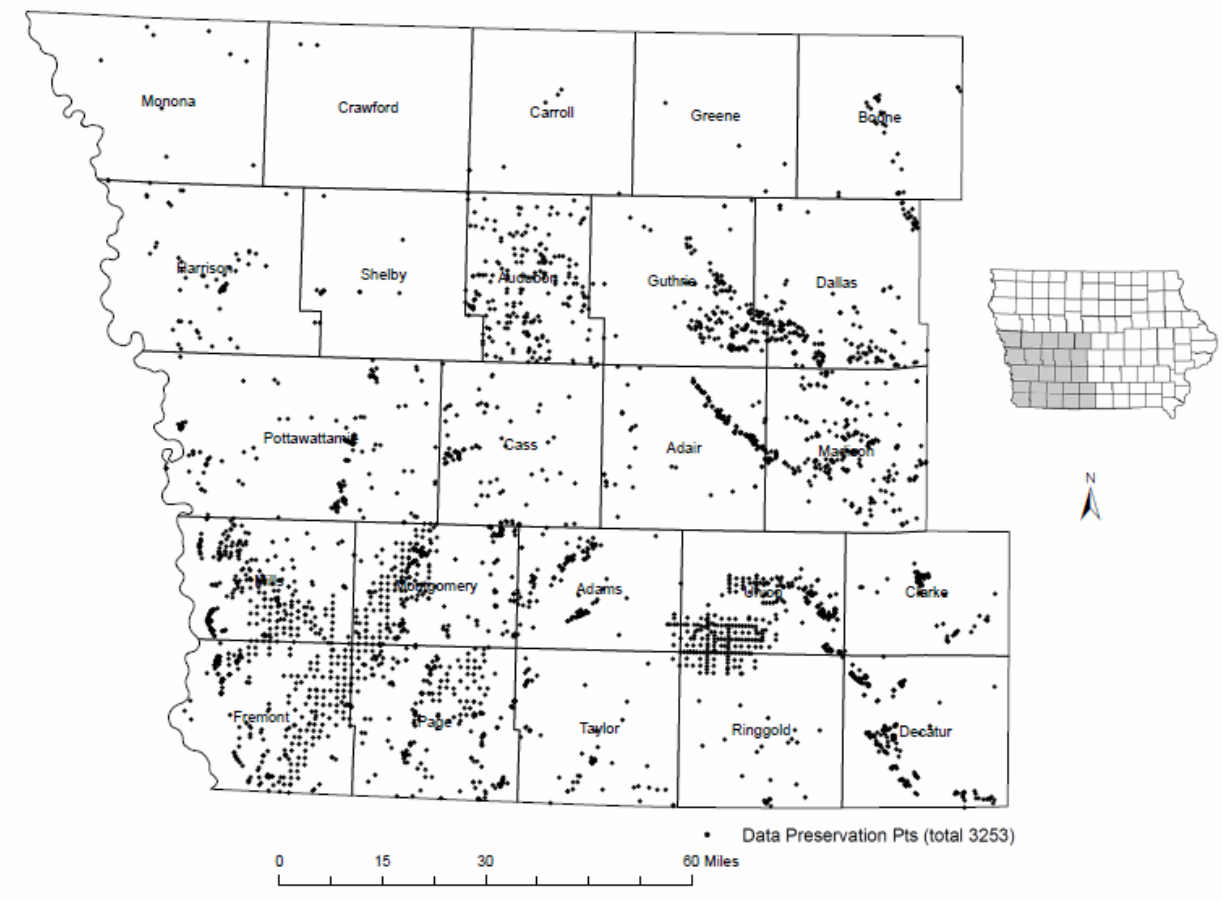


Figure 2. Southwest Iowa study area with dots showing location of new records added in fy 2009.

Table 2. Breakdown of site types and numbers of geologic sites entered into the GEOSAM database in fy 2009.

Site Type	Site Type Count	Pdf Available
Unknown	9	9
Drilled hole	1,040	1,039
Mine	27	27
Other	3	3
Outcrop	1,371	1,367
Quarry face or pit	406	404
Quaternary	259	259
Rain collector	1	1
Roadcut	137	137

· Data Preservation Pts (total 3253)

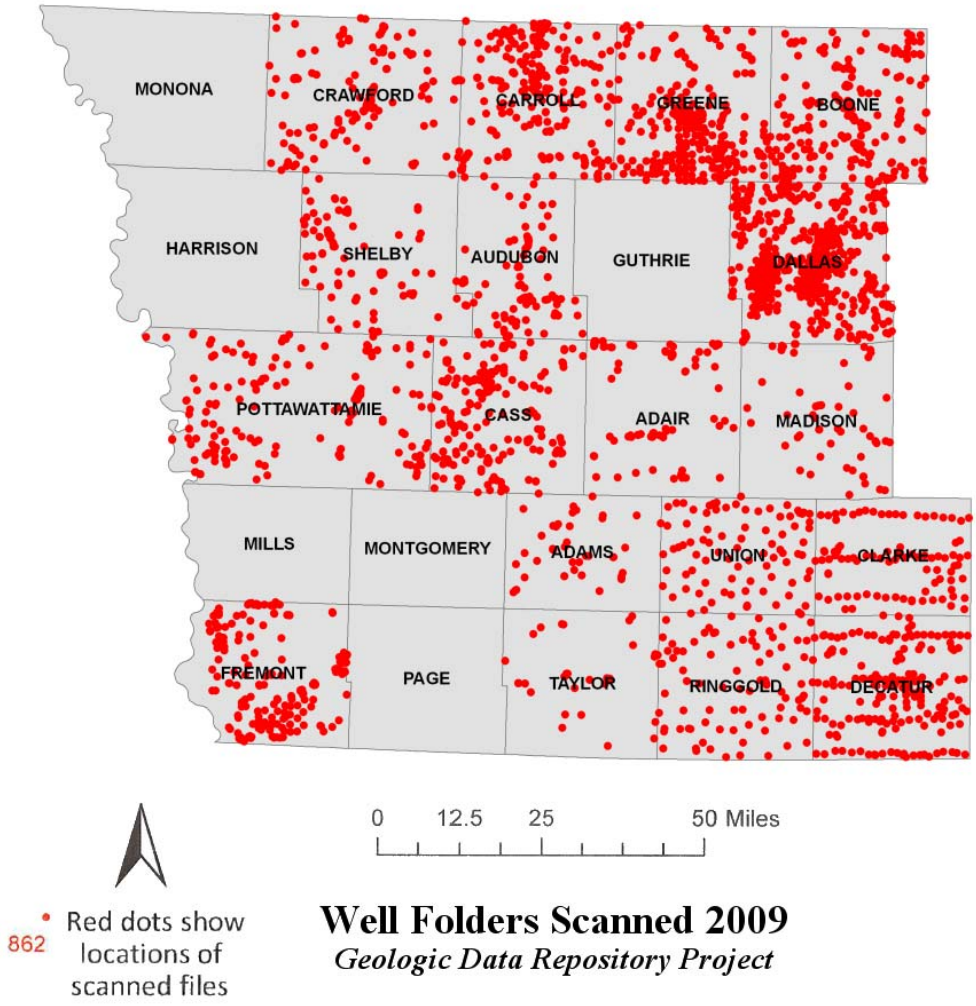


Figure 3. Map of southwest Iowa study area showing the location of 2,298 wells whose folders were scanned and pdfs attached to GEOSAM records during fy 2009.

Table 3. Large-format maps classified, scanned, georeferenced, and cataloged in fy 2009.

Map Type	Number Scanned	Number Georeferenced	Number Cataloged
Geologic	39	38	9
Structure	86	85	7
Thickness	51	50	6
Hydrogeologic	7	7	0
Geophysics	17	17	11
Other	32	22	3
Totals	**232**	**219**	**36**

Note: At the time of this report, 36 of the 232 scanned maps had been entered into GEOSAM.

Preparing Map Documents for Data Preservation

IGWS has a large collection of paper maps and other documents larger than page-size (8-1/2 x 11) developed during its history. Most of these documents were never intended for publication, but were used as working maps or reference maps by the geologists. Much of the content of the maps has been superseded by more recent work; however, the earlier maps are accurately viewed as work that contributed to currently accepted interpretations of a variety of geologic problems. Preservation of these maps and other documents is necessary for this reason.

During FY10 232 maps (Table 3) and other oversize documents (larger than 8½ x 11 in.) were prepared for preservation as digital images. Initially, the maps were reviewed by a staff geologist, assigned a unique identifier, and assigned to one of six thematic categories based on the content of the map.
Following the initial organizing and cataloging steps, the maps were scanned using a Contex wide-format scanner. The files were saved as uncompressed .tiff images as either 24-bit (full color) or gray scale images depending on the original document. During scanning, the authors' names, dates, scanning resolution, bit depth, were captured in a spreadsheet that was used to organize the scanning and georeferencing phases. Minor edits were performed on the scanned images including straightening and cropping. , and where appropriate georeferenced to the IGWS standard coordinate system. Thirty-five of the maps that included all or part of the current year's study area were added to an ESRI® unmanaged image catalog as jpeg2000 images. Following development of the image catalog, map centroids and bounding coordinates were calculated for use in preparing NGGDPP metadata. The map catalog will be published as part of the IGWS GIS library.

Metadata has been prepared and submitted for the thirty-five maps that include part of the southwest Iowa area.

Metadata Records from GEOSAM Added to the National Catalog

Metadata describing the information held in GEOSAM were loaded into the National Catalog in October 2010. This number of data items included in the metadata are summarized by collection as follows:

New Records (fy 2009)	
Thin Sections (fy 2009)	195 records*
New Records (fy 2009)	3,253 records
Amended Records (fy 2009)	2,298 records
Scanned Maps (fy 2009)	36 records
Total GEOSAM Records	72,982 records
Well Logs	59,377 records
Well Cutting Samples	34,951 records
Rock Cores	1,665 records
Scanned Drillers Logs	40,018 records
Scanned Strip Logs	21,482 records
Field Notes/Measured	4,407 records
Thin-Sections	195 records*
Hand Specimens	258 records

* 1,865 thin-sections in 195 records .

Note: These totals reflect holdings that previously existed in GEOSAM as well as the additions for the 24-county block of southeast Iowa.

Award 2009 Summary and Award 2010 Plans

The goals proposed for the 2009 project year were met with the minor changes described above. Metadata was created for most collection types described for Iowa. The IGWS sample management database underwent modifications to accept new data types.

The xml metadata has been uploaded and run through the USGS validator successfully.

We have requested all but the map data replace existing records in the NGGDPP data store. That will be the last step in the process and will complete our obligation for data submittal for the 2009 fiscal year.

For the IGWS 2010 award from NGGDPP, the focus on developing digital data from paper archives will shift to a block of counties in south central Iowa. Updated metadata will be developed near the end of the project period to replace the records provided in 2010. Work to improve integration of the new digital data types into GEOSAM will continue as needed.