FINDINGS SPRINGS IN THE FILE CABINET

Mason Johnson
Minnesota Pollution Control Agency, 18 Woodlake Drive Southeast, Rochester, Minnesota, 55904 USA, Mjohn206@uwsp.edu

Ashley Ignatius
Minnesota Pollution Control Agency, 18 Woodlake Drive Southeast, Rochester, Minnesota, 55904 USA, Ashley.Ignatius@state.mn.us

Abstract
The Minnesota Pollution Control Agency (MPCA), in partnership with other agencies, is currently undertaking comprehensive sub-basin assessments statewide over a ten-year period. Southeast Minnesota has over 17,500 kilometers of perennial and intermittent streams, making the task of comprehensive sub-basin assessment challenging; the task is further complicated by karst geology. In the summer of 2014, a pilot project began between the MPCA and Minnesota Department of Natural Resources (DNR) to digitally preserve paper documents which capture qualitative and quantitative data about the hydrology, water chemistry, geomorphology, biology, land use and karst features of southeast Minnesota streams. The paper documents in file cabinets were not in an accessible or easy-to-use format; as such, they were in a 'data silo.' The task was to preserve the documents so as to make the data usable by converting the documents into a digital format (Adobe PDF, GeoTIFF, ESRI Feature Class). To date, more than 4,000 documents (of an estimated more than 12,000) have been converted, made text-searchable, prepared for storage in a document management system, and made more accessible through a geographic information system (GIS). This previously inaccessible data is an important piece in understanding the karst region of southeast Minnesota. Within the documents scanned thus far, over 400 springs and other karst features have been identified, which are not currently recorded in Minnesota's Karst Feature GIS Database.

Introduction
The karst region of the "Driftless Area" in southeast Minnesota proves to be challenging for the Minnesota Pollution Control Agency (MPCA), which is tasked with assessing the streams of that area. Karst features, such as sinkholes and springs, can change the properties of a stream. For example, the presence of sinkholes can modify the contributing drainage area to a stream and a groundwater-fed spring can change temperature, chemistry and flow of a stream.

Stream assessment is conducted to determine if the stream supports its designated use, such as supporting aquatic life. Assessment is part of the Minnesota Water Quality Framework's Watershed Restoration and Protection Strategies (WRAPS) process — the comprehensive monitoring and assessment of Minnesota's 80 major watersheds (United States Geological Survey Watershed Boundary Dataset subbasins (HUC08)) on a ten-year cycle. This process includes identifying water quality impairments, sources of pollution, areas in need of protection and restoration, and strategies to achieve and maintain water quality standards and goals.

Stream assessment requires the review of existing information as well as the collection of new data. Of particular importance is the identification of karst features, because karst features can play a controlling role with respect to a stream's physical, chemical and biological properties. Existing information may come from previous work by the MPCA or, likely, from other agencies. These may be other state agencies, or may be federal, local or non-profit agencies. Existing information may also come from landowners or local land users (e.g., hunters, fishers, birders, fungi and plant collectors).

The opportunity for collaboration among these agencies is great; the challenge is lack of resources (monetary or time) or organizational barriers, which may inhibit proactive communication and collaboration. Simply stated, one agency may not know what data and research another organization has or is working on. When data is not proactively shared or accessible and searchable, it is not practically usable.

Such was the case with the Minnesota Department of Natural Resources (DNR) stream documents of southeast Minnesota. The MPCA and DNR have a strong history of collaboration, but the challenge was that the paper

documents in file cabinets were not in an accessible or easy-to-use format; as such, they were in a 'data silo.' In the summer of 2014, MPCA staff recognized that there was an immediate business need for the DNR's stream documents to be in a digital format. Therefore, a pilot project began between the MPCA and DNR to digitally preserve the stream documents which capture qualitative and quantitative data about the hydrology, water chemistry, geomorphology, biology, land use and karst features of southeast Minnesota streams. The MPCA recognized that this data was valuable and could be an important resource for stream assessment (e.g., identifying interrelated factors impacting a stream's biologic community), especially because these documents contained information about the location and characteristics of karst features.

By digitizing these stream documents, the data would become:

- Protected from physical damage or loss.
- Easily accessible, searchable, shareable.
- A collective resource enhancing one another, thus producing a more spatially and chronologically complete story of a particular stream.

Process

The task was to preserve the data contained in the documents so as to make the data usable. This was accomplished by converting the documents into a digital format (Adobe PDF, GeoTIFF, ESRI Feature Class). Proceeding by watershed, MPCA staff removed folders from the file cabinets one at a time. Each document within a folder was carefully prepared by removing staples and repairing any damage prior to scanning each to a computer in Adobe PDF format (these are herein referred to as the digital documents). In addition to the scanned copy saved on a computer, a backup copy was saved to an external hard drive. Documents were then reassembled and returned to their respective folders in the file cabinets. Finally, folders were marked to indicate that the scanning of all of the documents in each folder had been completed. A key part of this process was the document-naming scheme. Each digital document was named in a way to effectively "tag" it with keywords such that it could be seamlessly loaded into the MPCA's OnBase document management system for storage and retrieval. The digital document-naming scheme also allowed for the digital documents to be accessed through ESRI ArcGIS. Once scanned, Adobe Acrobat Professional software was used to make the digital documents text searchable and to create a searchable index. The indexed documents were organized and placed into folders by document name (e.g., stream assessments, survey reports, survey summaries, etc.).

To incorporate the digital documents into a GIS format, many steps were taken to ensure all documents were linked with their related stream feature class. After the scanned documents were indexed with Adobe Acrobat Pro, a log file was created of all the file paths and document names. The log file was imported into a Microsoft Excel Workbook where file paths were organized by theme, each theme a Microsoft Excel Worksheet (e.g., assessments, survey reports, summaries, etc.). Worksheets were then converted to an ESRI geodatabase table. In ESRI ArcMap, each theme table was related by kittle number (unique stream numbering system developed by the MN DNR) to a stream feature. As a result, users are able to use the Identify Tool to view PDF documents associated with a stream reach, simply by clicking on the feature in the ArcMap software.

In addition to the stream documents (which were generally typed or hand-written text documents including a few illustrations or maps), there were 147 large (greater than 21.59 cm x 27.95 cm) maps. These maps were generally USGS 7.5 minute series topographic map quadrangles, with hand-written annotation. On these maps were identified springs, seeps, electrofishing stations and other notes, such as water temperature and qualitative land cover descriptions, such as "good pheasant habitat." These maps were sent to a private contractor who was able to scan these larger size documents. The maps were returned to the MPCA as digital images (in Tagged Image File Format [TIFF]). Within the ArcMap environment, the digital images were georectified, or geographically referenced. Once georectified, the hand annotation could be digitized; in other words, points of interest such as springs in the digital image were made into a GIS feature class.

Discussion

Thus far, over 4,000 paper documents have been made digital, text searchable and indexed. In practical terms, a MPCA staff member could search for a certain word within a digital document or within the whole collection of digital documents, enabling a stream assessment team

to easily search for karst features mentioned within a stream's digital documents. Furthermore, in OnBase the digital documents serve as a back up to the paper document in the DNR's file cabinet. Within OnBase, documents can be found using keywords such as the DNR Stream Kittle Number, year (approximate date of creation of the original paper document) or document type (stream survey, stocking record, stream assessment, etc.). The storage of these digital documents in OnBase should be a more efficient way (rather than researching paper documents while physically sitting in the DNR office) to manage and share this data within the agency as well as outside the agency. Furthermore, the text searchable digital documents allow for copying and pasting text from the digital document to another type of document, such as a word processing document.

In ArcMap, the relationship between the stream feature class and the tables (tables that include the file paths to the digital documents), allow MPCA staff to use ESRI ArcMap's Identify tool to select a stream of interest and be provided with a list of digital documents associated with that stream. Using ESRI ArcMap's Hyperlink functionality, MPCA staff can click on a document listed within the Identify Tool window and open the digital document in Adobe Acrobat Reader — effectively digitally opening up the DNR folder, formerly only possible by visiting the physical office. The combination of accessing the digital documents from ArcMap and the text searchability of the digital documents allow MPCA staff to not only find springs mentioned within the digital document, but also to digitize that information as a spring feature (that is geographically referenced) in ArcMap, based on the information in the text. For example, a digital document's text that reads "spring, 0.5 cfs, 200 meters upstream from Bridge 2" may be sufficient to orient the reader to that digital document; however, when that spring location is captured as a geographically-referenced feature in a GIS feature class, it can be examined in context with other features on a sub-basin-wide scale. As these springs and other karst features are digitized through this process, they can be appended to the respective feature class.

From the 147 large format maps, digitizing was conducted to create a GIS feature class of springs from the hand-annotated locations on the maps. Next, the feature class was compared to the existing spring features within the Karst Feature Database created by the Minnesota Geological Survey and University of Minnesota. The comparison resulted in the determination that this process had identified 400 springs and other karst features not already in the Karst Feature Database. These 400 new features were shared with DNR, Minnesota Geological Survey and University of Minnesota staff in hopes of field verification and eventual inclusion within the database.

The great benefit of this pilot project is providing Minnesota Pollution Control Agency staff access to robust existing information for their stream assessment process; this information is particularly important for the complex karst region of southeast Minnesota. This was accomplished by liberating data about southeast Minnesota's streams and karst features from the 'data silo' that is a file cabinet of paper documents. The resulting digital documents, digital images, and GIS feature classes has made the data more manageable, accessible, and shareable.