STATE OF ALASKA
DIVISION OF
GEOLOGICAL &
GEOPHYSICAL
SURVEYS

Alaska
GeoSurvey News

http://wwwdggs.dnr.state.ak.us

Vol. 4, No. 2, June 2000

DGGS SCANNING PROJECT

by E. Ellen Daley

OVERVIEW

DGGS has long recognized that limited accessibility is a major impediment to the dissemination of geologic information, especially in Alaska where research libraries exist only in Juneau, Anchorage, and Fairbanks. Publications by DGGS and its predecessors contain a wealth of information about the basic geology of Alaska; mineral, oil and gas, and water resources; geologic hazards; land status and use; and even archaeological resources. Access to information of this type can be critical to private industry, government, and individuals for planning and making decisions that will enhance the health, safety, economic, and environmental well being of Alaska and Alaskans. Through funding provided by the USGS under the Minerals Data and Information Rescue in Alaska Project, DGGS is taking steps to make our publications widely available and the information contained in them easy to find.

Since July 1999, a concentrated effort has been underway at DGGS to convert several thousand of our publications and maps to electronic files and to make them text-searchable and available to the public via the Internet. Those who have visited the DGGS website lately (http://wwwdggs.dnr.state.ak.us) will notice that several recently issued publications can be downloaded at no charge. Publications (including text, data, maps, etc.) DGGS has released over the past few years were created as electronic files, so it is generally a simple matter to convert them for easy downloading from the Internet. These documents can be viewed on most computer operating systems (typically as Adobe Acrobat PDF files). Documents created in prior years, from the early 1900s through the mid 1990s, are currently available only in "hard copy" paper or Mylar formats. The task of creating usable electronic files from those documents is significantly more complex.

Our current estimate is that DGGS has approximately 2,400 documents in hard copy, consisting of around 65,000 pages of text and just over 3,000 maps.

Our current estimate is that DGGS has approximately 2,400 documents in hard copy, consisting of around 65,000 pages of text and just over 3,000 maps. These include publications in all of DGGS's document series and those of its predecessors (such as the Territorial Department of Mines).

Ultimately, DGGS envisions that our collection of documents will be fully searchable on line, and retrievable at no charge to view on screen, download, or print. While most publications will still be available for purchase from DGGS in hard-copy format, we anticipate that many of our customers will prefer to access them on line. Not only will this be more convenient for many customers, it will significantly reduce DGGS's future expenditures for printing and storing paper publications.

While significant progress has been made in converting the documents to electronic form, full implementation of web access, especially the advanced search features, will be a number of months down the road. However, you should begin to see selected documents on our website in the very near future.

The process of converting paper to bits and bytes is technically relatively straightforward, but logistically complex. The remainder of this article will give you a basic overview of the process and, more importantly, an idea of the format the information will be in and how to use it.

CONVERTING PAPER TO BITS AND BYTES

The first step in the process of producing electronic files from hard-copy documents is to scan the hard copies using flatbed or wide-format scanners. The next is to convert the image files to a format that will allow them to be used in most common computer operating systems while at the same time minimizing the size of the files. The final step, which is not discussed in detail in this article, is to design and construct the web access interface and the necessary storage system for the electronic library of documents.

The main obstacle encountered in this project has been deciding how best to minimize the sizes of the electronic documents so that they can be downloaded efficiently from the Internet, while at the same time preserving their legibility and fine detail. When a document is scanned, it becomes a "raster" image, which means that a uniform amount of information is recorded for each tiny area of each page or image, whether it contains information (such as text or graphics) or is

(continued on page 2)

simply blank space. Raster image files typically are very large compared to text or vector files: the size of a raster file is determined by the dimensions of the original, the resolution of the scan (typically in dots per inch) and the amount of information that is required about each "dot." Files become quite large when the document needs to be scanned in grayscale or color (rather than in black and white) because the amount of information required for each "dot" is greater.

Several approaches can be used to decrease the size of raster files. They can be downsampled to a reasonable file size, which in many cases grossly degrades the quality of the images. They can be compressed, which depending on the method can also significantly degrade the images. Or, in the case of the text portions of documents, they can be converted into "live text" through a process called Optical Character Recognition (OCR). We are currently using a combination of these methods.

Text Documents

The technology for scanning standard-sized documents (up to 11 by 17 inches) has been available for a number of years, and is a relatively simple, but labor intensive, procedure. Advances in automated document feeding and handling have made a task of this magnitude both possible and cost-effective.

DGGS has contracted this portion of the project to a firm in Anchorage, Alaska, that specializes in volume scanning and conversion of paper documents. DGGS collects, catalogs, and if necessary, repairs and unbinds the documents prior to shipping them to the contractor. The contractor scans them and converts them to Adobe Acrobat PDF files, OCRs them, and returns the electronic versions of two or more boxes of documents on a single CD.

The OCR software, as you might suspect, cannot recognize all characters, especially when the original document was faded, speckled, or is the photocopy of a photocopy of a photocopy, as some unfortunately are. Therefore, the files require varying degrees of correction before they can be released to the public. The two advantages of the OCR process are that the resultant text files are much smaller than the original raster files, and, even more important, that the documents are text-searchable. This means that users will be able to search every word of the entire collection and quickly focus on the documents that contain the information they seek.

Oversized Documents—Primarily Maps

There are several technical obstacles in converting oversized documents (primarily maps) to sufficiently small electronic files. In a typical geologic map, the smallest details must be preserved if the map is to be useful. These fine details often include symbols and text of a millimeter or less in size. In addition, some of DGGS's published maps are up to 4 feet wide and 6 feet long, so the combination of fine detail and large size can result in scanned images of over a gigabyte (1 million bytes). Not only is it currently futile to attempt to transmit a file of that size over the Internet, but also most personal computers are incapable of handling such a huge file. Significant downsampling of these files is generally unsatisfactory because of the amount of information lost. Converting the maps into "vector" files, where lines and polygons represent the image, would make them a fraction of their original size, but the current technology for "vectorizing" complex maps is inexact and prohibitively labor intensive.

Compression shows the most promise. Algorithms for compressing raster files to a fraction of their original size—without significant loss of information—have become available relatively recently. Compression ratios of 30 or more to 1 are achievable with a resultant image that is virtually indistinguishable from the uncompressed raster image at 1:1 scale. Compressed versions of very large color maps will still be very big (a 1 gigabyte file compressed at 30:1 will still be over 30 megabytes). Such files are impractical to download at this time, but they certainly can fit on a CD.

The maps will be released on DGGS's website as compressed raster images (a proprietary format from Lizardtech software called "MrSID"). Similar to PDF files, the compressed files will require a viewer. Free viewers are available for Macintosh, Windows, and UNIX operating systems—and our website will have a link to the site from which the viewers can be obtained. Although the viewers currently have limited printing capabilities, the image can be navigated on the screen very effectively, and can be exported to an uncompressed raster file at various resolutions.

Additional free plug-ins will allow certain GIS and imaging programs to use the files without decompression. According to GIS software developers, recent and future versions of some products will include native support for this format. Unfortunately, due to the labor required and the limited number of control points on many of the maps, few of these legacy maps, if any, will be georeferenced.

PROGRESS TO DATE

With the exception of the Territorial Department of Mines (TDM) documents, initial conversion of all of the available standard-sized publications was completed by late May. To date, more than 50,000 pages have been sent to the scanning contractor; 33,000 pages have been returned as electronic files and are in the process of being corrected. The TDM reports (approximately 12,000 pages) will be scanned by mid summer. Approximately one-third of the maps (1,000) have been scanned and are currently being processed.

NOTE: Mention of any company or brand name does not constitute endorsement by any branch or employee of the State of Alaska.

NEW PUBLICATIONS

PIR 2000-3. Technical review of the September 1999 groundwater disturbance near Ester, Alaska, by Jim Vohden, 68 p. $13.

RDF 2000-1. Major-oxide, minor-oxide, trace-element, and geochemical data from rocks collected in a portion of the Fortymile mining district, Alaska, 1999, by D. J. Szumigala, R.J. Newberry, M.B. Werdon, B.A. Finseth, D.S. Pinney, and R.C. Flynn, April 2000, 26 p., 2 sheets, scale 1:63,360. $28.60.

RDF 2000-2. Geochemical, major-oxide, and trace-element data from rocks collected in the Iron Creek area, Talkeetna Mountains B-5 Quadrangle, Alaska in 1999, by M.B. Werdon, J.R. Riehle, J.M. Schmidt, R.J. Newberry, and G.H. Pessel, x p., 2 sheets, scale 1:63,360.

ORDERING INFORMATION

For each publication ordered, include both the publication title and number. Mail orders are payable in advance. Make check or money order in U.S. currency and payable to the **State of Alaska**. Credit cards are not accepted. Telephone orders are accepted by the Fairbanks office between 8 a.m. and 5 p.m. Alaska time. Fax and email requests are accepted any time; these orders will be invoiced. If you would like to leave your order on voice mail, this can also be done 24 hours a day and you will be invoiced.

SHIPPING & HANDLING

- Domestic postage - $1.00/copy of each report
- Canada and Mexico - $1.50/copy of each report
- All other international - $2.50 surface
  $5.00 air/copy of each report
- For rolled-map orders requiring mailing tubes, add an additional $3.50.

WHERE TO ORDER

Publications of the Division of Geological & Geophysical Surveys are available over the counter, by mail, phone, fax, or email from the DGGS Fairbanks office:

ATTN: Geologic Communications Section–Sales
Alaska Division of Geological & Geophysical Surveys,
794 University Avenue, Suite 200
Fairbanks, AK 99709-3645
(907) 451-5020 Fax (907) 451-5050
Email: dggspubs@dnr.state.ak.us

Prices of DGGS publications are subject to change. Increases in costs make it necessary to raise the selling prices of many publications offered. It is not feasible for DGGS to change the prices stated in previous announcements and publications in stock, so the prices charged may differ from the prices in the announcements and publications. Overpayments of $2 or less will not be refunded.

Please send address corrections to:

Newsletter, Alaska Division of Geological & Geophysical Surveys
794 University Ave., Suite 200, Fairbanks, AK 99709-3645
fax (907)451-5050
email: joni@dnr.state.ak.us
http://wwwdggs.dnr.state.ak.us

If you have access to the Internet and would like to receive this newsletter via email, please drop us a line at:
dggsnews@dnr.state.ak.us

Alaska Department of
NATURAL
RESOURCES

State of Alaska
Department of Natural Resources
Division of Geological & Geophysical Surveys
794 University Avenue, Suite 200
Fairbanks, AK 99709-3645

Dear Readers:

The DGGS legacy-document scanning project represents an important but modest beginning to solving a larger problem. Each year more of our customers have the capability and need to electronically extract, analyze, and recombine geologic data enhanced by the value of their intellectual insights. Both our customers and DGGS face expectations of faster responses to ad hoc inquiries from people that need geologic information for making better decisions. These expectations cannot be met if access to the body of Alaska geological and geophysical data is limited to conventional, paper-based documents. DGGS is exploring ways to go beyond the Internet-accessible electronic publications described in this issue of *GeoSurvey News* to eventually provide on-line map and data files configured for use in popular statistical and geographic information system (GIS) software. In the interim, the collection of electronic documents made available by the present project extends to all of our customers the same access to DGGS geologic data that exists in our offices.

Sincerely,

Milton A. Wiltse

Milton A. Wiltse
Director and State Geologist

Visit our web page at http://wwwdggs.dnr.state.ak.us

STATE OF ALASKA
DEPARTMENT OF NATURAL RESOURCES
DIVISION OF GEOLOGICAL & GEOPHYSICAL SURVEYS

Mail order to:

Division of Geological & Geophysical Surveys
Attn: Geologic Communications Section
794 University Avenue, Suite 200
Fairbanks, Alaska 99709-3645

E-mail dggspubs@dnr.state.ak.us
http://wwwdggs.dnr.state.ak.us
Phone: (907) 451-5020
Fax: (907) 451-5050

REPORT NO.	TITLE	QUANTITY	UNIT PRICE	POSTAGE	TOTAL
				Total amount due	

Domestic *postage $1.00 per copy of each report ordered*
Canada *and* ***Mexico****, $1.50 per copy of each report ordered*
All other international*, $2.50 surface, $5.00 air mail per copy of each report ordered*

Ship publication(s) to:

Name _____

Organization _____

Address _____

City, State _____ Zip _____

Payment must accompany your order. Make check or money order payable to **STATE OF ALASKA**. We cannot accept credit cards.