

# GanFeld: Geological Field Data Capture

By Guy H.D.P. Buller

Geological Survey of Canada  
601 Booth St.  
Ottawa, Ontario, Canada K1A 0E8  
Telephone: (613) 947-1889  
Fax: (613) 947-9518  
e-mail: [gbuller@nrcan.gc.ca](mailto:gbuller@nrcan.gc.ca)

## OVERVIEW

GanFeld:  
gan – Old English meaning Open  
feld – Old English meaning field

The collection of geological information has been an ongoing process in Canada even before the inception of the Geological Survey of Canada (GSC) by William Logan 162 years ago. Much of the geological information collected was based on the need to discover new mineral resources for a new country and to further understand geological phenomena. Geoscientists go out to the field, collect samples to examine more closely and model a geological process that often leads to a report on their findings. The difference between then and now is the speed at which things happen as well as the precision of the location, the accuracy of the results, and how information is disseminated to the world.

The importance of geological information in the past 40 years has expanded beyond the focus of pure research or mineral resource discovery. With this change in scope there is the need to share accurate, up to date information between different groups and disciplines, whose demands continue to grow. Geological surveys around the world acknowledge that they house a wealth of information that can be critical to the enhancement of business opportunities, the environment, and the citizens of their respective nations. Many of these organisations also recognize that the information that is easily available often is not current, or failing that, is difficult to discover, and is often in formats that are archaic to modern technology. These hurdles limit easy access and are not sympathetic to the demands of today's rapid and potentially critical decision-making processes. In response to these issues, a concerted effort has taken place to find solutions that make the products and complete data of geoscientists' work more accessible and of an assured high quality for today's needs.

To meet the above challenges and to streamline efforts, organisations are looking at the collection, processing and analysis, and final dissemination of informa-

tion from a business perspective. The different steps that are involved in making a geological report have been examined as critical parts of a whole process, rather than being end results in themselves. This 'holistic' view of the information process has shown where and how there can be immediate improvements in data quality and a reduction in the time it takes to publish final results.

A significant improvement in this information processing can be achieved by reducing our reliance on paper formats at various phases of geological research; this finding has been well documented by the British Geological Survey. By evolving from paper field notes to electronic field data-capture, we reduce the likelihood that transcription or scientific errors will enter the data collection process. Collecting information in this way also increases the ability to search and manipulate the field data. By using such electronic systems, quality assurance and quality control (QA/QC) begin at the inception of data collection.

## FIELD DATA CAPTURE

Many different computer applications are being used to capture field data electronically, and several groups both nationally and internationally (e.g., <http://www.bgs.ac.uk/dfdc/home.html>) are working hard to make data collection applications fit their specific set of business rules. In almost all cases, field applications have been developed 'in house' by manipulating existing software applications to fit into the requirements that geoscientists demand for data capture. Some of these applications are full geological mapping systems that "capture" nearly all the geologic information for a single point of interest, whereas other applications limit their information to XY coordinates with a few electronic notes and then compliment this electronic data with hand written notes in field journals. In the case of two full scope well-known field applications, *FieldLog* (Brodaric, 2004) and *GeoMapper* (Brimhall and Vanegas, 2001), a laptop computer is used to gather and hold information electronically; these two systems have had broad acceptance by researchers and have been very successful in their many varied

field deployments. Yet, the use of a laptop computer often means that field data gathering systems still rely on the use of paper forms and the manual transfer of this data to an electronic type format. Alternately, if true site-to-site specific information is to be captured using a laptop computer, a vehicle of some description (either a truck or a 4-wheeler) is required to transport the computer to each location due to the weight and size of laptop computers. This can also mean that special vehicle mounts are required to hold the computer in place on board the vehicle.

## Handheld Computers

With the advances of technology, computers have become smaller and personal digital assistants (PDAs) have become more powerful, more rugged, and more conducive to being used while on traverse. An initial trial in Canada of electronic data capture using PDAs was researched and developed by Gilbert, Parlee and Scott (2001) and further extended by Celine Gilbert and Edward Little (GSC, oral communication) using the Palm handheld devices that leverages the power of Microsoft Access through software known as Pendragon Forms. The continued success of these data gathering devices in arctic terrains have prompted their use in other areas of Canada and have proved themselves as reasonably inexpensive mobile field data gathering systems. The idea of using a truly relational data structure for capturing data has been extended further with the goal of loading information directly into a data base structure that mimics the data structure found at the corporate level (Buller, 2002). This single database structure concept would greatly facilitate the transfer of information to a corporate data holding and reduce the data manipulation needed to facilitate the transfer.

While many of these systems work well for an individual project or survey, extending their functionality to other projects or groups often means substantial redevelopment of the application. In real terms, this means that there has been no improvement in the level of data access or sharing, because data created with the various different applications is not easily exchanged; in reality, we have only altered the format, from paper to various, less accessible electronic formats. Furthermore, the information captured or the terminology being used for a project may be specific to a single researcher and thus only may have a life span equal to the length of time that the geologist is employed at the organization. This lack of interoperability is well known amongst researchers and organisations, and much work has been done to 'translate' information into different formats to enhance communications between these different groups (Brodaric, 2004). These data translation activities have met with limited success and are recognised as large consumers of time and resources.

## Searching for Solutions

The Canadian federal government's *Government On Line* initiative ([http://www.gol-ged.gc.ca/index\\_e.asp](http://www.gol-ged.gc.ca/index_e.asp)) intends to have most government information available on line by 2005. This initiative has been an incentive to put geological data into electronic formats and to increase the accessibility of this information via the Internet. To achieve these ends, the overall business of geological information collection and distribution needed to be examined.

At the Geological Survey of Canada (GSC) Terrain Sciences Division in Ottawa, there has been a concerted effort over the past few years to streamline the workflow of sample processing. As a result, the GSC has developed a laboratory information system (LIMS). The LIMS has been effective in improving the quality of the analytical results, and assists in expediting quality information transfer to the final publication process (R. Laframboise, oral communication, 2002). As the LIMS was being finalized, steps were being taken to integrate geochemical data across three divisions of the GSC into a common geochemical data structure. These developments were seen as part of the foundation that would facilitate a geochemical mapping web presence using ESRI's MapObjects, thus making large gains towards the government's on-line initiative.

The MapObjects application requires ESRI Shapefiles to deliver maps to clients via the web. In the case of many raw data gathering systems information is captured in formats that are not visual and therefore are not spatially referenced. This difference in data formats is similar to the translation problem mentioned earlier, in that there is a need to manipulate raw information to develop Shapefiles for use in GIS systems. This activity of altering data formats may not be problematic but it splits the raw information into a spatial file and a data file. This division of information is counter productive and it is felt that capturing map data directly in the field on a station-to-station basis would be an effective way of streamlining the information process while at the same time capturing vital geological information. The challenge to capture this sort of map data as well as other information was met with the development and release of ESRI's ArcPad handheld mapping application (<http://www.esri.com/software/arcpad/index.html>).

## ArcPad

The ArcPad application has essentially put a GIS in the geologist's pocket, giving them the ability to plot a variety of map information (polygon, line and point) and to directly capture point information from a GPS receiver. By setting up the Shapefile data table (a DBF file) to

capture a wide range of other information in addition to spatial data, it is possible to have the best of both worlds, a digital data capture system as well as a visual display of the map data while out in the field. A further advantage with having a map interface is that other map information, such as gravimetric data or geological feature sets (polygon data, outcrop delineation sets, etc.) can be accessed using the same device, and viewed in the field with newly captured map data. This combination of spatially-related raw data and maps effectively means that researchers at the end of a field season have a preliminary map that is available for publishing, as well as an easily searchable data set that is geographically referenced. ArcPad has allowed us to reach a main goal, which is to better help the geologist in their work.

Though there is some resistance toward the new way of collecting data it must be kept in mind that, at one time, the use of paper forms in the field was considered as an inconvenience to the geologist but are now often seen as an indispensable aid to the systematic capture of information in the field. Creating electronic forms allows the geologist to retrieve, share, and examine data more easily, and in turn allows the geologist to think about geology rather than be concerned with the input of raw information into computer systems. As an additional bonus, any functional coding to customize ArcPad is done using VBScript, meaning that web developers who build active server pages (ASPs) can transfer their expertise directly to the development of ArcPad applications.

## Field To Curatorial Project

Over the past year there have been far-reaching changes at the GSC. These changes have instituted a project-based system, and the *Field to Curatorial Project* (FTC) is one of these projects. This project's main goal is to track a physical sample from initial collection, through processing and analysis, to the archive. The desire to have this broad spectrum of information accessible at all times was seen by management as an important goal. Under these guidelines the FTC project has three distinct modules (field, laboratory, and archival). The need for continued refinement and development of a robust field data-gathering application became clear, as it is the first step to making field data more interoperable between groups and the project modules.

To meet the goals of the FTC project it was necessary to look at the work that geologists do and to consider this work from a business model perspective. This modelling was extended to the field module and at the earliest stages of the project, it was recognised that although geologists do similar activities they do not always use the same language for describing these activities or common things in the field. Therefore, one of the first steps for the field

module was to try to standardise some of the common terms and expressions that are used in three different phases of geological field research. These strictly field related phases of geological research are an arbitrary breakdown and have no set time period and are recognised as pre-field research, fieldwork, and post-field data manipulation or research.

During the first phase there can be a considerable amount of paper research for an area that may include interdisciplinary discussion. Often, common terms are used in several different ways and are mostly non-scientific words that are used in daily speech. These words are often applied to items or actions specific to the discipline or to the researcher during the various phases of geological research activities. Between researchers the 'translation' of words to gain meaning is not a problem as it is an organic process that humans use in everyday conversation, but because there is no such intrinsic translation between computers, this level of ambiguity causes havoc when dealing with databases. In order to facilitate the input of data into relational database systems, terms that have specific definitions must be agreed upon by a variety of users.

These words have been developed by consulting different ISO publications, considering other developments within the GSC as well as soliciting information from a number of researchers. The original set of words were then reviewed by a number of people inside and outside the project, and have become the starting point of a lexicon to be used for information collection standards. Table 1 shows an example of some of the developed lexicon.

The set of common definitions in turn determines the minimum information required for any geological project, and gives more consistent information sets between different field projects. Through the use of a questionnaire, geologists were surveyed to gather information about the various aspects of fieldwork carried out in different parts of the GSC. These questionnaires were followed up with discussions and meetings as a way of clarifying the modeled work process and determining clear definitions for terms that an individual researcher uses. It must be kept in mind that the common word set is not a static entity, as new words and definitions will be added over time. This type of development, which builds a common word system, is similar to other efforts (e.g., ISO Standards) in attempting to standardise an activity that is carried out by many individuals in a branch of research. By following such a standard system, an individual's specific words can be matched to the common word set and, subsequently, field project-specific data can easily be transferred and stored in a relational database along with other inter-department survey data. This common word set accomplishes two goals,

- the facilitation of data input to a common model, and
- a reduction of data ambiguity both now and in the future.

Over time, the accumulated information will become more of an asset to the pre-field research phase and will allow researchers to more easily share post-field information. Furthermore, general field information, such as observations about vegetation or morphological descriptions outside of the research scope, can also be extended to web applications or web services to further promote geological works and to demonstrate to managers and general public, in a timely manner, of the work that is being completed by a research group. This up-to-date information becomes more important when dealing with regions that have surface-access-rights issues or areas of a high sensitivity (i.e. environmentally protected areas) and gives a clear indication to the general public as to the extent of the work being done by a party.

Part of the challenge faced by the FTC project was the introduction of a new business paradigm by management that has focussed on stronger accountability of expenses and more intra-department collaboration of activities. Thus, in preparation for this change in business

process and before any of the software development began, measures were taken by the FTC project Information Management (IM) coordinator (Richard Laframboise) to extensively plan the project's time line. Over the past year the use of business requirements analysis has been our main focus in an effort to both document the development of the project and also give direction and communication to the various groups within the GSC that are distributed throughout the country.

This business-centric effort has placed a large emphasis on the planning stages of this project and has used the Zachmann framework model and business requirements analysis as demonstrated by Hay (2003). These planning and analysis activities have been invaluable in understanding the scope of the application and the roles of the different individuals who are involved in the FTC project. In the past, this type of planning activity has had limited use because many previous projects have had impacts that were limited to very small groups or individual researchers. As there is a desire by management to expand the extent of web accessible information and to have more accountability, a project of this kind needs to extend its contacts to as many groups as possible in order for the long term planning phase to be successful. The use of the planning tools have been invaluable for focusing the

**Table 1.** A working example of the lexicon that has been developed at the Geological Survey of Canada's Field To Curatorial Project, for the collection and storage of field information. (Note: Information in *bold italics* indicates an edit to the lexicon that has yet to be considered as accepted).

Word	Definition	Source
<b>Sample</b>	1) Portion of material selected from a larger quantity of material. 2) The raw material collected in the field and shipped back to the lab.	1) ISO 11074-2 2) Geochemistry Database concept
<b>Specimen</b>	Specifically selected unit/portion of a material taken from a dynamic system and assumed to be representative of the parent material at the time it is taken.	ISO 11074-2
	NOTE 1: A specimen may be considered as a special type of sample, taken primarily in time rather than in space.	
	NOTE 2: The term "specimen" has been used both as a representative unit and as a non-representative unit of a population, usually in clinical, biological and mineralogical collections.	
<b>Activity</b>	<i>An action carried out by a field party at a specific station that in some way gathers information about that specific field station. Examples: observations (including no activity), picture, drawing, sampling.</i>	<i>Guy Buller</i>
<b>Sampling</b>	Process of drawing or constituting a sample (ISO 3534-1:1993).	ISO 11074-2
	NOTE: For the purpose of an investigation, "sampling" also relates to <i>in situ</i> testing carried out in the field without removal of material.	



development plan for the present fiscal year and helping the individual developers recognise how they fit within the project itself.

## CONCLUSION

Computerized mapping is finally being widely adopted for field use. Continued advancements in technology will make data collection systems commonplace and will be an even greater asset to the geologist in the future. As the costs of running a field camp increase, no longer do we have the luxury of letting field data collections languish in obscurity by having data in a multitude of formats that are not interoperable or easily accessible. The data collected by the geologist at an individual station has importance, as do observations and 'thoughts'. These thoughts, at the time of data capture, can give clarity to a geological model when contrary models are introduced. Furthermore, field data that today may seem unimportant may in the future become extremely useful.

Any data gathering system that is developed, regardless of operating platform, must have interoperability of data as a main goal. This means that some of the focus of any development has to be centered on a data storage structure that allows researchers to use any data capture tool available. This data storage needs to be able to extend the availability of data to researchers and also allow for single queries to access multiple, seemingly disparate data sets.

The planning stages for the Field to Curatorial project have been most helpful in understanding the scope of the project. It is hoped that by following such a stringent planning stage, others who intend to develop a similar system can learn from the process that is being documented. The planning stage is critical to determining the actual needs of the business and the process to meet those needs.

To simply capture field information for the specific use of a single researcher limits the sharing of data and ultimately does not advance the scientific process. The information must be accessible by others, now and in the future, in order to serve the public and the science.

## REFERENCES

- Brodaric, Boyan, 2004, The design of GSC FieldLog: ontology-based software for computer aided geological field mapping: *Computers & Geosciences*, v. 30, issue 1, p. 5-20, accessed at <http://www.sciencedirect.com/science/article/B6V7D-4B1SFSX-1/2/e76e0967fde762328b7080f865c7fbc8>.
- Brimhall, George, and Vanegas, Abel, 2001, Removing Science Workflow Barriers to Adoption of Digital Geological Mapping by Using the GeoMapper Universal Program and Visual User Interface, *in* Soller, D.R., ed., *Digital Mapping Techniques '01—Workshop Proceedings: U.S. Geological Survey Open-File Report 01-223*, p. 103-114, accessed at <http://pubs.usgs.gov/of/2001/of01-223/brimhall.html>.
- Buller, Guy, 2002, Ganfield: data integrity from field to final product, *in* *Capturing Digital Data in the Field Workshop*, British Geological Survey, Nottingham, England 25 & 26 April 2002, accessed at <http://www.bgs.ac.uk/dfdc/gbuller.html>.
- Gilbert, C., Parlee, K., and Scott, D.J., 2001, A Palm-based digital field-data capture system: Geological Survey of Canada, *Current Research 2001-D23*, 10 p.
- Hay, D.C., 2003, *Requirements Analysis: From Business Views to Architecture*: Prentice Hall PTR, Upper Saddle River, New Jersey, 496 p.

## HARDWARE AND SOFTWARE CITED

### Operating System

PalmSource, Inc., 1240 Crossman Ave., Sunnyvale, CA 94089, (408) 400-3000, Fax: (408) 400-1500, accessed at <http://www.palm.com/us/>.

### Hardware

palmOne, Inc. Corporate Headquarters, 400 N. McCarthy Blvd., Milpitas, CA 95035, (408) 503-7000, Fax: (408) 503-2750, accessed at <http://www.palm.com/us/>.

Hewlett-Packard Company, 3000 Hanover St., Palo Alto, CA 94304-1185, (650) 857-1501, Fax: (650) 857-5518, accessed at <http://welcome.hp.com/country/hk/en/welcome.html>.

### Software

ESRI, 380 New York St., Redlands, CA 92373-8100, (909) 793-2853

Pendragon Software Corporation, 1580 S. Milwaukee Ave., Suite 515, Libertyville, IL 60048, (847) 816-9660, Fax: (847) 816-9710, Sales Email: [info@pendragonsoftware.com](mailto:info@pendragonsoftware.com), Support Email: [support@pendragonsoftware.com](mailto:support@pendragonsoftware.com), accessed at <http://www.pendragon-software.com/index.html>.