A Conceptual Approach to the Development of Digital Geological Field Data Collection

By Guy Buller

Northern Division
Geological Survey of Canada
255-601 Booth Street
Ottawa, Ontario K1A 0E8
Telephone: (613) 947-1889
Fax: (613) 947-9518
e-mail: gbuller@nrcan.gc.ca

ABSTRACT

Members of the Geological Survey of Canada (GSC) have for several years been successfully developing digital systems that aid geologists in the capture of field data. In the past, development has been completed, or driven, by an individual researcher on a per project basis and, therefore, systems have been specific to that geologist's work. This sort of application development has often meant that the work takes place in virtual isolation and the resulting application can be very limited in scope or usability for other researchers.

Due to the demands of business re-alignment in the GSC over the past few years, there has been an attempt to work toward a single system that could be used by a variety of researchers for the collection of field information. To accomplish this broad-spectrum development, the work has been conducted in coordination with many mapping projects; this has proven advantageous in coordinating development between many projects across the organisation. By following this strategy, the GSC is attempting to bring consistency to the data-gathering efforts, and thereby also minimize the isolation of projects, which was a problem in the past.

THE VIEW

Fieldwork and data gathering processes that are carried out using pencil and paper are in no way flawed, but the raw data from the fieldwork can appear cryptic because of an individual's unique techniques or terminologies, or the specific goals of the project. Furthermore, because there are repetitive aspects to fieldwork, the mapper commonly develops an individualized note-taking style that includes various abbreviations and other "short hand" techniques that provide short cuts to limit the amount of writing that is necessary. Often, the data collection location is not as idyllic as shown in Figure 1, and the amount of short hand can be dependent on the amount of time available at each site or proportional to the number of biting insects (Figure 2) or the temperature. These short cuts are easily understood because they are in context with what was recorded in previous day's work (e.g., "SOS" may mean "Same Old Stuff"), but this ambiguous information, over time, loses its meaning.

Short cuts are most used by people who are trying to solve problems or make progress under tight time constraints (Shalloway and Trott, 2004). However, they are impossible to process electronically, because the context cannot necessarily be captured and the number of short cuts and their meanings are unlimited. The use of the term "24-7" has come to mean "all the time"; although the context is not present, it is 'understood' by nearly everyone. On the other hand, project-specific short cuts are seldom as widely understood. Attempts to interpret the meaning of such short cuts may result in information or field observations that may not have been the intent of the original researcher. This may not be a problem if the individual is available to resolve any ambiguity, but with the passage of time, the researcher will become unavailable, to put it gently. As a result, attempts to convert old information into a database can introduce significant errors in data and scientific interpretations.

In the past, the principal developers of field data collection systems were the individuals who conduct the scientific research, and they addressed the data collection issues of their own project. By using a variety of software applications, data gathering systems have been developed, with inherent, project-specific short cuts. This sort of development has been very effective, because the person who controls how the information is to be collected or interpreted also can make any changes to the application that may be required. In some cases, however, these applications can imbue the data with a regional or research specific flavour that may be rather unique, even though geologic principles and observations are the same for any project.

Figure 1. Quaternary data collection in a pipeline trench (photograph by A. Plouffe, 2005).

Figure 2. Data collection in North Western Alberta (photograph by A.Plouffe, 2005).

These unique approaches to data capture can also be a product of application development, as a researcher is faced with a short preparation time prior to the field season. If a suitable beta application is developed and successfully meets the immediate needs of the project in the first season, it is probable that with each successive year the application will gain more functionality. With each subsequent year of use, the application becomes more entrenched into a specific data collection format and subsequently becomes less accessible by researchers outside of that specific project.

The building of applications to meet individual project needs has worked well in the past but more multidisciplinary, cooperatively-driven projects with diverse expertise have come into vogue, and systems that are developed as described above do not easily transfer to these larger, more complex projects. In order to facilitate another group's use of an application, a redesign of some sort must take place and the existing application is often patched to address the needs of the project. Changes to the application commonly are determined by decomposing it into its functional parts and if problems are found, those specific parts are modified. This functional decomposition (Shalloway and Trott, 2004) is a natural way for people to understand very large systems, but sometimes these modifications cause other parts of the system to be adversely affected. Further changes can mean that the modified parts become even more bound to certain field-specific functions and there is no "graceful evolution" (Shalloway and Trott, 2004) toward a deployable solution or to any new requirements that may occur.

Due to this project-specific development, there now exist many different applications that do not communicate well with one another. Furthermore, the maintenance of such systems over the long term becomes onerous for the researcher; time is spent tending the application rather than concentrating on their science. In some situations the entrenchment of specific systems is so strong that there is reluctance to change to newer systems, which in turn creates a certain unwillingness toward sharing and storing data within a corporate system. Yet, no matter the process of collection, all of the data that researchers accumulate is important to both the organisation as well as the scientist.

MOVING FORWARD

Many individuals recognize that much of the geological information gathered in different projects is virtually

the same in terms of content, although the style of reporting the observations, including format and terminology used, may differ greatly. This recognition has been an impetus to develop field data-capture and data-storage standards that are based on broadly accepted international standards. Some standards development had begun at the GSC (Buller 2004) but was very closely tied to the activities of an individual division at the GSC rather than being generic to the discipline of geology. To achieve a better model and to adhere to international standards it was decided to follow the work coordinated by the Open GIS Consortium (OGC). The OGC's release of a paper addressing Observations and Measurements (Cox, 2003) has helped to advance the conceptual modelling for electronic data collection, and the OGC model fits well with the way the GSC scientists collect information. Work is being done to formulate a physical model that can be applied to actual field activities and will form the foundation for the new modular approach of the latest version of the digital field collection application.

As projects move to an electronic system of data capture, they must rely more heavily on IT professionals, who serve as the bridge between the final corporate database and the users of field systems. The ability to understand the needs of the researcher is paramount as they are, for our field system, the end users; as such, they provide essential guidance on how the system should be designed. At the same time, objectivity during development must be maintained because there is a strong tendency to focus on the needs of a single (perhaps dominant) client, thereby risking the possibility of making development too project-specific. With lack of objectivity there is no change from the existing development process but rather simply a change in who does the development work.

The advantage of a non-project-specific development group is that there is no vested interest in any existing system; instead, the group focuses on the needs of all researchers. By being the 'interpreters' of the existing diverse collection systems, the development group is able to mingle concepts together and develop a unique, customized view of the data collection system that is based on an all-encompassing generic framework. This approach is similar to having a personalized desktop on an office computer while running on a common network. For any new system development however, both the developers and end-users must be able to approach the process in a cooperative manner and accept that the application should extend beyond the specifications of any one project. It must be recognized that a complete analysis of field data gathering techniques needs to be undertaken in order to understand how to develop systems that meet the needs of the whole organisation.

REQUIREMENTS

"One thing you will never hear (from developers)

is, 'not only were our requirements complete, clear and understandable, but they laid out all of the functionality we were going to need for the next five years!"

As Shalloway and Trott (2004, p. 6) pointed out in this quote, initial requirements are not written in stone. When clients are presented with an application having some broad level of capabilities, they quickly can envision many other possible uses for the device, and so demands for future editions are soon developed. This means that requirement analysis is an on-going activity, and everyone can expect that changes are inevitable. To help consider these design changes, systematic business analysis using a commonly accepted approach, such as the Zachman Framework (Hay, 2003), allows programmers and users a long-term view of the development life cycle that clearly demonstrates the steps needed to meet existing requirements of a project. Business analysis is an iterative approach, and such an approach can yield better design criteria and flexibility to adapt to changing requirements. By using this approach it is expected that activities that are overly project-specific will be winnowed out and only common categories will remain to allow for generic object modelling. This modelling will be important in the development of precise object classes to facilitate the transfer of data to corporate databases.

Though the analysis approach is complete in its understanding of systems, it can sometimes run counter to project objectives that have specific mandates to produce something tangible in a limited time frame. There is always an implicit desire for a development team to have a final product that will be useful for many years, but in order to meet the short-term goals of a project these long-range plans often are sacrificed. To further exacerbate development barriers, resources often are extremely limited and yet the expectations of end users are very high.

The barriers to meeting long-term corporate goals can be overcome, but it must be understood by managers that system analysis in many organizations has not reached maturity and the learning curve for understanding and implementation is steep. The analysis activity simply produces the blue prints to the application and only models how a solution will be developed based on the requirements discovered. For people who require "real answers" and are not familiar with requirements analysis, this stage of development can sometimes be thought of as non-work, as it only gives a path toward the solution, rather than the solution itself. There is a distinct need for application architects and software developers to be able to muster management support and understanding for this critical process of application planning. With proper analysis, the final product will be better suited to the needs of the user and will be developed through fewer iterations. Over the long term, a well-designed application will be easier to maintain, will be expandable and will cost less in development for any future modifications.

A SOLUTION

The first iterations of the collection system organized all the field observations into one or two Shapefiles. The linked dbf files held the collected information that was supported by a single multi-line (1000+ lines of code) script that controlled the user interface. The single, large script quickly became unmanageable, as developers, in two different areas of the country, were required to make rapid user-defined functional changes to the application. In a very short time, multiple iterations of the same application were available, with no concrete way of addressing the variety of "wish lists" that were being submitted by users.

To solve the problems, steps were taken to reformulate the script and make the system more modular by relating individual field activities to individual Shapefiles. This objectification simply models common activities of a researcher that take place at the various stations that are visited during a day of fieldwork. Activities for most geologists are very similar (Figure 3) where certain activities are followed by other dependent activities. For example, all activities must be related to a station, and a sample must be related to an earth material.

By functionally decomposing the work, the information-gathering process becomes compartmentalized. In terms of a final data product, rather than a traditional spreadsheet comprised of sixty or more columns, we consider the information to be the attributes of geo-
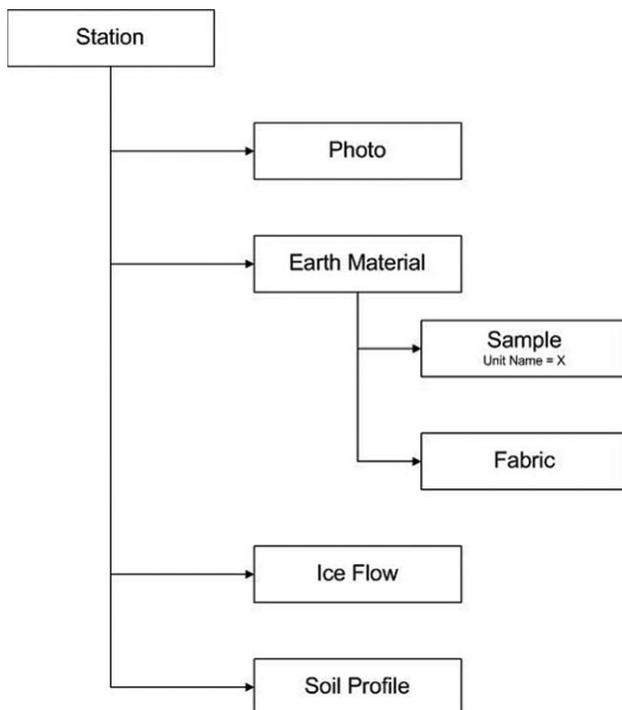

Figure 3. Field activity model.

referenced points placed directly on an electronic map. By dividing up the different activity sets (as shown in Figure 3) into distinct layers within a GIS, independent information can be collected in a loose relational format that leverages the GIS capabilities of the platform application. The extra number of layers does increase the number of files to be handled by the system, but it is a design trade off to allow for more functionality for the user. This latest development also tries to implement the recent OGC specification for field data capture, but does so without the burden of specifying any particular relational database. This 'distributed' spreadsheet flat file data holding can be easily transferred into a relational database system where the power of the database can be brought to bear on the collected information. This transfer is made feasible by the fact that the flat files have built-in relationships to the associated activity that has been previously captured (e.g., samples <u>must</u> have an earth material and earth material <u>must</u> have a station).

By applying requirements analysis during a planning stage and by relying on the experienced gained from previous year's development, changes to the application were kept in line with the goal of data transfer to a corporate system. With solid communication between developers and good team procedures along with individual component development that was tightly focused, changing requirements to suit our user's needs were easy to administer. In this way, the various parts of data collection are treated as discrete objects having specific attributes and properties and a single platform is able to have a multitude of functionality that can easily evolve over time. Furthermore, compartmentalized coding has helped to isolate any glitches within specific components, thus making them easier to discover and correct. Also, if one of the data collection modules does not operate properly, then only that component is unavailable, rather than breaking the entire application.

The development team has found that the length of development has become shorter since the implementation of the modular approach. The focus of development is on the module to be added, rather than on determining how it will fit with or affect the rest of the application. What used to take a couple of weeks of intense coding can now be shortened to a few days, depending on the complexity of the requirements for the new module. Most importantly, this means that a successful application module is not patched together, but instead is built upon the existing standardized business format and maintains the common end-user interface. Since the field system (which is called Ganfeld; see Buller, 2004) is a visual interface that leverages the GIS functionality of ArcPad, there was the need to develop a system from the ground up using a different design approach than had been used by designers of data capture systems in the past. By not focusing on past design, the development team was able to let go of the old applications and allow the new design processes to

advance more freely. This has resulted in a more flexible system that is easily adaptable to a variety of different foci of research and may even possibly be extended beyond geo-science projects, because the field activity model can be applied to any spatially related fieldwork.

IN CONCLUSION

It has become clear that the development of data collection applications cannot proceed in a non-systematic fashion. The ability to step back from an existing design and examine all possibilities allowed the development of a set of interrelated components. Also, the ability for developers to write and modify these components of the system without interfering with the whole application allowed for parts of the application to be delivered in time for the 2005 field season.

Business planning activities such as requirements analysis have not yet become mainstream; however, by following best practices in design we have been able to complete many of the goals that we set for ourselves. The need for an easily maintained system that can contain much of the scientific data collected is intrinsic to the many goals of an organisation. Also, the development of a field data repository that follows internationally accepted standards is required to ensure the preservation and access to all the information collected in the field.

It is expected that over time the various application modules that have been developed for field data capture will be altered to the point where certain modules will become a standard set of modules while more unique items turn into interchangeable components. As there is an existing modular design, these changes will be easily accomplished while still maintaining a level of base functionality.

More work is planned that will smooth out the rougher edges of the application, and procedures will be put in place to more easily consider changes in requirements. Further development of the field objects as per the OGC specification will continue, as will the realignment of data storage systems to contain this information.

REFERENCES

Buller, Guy H.D.P., 2004, Ganfeld: Geological Field Data Capture, in D.R. Soller, ed., Digital Mapping Techniques '04—Workshop Proceedings: U.S. Geological Survey Open-file Report 2004-1451, p. 49-53, accessed at http://pubs.usgs.gov/of/2004/1451.

Cox, Simon ed., 2003, Observations and Measurements Version 0.9.2; Open GIS Interoperability Program Report: Open GIS Consortium Inc., 4 April 2003, reference number OGC 03-022r3.

Hay, D.C., 2003, Requirements Analysis From Business Views to Architecture: Prentice Hall PTR, Upper Saddle River N.J.

Shalloway, Alan, and Trott, J.R., 2004, Design Patterns Explained: A New Perspective on Object Oriented Design, 2nd Edition: Software Patterns Series, Addison-Wesley Professional.