The Paleontological Stratigraphic Interval Construction and Analysis Tool

by

Joshua Allen Reed

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Human Computer Interaction

Program of Study Committee:
Cinzia Cervato, Major Professor
James Oliver
Chris Harding
Douglas Fils

Iowa State University

Ames, Iowa

2007

UMI Number: 1446141

UMI Microform 1446141

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Core description diagrams are the primary record of the cylindrical rock samples that result from the scientific drilling process. Typically, these diagrams are drawn by hand in field books and then drafted up in a graphics program for publication. Very rarely are the actual data encoded in the diagrams, e.g., depth in core, grain size, and lithology, captured in a format that can be manipulated and analyzed. This thesis introduces the Paleontological Stratigraphic Interval Construction and Analysis Tool (PSICAT), an interactive, cross-platform environment for creating, viewing, and editing core description diagrams, and discusses the design and implementation of its extensible software architecture and data model which allows it to seamlessly capture and visualize core description data. PSICAT was used to log nearly 1300 meters of sediment core drilled during ANtarctic DRILLing (ANDRILL) project's McMurdo Ice Shelf expedition.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

Core description diagrams are the primary record of the cylindrical rock samples that result from scientific drilling. As the cores come out of the ground, they are examined and described by sedimentologists—scientists who specialize in identifying and interpreting the "story" of the sediments in the sample. A few of the many things that the sedimentologists are looking for when examining the sample include: what type of sediment is present, e.g. is it sand? mud?; where does the type and size of the sediment change; are there special structures like bedding or faults presents; and are there fossil traces present. All of these features are represented pictorially on the core description diagram. This description process and the resulting diagrams are very important as they provide the scaffolding on which all further analysis is built.

On past drilling expeditions, the sedimentologists would draw these diagrams by hand, usually including four meters of core per page, and then pass them off to a dedicated drafts-person who would draft them up in a graphics application like CorelDRAW (Barrett et al., 1998; C. Fielding, Personal Communication, May 2005). This process worked, and it allowed scientists to create nice diagrams, but it had several shortcomings. Because of the complexity of the graphics software and of the information to be captured, it required a person who was dedicated to drafting up the diagrams. It also required the diagrams to be created at least twice—once by hand and then a second time in digital format. Once digitized, the diagrams were static images. This was a problem when changes needed to be made. Small, local changes to individual diagrams could be applied fairly easily, but making changes spanning multiple diagrams was a burden. Truly substantial changes, like changing the symbol that represents a type of rock, would have required re-drafting every diagram—an onerous task when there are diagrams for 1000 meters of core to update. Static images also complicated the task of searching for specific features in the information-dense diagrams because it required searching each of the individual diagrams for the features of interest.

A handful of specialized commercial software packages, e.g., Strater[1], WellCAD[2], and LogPlot[3], exist to aid the core description process. The main focus of these commercial packages is the final production of the core description diagrams. They require the description data to have already been manually captured in spreadsheets or databases. None provide an environment for simultaneously capturing and visualizing core description data.

To address the limitations of the traditional core description process, I have developed a tool

1 Golden Software: Strater Product Description. Golden Software. http://www.goldensoftware.com/products/strater/strater.shtml
2 WellCAD. Advanced Logic Technology. http://www.alt.lu/wellcad.htm
3 LogPlot 2005. RockWare, Inc. http://www.rockware.com/catalog/pages/logplot.html

called PSICAT, the Paleontological Stratigraphic Interval Construction and Analysis Tool. PSICAT is an interactive, cross-platform environment for creating core description diagrams that also captures the data represented in the diagrams. What distinguishes PSICAT from the existing commercial offerings is the novel way that it seamlessly integrates data capture and data visualization in an interactive environment. It mimics the traditional core description approach in that it provides the user with a drawing environment that she is familiar with, but as she draws, PSICAT is analyzing the diagram and extracting the data, e.g. depth, grain size, and lithology. This data is stored on the user's computer and PSICAT uses it to generate a visual representation that reproduces what the user had originally drawn. This whole process takes place in the background and is so rapid that it is imperceptible to the user. PSICAT is able to offer the best of both worlds: an environment that is familiar to the user so she does not have to drastically change the way she describes core and all the benefits of actually capturing the core description data in a form that can be analyzed and searched.

PSICAT is designed to support the complete core description process: 1. data capture; 2. visualization; 3. analysis; 4. collaboration; 5. revision; and finally 6. publication. It is also designed to be used by the broader geoscientific community, which includes multiple drilling projects, each with their own specific requirements and customizations. Since this is an ambitious and complicated undertaking involving many different stakeholders, types of data, interactions, visualizations, and analyses, PSICAT required a flexible and extensible software architecture and data model. When designing and developing PSICAT, I departed from the standard approach of developing a single, monolithic piece of software and instead embraced the idea of developing PSICAT as a series of individual components that could be assembled into a coherent application. This allowed me to re-use components that provided common functionality while still allowing for the development of one-off, custom components for specific drilling groups or tasks. As such, PSICAT is a platform for core description rather than simply a single piece of software.

PSICAT was not developed as a purely academic exercise; it was designed, from the beginning, for real world use. The first large-scale deployment of PSICAT took place in October 2006-January 2007 where it was used on-ice in Antarctica to log the nearly 1300 meters the sediment core drilled as part of ANtarctic geological DRILLing (ANDRILL) project's McMurdo Ice Shelf expedition. I accompanied the expedition to observe the users interacting with PSICAT and to fix any problems encountered. The opportunity to observe PSICAT in use under real world conditions and get feedback directly from the users was invaluable. Antarctica was the ideal place for this field testing because living and working in such close quarters with the users meant that there was no place to hide when things went wrong, providing ample motivation to really listen to the users and keep them happy.

Thesis Organization

The rest of this thesis is as follows: Chapter 2 is a paper titled "CHRONOS's PSICAT: Core Logging And Data Capture In One Tool" that I submitted in April 2007 along with Cinzia Cervato and Doug Fils to the special issue of Computers & Geosciences on Geoscience Knowledge Representation for Cyberinfrastructure. This paper provides a technical overview of PSICAT's software architecture and data model, and discusses the implementation and field testing of the software. It is currently under review for publication.

Chapter 3 is a paper titled "PSICAT: A New Open-Source Core Description Application" in preparation for submission to the Geological Society's online journal, Geosphere. It focuses on introducing PSICAT, including some of its noteworthy features, to the geoscience community and provides a discussion of the field testing experience and results. Geosphere was chosen because its publication format allows for the inclusion of full color screenshots and animations.

Chapter 4 provides a general conclusion to this thesis, including a discussion of results and suggestions for future research.

The Appendix is an article titled "How I Spent My Summer: Three Months in Antarctica with Eclipse and Java" which was published as the feature article in the December 2006 issue of Enterprise Open Source Magazine. The article discusses some of the technical aspects of the design and implementation as well as some aspects of the field testing that were not covered in the previous two papers. Enterprise Open Source Magazine is not a scholarly journal, so the text of the article was not included in the body of the thesis.

References

Barrett, P.J., Fielding, C., Wise, S.W. (Eds.), 1998. Initial Report on CRP-1, Cape Roberts Project, Antarctica. Terra Antarctica 5, 1, pp. 141.

CHAPTER 2. CHRONOS'S PSICAT: CORE LOGGING AND DATA CAPTURE IN ONE TOOL

A paper submitted to Computers & Geosciences Special Issue: Geoscience Knowledge Representation for Cyberinfrastructure

J. Reed[4], C. Cervato[4], and D. Fils[4]

Abstract

The next generation geoscientific research is increasingly focused on the capture and analysis of data. One area that has traditionally lagged behind in data capture is the core description process. Typically, core description diagrams are drawn by hand in field books and then drafted up in a graphics program for publication. However, very rarely are the actual data encoded in the diagrams, e.g., depth in core, grain size, and lithology, captured in a format that can be manipulated and analyzed. In this paper we introduce CHRONOS's Paleontological Stratigraphic Interval Construction and Analysis Tool (PSICAT) and its novel approach to capturing and visualizing core description data. PSICAT is a graphical editing tool designed to support the core description process from the initial data capture, to visualization, to data analysis. It was field-tested on the ANtarctic DRILLing (ANDRILL) project's McMurdo Ice Shelf expedition where it was used to log nearly 1300 meters of sediment core drilled during the October 2006 – January 2007 field season. PSICAT performed well, offering many time-saving and science-enabling advances, and will be used again in Antarctica on ANDRILL's Southern McMurdo Sound expedition in 2007-2008.

Introduction

Increasing importance is placed on the capture of data as part of the next generation geoscientific research projects. The visual core description process, an integral part of any drilling project, has traditionally lagged behind in the data capture area. Typically, core description diagrams are drawn by hand in field books and then re-drafted in a generic graphics application like CorelDRAW (Barrett et al., 1998; C. Fielding, Personal Communication, May 2005) or a specialized application like AppleCORE[5]. The generic graphics application is not specifically designed for drafting core description diagrams, so it is not well suited to creating standard, consistent diagrams. AppleCORE, which was used by the Ocean Drilling Program, no longer appears to be actively developed or supported.

---

4 Department of Geological and Atmospheric Sciences, Iowa State University

5 AppleCORE – Borehole Log Production. GeoMEM. http://dspace.dial.pipex.com/town/place/vy12/products/ranger/applcore.html

A handful of specialized commercial software packages, e.g., Strater, WellCAD, and LogPlot, exist to aid the core description process. The main focus of these commercial packages is the final production of the core description diagrams. They require the data to have already been manually captured in spreadsheets or databases. None provide an environment for simultaneously capturing and visualizing core description data.

PSICAT, the Paleontological Stratigraphic Interval Construction and Analysis Tool, is an interactive, cross-platform environment for creating, viewing, and editing core description diagrams. As the user draws, PSICAT analyzes the diagram and extracts the data, e.g., depth, grain size, and lithology. PSICAT stores these data on the user's computer and then generates a visual representation of the data that reproduces what the user had originally drawn. This whole process takes place in the background and is so rapid that it is imperceptible to the user.

PSICAT is designed to support the complete core description process: 1. data capture; 2. visualization; 3. analysis; 4. collaboration; 5. revision; and finally 6. publication. Since this is an ambitious and complicated undertaking involving many different types of data, interactions, visualizations, and analyses, PSICAT requires a flexible and extensible software architecture and data model. This paper describes the software architecture and data model that PSICAT uses and outlines the features implemented in the current version (1.0.40). Finally, it discusses the integration of PSICAT within the ANtarctic geological DRILLing (ANDRILL) project and the future directions of PSICAT's development.

Software Architecture

Core description is a complex process involving the capture of a wide array of data using various methods, e.g., drafting, taking notes, and integrating external datasets. To facilitate this, PSICAT has been designed around a flexible, extensible software architecture that allows it to adapt to a variety of data types and tasks.

Pure Plug-in Architecture

PSICAT is designed around the concept of a pure plug-in architecture (Birsan, 2005). The concept of plug-ins and plug-in architectures is not a new one. Most computer users have utilized software, such as a Web Browser, that are built on a traditional plug-in architecture. This consists of a host application, which provides a set of well-defined interfaces and extension mechanisms that developers can "plug into" and add new functionality without changing the host itself. The host application is fully functional; the plug-ins simply augment it with new features. On the other hand, in a pure plug-in architecture the host application contains no end-user functionality; it simply acts as an engine for running plug-ins. All of the end-user functionality (i.e., the application in the traditional sense) is implemented as a collection of collaborating plug-ins.

Using a pure plug-in approach offers many advantages, including a high-degree of flexibility and re-use. The final application can be easily customized to a specific user group or task simply by virtue of which plug-ins are included or removed. If a plug-in already exists to perform a particular function, the application developer can re-use it as is or modify it as needed to address the group's requirements. This results in an tightly focused application that does exactly what the users need without cluttering it with functionality that won't be used.

The pure plug-in architecture has also the advantage of allowing the application to be extended in ways not initially planned for. With a traditional architecture, developers are limited to extending the application via the predefined extension mechanisms provided by the host application. On the other hand, applications composed of many collaborating plug-ins offer more opportunities for extension.

PSICAT does not implement its own plug-in engine, instead it utilizes Equinox[6], the plug-in engine provided by the Eclipse project[7]. This offers several advantages: 1. Equinox implements the well-defined, community developed OSGi (Open Services Gateway Initiative) Release 4 specification (Marple and Kriens, 2001; OSGi Alliance, 2005), which defines how plug-ins depend on and interact with each other; 2. the engine is mature and stable, providing the foundation for several large, enterprise-class applications, including the Eclipse Integrated Development Environment itself (Gruber et al., 2005); and 3. several hundred plug-ins[8], from both Eclipse and other groups, already exist and can be used with Equinox.

Plug-in Layers

The many plug-ins that make up the PSICAT application can be organized into five distinct layers as shown in Figure 2.1. Each layer builds on the functionality provided by the layers below it and provide new functionality to the layers above it.

The Core Layer provides functionality common to all applications. It includes the Equinox plug-in engine, described above, as well as the Eclipse Rich Client Platform (RCP) and the Eclipse Graphical Editor Framework (GEF). The Eclipse RCP provides a framework for building applications that consist of views, editors, wizards, etc. It implements these common concepts in a re-usable way so PSICAT does not have to implement them itself. The Eclipse GEF provides a framework for building graphical editors on the RCP. Using these existing Eclipse technologies allowed the development of PSICAT to focus on solving the core description problem.

The CHRONOS Modeling Framework (CMF) Layer builds on the Core Layer and defines

6 Equinox. The Eclipse Foundation. http://www.eclipse.org/equinox/
7 Eclipse. The Eclipse Foundation. http://www.eclipse.org
8 Eclipse Plugin Central: The Information Portal for the Eclipse Ecosystem. Eclipse Plugin Central. http://www.eclipseplugincentral.com/

the data and diagram models used by PSICAT. This layer also provides generic services to manage these models as well as diagrams, resources, and configuration. The data model is discussed in Section 3.

The PSICAT Application Layer provides the basic views, editors, and wizards that make up the PSICAT application. It also implements services to manage common tasks such as user authorization and depth in the core. This layer, along with the previous two layers, provides the scaffolding for data capture and visualization without the specialized knowledge of the core description process implemented in the next layer.

The PSICAT Domain Layer is where the functionality targeted at the end user is implemented. It consists of two basic types of plug-ins, service plug-ins and column plug-ins. Service plug-ins add new features, such as image export and searching, to the application. Column plug-ins add the code for capturing and displaying new types of data. Each column plug-in generally consists of models for the data being captured, controllers for managing and editing the models, and figures for displaying the models.

The Customization Layer adds project-, group-, or task-specific customizations to PSICAT. This includes supplying necessary configuration data and resource for the plug-ins in the previous layers. Project-specific column and service plug-ins are also included in this layer.

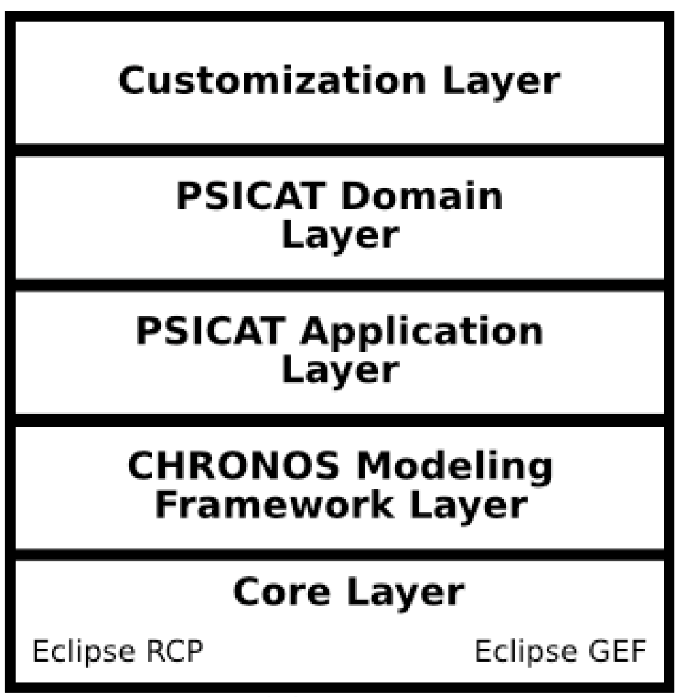


Figure 2.1: The plug-ins that make up the PSICAT application can be organized into five distinct layers, each building on the functionality provided by the layers below it and providing new functionality to the layers above it.

Data Model

The design phase identified numerous types of data to be captured. Many of these data types share common properties; for example, intervals, beds, and stratigraphic units are all defined over a depth range, i.e. each has a top and a base depth associated with it. However, each has also unique properties. Stratigraphic units have an identifier associated with them, whereas intervals and beds do not. Similarly, intervals have lithologies associated with them and stratigraphic units do not. To accommodate these various data types and also those that have not been identified yet, PSICAT required a flexible data model and representation scheme for the data.

Model Objects

Every high-level data type, such as intervals, beds, stratigraphic units, and sedimentary structures, in PSICAT is represented by a Model object. At its heart, each Model object is simply an associative array of key-value pairs called properties. The associative array structure was chosen because most, if not all, high-level data types can be represented as a collection of named properties. For example, an interval is simply an object that has a top depth property, a base depth property, a grain size property, etc. The idea of using an associative array to build higher-level data structures is not unique to PSICAT; the Lua programming language takes the same approach (Ierusalimschy, 2006). The associative array data structure was also chosen because it is well understood and most programming languages offer an implementation, either natively in the language (e.g., in Javascript and Python), or through libraries (e.g., in Java's Map and C++'s Standard Template Library).

Every Model object has two implicitly defined properties: an `id` and a `type`. The `id` property is the unique identifier of the Model and the `type` property is the specific data type of the Model, e.g., `Interval`, `Bed`, or `StratigraphicUnit`. Beyond these two implicit properties, the Model can have an arbitrary number of other properties defined for it. Models may also have child Models associated with it. Child Models are standard Model objects that have been nested to represent a parent-child relationship. For example, a "Bed" Model may be nested inside an "Interval" Model indicating that the bed occurs within a larger interval. Nesting allows the building of complex, hierarchical trees of Models. To ease loading and persisting Models to external data representations, all property names and values must be character strings. While at first this may seem unnecessarily limiting, in practice, as described in the next section, it is a non-issue.

Model Mapping

Model property names and values are limited to character strings. This simplification can cause problems when the property value represents another data type, e.g., a number. There is no way to numerically compare the character string "8.92" with "12.71" since they are both just characters and not numbers. Similarly, since the Model object is not aware of property value types, the value of "number" is as valid for a depth property as "8.92".

To address this, PSICAT applies a mapping action where it maps the generic Model object to a specialized subclass of Model based on the value of the `type` property that every Model must define. This specialized Model object is still backed by the associative array but it can define specific functions for accessing the properties. For example, the `AbstractDepthAwareModel` subclass of Model defines the property accessors `double getBaseDepth()` and *`void setBaseDepth(double depth)`*. The `setBaseDepth` method takes an actual number, converts it to a character string, and stores it as the base depth property value. Similarly, the `getBaseDepth` method converts the character string representation of the base depth property value and returns it as an actual number.

Data Representation

The data model described in Section 3.1 is flexible enough to be serialized to many different representations. During the course of its development, PSICAT has used three different file representation schemes and can load and store the data model in a relational database. Currently PSICAT uses a single eXtensible Markup Langague (XML) file per project to persist the data model:

```
<models>
    <model id="0" type="Project" parent="">
       <property name="name">Sample Project</property>
    </model>
    <model id="1" type="Interval" parent="0">
       <property name="depth.top">0.00</property>
       <property name="depth.base">0.58</property>
          ...
    </model>
    <model id="2" type="Interval" parent="0">
       <property name="depth.top">0.58</property>
       <property name="depth.base">1.12</property>
          ...
    </model>
</models>
```

This example represents a `Project` with the name "Sample Project" that has two `Interval`s as its children. Each Model provides a `parent` attribute that lists the identifier of its parent so that the Model hierarchy can be reconstructed. Since XML supports nesting of elements, the `Interval` Models could have been nested within the `Project` Model. However, stack overflow errors can occur from recursing too many times while building deeply nested Model trees. With this format, PSICAT parses all of the Models in the file in one pass and then constructs the Model tree.

Implementation

An initial implementation of PSICAT, version 1.0.40, is currently available at the CHRONOS portal[9]. It contains the functionality to capture and display various types of core description data, export those data in multiple formats including as core description diagrams at various scales and page settings for publication, and analyze the data to generate summary diagrams and spreadsheets. Below we illustrate how the user would use the majority of the features in the initial implementation of PSICAT.

Before describing any core with PSICAT, the user must perform two preliminary steps. The first is to create a Project to store all of the captured core description data. PSICAT allows the user to have multiple Projects in the workspace and to choose the one she wants to work with. The second task is to define one or more Diagrams. Diagrams provide a view of the data by defining which columns appear on the core description diagram and in which order. Each column is responsible for generating a visual representation of a particular type of data, e.g., drawing a polygon for each stratigraphic interval or plotting symbols denoting the type and depth of fossil occurrences. PSICAT currently includes columns for capturing and displaying the following types of data: geological ages, stratigraphic units, drilling disturbance, core boxes, intervals and beds, facies, depositional environment, bioturbation, symbols (e.g., sedimentary structures, fossil occurrences), clasts, and written descriptions. In addition, external datasets can be plotted as line graphs and histograms. The user can include a column multiple times in the same Diagram.

Once the user has created a Project and at least one Diagram, she is ready to start describing the core. When she opens the Project and selects the Diagram to use, PSICAT opens a graphical editor displaying the Diagram and any saved data in the Project. The main area of the graphical editor is the drawing area where the diagram is displayed. A palette displays the drawing tools available for the columns in the open Diagram. As the user draws in the diagram area, PSICAT automatically captures the data of what is drawn and adds them to the Project. The depth scale of the Diagram can be dynamically changed to vary the detail of the data captured. The user can also directly edit the properties of the Models, simultaneously updating the diagram.

During the core description process, a Summary Diagram wizard can be used to automatically generate data summaries based on user-specified parameters. The new summary Project can then be visualized by PSICAT. The user can export the core description data as Excel spreadsheets or the diagrams as images for publication using the Image Export wizard. The Image Export Wizard allows her to select the Project and Diagram, the scale and page settings, and the image format of the output.

Finally, many of PSICAT's current features can be easily customized. For example, PSICAT comes lithologies and limited sedimentary structure symbols based on U.S. Geological Survey's standards, but new schemes can easily be added. With some programming knowledge, new column

9 CHRONOS Portal. CHRONOS. http://portal.chronos.org

types can be also integrated into PSICAT in a matter of hours.

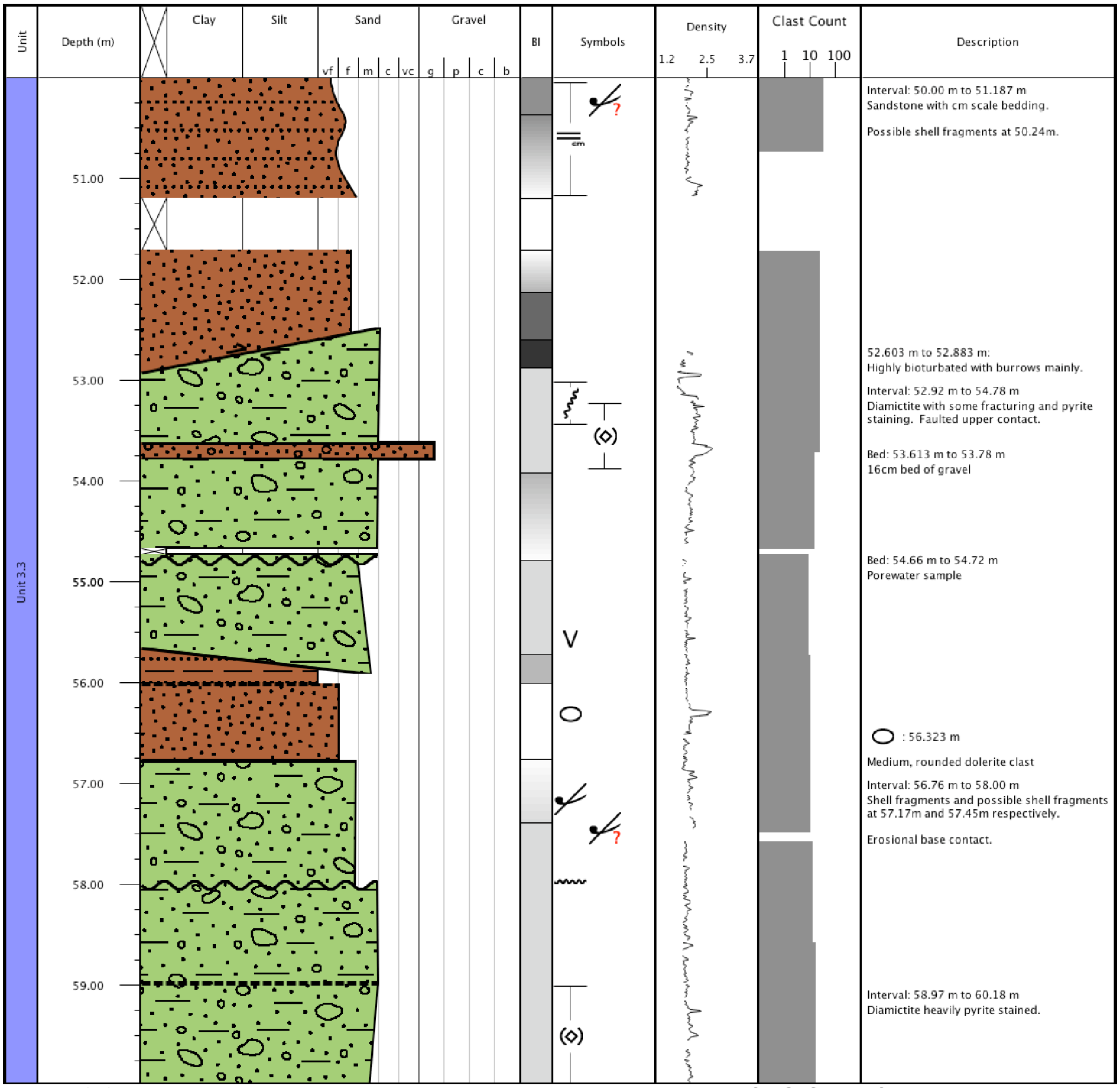


Figure 2.2: An example core description diagram illustrating many of PSICAT's features. Stratigraphic units are labeled in the Unit column, lithostratigraphic intervals are drawn in the third column, bioturbation is indicated by varying levels of shading in the BI column, sedimentary structures and fossil occurrences are indicated with symbols, a histogram of clast frequency is indicated in the Clast Count column, and written descriptions are included in the Description column.

Field Testing

The first large-scale field testing of PSICAT took place in October 2006-January 2007. PSICAT was used on-ice in Antarctica to log all of the sediment cores drilled as part of ANDRILL's McMurdo Ice Shelf (MIS) expedition. It replaced the previous process of drawing the core description diagrams by hand and then drafting them in CorelDRAW. PSICAT received extensive testing, being used 12 hours a day, 7 days a week, to log the nearly 1300 meters of core drilled. Overall, PSICAT performed exceedingly well. Development on PSICAT continued on-ice to promptly fix the inevitable bugs and to add new features needed by the users.

In addition to successfully replacing the previously used core description process, PSICAT offered ANDRILL numerous advantages that saved time and enabled better science. The most significant time-saver was the ability to automatically generate summary diagrams. In the past, they were manually compiled by the sedimentology team at the end of the drilling expedition. With PSICAT, the diagrams can be quickly generated at will. This means that they are generated every couple of days, allowing the science team to visualize emerging trends while drilling.

Since PSICAT worked with data instead of images, it allowed the science team to revise the core description data and instantaneously see the results of their revisions. This enabled the science team to freely explore different interpretations of the core to produce the best science possible. Two examples of this was the team's decision to highlight volcanically derived material on the description diagrams and how the facies and depositional environment schemes evolved during drilling. Several hundred meters into the drilling, the science team noticed a prevalence of volcanically derived material in the core but it was hard to distinguish it from, e.g., sandstone, because the only difference was the presence of 'V's in the lithology pattern of volcanically derived sandstones. The science team decided to change the color of volcanic materials in the core description to orange so that they stood out more. For PSICAT, this was a simple change and all of the diagrams were instantly updated. In the past, each of the diagrams would have had to be updated by hand. The ability to easily revise was used extensively when the science team developed the facies and depositional environment schemes. PSICAT allowed them to propose a facies scheme, look at the facies within the context of the core, and then revise the scheme several times before they found one that they were confident in. If the diagrams were drafted by hand, then the focus would have been to agree on one facies scheme before drafting up the core diagrams to avoid having to re-draw them.

Finally, PSICAT enabled the science team to quickly identify specific areas of interest in the core and extract the core description data in formats, e.g., Excel spreadsheets, that allowed them to perform further analysis. The ability to search the core description data was a big time saver. With static images, the science team would have to look at each of the several hundred diagrams to identify the features of interest. PSICAT allowed them to enter a few constraints that described which features they were interested in, and quickly returned the sections that matched what they were looking for.

Conclusions

PSICAT is a free, open-source tool for creating, viewing, and editing core description diagrams that works across all platforms. It was designed to improve upon the traditional core description process (hand-drafting the core diagram in the field book and then creating a digital version of it with a graphics software like CorelDRAW) and the current commercial offerings such as Strater, LogPlot, and WellCAD, which require the user to manually capture the core description data in a spreadsheet or database before it can be visualized. PSICAT seamlessly integrates data capture and data visualization by providing an environment where the user can draw the core description diagrams while the data, e.g., depth, grain size, and lithology, are automatically extracted. Capturing the actual core description data as opposed to just capturing the core description diagrams offers many advantages. PSICAT can use the data to automatically generate summary core description diagrams and to help the user identify areas of interest within the core. Also, the data can be stored in a database, making it available for further analysis.

The merit of PSICAT's approach to the core description process was proven during the October 2006-January 2007 Antarctic field season where it was used to log the nearly 1300 meters of core drilled as part of the ANDRILL McMurdo Ice Shelf expedition. PSICAT performed well under continuous daily use in the field. It successfully replaced the process used in the previous Antarctic drilling expedition that required the core description diagrams to be first drawn by hand and then re-drafted in CorelDRAW. It also offered many time-saving and science-enabling advantages that the previous process, including automatic generation of summary diagrams and the ability to quickly and easily revise the diagrams.

PSICAT's extensible, pure plug-in software architecture and flexible data model allow it to support the complete core description process from data capture to analysis, to publication. They allow PSICAT to capture and visualize a wide variety of data types, and to easily adapt and customize to specific groups of users or tasks.

The development of PSICAT is on-going. It will be used again in Antarctica on ANDRILL's Southern McMurdo Sound expedition during the September 2007-December 2007 field season. We are focusing on improving the user experience with PSICAT based on observations made during the first field testing. We also plan to offer more import and export options, more data column types, and improved search functionality in the next major release of PSICAT (v. 2.0).

The current version, 1.0.40, of PSICAT can be freely downloaded from the CHRONOS portal at http://portal.chronos.org/psicat-site/

Acknowledgments

We wish to thank Chris Fielding for his assistance during the design and development of PSICAT. We would also like to thank the ANDRILL MIS team and specifically the Science Management Office and the MIS Sedimentology team, Larry Krissek, Ellen Cowan, Gavin Dunbar, and Thom Wilch for their insightful feedback during field testing. Funding for PSICAT's development was provided by the National Science Foundation through awards to CHRONOS (EAR-0315216 and 0524285) and ANDRILL (ANT-0342484).

References

Barrett, P.J., Fielding, C., Wise, S.W. (Eds.), 1998. Initial Report on CRP-1, Cape Roberts Project, Antarctica. Terra Antarctica 5, 1, pp. 141.

Birsan, D., 2005. On plug-ins and extensible architectures. ACM Queue 3, 2, pp. 40-46.

Gruber, O., Hargrave, B.J., McAffer, J., Rapicault, P., Watson, T., 2005. The Eclipse 3.0 platform: adopting OSGi technology. IBM Systems Journal 44, 2, pp. 289-299.

Ierusalimschy, R., de Figueiredo, L. H., Celes, W., 2006. Lua 5.1 Reference Manual. Lua.Org, 113 pp. ISBN 85-903798-3-3.

Marples, D., Kriens, P., 2001. The Open Services Gateway Initiative: An Introductory Overview. IEEE Communications Magazine 32, 12, pp. 110-114.

OSGi Alliance, 2005. About the OSGi Service Platform, Technical Whitepaper, Revision 4.1. OSGi Alliance, 20 pp.

CHAPTER 3. PSICAT: A NEW OPEN-SOURCE CORE DESCRIPTION APPLICATION

A paper for submission to the Geological Society of America's online journal Geosphere

J. Reed[10], C. Cervato[10], and D. Fils[10]

Abstract

Core description diagrams are an elegant way of representing and conveying the vast amount of data captured during the visual core description process. Typically, these diagrams are drawn by hand in field books and then drafted up in a graphics program for publication. Very rarely are the actual data encoded in the diagrams, e.g., depth in core, grain size, and lithology, captured in a format that can be manipulated and analyzed. In this paper we introduce CHRONOS's Paleontological Stratigraphic Interval Construction and Analysis Tool (PSICAT) and its novel approach to capturing and visualizing core description data. PSICAT is a graphical editing tool designed to support the core description process from the initial data capture, to visualization, to data analysis. It was field-tested on the ANtarctic DRILLing (ANDRILL) project's McMurdo Ice Shelf expedition where it was used to log nearly 1300 meters of sediment core drilled during the October 2006 – January 2007 field season. PSICAT performed well, offering many time-saving and science-enabling advances, and will be used again in Antarctica on ANDRILL's Southern McMurdo Sound expedition in 2007-2008. Loaded with USGS's standard lithology patterns and colors, PSICAT's modular design makes it easily adaptable for use with land-based sections and a broad range of lithologies.

Introduction

Data capture is becoming increasingly important in next-generation geoscientific research. One area that has traditionally lagged behind in data capture is the visual core description process, an integral part of any scientific drilling project. Core description diagrams are an elegant way of representing and conveying a vast amount of data about drill cores (e.g., Majors, et al., 1998; Michels, et al., 1998; Belsher and Harris, 2003; Rothwell and Rack, 2006). Unfortunately, the traditional approach to core description is to sketch the diagrams in field books and then re-draft them in a generic graphics application like Adobe Illustrator(TM) or CorelDRAW (TM) as was done on the Cape Roberts drilling expeditions (Barrett et al., 1998; C. Fielding, personal communication, May 2005). This approach produces publication quality diagrams, but does not provide a way to answer questions like "How much of this hole is diamictite?" and "Where are all the shell fragments?" without resorting to visually reviewing each diagram and manually recording the features of interest.

10 Department of Geological and Atmospheric Sciences, Iowa State University

To address the limitations of the traditional core description process, we have developed a tool called PSICAT, the Paleontological Stratigraphic Interval Construction and Analysis Tool. PSICAT is an interactive, cross-platform environment for creating core description diagrams that also captures the data represented in the diagrams (Reed et al., submitted 2007). The idea of capturing the core description data is hardly new or unique. A handful of commercial software products, including AppleCORE, Strater, and WellCAD, exist to produce core description diagrams from data that has been previously captured in spreadsheets.

What distinguishes PSICAT from these commercial offerings is the novel way it seamlessly integrates data capture and visualization in an interactive environment. It mimics the traditional core description approach by providing the user with a drawing environment similar to the drafting software she is familiar with (see Figure 3.1), but as she draws, PSICAT analyzes the diagram and extracts the core description data, e.g. depth, grain size, and lithology, into a file. PSICAT then generates a visual representation of the captured data that reproduces what the user is drawing. This whole process takes place in the background and is so rapid that it is imperceptible to the user. With PSICAT we are able to offer the best of both worlds: an environment that is familiar to the user, so she does not have to drastically change the way she describes core, and all the benefits of actually capturing the core description data in a format that can be analyzed and searched.

Noteworthy Features

Capturing the core description data offers many advantages over simply drafting core description diagrams. Below are some features of PSICAT that are only possible because it works with data instead of only images.

Extract data in common formats for further analysis

The first, most obvious advantage of capturing the core description data is that they are available for further analysis. PSICAT allows the user to export all or a subset of the captured data as an Excel spreadsheet as in Figure 3.2. In this format, the data are available for further analysis and plotting, allowing the user to answer questions like: “What percentage of the hole is diamictite?”

Search the core description for areas of interest

PSICAT also includes sophisticated searching capabilities so the user can quickly find areas of interest in the core without having to resort to looking at each diagram. PSICAT currently provides a search interface that lets the user specify areas of interest with natural language queries such as “sections of diamictite containing symbol pyrite in 0-500m” and “sections of symbol shell or symbol fragmented shell”. Work is currently underway on a "Google-like" search interface of the written description data.

Integration of external data sets

Often it is useful to plot external datasets, such as physical core properties captured by a multi-sensor core logging system, alongside the core description. PSICAT allows the user to import external datasets and integrate them directly into the core description diagrams as shown in Figure 3.3.

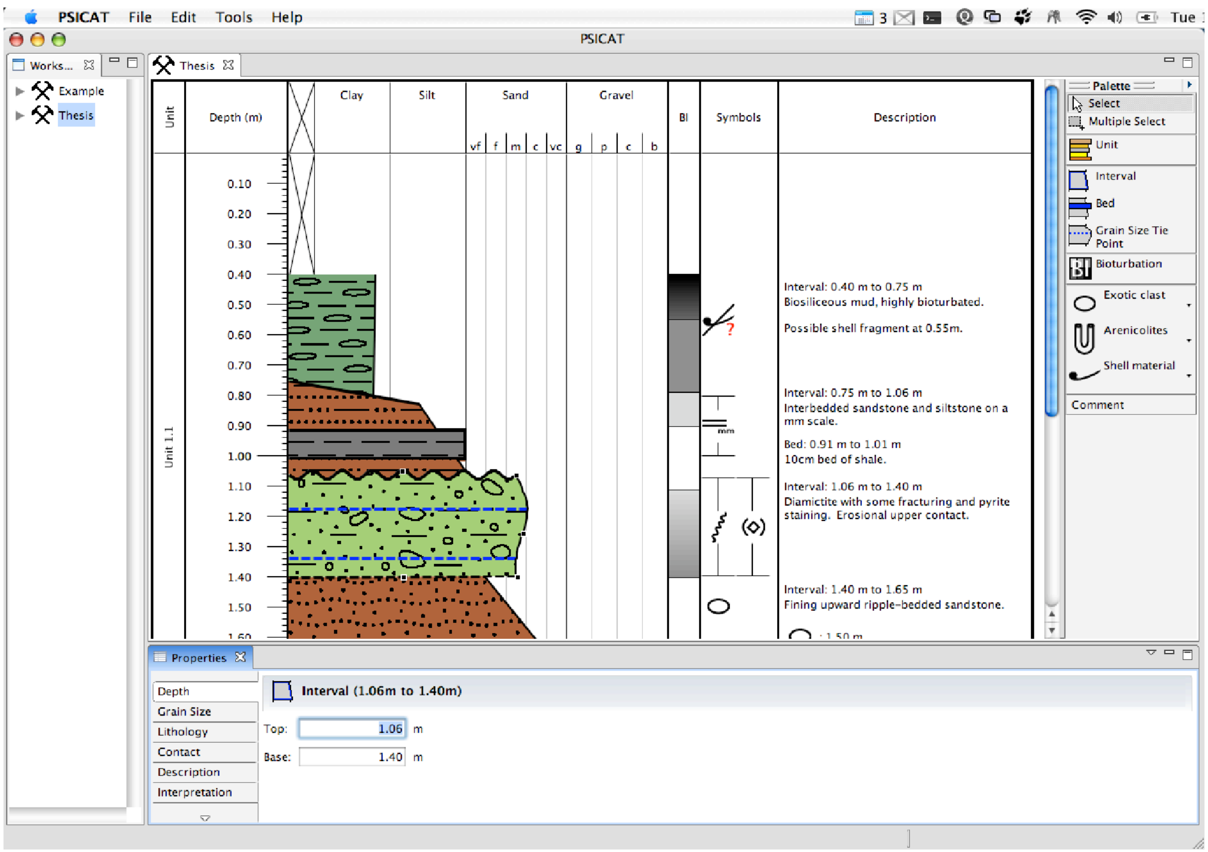
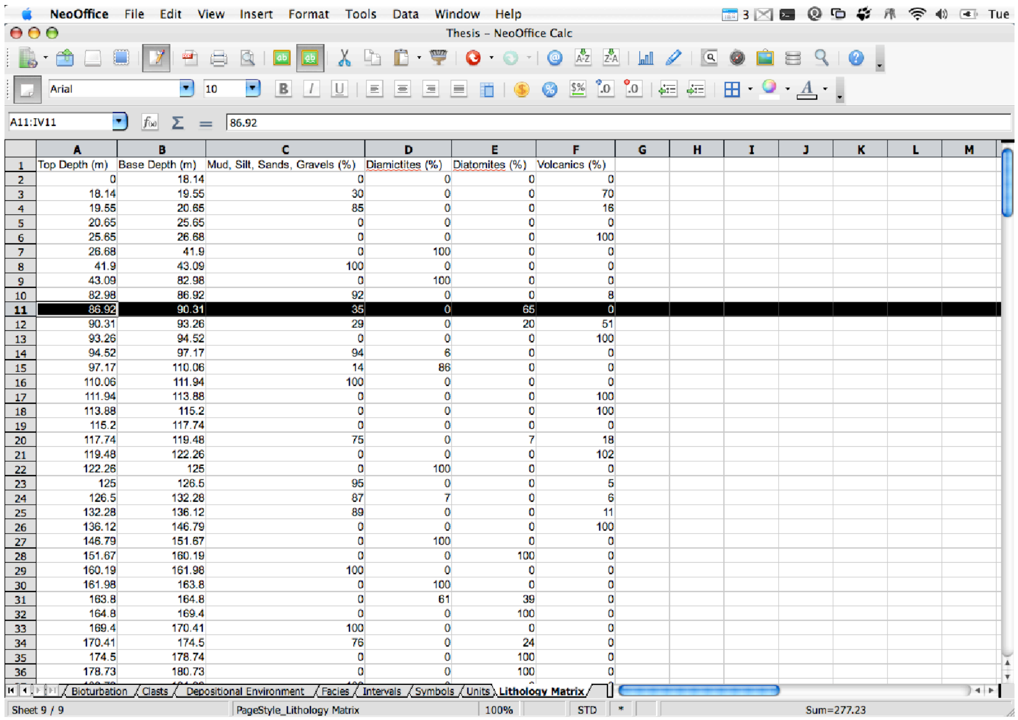


Figure 3.1: Editing a core description diagram in PSICAT. The user's projects are listed in the Workspace pane at the left. The diagram that is currently being edited is displayed in the center. The graphical tools available to the user are listed in the palette to the right. The properties of the selected object are displayed at the bottom.

Quickly update and re-plot diagrams

Imagine having to change the pattern that represents “sandstone” in all of the diagrams for several hundred meters of core with a graphics application. Each diagram would have to be edited by hand—no small task if the core contained a lot of sandstone. With PSICAT, updates like this are trivial because it generates the core description diagrams from the data. The user simply has to change the pattern associated with sandstone once and all diagrams are automatically updated.



	A	B	C	D	E	F
1	Top Depth (m)	Base Depth (m)	Mud, Silt, Sands, Gravels (%)	Diamictites (%)	Diatomites (%)	Volcanics (%)
2	0	18.14	0	0	0	0
3	18.14	19.55	30	0	0	70
4	19.55	20.65	85	0	0	16
5	20.65	25.65	0	0	0	0
6	25.65	26.68	0	0	0	100
7	26.68	41.9	0	100	0	0
8	41.9	43.09	100	0	0	0
9	43.09	82.98	0	100	0	0
10	82.98	86.92	92	0	0	8
11	86.92	90.31	35	0	65	0
12	90.31	93.26	29	0	20	51
13	93.26	94.52	0	0	0	100
14	94.52	97.17	94	6	0	0
15	97.17	110.06	14	86	0	0
16	110.06	111.94	100	0	0	0
17	111.94	113.88	0	0	0	100
18	113.88	115.2	0	0	0	100
19	115.2	117.74	0	0	0	0
20	117.74	119.48	75	0	7	18
21	119.48	122.26	0	0	0	102
22	122.26	125	0	100	0	0
23	125	126.5	95	0	0	5
24	126.5	132.28	87	7	0	6
25	132.28	136.12	89	0	0	11
26	136.12	146.79	0	0	0	100
27	146.79	151.67	0	100	0	0
28	151.67	160.19	0	0	100	0
29	160.19	161.98	100	0	0	0
30	161.98	163.8	0	100	0	0
31	163.8	164.8	0	61	39	0
32	164.8	169.4	0	0	100	0
33	169.4	170.41	100	0	0	0
34	170.41	174.5	76	0	24	0
35	174.5	178.74	0	0	100	0
36	178.73	180.73	0	0	100	0

Figure 3.2: The core description data exported from PSICAT as a spreadsheet and opened in Excel.

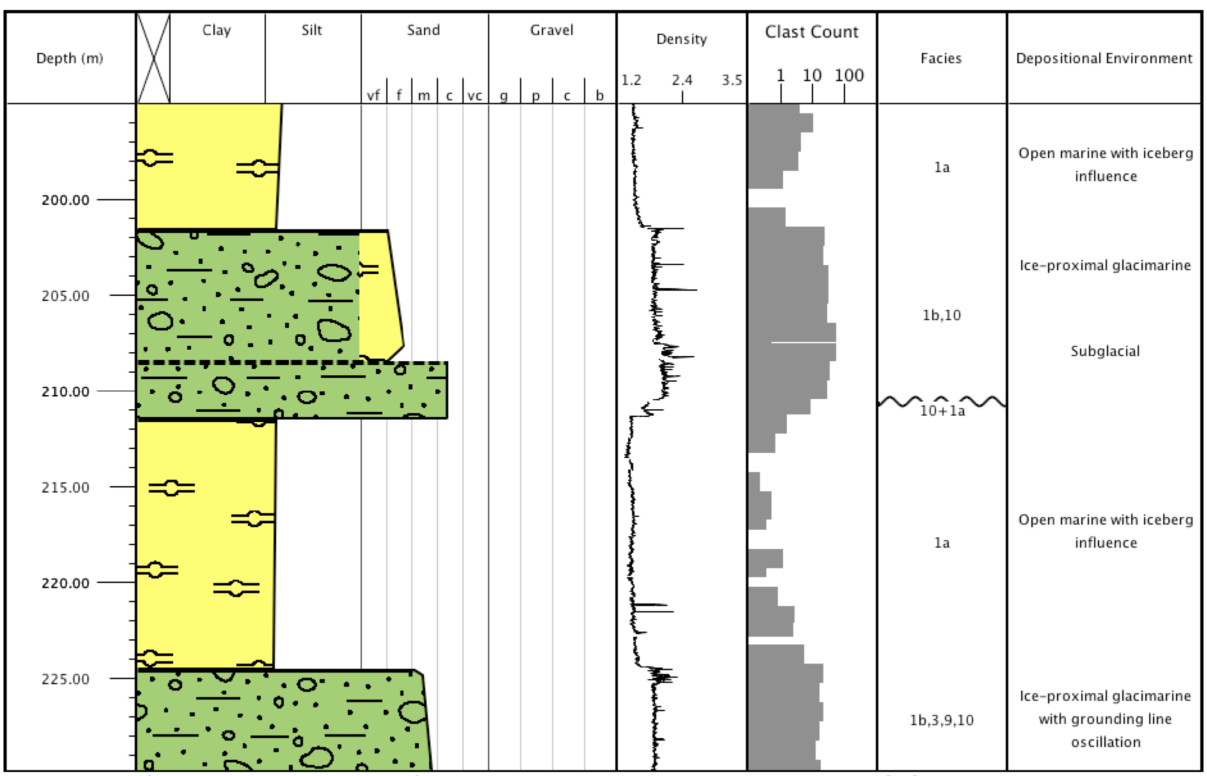


Figure 3.3: Core density and clast frequency, both captured externally to PSICAT, integrated into the core description diagram created by PSICAT.

Automatic summarization

The core description process often involves more than just the initial characterization of the core. It also includes derived descriptions, such as summary diagrams, which show the general trends. Creating these summary diagrams basically requires re-logging the whole core at a less-detailed scale to show only the important features. To address this, PSICAT includes a feature which processes the core description data with a sophisticated set of rules to automatically summarize it in a matter of seconds as shown in Figure 3.4. The summarized data can be edited, analyzed, searched, and plotted just like any other core description data in PSICAT.

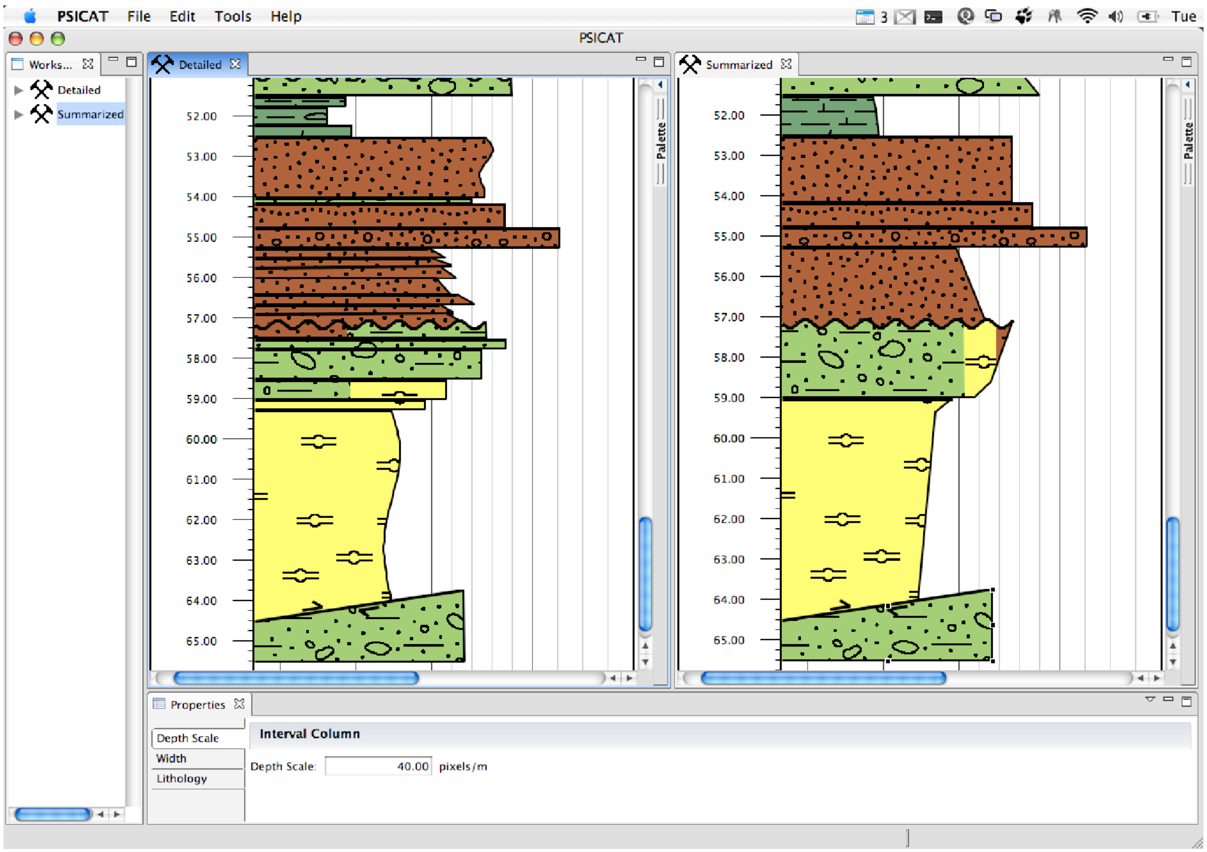


Figure 3.4: Side by side comparison of a detailed diagram (left) and the summary diagram (right) automatically generated by PSICAT.

Design and Development

PSICAT was designed to address a specific, real world need identified by the geoscientific community. We partnered with the ANtarctic geological DRILLing (ANDRILL) project to develop a core description solution for their 2006-2007 McMurdo Ice Shelf (MIS) drilling expedition. PSICAT was designed to exactly address ANDRILL's needs but with the broader community needs in mind. We took advantage of the close collaboration with ANDRILL sedimentologists during the design and development phase. The MIS expedition was an invaluable opportunity to see how PSICAT worked under real world conditions and to gather user feedback.

Design on PSICAT began in June 2005. We worked closely with Chris Fielding, a sedimentologist involved in ANDRILL, to understand the core description process and gather required features. We also talked with other members of the scientific community to understand how their core description needs differed from ANDRILL's. This interaction drove the design of PSICAT to meet ANDRILL's needs but also allow it to be customized for use by other groups.

Since we wanted PSICAT to be as broadly used as possible, we made two major decisions when we began the development. First, it needed to work equally well on Mac, Windows, and Linux, so we chose to develop it in the Java programming language. The other decision was to depart from the standard approach of developing a single, monolithic piece of software because this approach was not flexible or adaptable enough to meet the needs of the broader community. Instead we embraced the idea of developing PSICAT as a series of individual modules that could be assembled into a coherent application. This would allow us to re-use modules that provided functionality common to all drilling groups while still supporting the development of custom modules for specific groups or tasks. This makes PSICAT a platform for core description rather than simply a single piece of software.

Field Testing

The first large-scale field testing of PSICAT took place in October 2006-January 2007 when it was used on-ice in Antarctica to log the nearly 1300 meters of sediment core drilled during ANDRILL's MIS expedition. The lead author accompanied the expedition to observe the users interacting with PSICAT and to provide technical assistance. The field testing focused on observing and gathering feedback from two user groups: sedimentologists who create the core description with PSICAT and the other scientists who explore and use the core description to do their science. Observations of the former group were used to assess the ease of use and completeness of PSICAT. Feedback from the latter group was used to assess whether the diagram and data produced by PSICAT were able to support the science being done.

The majority of time on-ice was spent observing and gathering feedback from the sedimentology team who were using PSICAT to produce the core description. The majority of feedback focused on usability issues such as finding a simpler way to enter all of the clast information. The testing also identified some features that were not anticipated and the occasional software bug. Development of PSICAT continued while on ice to fix the bugs and add new features, resulting in the release of over 30 updates to PSICAT.

Considerable time was also spent working with the science team to produce the diagrams and data they needed, both on ice and after the expedition. Requests like plotting physical property data on the core logs or finding areas of interest in the core resulted in the development many of the features described above. Development of PSICAT continues in preparation for the 2007-2008 drilling season.

Conclusions

Overall, PSICAT performed exceedingly well during its first field testing. It was able to produce the same quality of diagrams that would have been produced if they had sketched the diagrams by hand and drafted them up in CorelDRAW. As such, PSICAT can be considered a drop-in replacement for the previous process. PSICAT also offered ANDRILL numerous advantages that saved time and potentially enabled better science. The most significant time-saver was the ability to automatically generate summary diagrams. In the past, these diagrams were manually compiled by the sedimentology team at the end of the drilling expedition. With PSICAT, the diagrams could be generated at any point in time during the drilling. This meant that as soon as the detailed description was done, so were the summary diagrams. It also meant that the science team did not have to wait for the end of the expedition to see the general trends in the core.

PSICAT also allowed the science team to revise the core description data and instantaneously see the results of their revisions. This enabled the team to test various models of interpretation of facies and depositional environments as drilling was ongoing. PSICAT was also used to quickly identify and highlight specific areas of interest in the core and extract the core description data in formats, e.g., Excel spreadsheets, that allowed the scientists to perform further analysis. This was done by entering a few constraints about certain features and PSICAT returned the sections that matched the requirements.

Future development is focusing on: improving data capture, adding more analysis and visualization features, tool integration, and broader applications of PSICAT, such as for logging outcrops and land-based sections. In the data capture area, we are focusing on a few large-scale changes identified during this field season and adding more data column types. In the analysis and visualization area, the development is focused on more import and export options, improving integration of external datasets, providing alternate data visualizations, and improving the search functionality. On the tool integration front, we are working with the CoreWall project to integrate PSICAT output into their Corelyzer visualization software (Rao, et al. 2004). We are also working with the community to develop educational activities for undergraduate students and we are soliciting feedback from other projects interested in incorporating PSICAT into their core and section description process.

The latest version of PSICAT is freely available from the CHRONOS portal: http://portal.chronos.org/psicat-site/

Acknowledgments

We wish to thank Chris Fielding for his assistance during the design and development of PSICAT. We would also like to thank the ANDRILL MIS team and specifically the Science Management Office and the MIS Sedimentology team, Larry Krissek, Ellen Cowan, Gavin Dunbar, and Thom Wilch for their insightful feedback during field testing. Funding for PSICAT's development was provided by the National Science Foundation through awards to CHRONOS (EAR-0315216 and 0524285) and ANDRILL (ANT-0342484).

References

Barrett, P.J., Fielding, C., Wise, S.W. (Eds.), 1998. Initial Report on CRP-1, Cape Roberts Project, Antarctica. Terra Antarctica 5, 1, pp. 141.

Belsher, T.W., and Harris, P.M., 2003.Facies and fracture variation in a deep shelf carbonate; core and log comparison. Annual Meeting Expanded Abstracts - American Association of Petroleum Geologists, 12, p.12.

Major, C.O., Pirmez, C., and Goldberg, D., 1998. High-resolution core-log integration techniques; examples from Ocean Drilling Program. In: Harvey, P. K. and Lovell, M. A. (Eds.), "Core-log integration", Geological Society Special Publications, 136, p. 285-295.

Michels, K.H., Kudrass, H. R., Huebscher, C., Suckow, A., and Wiedicke, M., 1998. The submarine delta of the Ganges-Brahmaputra; cyclone-dominated sedimentation patterns. Marine Geology, 149, p.133-154.

Mithal, R., and Becker, D., 2006. The Janus data base; providing worldwide access to ODP and IODP data. In: Rothwell, R.G. (Ed.), "New techniques in sediment core analysis", Geological Society Spec. Pub., 267, p. 253-259.

Rao, A., Kamp, B., Noren, A., Ito, E., Schurrenberger, D., Morin, P.J., Leigh, J., Johnson, A., Renambot, L., and Rack, F., 2004. CoreWall; a visualization environment for the analysis of lake and ocean cores, Abstracts with Programs - Geological Society of America, 36, 5, p. 273

Reed, J.A., Cervato, C., and Fils, D., submitted 2007. CHRONOS's PSICAT: Core Logging and Data Capture in One Tool. Computers and Geoscience Special Issue: Geoscience Knowledge Representation for Cyberinfrastructure.

Rothwell, R.G., and Rack, F., 2006. New techniques in sediment core analysis; an introduction. In: Rothwell, R.G. (Ed.), "New techniques in sediment core analysis", Geological Society Spec. Pub., 267, p. 1-29.

CHAPTER 4. GENERAL CONCLUSION

Discussion

PSICAT is a free, open-source tool for creating, viewing, and editing core description diagrams that works across all platforms. It was designed to improve upon the traditional core description process (hand-drafting the core diagram in the field book and then creating a digital version of it with a graphics software like CorelDRAW) and the current commercial offerings such as Strater, LogPlot, and WellCAD, which require the user to manually capture the core description data in a spreadsheet or database before it can be visualized. PSICAT seamlessly integrates data capture and data visualization by providing an environment where the user can draw the core description diagrams while the data, e.g., depth, grain size, and lithology, are automatically extracted. Capturing the actual core description data as opposed to just capturing the core description diagrams offers many advantages. PSICAT can use the data to automatically generate summary core description diagrams and to help the user identify areas of interest within the core. Also, the data can be stored in a database, making it available for further analysis.

The merit of PSICAT's approach to the core description process was proven during the October 2006-January 2007 Antarctic field season where it was used to log the nearly 1300 meters of core drilled as part of the ANDRILL McMurdo Ice Shelf expedition. PSICAT performed well under continuous daily use in the field. It was able to produce the same quality of diagrams that would have been produced had ANDRILL used the process of sketching the diagrams by hand and drafting them up in CorelDraw. As such, PSICAT can be considered a drop in replacement for the previous process. It also offered many time-saving and science-enabling advantages that the previous process, including automatic generation of summary diagrams and the ability to quickly and easily revise the diagrams.

Development of PSICAT is ongoing. It will be used again in Antarctica on ANDRILL's Southern McMurdo Sound (SMS) expedition during the September 2007-December 2007 field season. Future development is focusing on three main areas: improving data capture, adding more analysis and visualization features, and raising awareness of PSICAT. In the data capture area, we are focusing on a few large scale changes identified this field season that could not be implemented on ice and adding more data column types. In the analysis and visualization area, the development is focused on more import and export options, improving integration of external datasets, providing alternate data visualizations, and improving the search functionality. On the awareness side of things, we are working with the community to develop educational activities for undergraduate students and we are soliciting feedback from other drilling projects interested in incorporating PSICAT into their core description process.

APPENDIX

Schwartz Communications, Inc.

ENTERPRISE OpenSource MAGAZINE
WWW.OPENSOURCE.SYS-CON.COM

December 2006

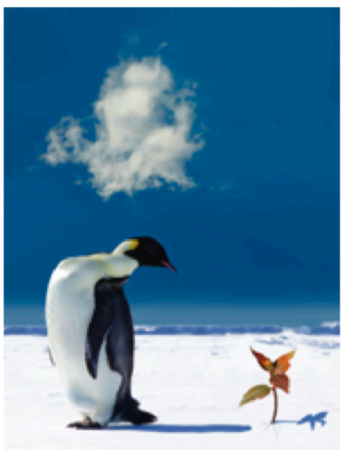
practical application

How I Spent My Summer

Three months in Antarctica with Eclipse and Java

by Josh Reed

I recently came home from three months in Antarctica. As a software engineer, Antarctica was the last place I expected to be. I was there as part of the ANDRILL project, a multinational collaboration comprised of scientists, educators, students, technicians, drillers, and support staff from Germany, Italy, New Zealand, and the United States.



Their goal is to uncover the elusive geological history of Antarctica by drilling and recovering sediment core samples from under the Antarctic ice shelf and sea-ice in hopes it will reveal how the ice on and around Antarctica reacted in past periods of climate change. They were using PSICAT (pronounced sigh-cat), a Java-based graphical editing tool I developed, to describe the core samples they were drilling. Since the expedition's fate depended on this data, and because PSICAT was largely unproven, I accompanied the project to ensure that PSICAT performed as expected. This gave me the unique opportunity to be fully involved with the people who were actually using the software — to see what worked, what didn't, and improve the quality of the software and user experience on a daily basis. The journey was long, with plenty of trials along the way, but in the end PSICAT performed better than anyone expected and, most importantly, the people who actually had to use it were still talking to me at the end of the expedition. This article will discuss the design and implementation of PSICAT for use on the harshest continent in the world.

Background

PSICAT, which stands for Paleontological Stratigraphic Interval Construction and Analysis Tool, is a standalone Java-based graphical editing tool for creating and viewing core description diagrams. Core description diagrams are the primary record of the cylindrical rock samples that result from scientific drilling. As the cores come out of the ground, they are examined and described by sedimentologists — scientists who specialize in identifying and interpreting the "story" of the sediments in the sample. Some of the many things that sedimentologists look for when examining a sample include: what type of sediment is present – for example, is it sand? mud? – where does the type and size of the sediment change; are there special structures like bedding or faults presents; and are there fossil traces. All of these features are represented pictorially on the core description diagram. This description process and the resulting diagrams are very important since they provide the scaffolding on which all further analysis is built.

On past drilling expeditions, the sedimentologists drew these diagrams by hand, usually including four meters of core per page, and then passed them off to a dedicated draftsman who would draft them up in a graphics application like CorelDraw or Adobe Illustrator. Although the process worked and allowed scientists to create nice diagrams, it had several shortcomings.

Because of the complexity of the graphics software and of the information to be captured, it required a person dedicated to drafting the diagrams. It also required the diagrams to be created at least twice — once by hand and then a second time in a digital format. Once digitized, the diagrams were static images, which was a problem when changes had to be made. Small local changes to individual diagrams could be made fairly easily, but making changes spanning multiple diagrams was a burden. Truly substantial changes, like changing the symbol that represents a type of rock, would have required re-drafting every diagram — an onerous job when there are diagrams for 1,000 meters of core to update. Static images also complicated the task of searching for specific features in the information-dense diagrams because it required searching each of the individual diagrams for the features of interest.

Realizing the shortcomings of the traditional approach, when it came time for the ANDRILL project to decide on the core description process for their upcoming Antarctic drilling expeditions, they decided to explore other options. Being a new project, they had the chance to modernize their core description process without having to support legacy approaches. They evaluated several commer-

About the Author

Josh Reed is a masters candidate in the Human Computer Interaction program at Iowa State University and was the IT and data manager for the ANDRILL project's 2006-2007 McMurdo ice shelf drilling expedition. He has over five years of Java programming experience working on a variety of Web and rich client applications. He will finish his degree in June after which he hopes to continue solving complex IT problems.

email?

cial core description applications, but none offered the level of detail and flexibility that ANDRILL required.
Around that time, I was doing graduate work with the CHRONOS project at Iowa State University. CHRONOS specializes in providing IT resources and know-how to the geoscience community. My advisor, Dr. Cinzia Cervato, heard about ANDRILL's search for a core description solution and approached them about developing some software to meet their specific needs. Thus PSICAT was born.

Design

The crux of the problem with the traditional approach is that the sedimentologist is drawing a diagram that represents some data without actually capturing the data itself. It's like drawing a bar chart without keeping the data being charted. However, for scientists, capturing the data is as important as displaying it. Having the data makes searching for specific features easy: one simply specifies the constraints of what is interesting – for example, sections of core that are at least two meters in length and contain no pyrite – and the software crunches the data to find the areas that satisfy those constraints. Working with data also has the advantage that if diagrams can be generated from the data then changes at any scale become simpler — just update the data and regenerate the diagrams.

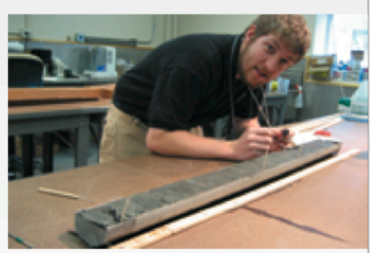
A data-centric approach opens up many other interesting possibilities. The core samples are initially described in a high level of detail. Often, however, it's useful to see the general trends in larger sections of the core. Without the data behind the diagrams, summarization requires going back through the original detailed diagrams and pulling out the key features by hand. By capturing the data, this summarization can be done programmatically. Decoupling the data from the diagram allows it to be visualized in different ways. For example, one user may want to see each individual occurrence of a specific trace fossil. Another may want to see a histogram of all trace fossil occurrences per meter. Since they both access the same data, both these visualizations are possible and easily done.

In the traditional approach used by previous Antarctic geologic expeditions, no core description data was collected in a digital format.
The challenge was to collect the data without drastically modifying the core logging routine that the scientists were accustomed to. PSICAT takes the innovative approach of automatically capturing the data as the user drafts the digital core description diagram. It provides a drawing environment similar to other graphics software, but it's customized to draw core description

PSICAT in Antarctica

When I finally deployed to Antarctica to support PSICAT, I did so with some trepidation. I expected that there would be the inevitable bugs to fix and new features to implement. I was reasonably certain that PSICAT would not be a complete failure, but it was a little nerve-wracking to have a $30 million project depending on my unproven software. During my three months in Antarctica I was stationed in McMurdo where I shared cramped sleeping and living quarters with the other members of the ANDRILL team and worked 14-hour shifts, seven days a week. Living and working in such close quarters with the users meant that there was no place to hide when things went wrong. This situation provided ample motivation to really listen to the users and keep them happy. After the first few weeks of intense interaction with the users and bug fixing, PSICAT's performance improved to the point that I came to believe that it was really going to work. The habit of frequent releases helped me to address the users' requests quickly. In the three months on the ice, I released 30 updates with bug fixes, feature enhancements, and new features, an average of one update every three days. I was able to address most requests in one or two days. The ones that I couldn't address right away usually involved larger-scale modifications to make existing features easier or more intuitive to use. Since I wanted to undertake large changes after giving proper thought to the design and implications of these changes, I made a note of suggestions for investigation and inclusion in future releases.

In the end, PSICAT's performance exceeded everyone's expectations and the additional benefits of using it contributed significantly to the success of ANDRILL's first drilling expedition. Since PSICAT provided instant access to the core description diagrams, it allowed the core to be presented and discussed on the same day it was described by the sedimentology team. Gone were the days of crude photocopies of hand-drawn description diagrams and waiting a day or two for the initial digital versions to be available. Being able to present the diagrams at the morning meetings allowed the scientific team to discuss the core in detail and develop a sampling strategy for further scientific analysis. Another key benefit of using PSICAT was the ability to summarize the data automatically. In the past, the sedimentology team would spend many hours drafting summary diagrams by hand at the end of the expedition. Now summaries could be generated instantly and on-demand. This allowed the science team to see the emerging trends in the core as the drilling progressed, an enormous advantage for understanding the geological evolution of the core. Finally, I observed that PSICAT enabled better science by allowing the science team to revise the data freely to improve its accuracy. In the past, changing some aspect of the core description half-way through the drilling would have required going back and re-drafting every diagram and was obviously not routinely done. Instead, with PSICAT these changes could be executed in seconds resulting in more accurate diagrams.



diagrams. This simplifies the software since the user only sees the tools that she is likely to use. She selects one of the tools available from the palette and begins drawing the diagram. Behind the scenes, PSICAT is capturing the data — depth, grain size, type of rock, symbols — that the user is drawing and simultaneously generates a visual representation of those data on the screen. In addition to capturing data through what the user is drawing, PSICAT also provides a property area to enter data directly. This lets the user enter data that can't be drawn, such as written descriptions and lithology definitions, and specify exact values for properties like depth and grain size, which can be difficult to draw exactly with a mouse. When PSICAT is closed and reopened, it completely regenerates the diagram from the saved data. The user can then continue to add data or edit previously drawn objects, and the additions and changes are immediately reflected on the screen. When the user is finished, she can export the diagram, or a subset of it, to various graphic formats as a single image or multiple (e.g., two meters/page) images.

How does PSICAT do this complex task? The first step is to isolate the various components of a core description diagram and analyze each one to identify which "data" it's encoding. This analysis is used to create what PSICAT calls a "model." Each model defines the data it collects and the constraints on how it can be used. For example, a model may have a depth data element that is defined as a floating point number. The model can enforce this constraint by preventing the depth from being set to a non-number. It's vital to properly define the models and their constraints so that PSICAT can extract the data from what the user has drawn.

Since my background is in computer engineering and not geology, this step required me to work closely with scientists like Dr. Chris Fielding, a sedimentologist at the University of Nebraska working with the ANDRILL project. One of the most critical aspects for this interaction to be successful was to find a common language for communication. Each of us has specialized language and domain-specific concepts that we had to make sure we explained to the other. I took this challenge as a great opportunity to learn about a science that I had little experience with. By the end, although not quite an expert, I had picked up enough of the vocabulary and concepts to use them correctly and convince casual observers that I actually had a background in geology. This was invaluable later in the design phase because I was able to say, "This looks a lot like X does it work the same way?"

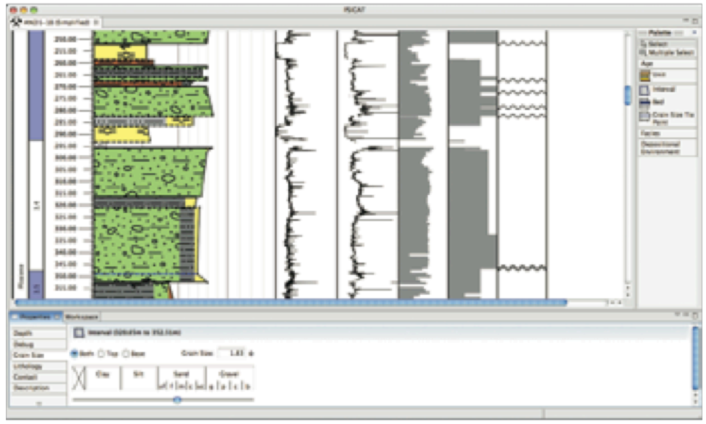


Figure: PSICAT screenshot

The modeling step resulted in the creation of a number of models and in two significant realizations. The first realization was that there were many models but they were all essentially independent. An interval of sandstone in the core may contain areas of horizontal laminations, but on the diagram they are represented separately on different parts of the diagram. The interval model doesn't need to know anything about the sedimentary structure models that the user has defined, even if they overlap in depth. The second realization was that different projects, different groups within the same project, and different tasks, all required different features and views of the data. For example, when trying to decide where to sample the core, the group working with microfossils may be interested in seeing the rock type and where trace fossils were found in the core but not necessarily the detailed written descriptions of the core. Later, when interpreting their results, this same group may be interested in seeing all of the data that's been collected, including the descriptions.

It's important to note that while I worked closely with the ANDRILL project to make sure PSICAT would fulfill its needs, it was my intention from the very beginning to create a tool that would be useful to the broader geoscience community. Thus, when designing PSICAT I focused on developing the features that would be useful to the community while still supporting project-, group-, and task-specific customizations. This allows me to reuse most of the code and only replace or customize it where necessary to meet each group's specific needs.

Implementation

The realizations described above factored heavily into my decision to implement PSICAT as an Eclipse Rich Client Platform (RCP) application. The modular plug-in-based architecture of the RCP lent itself well to the independent nature of the models in PSICAT. Each model and its supporting classes could be implemented and tested as an independent set of

> Living and working in such close quarters with the users meant that there was no place to hide when things went wrong. **This situation provided ample motivation to really listen to the users and keep them happy.**

plug-ins. The RCP also addressed the customization/multiple configuration issue. I could easily create multiple versions of PSICAT, each customized to specific groups or tasks, simply by controlling which plug-ins were included with each version. Another huge advantage of using the Eclipse RCP was the ability to leverage other Eclipse technologies. PSICAT builds extensively on the functionality provided by Eclipse's Graphical Editing Framework (GEF), which allowed PSICAT to get off the ground quickly. I also used the Eclipse Update Manager to distribute updates and new features, something that proved to be immensely useful when PSICAT was deployed in Antarctica.

PSICAT is implemented as four layers built on a core of Eclipse's RCP, GEF, and Update Manager. The first two layers provide the base application and foundation on which the other layers are built. The last two layers implement all of the end-user functionality and project-specific customizations. Each layer is described in more detail below:

- The first layer, closest to the core, is the CHRONOS Modeling Framework. This layer defines the object model and diagram model used by PSICAT and provides services to manage the models, diagrams, resources, and configuration. It contains nothing specific to geology so it can be reused elsewhere.
- The next layer provides the actual PSICAT application and various PSICAT-related services. These services manage depth in the core and user authorization, and provide some basic models that implement common, reusable functionality. This layer also provides the views, editors, and wizards that make up the PSICAT application.
- The third layer consists of two basic types of plug-ins: service plug-ins and column plug-ins. Service plug-ins add new features to the application such as image export functionality and subversion integration. Column plug-ins add the code for capturing and displaying new types of data. Each plug-in generally consists of models for the data being displayed, EditParts for managing the models, figures for displaying the models, and PropertySections to edit the properties of the models.
- The final layer provides project-specific customizations to PSICAT. This includes supplying configuration data and resources for the plug-ins in layer three and implementing any new project-specific column plug-ins.

The current version of PSICAT uses about a dozen Eclipse plug-ins and adds another 40 plug-ins across the four layers. During the development of PSICAT, I adopted the strategy of delivering frequent releases, every two or three weeks, to my test users. Each release was limited in scope, usually including only a few new features or updates to existing features, but the frequent feedback from the scientists made sure that I was headed in the right direction. In the meantime, I could take advantage of the modular nature of the RCP and continue developing new features.

Conclusions

Looking back over the whole experience, from the initial conception of the software, through the design phase, to spending three months at the bottom of the world, it has been a truly amazing journey. Working on PSICAT has been challenging from both a programming standpoint as well as from a user interaction perspective. Being able to experience the whole spectrum of software development, from gathering requirements to delivering an application and seeing how the users interact with it, has been extremely instructive and rewarding as a software engineer. PSICAT has transformed the core description process and the interactions between scientists by making the whole process much more efficient and effective. I'm excited to see how it will further enable the future interaction of the ANDRILL scientists with the core data. PSICAT will be returning to Antarctica again in October 2007 for ANDRILL's second drilling expedition.

Resources

- *ANDRILL:* http://andrill.org
- *CHRONOS:* http://chronos.org
- *Eclipse:* http://www.eclipse.org
- *PSICAT:* http://portal.chronos.org/psicat-site/
- *Josh's blog from Antarctica:* http://josh-in-antarctica.blogspot.com

About ANDRILL

ANDRILL (ANtarctic geological DRILLing) is a multinational collaboration comprised of more than 200 scientists, students, and educators from five nations (Germany, Italy, New Zealand, the United Kingdom and the United States) to recover stratigraphic records from the Antarctic margin using Cape Roberts Project (CRP) technology. The chief objective is to drill back in time to recover a history of paleoenvironmental changes that will guide our understanding of how fast, how large, and how frequent were glacial and interglacial changes in the Antarctica region. Future scenarios of global warming require guidance and constraint from past history that will reveal potential timing frequency and site of future changes.

Acknowledgements

I would like to take this opportunity to thank everyone involved in making PSICAT a success. First and foremost, I'd like to thank my advisor, Dr. Cinzia Cervato, for finding me such a challenging and interesting problem to work on, and for keeping me in the graduate student lifestyle I've grown accustomed to. On the geology side of things, I'd like to thank Dr. Chris Fielding for his help in explaining the intricacies of the core description process, and on the technology side of things I'd like to thank Doug Fils and Xiaoyun Tang for their insightful guidance in designing PSICAT. And finally, I'd like to say thanks to the ANDRILL project for putting their faith in PSICAT, and a special thank you to the MIS sedimentology team. I can't imagine a better group of people to spend three months in Antarctica with.