

Knowledge discovery via a virtual geological observatory

Thomas Landgrebe¹, Dietmar Müller², Simon Williams³

EarthByte Group, School of Geosciences, The University of Sydney, Sydney, Australia,
{¹thomas.landgrebe, ²dietmar.muller, ³simon.williams}@sydney.edu.au

INTRODUCTION

Earth Science has a history of being data rich and information poor, an imbalance which is growing year by year, especially due to a flood of remotely-sensed data. A major problem now faced by geoscientists is how to amalgamate data and to connect them to analysis and process modelling tools. This is particularly relevant to the study of mineral and petroleum resources, which typically form over time periods of hundreds of millions of years. In order to make digital Earth data sets suitable for resource exploration, one needs to be able to restore the geographic positions of all data back through geological time, and link data to models of the processes by which resources are formed. Existing GIS tools fail to model or analyse the inherent time-dependency that exists, making it virtually impossible to analyse in a plate tectonic context. A "Plate Tectonic GIS" in which all data are attached to moving tectonic plates is a fundamental goal toward successful resource exploration and understanding Earth processes through geological time.

To take full advantage of the unprecedented amount of digital geodata now available through emerging electronic networks and the rapidly growing availability of open-source software, radically different methodologies are needed. The Virtual Geological Observatory (VIRGO) project is designed to fill a fundamental technological gap. Through VIRGO we are beginning to exploit the enormous potential of digital information on the World Wide Web to organise the research community, and provide access to a novel data analysis system that facilitates 4D data synthesis through space and geological time. With an open-source platform, universities, government and industry will be able to harness the power of VIRGO to work on their own geological problems, using a mix of proprietary and publicly available data.

MANIFESTATION OF VIRGO

VIRGO is an evolution of the GPlates software concept [2], designed to develop a formal knowledge discovery platform that connects testable ideas, concepts and models regarding plate tectonic and geodynamic processes with reconstructable digital geodata. A GIS-like framework with temporal awareness forms the core of the information model, interoperating with various GIS formats. As such it is able to reconstruct point, line, polygon and raster (gridded) data in accordance with past plate configurations, and model other time-dependent attributes in a coherent fashion. This is fundamental toward understanding the causes and effects of lithosphere-mantle interaction, associated with the practical aim of successful resource exploration. Refer to Figure 1 for an illustrative example of VIRGO simultaneously analysing several reconstructed geographical datasets at 112 million years in the past.

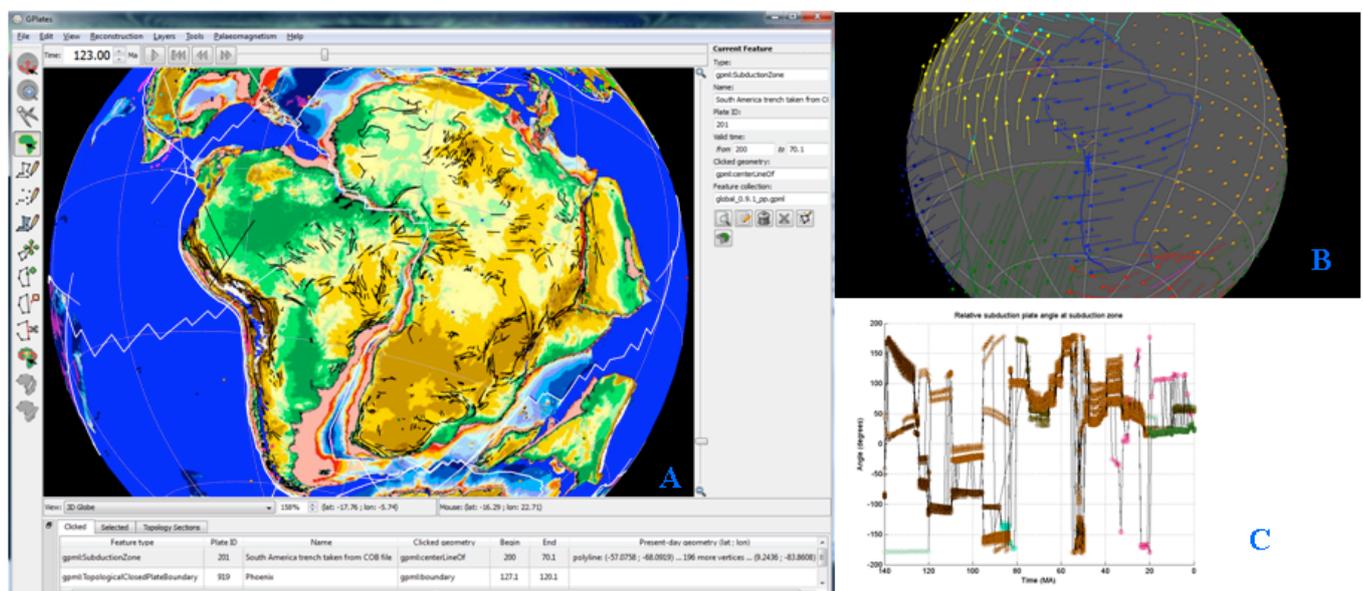


Figure 1: Illustrating VIRGO simultaneously analysing several datasets reconstructed to 123 million years in the past, and example quantitative analyses shown. (A) shows a number of datasets overlaid, followed by the analysis of convergence velocities between Andean ore deposits relative to a neighbouring subducting plate (B), (C).

BENEFITS FOR GEOSCIENCE AND EXPLORATION

Topical application areas include studying the role that tectonic plates play in the nature and time-dependence of convection in the liquid mantle, and complex relationships that exist between plate dynamics and the manner in which ore deposits and hydrocarbons form. Both hydrocarbon and mineral resources commonly have a clear association with

plate kinematic and geodynamic phenomena [1], namely close proximities (with time) to different plate boundaries, to hot mantle upwellings (plumes) or to downwellings associated with subduction and particular plate tectonic processes. An additional complexity introduced is the need to represent plate deformation. Whilst the majority of the continents can be treated as rigid plates, the continental margins (where the majority of hydrocarbons are located) have experienced significant deformation through timescales of the order 100 million years. VIRGO will provide a means to represent and model this deformation through time.

The existence of complex time-dependent relationships and phenomena are well known by practitioners and experts, but the sheer complexity of the types of analyses involved place practical limits on current methodologies and computer-aided tools. VIRGO intends on providing a sophisticated spatio-temporal geoscientific knowledge discovery tool, with important capabilities such as interoperability with important data formats, spatio-temporal processing, interactive user interfacing and a data mining capability all integrated into a coherent software entity.

KNOWLEDGE DISCOVERY VIA VIRGO

The VIRGO concept provides users with a framework facilitating knowledge discovery in this domain by bringing together the important components that are traditionally dealt with independently. Two key aspects play a role, namely providing access to a wide variety of datasets across many domains, and a flexible, interactive data-exploration capability manifested via visualisation-based and quantitative-analysis tools within a spatio-temporal environment. Such a setup places VIRGO in a unique position to systematically study a wide range of phenomena in this domain.

From a data access perspective, VIRGO interoperates with a variety of data formats, such as ESRI shapefiles, NetCDF rasters, common image formats, the GML (Geography Markup Language)-based protocols GPML [3] and GeoSciML [4], and specific plate tectonic formats. This makes it convenient to export data from other tools such as geodynamic simulation software, into GPlates. VIRGO will typically simultaneously access several datasets, as a precursor to subsequent data coregistration and analysis.

The VIRGO provides new protocols for storing, accessing and simulating data in a palaeogeographic framework, thus allowing the user to access this data in a real-time, temporally-coherent fashion, and link these data to process models. The core of VIRGO provides an enabling technology for a new generation of coupled plate tectonic/mantle convection process models by allowing users to create and export geodynamic model boundary conditions, track observations through geological time and visualise reconstructed data and dynamic model outputs. VIRGO enables the user to follow any observations attached to geographic coordinates through geological time, from supercontinental assembly and breakup to the present-day Earth with widely dispersed land masses.

An important capability provided by VIRGO is the provision for quantitative spatio-temporal data analysis and data mining. Once data has been imported, VIRGO will allow for a customised spatio-temporal *co-registration* setup. In this step, various pairs of datasets and attributes are combined together. The co-registration considers spatial relationships over time (e.g. two features may only be associated for a limited period of time), and can handle any geometry types. The concept of a co-registration *operator* has been proposed, serving as an abstract notion representing the manner in which two feature types should be associated. The operator allows for a wide variety of (potentially complex) combining options to be customised by the user, thus configuring an analysis that makes sense for the problem/data at hand. Co-registration is the formal mechanism for transforming disparate datasets into a homogenous data structure, that can serve as an input to subsequent analysis and data mining tools.

Knowledge discovery in VIRGO will be multi-faceted to cope with the complex, dynamic nature of geoscientific research, encompassing the following: discovery of new relationships; confirming hypotheses; relating phenomena and processes; performing complex spatio-temporal analyses (see [5] for a more general discussion). A suite of data mining and analysis tools will be incorporated into VIRGO, allowing for the customisation of workflows to cope with issues such as high attribute dimensionality, redundancy, limited training data, computational complexity, and the spatio-temporal nature of the data. Analyses will be partitioned into both *supervised* and *unsupervised* paradigms, the former more suited to relating phenomena and confirming hypotheses, whereas the latter is suited to finding new undiscovered relationships. Analysis aids in the form of interactive plots, charts and statistics will be an integral part of the process, helping to convey information from machine learning tools such as cluster analyses, data projections, and classifications.

REFERENCES

1. Robb, L., *Introduction to Ore-Forming Processes*. Wiley-Blackwell, 2004. ISBN: 978-0-632-06378-9.
2. GPlates. Available from: <http://www.gplates.org/>, accessed 28 June 2010.
3. GPML – GPlates Markup Language. Available from: <http://www.gplates.org/gpml/>, accessed 28 June 2010
4. GeoSciML. Available from: <http://www.geosciml.org/>, accessed 28 June 2010
5. Weldon, J.L., *Data Mining and Visualization*. Database programming and design, 1996. 9(5).