GHGT-10

Geologic Sequestration Software Suite (GS³): a collaborative approach to the management of geological GHG storage projects

Alain Bonneville*, Gary D. Black, Ian Gorton, Peter Hui, Ellyn M. Murphy, Chris J. Murray, Mark L. Rockhold, Karen L. Schuchardt, Chandrika Sivaramakrishnan, Mark D. White, Mark D. Williams, Signe K. Wurstner

Pacific Northwest National Laboratory, Richland, WA 99354, U.S.A.

Abstract

Geologic storage projects associated with large anthropogenic sources of greenhouse gases (GHG) will have lifecycles that may easily span a century, involve several numerical simulation cycles, and have distinct modeling teams. The process used for numerical simulation of the fate of GHG in the subsurface follows a generally consistent sequence of steps that often are replicated by scientists and engineers around the world. Site data is gathered, assembled, interpreted, and assimilated into conceptualizations of a solid-earth model; assumptions are made about the processes to be modeled; a computational domain is specified and spatially discretized; driving forces and initial conditions are defined; the conceptual models, computational domain, and driving forces are translated into input files; simulations are executed; and results are analyzed. Then, during and after the GHG injection, a continuous monitoring of the reservoir is done and models are updated with the newly collected data. Typically the working files generated during all these steps are maintained on workstations with local backups and archived once the project has concluded along with any modeling notes and records. We are proposing a new concept for supporting the management of full-scale GHG storage projects where collaboration, flexibility, accountability and long-term access will be essential features: the Geologic Sequestration Software Suite, GS³.

© 2010 Elsevier Ltd. All rights reserved

Keywords: Geological storage of CO2; CCS; Project management; CO2 injection; reservoir modeling

1. Introduction

Advanced multi-phase flow and transport modeling codes coupling physical, mechanical, chemical, and biological processes are expected to play a crucial role in understanding/evaluating the feasibility and long term effects of sequestering CO2 in large-scale deep geologic reservoirs. However, the process of managing and interpreting raw data, building the conceptual model of the subsurface domain from that data, and finally transforming/ scaling the conceptual model into the numerical model such that it can be evaluated by these types of simulation codes is currently a tremendous challenge even for experienced modelers. The number of data sets to prepare and the number of tools that can be applied under different conditions is extensive and difficult for modelers to manage. More and more, this process involves an interdisciplinary team of geologists, hydrologists, and

* Corresponding author. Tel.: +1-509-371-7263; fax: +1-509-371-7150.
E-mail address: alain.bonneville@pnl.gov

doi:10.1016/j.egypro.2011.02.318
geochemists - where each may individually maintain information on source data. Additionally the process may involve the development and evaluation of several conceptual and numerical models leading to tens to thousands of simulations being run for just a single site. Organizing, managing, and tracking the data used during all of these stages is critical to the integrity, verifiability, and repeatability of the analysis.

The Geological Sequestration Software Suite (GS³) provides a process by which teams can accomplish the simulation process, while the system automatically manages the data, data translations, versions of conceptual and numerical models, captures provenance and user annotation, and in doing so, vastly reduce the burden on modelers for manually organizing and tracking information throughout the modeling process. GS³ incorporates both off-the-shelf and custom tools that many modelers are already familiar with and allows modelers to use their own commercial software during the modeling process. A key aspect of GS³ is that the user can choose among common simulators to create an input file.

2. General Concept

GS³ is web-based so team members can collaborate regardless of their geographic location. In addition, GS³:

- Streamlines geologic model development and provides access to a growing subsurface database
- Stores all the site data in one location where it can be updated and protected if proprietary; it also stores a history of the geologic models and previous simulations
- Keeps track of data input from different team members working on the same site and tracks models through the simulations process
- Allows application users access to the most advanced science-based simulators available for modeling a site
- Provides simulator developers a platform to quickly validate changes and improvements in the scientific simulators, and regularly benchmark the different simulators
- Supplies a development site for a new community subsurface model.

![Figure 1. GS³ Platform for modeling process and data management. From a User's perspective, here is the development process based on the natural progression of events that occur during the modeling lifecycle.](image-url)
The subsurface model development and simulator execution process required to support geologic sequestration studies is complex and inherently iterative. Figure 1 shows the key aspects of subsurface modeling and their relationship to one another. Although the modeling steps are presented as being part of a cycle, the process is not simple, linear or one-directional. Each element represents a complex process that requires considerable expertise from various members of the project team. At each stage, the user is able to access software tools on their own workstation to support these processes. These tools range from the general purpose such as Excel and shell scripts, to highly specialized solid earth modeling graphical user interface environments like Petrel® or EarthVision®, and simulation codes such as STOMP [1] or TOUGH2 [2].

The overall framework needs to support the development of multiple conceptual models and multi-scale numerical models. At any stage, a modeler may re-enter the process and refine their model or assumptions. For example, upon running a simulation, the need for a revision to the numerical model grid resolution may be recognized. Also not depicted in this figure are the wide range of both data sets and tools necessary to manage the site modeling lifecycle. The number and complexity of the tools and data sets make the process challenging even for experienced modelers.

![Figure 2. GS3 Platform Architecture from a component perspective. If we look on the computer side, we have the architecture of the platform developed on open source. The framework is founded on a unique linking of Semantic MediaWiki® and Subversion®/SVN®.](image)

GS3 will connect two user communities: application engineers, who are modeling a specific site for geologic sequestration; and development engineers, who want to advance the scientific simulators. As the

---

1 http://subversion.tigris.org/
geologic sequestration technology matures through laboratory experiments and field demonstrations, GS³ allows the testing of new theoretical ideas against recognized standards and field observations. With widespread use, GS³ will answer important questions such as the impact of multiple injection sites in the same reservoir or regional basin at resolutions not currently available.

3. Architecture and functionalities

The GS³ Framework integrates the data, tools, and models into a coherent modeling environment focused specifically on carbon sequestration. Whereas the domain user would view GS³ from the modeling process diagram in Figure 1, a computer scientist would view the engine under the hood driving this system as that depicted in Figure 2. The core of GS³ is comprised of the semantic Mediawiki² environment with extensions for scientific data management and processing. The outer ring of the architecture consists of representative functions and capabilities tailored specifically to the geologic sequestration community, the main ones are described below.

3.1. Core Architecture and Simulation Framework

The GS³ framework is an open-source, web-based modeling platform that integrates seamlessly with data management capabilities, including provenance capture and user annotation [3]. Collaboration through data and knowledge sharing is a driving force behind our architecture, as is the goal of providing infrastructure that enables a self-sustaining community. Because wikis provide a shared, collaborative content development space, they are an attractive technology for GS³. Most wikis provide a rich web application development platform and a mechanism to add plug-in modules that operate on the wiki content. Of particular interest for GS³ are the semantic extensions that enable a range of dynamic views of the content.

Figure 3 shows a conceptual view of our architecture. The core is comprised of the semantic mediawiki environment with our extensions for scientific data management and processing. The outer ring of the architecture consists of representative functions and capabilities tailored specifically to the geologic sequestration community.

Wikis also support upload of data files such as images for presentation and document sharing. Their main weakness, in terms of our requirements, is the lack of a rich content management environment. Documents, as opposed to pages, are not versioned and mechanisms for uploading, organizing, and browsing collections of data are limited. However, the richness of the semantic wiki environment led us to base our architecture on this technology and we have developed the required extensions for handling scientific data and processes.

Content Management Systems (CMS) provide a collaborative file sharing capability with features such as access controls, workflow triggers, content versioning, metadata management and searching. They have been successfully applied to a variety of production scientific data management applications [4]. The separation of content and metadata, support for arbitrary content, and an extensible metadata model with the ability to reference externally located data are crucial for handling diverse science data sets. We are therefore integrating Subversion®³, a version control system used for managing software development and possessing many fundamental CMS features. The integration includes support for bulk data upload, automatic metadata extraction, wiki-based browsing of files and collections in Subversion, and semantic markup through page facades to support

---

² http://semantic-mediawiki.org
³ http://subversion.tigris.org/
dynamic views. It is crucial to support both wiki based data access and access via command line utilities and other tools to support the way that scientists work.

Provenance is a special type of metadata that provides a summary of the history and context of the data and a causal graph that shows how a result was computed. In the GS³ context, provenance can be used to track the relationships between simulations and their inputs and output, conceptual and numerical models with simulations, and simulations with analyses. In the context of benchmarking and validation, it can additionally support the documentation of relationships between defined experiment definitions, validation data, and model version information.

The scientific data pipeline component is a data ingest pipeline that accepts published documents (PDF files being the dominant format) and analyzes the content against a taxonomy of important terms for geologic sequestration. The pipeline associates quantitative or qualitative values with those terms by applying natural language processing, and adds the semantic markup to the wiki knowledge base. The taxonomy can be maintained and extended by scientists through normal wiki editing mechanisms. This pipeline provides an automated knowledge extraction capability, with the results available for searching, browsing, and inclusion.

A number of mechanisms will be supported by GS³ for extending functionality by the user community. Heavy reliance on scripting languages is being made wherever possible including Python, Jython, Perl, and PHP (native wiki scripting language).

3.2. Data Assimilation Tools for CO2 Reservoir Model Development

This set of tools allows scientists and engineers to efficiently assemble, assimilate, and track large volumes of disparate data for building geologic models of CO₂ sequestration. Open-source, proprietary and commercial developed reservoir characterization and visualization software tools are available through GS³ for data assimilation and geologic model building in support of carbon sequestration modeling. Resulting geologic models form the basis for numerical modeling of reservoir processes at multiple scales.

The tools are composed of programs, utilities, and scripts for:

- evaluating, preparing, and pre-processing subsurface characterization data
- facilitating import, export, and linkage to external geologic sequestration-specific modules
- assimilating large, diverse data for building and parameterizing conceptual-mathematical models, emphasizing data and processes that are specific to geologic sequestration
- tracking authorship, data transformations, and analysis decisions used in calibrating and integrating diverse hard and soft data to ensure efficient updating of auditable, fully traceable, and defensible models.

Data are maintained in flat files rather than some form of database management system. This also lowers the barrier for end users to extend the capability of GS³. For defining standardized data formats, GS³ is first relying on existing domain file standards such as LAS, SEGY, and RESCUE. But, when these formats are not sufficient, self-describing file formats such as XML and NetCDF will be used.

3.3. Numerical model development

A java-based application is being developed for defining and revising numerical models that will support multiple simulators (e.g. STOMP[1], TOUGH2[2]). This application will assist in developing multi-scale numerical models and will support multiple conceptual models. The software provides forms for the inputs required for defining a numerical model with the data stored in an XML file. While options are provided for the user to define relatively simple grids and material property zoning, the system also provides interfaces via files for accommodating custom or commercial programs/scripts such as grid generation software or sold earth models (e.g. Earth Vision or Petrel). The model definition software has options for defining features on the numerical grids based on geometric objects, which reduces the need for the user specifying node/element numbers (which change when the
grid is revised). The software also provides three-dimensional previews of the numerical grid and material property zonations early in the model development process. The model definition can be revised and simulator-specific input files are generated or regenerated. As part of the GS³ platform, the provenance and dependencies for data files used in the numerical model definition will be tracked. In addition to this metadata providing for configuration control for a specific simulation, the user will be notified when simulation input files would be impacted by changes in the site data.

The suite of tools developed support input file generation for multiple simulators, parallel computing platforms, stochastic simulations, and inverse modeling and parameter estimation. These tools allow researchers to rapidly develop and run multi-scale simulations based on one or more conceptual models using various simulators in a parallel computing environment. Ultimately, users will be able to better understand the efficacy of a particular sequestration approach given the uncertainties of the natural environment.

3.4. Visualizing Uncertainty in Conceptual and Numerical Models for Geologic Sequestration

Scientific visualization and visual analytics are two classes of fundamental strategies that can provide insight into the complex events and processes modeled by numerical simulations. Such strategies are integral to the process of verifying the feasibility and identifying the impact of long term greenhouse gas storage in deep underground reservoirs.

To address the challenges associated with the increasing complexity and expanding scale of numerical simulations, GS³ will incorporate a suite of visualization software tools that will: 1) support cross-model comparisons and analysis based on a collection of simulated data; 2) visually identify salient trends in numerical model uncertainty and sensitivity; 3) provide distributed visualization support for large-scale data; and 4) help track and maintain provenance data for numerical models and simulations. Components from this software will integrate into the GS³ framework and visually create an information feedback loop to accelerate the process of developing, analyzing, and verifying the integrity of a given numerical model, or a suite of models.

3.5. Intercomparison of Numerical Simulators for Geologic Sequestration

Important advancements have been made in numerical simulation codes for geologic sequestration since the last intercomparison of simulators was conducted in 2004 [5]. The current standard for numerical simulator includes capabilities for predicting hydraulic trapping, nonisothermal conditions, transitions to subcritical conditions, ground-surface interactions, injection wells, geochemistry, coupled hydrology-geochemistry-geomechanics, heterogeneous basin-scale domains, and wettability transitions. These capabilities, contained in a suite of science-based simulators, are continuously being updated to analyze sequestration processes in deep saline reservoirs. This increasing complexity in numerical simulators requires frequent validation and intercomparison of simulators and should be a critical component of the code development process.

A computational environment, operating within the GS³ is being designed, developed, populated with data, and demonstrated for the diagnostics, validation, and intercomparison of numerical simulators for geologic sequestration of greenhouse gases in deep saline reservoirs. The following five products will be completed: 1) computational framework for diagnosis and intercomparison; 2) a suite of validation problems and solutions; 3) observational data sets for model validation; 4) standardized documentation of code features and characteristics; and 5) systematic accounting of disagreements for validation problems. The simulator development community at other US national laboratories and universities will be involved in the design and validation problem set.

4. Using GS³

Access to GS³ can be from computers running any standard operating system including Microsoft Windows, Macintosh OS X, and Linux/UNIX. Two approaches could be used as a long-term host of GS³ for external users, 1)
a separate dedicated computer cluster acting as a server; 2) distinct sites with their own GS³ environment through a “virtual machine” distribution of the software.

Recently, Pacific Northwest National Laboratory (PNNL) has installed a cluster to support the development and internal testing of new GS³ software and to house the software for invited users outside PNNL. This gives a good example of the type of multi-purpose configuration needed to run GS³. The first purpose of the cluster is to provide high capacity, reliable, and high performance storage of large volumes of data as will be needed to manage hundreds of ongoing carbon sequestration reservoir modeling projects. This is accomplished by 8 terabytes of RAID 5/6 mirrored disk storage. The second purpose is to provide the ability to efficiently run parallel processor optimized reactive transport simulators used for carbon sequestration modeling. A 4 node, 32 core processor configuration with 96 gigabytes of local memory will initially fulfill this purpose. The final purpose of the cluster is to run the underlying services for the wiki-based web server front-end and high performance content management system employed by the GS³ environment. A 2 processor, 48 gigabytes of memory node with mirrored 1 terabyte disks is dedicated to running the interactive GS³ services separate from the 4 compute cluster nodes.

Although GS³ is a collaborative environment where sharing of work is encouraged and facilitated through the system, it is understood that protecting intellectual property and other proprietary data is a strict requirement dependent upon the users of GS3 and how the modeling being performed is being managed in terms of contracts and other sensitivities. Thus GS³ will employ a multi-level approach to security in order to protect the data maintained within the system. A top-level level of security in GS³ denies access to any of the data within the system except to those with approved accounts. For those with account access, GS³ must provide access control to the data that is maintained both within the wiki and the content management system components.

5. On-going projects

- Starting in 2010, GS³ will be used as a common platform for Sim-SEQ, a multi-year collaborative initiative aiming to objectively evaluate the modeling efforts of different research groups as they are applied to geologic carbon sequestration field tests in the United States [6]. Note also that GS3 will be used by the Southwest Regional Carbon Sequestration Partnership in one of its projects to collect site data and make it available to the large group of collaborators and the DOE program managers. Several other US CS projects are considering using GS³ in a near future.

- In the Fall 2010, a joint research project on geological storage of CO2 in a saline aquifer between China and US researchers will use GS³ as the common platform for collaboration in the framework of a joint program between PNNL, the US National Energy Technology Laboratory (NETL) and the Chinese Academy of Science (CAS). This long-distance collaborative research will use the new GS³ cluster server installed at PNNL (see above).

6. Conclusion

Developed in open source, the Geologic Sequestration Software Suite, GS³, is a computational platform that will enable national and international collaborations. It provides collaborators dynamic access to evolving field and laboratory datasets, tools to facilitate the modeling process, and the ability to rapidly incorporate and validate new science into subsurface simulators. GS³ is designed to reduce modeling process burdens on scientists and engineers by encouraging interdisciplinary teams, making computational tools more accessible, and automatically tracking data provenance.

An extensible, dynamic and integrated computing environment, GS³ stewards users through the modeling process, from data mining, to creating the site geologic model, to launching scientific simulators, to analyzing simulated data. The GS³ framework creates rigor in this modeling process while providing scientists and engineers the flexibility to use their own favorite tools and software. Through the use of GS³, direct comparisons of best modeling practices can be made to accelerate the permitting of new geologic sequestration sites.

A first operating version of GS3 is already available and the version with full functionalities is expected to be released by the end of 2011.
Acknowledgement

GS3 is developed in the framework of the Carbon Sequestration Initiative (csi.pnl.gov), a program supported by the Pacific Northwest National Laboratory in its efforts to accelerate the development and deployment of integrated emissions capture and storage solutions. In addition to work funded by the U.S. Department of Energy and collaborative efforts with China, the Laboratory is investing $50 million in emissions management R&D.

References