

Software aspects of the Geant4 validation repository

**Andrea Dotti¹, Hans Wenzel², Daniel Elvira², Krzysztof Genser²,
Julia Yarba², Federico Carminati³, Gunter Folger³,
Dmitri Konstantinov^{3,4}, Witold Pokorski³ and Alberto Ribon³**

¹ SLAC National Accelerator Laboratory, 2575 Sand Hill Rd, Menlo Park, CA 94025 ,USA;

² Fermilab PO Box 500, Batavia, IL 60510, USA; ³ CERN CH-1211 Geneva 23, Switzerland;

⁴ State Research Center of Russian Federation, Institute for High Energy Physics, Protvino, Russia

E-mail: adotti@slac.stanford.edu

Abstract. The Geant4, GeantV and GENIE collaborations regularly perform validation and regression tests for simulation results. DoSSiER (**D**atabase of **S**cientific **S**imulation and **E**xperimental **R**esults) is being developed as a central repository to store the simulation results as well as the experimental data used for validation. DoSSiER is easily accessible via a web application. In addition, a web service allows for programmatic access to the repository to extract records in JSON or XML exchange formats. In this article, we describe the functionality and the current status of various components of DoSSiER as well as the technology choices we made.

1. Introduction

Software collaborations in high energy physics regularly perform validation and regression tests for simulation results. It is the case for groups related to Geant4 [1] and GeantV [2] which are tool-kits for the simulation of the passage of particles through matter and in the case of GENIE [3] which is a Monte Carlo event Generator for modeling neutrino interactions. A validation test compares simulation results with data obtained by various experiments while a regression test compares results of two or more versions of the software for an observable. As the areas of application expand, tests are improved and new tests are developed to insure that the physics processes are modeled correctly. As the number of regularly performed validation and regression tests increases and the collection of results grows, storing them and making them available to members of participating collaborations and the user community in general becomes a challenge. Therefore DoSSiER (**D**atabase of **S**cientific **S**imulation and **E**xperimental **R**esults) is being developed to collect and organize the materials in one central repository and to make this data easily available via a web application and web service. DoSSiER should be of general interest to collaborators and the user community. For example it should help experimenters to find answers to questions like:

- How well does the simulation describe the data of interest to the experiment?
- What data are actually used to validate the implemented physics?
- Which of the various models best describe the data of interest to the experiment?
- What are the benefits of switching to the latest improved version of the software?

2. Requirements

Below we briefly list the requirements and the desired functionality.

- It should be possible to store multidimensional experimental and simulation data used for validation in the form of arrays of data points and as histograms with meta-data. In addition, it should be possible to store complete plots with meta-data.
- Provide an API (Application Program Interface) to the database.
- Provide a web service which allows programmatic access. This allows the retrieval of records using any programming language that can handle http requests and can parse the returned JSON/XML strings. A C++ API is of particular interest since it allows access to DoSSiER records directly from C++ validation jobs.
- The web application should provide an easy-to-use web-interface providing search engines and easy-to-navigate menus with default values, and should allow the user to select and overlay tests with experimental data. The actual plots are created on the fly.
- The web application and service should be based on modern Internet technologies and industry standards. It should also provide a fresh modern look and adhere to all security requirements set by the hosting laboratories.

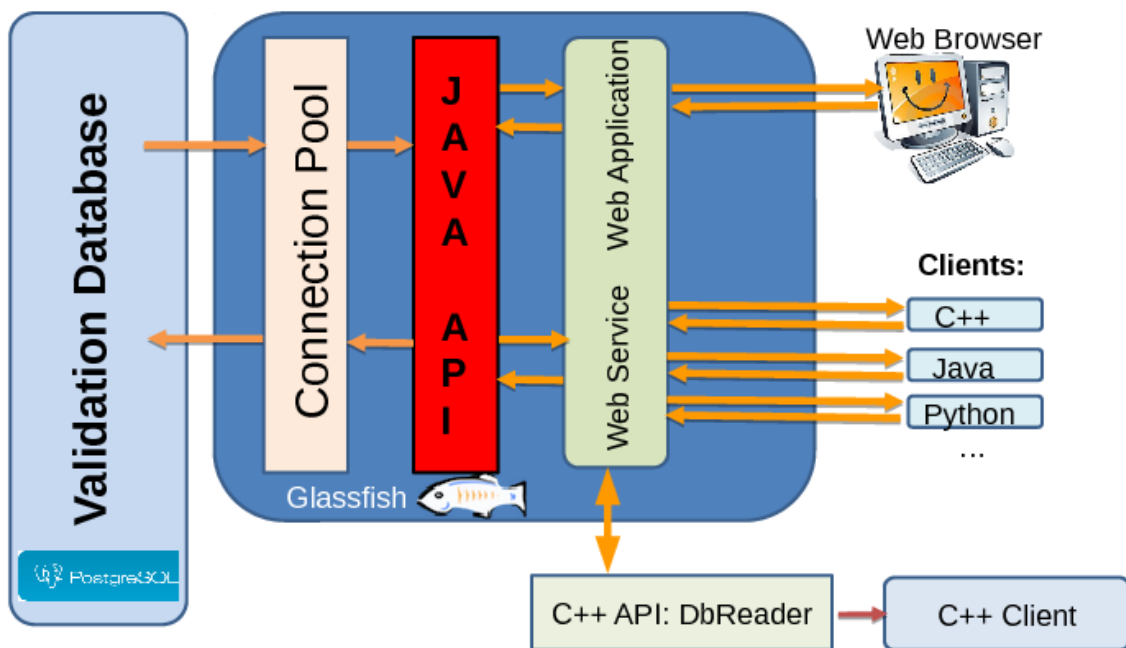


Figure 1. Software components of DoSSiER.

3. Software Components

DoSSiER is comprised of the software components shown in Figure 1. Each component will be described in detail in the following part of this section. All the technologies that we use are very well integrated in the IDE (integrated development environment) of our choice, NetBeans [4]. This greatly simplifies the development effort.

- **Validationdb** is a relational PostgreSQL database [5] that stores data in the form of images with meta-data, multidimensional histograms, and arrays of data points. Images of final

plots (gif, jpeg, etc.) are stored as blobs (binary large objects) while histograms and data points are serialized into arrays in the database. The meta-data provides a description and lists literature references to the experimental data and other parameters that describe the test/experimental result (e.g. Geant4 version, beam particle, beam energy/momentum, reaction, target material, secondary particles etc.). Besides the mandatory parameters one can supply additional tags to provide more information. For literature references we make use of the INSPIRE High-Energy Physics Literature Database [6]. To facilitate fast access, we keep a copy of the relevant information of the referenced articles in DoSSiER, but a link to INSPIRE is provided in case that one wants to access additional information (e.g. retrieve a copy of the full article).

- **ValidationLib:** is a multilayered Java API based on the data access object (DAO) [7] software design pattern. It provides an abstract interface to the database without exposing details of the database. To provide a format for data exchange and/or for uploading results into DoSSiER we use the Java Persistence API [8] or gson [9] which are Java libraries for serialization and deserialization of Java Objects into and from XML/JSON strings, respectively.
- **DoSSiER Web Application and Web Service [10]** offers menus that are easy to navigate, allow for interactive selection and overlay of compatible results. Data is compatible when observable and measurement variables reasonably agree. The application logic is such that only compatible data can be selected via the provided menus. The web application also provides security and authentication to grant access to groups of functions and data that are internal to the participating collaborations, e.g. viewing results from development releases, upload of new test results and modification of selected data. The Web Application is based on Java Platform, Enterprise Edition (Java EE) [11] and utilizes the PrimeFaces [12] open source JSF (Java Server Faces) component suite. The database is accessed using the classes provided by the Java API. It is deployed on a GlassFish application server [13] running on the FermiCloud service [14]. The DoSSiER Web Service is a RESTful [16] web service based on JAX-RS [17] that allows programmatic retrieval of data e.g. by validation jobs. Currently the functionality is relatively limited in that it only allows for retrieving results or the retrieval of the dictionaries describing the meta-data via their primary database key, but we plan to extend the API with e.g. search commands.
- **DbReader** is a C++ API providing access to data stored in DoSSiER directly from C++ clients. This is used by selected Geant4 validation jobs to directly retrieve the experimental data for comparison from DoSSiER.
- **Python tools** Not shown in Figure 1 are the Python-based ancillary tools that we are using to read and convert histograms and collections of data points from/to different data formats including ROOT, ASCII (CSV) and the JSON format used by DoSSiER. The files in JSON/XML format can then be used to download from and upload to DoSSiER. Other functionality provided by this tool are the ability to inspect and interact with histograms (matplotlib) and script integration with the command line tools(CLI).

4. DoSSiER Views

In this section we highlight a few selected views currently provided by DoSSiER. Figure 2 shows the DoSSiER entry page. It gives access to viewers of experimental data and Geant4 simulations. At the moment the GENIE and GeantV buttons are placeholders as adding data is still in progress. In addition the menu on the right allows access to database statistics, the dictionary browser and provides access to the RESTful web service.

Figure 3 shows the DoSSiER dictionary browser which allows to navigate the references, test descriptions and the meta-data dictionaries used to describe the data stored in DoSSiER.

Figure 4 shows the display (blue squares) of a Geant4 simulation estimating the total cross section of π^- on a Carbon target as a function of kinetic energy. On the top of the view we find information describing the Geant4 test and the publications from which the data for physics validation was extracted. The menus on the top of the graph are automatically created based on the contents of the database. In this case, one can select results based on target material (Pb, Cu, S, C, Au, Ag, Sn, Ca, Fe) and beam particle (π^+ , π^-). If there are more variables to choose from, then additional menus will be created. Each test comes with a set of defaults and one automatically gets the default plot when initially selecting a test. When a user selects the pull down menu for a given variable, the fields of the preselected defaults are highlighted to guide further refinement of the selection. In addition, if the overlay option is selected, DoSSiER will search the database for any experimental results matching the meta-data of the test selection. In this case the search finds measurements from three different publications (orange, green, red curves) and overlays them on the simulation result. For plotting we use the plotly.js [15] JavaScript library.

Figure 5 shows a display of experimental data as a table, which shows values of the differential cross section of deuteron production induced by a neutron beam on a Cu target. The header of each table lists the meta-data (beam, reaction, target, secondary, observable ...). DoSSiER allows the data to be downloaded in various formats (EXCEL, PDF, CSV, XML, JSON).

5. Conclusions and Outlook

DoSSiER provides a repository to collect and organize validation and regression test results and the experimental data used for physics validation tests. The web application component of DoSSiER allows easy access to the information for members of simulation collaborations and the general user community. The web service component allows for programmatic access to the data stored in DoSSiER. We are continuously adding new tests and experimental data to the repository. Originally we concentrated on Geant4 tests of hadronic and electromagnetic physics that are of interest for high energy physics experiments, but in the future we plan to include tests that are of interest in other areas such as medical and space science. Also we are in the process of adding GEANIE and GeantV data in cooperation with the respective collaborations.

References

- [1] Allison J et al. 2016 *Nuclear Instruments and Methods in Physics Research A* **835** (186-225).
- [2] <http://geant.cern.ch/>.
- [3] The GENIE Neutrino Monte Carlo Generator, arXiv: 0905.2517.
- [4] *Welcome to Netbeans*, <https://netbeans.org/>.
- [5] PostgreSQL, <http://www.postgresql.org/>.
- [6] High-Energy Physics Literature Database, <http://inspirehep.net/>.
- [7] <http://www.oracle.com/technetwork/java/dataaccessobject-138824.html>.
- [8] http://www.oracle.com/technetwork/java/javaee/tech/persistence-jsp-14004_9.html.
- [9] <https://github.com/google/gson>.
- [10] <http://g4validation.fnal.gov:8080/DoSSiER/>.
- [11] <http://www.oracle.com/technetwork/java/javaee/overview/index.html>.
- [12] PrimeFaces Ultimate JSF Framework, <http://www.primefaces.org/>.
- [13] Glassfish Server, <https://glassfish.java.net/>.
- [14] FermiCloud, <http://fclweb.fnal.gov/>.
- [15] The open source JavaScript graphing library that powers Plotly, <https://plot.ly/javascript/>.
- [16] Representational state transfer (REST) https://en.wikipedia.org/wiki/Representational_state_transfer.
- [17] Java API for RESTful Services (JAX-RS), <https://jax-rs-spec.java.net/>.

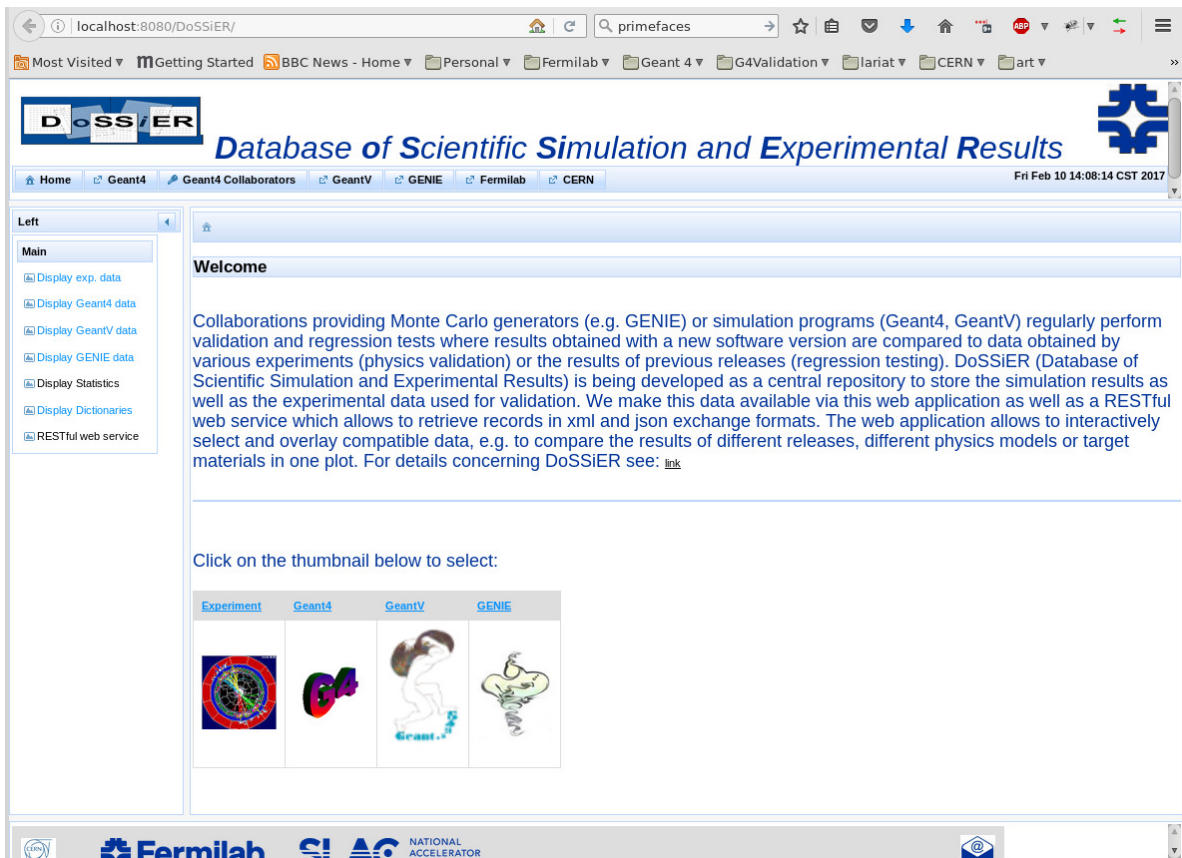


Figure 2. The DoSSIeR entry page. It gives access to viewers of experimental data and Geant4 simulations. In addition the menu on the right allows access to database statistics, the dictionary browser and provides access to the RESTful web service. At the moment the GENIE and GeantV buttons are placeholders as adding data is still in progress.

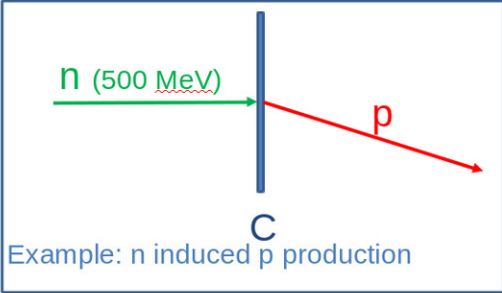
DoSSiER Database of Scientific Simulation and Experimental Results

Home Geant4 Geant4 Collaborators GeantV GENIE Fermilab CERN Fri Feb 10 16:05:07 CST 2017

Left

- Main
 - Display exp. data
 - Display Geant4 data
 - Display GeantV data
 - Display GENIE data
 - Display Statistics
 - Display Dictionaries
 - RESTful web service

Dictionary look up



Example: n induced p production

Beam: mono energetic neutrons
Target: Carbon.
Secondary: protons.
Observable: differential cross section
Reaction: particle production
MC detail: e.g. BIC, geant4.10.1.p02

In addition parnames/parvalue pairs can be added

The meta-data (Beams, Materials, Particles, Observable, Reaction, Reference, Monte Carlo details...) describing the results is stored in dictionaries. The meta-data is used to select and search for a specific result. In addition it is used to match experimental and simulated results. Select the dictionary you want to browse from the menu below.

Name	Description
Access	describes the access level to the data e.g. some data is open to everyone (public) while other results can only be viewed by members of the collaboration
Beams	describes the Beams. Complicated Beams (e.g. neutrino flux files, test beams consisting of many particles can be described by the schema).
Datatypes	describes in what form data is stored e.g. 1D/2D histograms, Dataset Points etc.
Materials	contains all the different target materials.
Mcdetails	contains information about the MonteCarlo tool used like version tag and the physics model used.
Mctools	lists the different Monte Carlo tools that contribute to results in the data base.
Observables	lists the different observables.

Fermilab NATIONAL ACCELERATOR

Figure 3. The DoSSiER dictionary browser allows to navigate the references, test descriptions and the meta-data dictionaries used to describe the data stored in DoSSiER.



Figure 4. Display of a Geant4 simulation (blue squares) estimating the total cross section of π^- on a Carbon target as a function of kinetic energy. On the top we find information describing the Geant4 test and the publications from which the data for physics validation was extracted. The menus on the top of the graph are automatically created based on the contents of the database. In this case, one can select results based on target material (Pb, Cu, S, C, Au, Ag, Sn, Ca, Fe) and beam particle (π^+ , π^-). If there are more variables to choose from, then additional menus will be created. Each test comes with a set of defaults and one automatically gets the default plot when initially selecting a test. If the overlay option is selected, DoSSiER will search the database for any experimental results matching the meta-data of the test selection. In this case the search finds measurements from three different publications (orange, green, red curves) and overlays them on the simulation result.

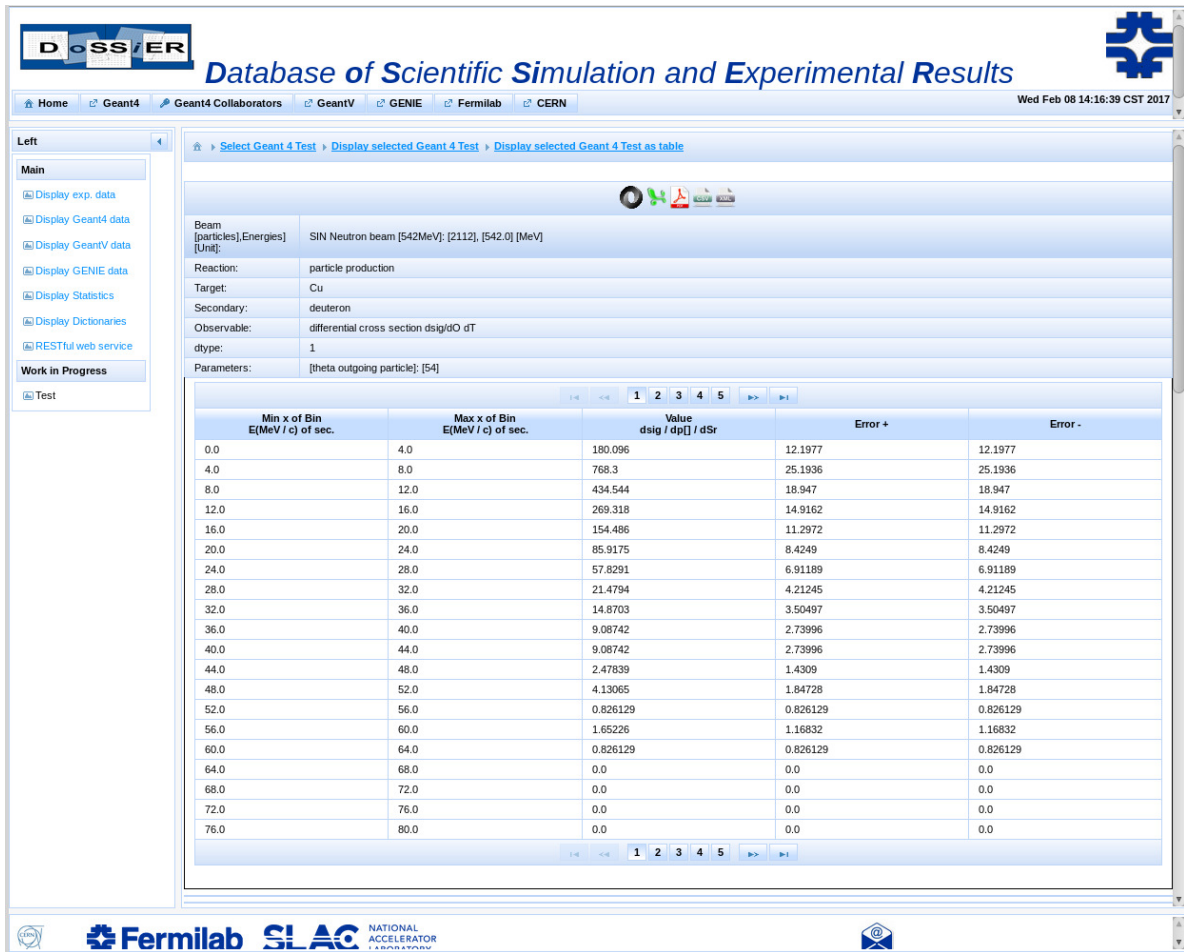


Figure 5. Display of Geant4 simulation data as a table. Shown are differential cross section values of deuteron production induced by a neutron beam on a Cu target. The header of each table lists the meta-data (beam, reaction, target, secondary, observable ...). DoSSiER allows to download the data in various formats (EXCEL, PDF, CSV, XML, JSON).