Po**S** PROCEEDINGS OF SCIENCE

DoSSiER: Database of Scientific Simulation and Experimental Results

Hans Wenzel[*]**, Julia Yarba, Krzysztof Genser, Daniel Elvira**
Fermilab
E-mail: `wenzel@fnal.gov`

Witold Pokorski, Federico Carminati, Dmitri Konstantinov,[†] **Alberto Ribon, Gunter Folger**
CERN

Andrea Dotti

SLAC

The Geant4, GeantV and GENIE collaborations regularly perform validation and regression tests for simulation results. DoSSiER (**D**atabase of **S**cientific **Si**mulation 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 can be easily accessed 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.

38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA

---

[*]Speaker.

[†]also Institute For High Energy Physics (IHEP), Protvino

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 toolkits for the simulation of the passage of particles through matter and in the case of GENIE [3] which is a MC 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 **Si**mulation 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. This application should be of general interest to collaborators and the users community. For example it should help experimenters to find answers to questions like:

- What data are actually used to validate the implemented physics?

- How well does the simulation describe the data of interest to the experiment?

- Which of the various models describes the data of interest to the experiment the best?

- 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 form arrays of data points and as histograms with metadata. In addition, it should be possible to store complete plots with metadata.

- Provide an API (Application Program Interface) to the database.

- Provide a web service which allows programmatic read-only access. This allows to retrieve 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 to access DoSSiER records directly from C++ validation jobs.

- The web application should provide an easy to use web-interface providing search engines, easy to navigate menus with default values and should allow the user to select and overlay tests with experimental data. In this case 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.
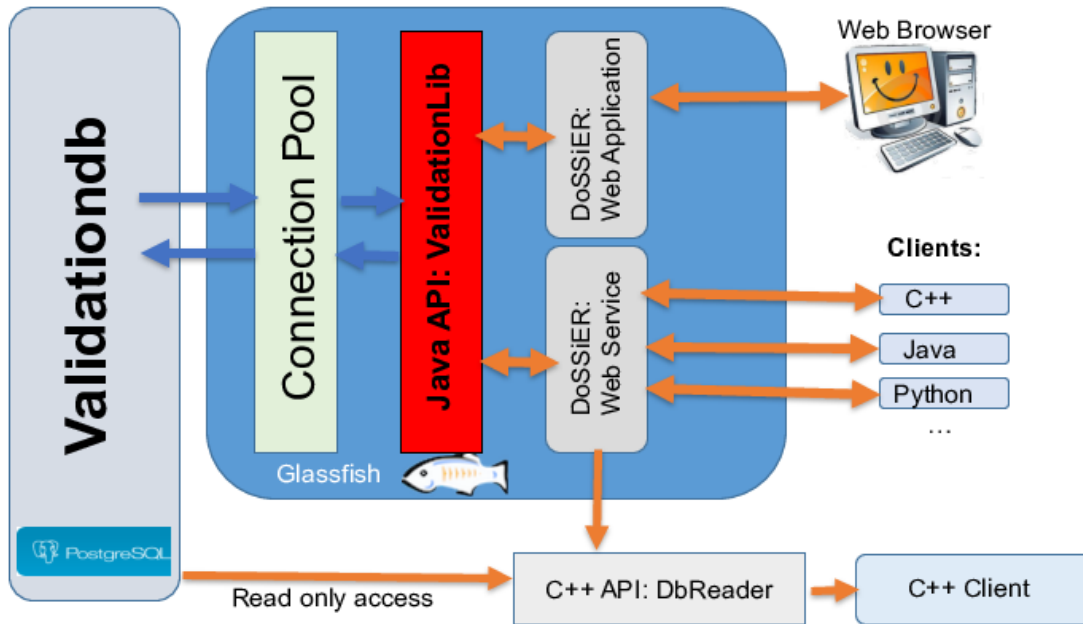
Figure 1: Software components of DoSSiER.

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 form of images with metadata, 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 metadata provides a description, lists literature references to the experimental data, as well as 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 are 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 guarantee 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 [10]:** offers easy to navigate menus and to interactively select and overlay compatible results. Data is compatible when observable and measurement variables 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 tests 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].

  Figure 2 shows the display of the data from an experiment measuring the total cross section of charged pions as a function of kinetic energy. The menus on the top are automatically created based on the contents of the database. In this case, one can select by target material (S, Be, C, Li6, Li7, He) and beam (π^+, π^-). If there are more variables to choose from additional menus will be created. Each experiment comes with a reasonable set of defaults and one automatically gets the default plot when initially selecting an experiment. When the 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 by the user. Figure 3 shows a display of experimental data as a table, which shows values of the differential cross section of proton production induced by a neutron beam on a Cu target. The application allows to download the data in various formats (excel, pdf, csv, xml, json).

- **DoSSiER Web Service:** is a RESTful [15] web service based on JAX-RS [16] 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 metadata via their primary database key respectively, 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.

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 functionalities provided by this tool are the ability to inspect and interact with histograms (matplotlib) and script integration with the command line tools(CLI). It can serve as a PYTHON API to be used in python programs.

4. Conclusions and Outlook

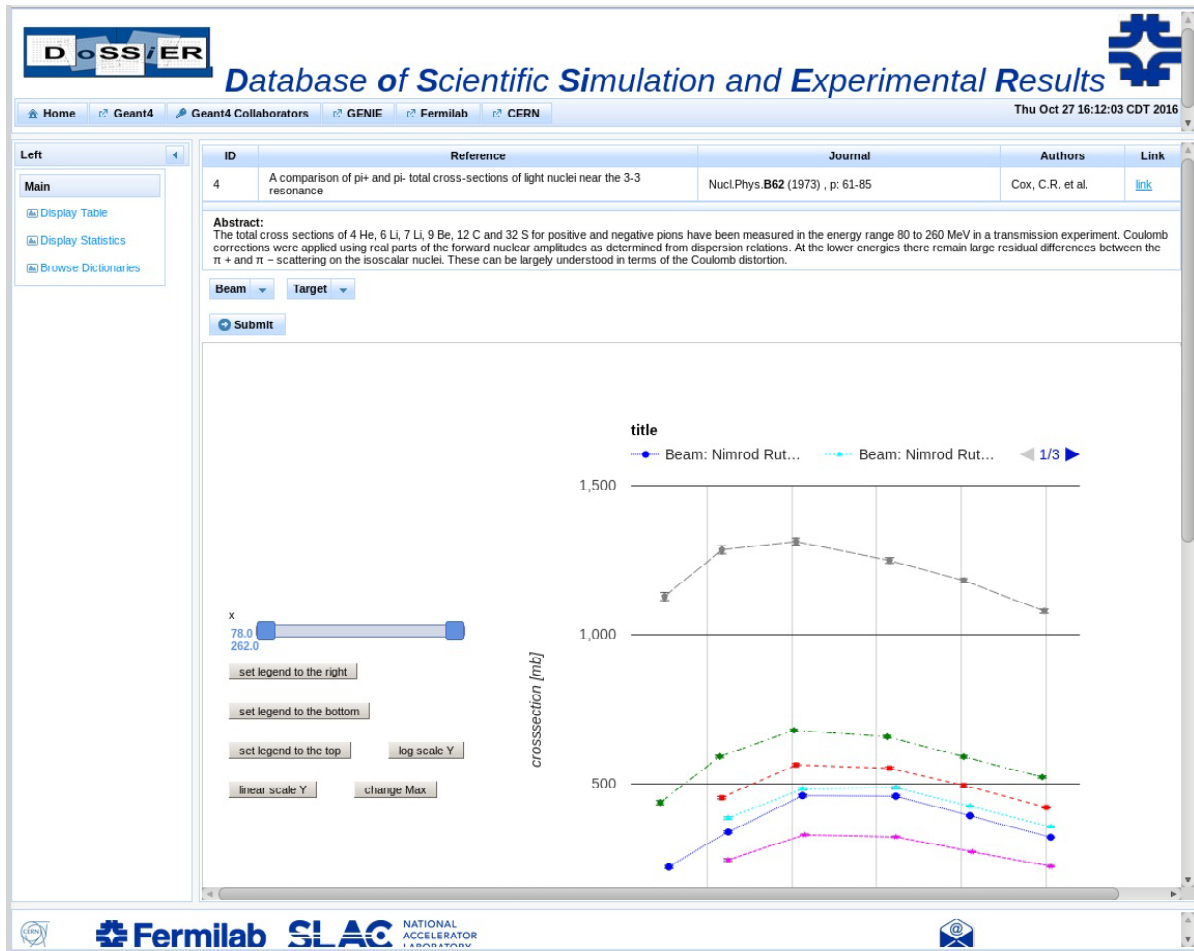A repository to collect and organize validation and regression test results as well as the experi-

Figure 2: Display of the data from an experiment measuring the total cross section of charged pions as a function of kinetic energy. The menus on the top are automatically created based on the contents of the database. In this case one can select by target material and beam. If there are more variables to chose from additional menus will be created. Each experiment comes with a reasonable set of defaults so one automatically gets the default plot when initially selecting the experiment.

mental data used for validation has been designed and implemented. A web application allows easy access to the information for members of simulation collaborations and the general user community. A web service allows for programmatic access to DoSSiER. Additional tests and experimental data are committed to the repository frequently. Originally we concentrated on 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.

References

[1] J. Allison et al., "Recent developments in Geant4", Nuclear Instruments and Methods in Physics Research A 835 (2016) 186-225 and references therein.

[2] "GeantV home page", http://geant.cern.ch/.
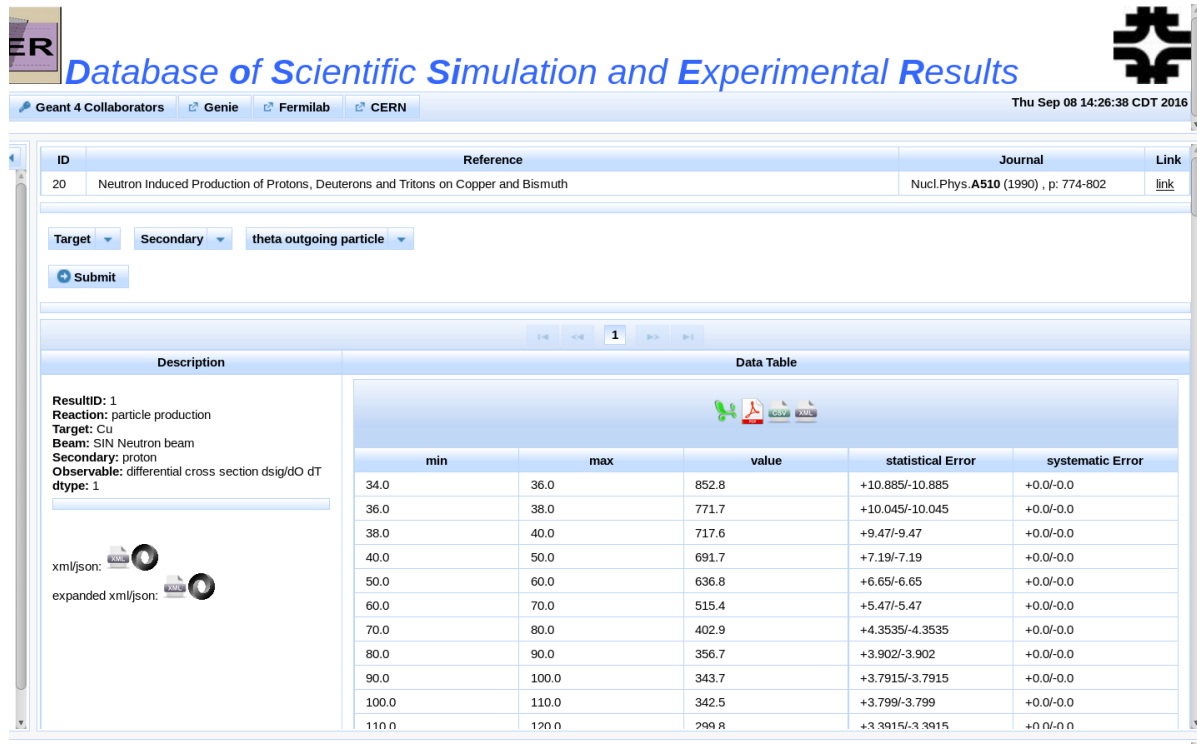
Figure 3: Display of experimental data as a table, which shows values of the differential cross section of proton production induced by a neutron beam on a Cu target. The application allows to download the data in various formats (excel, pdf, csv, xml, json).

[3] "GENIE home page", `http://www.genie-mc.org/`.

[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-140049.html".

[9] `https://github.com/google/gson`.

[10] `http://g4devel.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] "Representational state transfer (REST)"
`https://en.wikipedia.org/wiki/Representational_state_transfer`.

[16] "Java API for RESTful Services (JAX-RS)", `https://jax-rs-spec.java.net/`.