Database of Scientific Simulation and Experimental Results

Outline:

- What is DoSSiER?
- Why do we need a repository for experimental data?
- Requirements
- Components
- Choice of technologies
- Steps to create a new test
- Summary and Conclusions.

What is DoSSiER?

- Name says it all: **D**atabase **o**f **S**cientific **Si**mulation and **E**xperimental **R**esults
- Collection of experimental data that can be used to validate all aspects of Geant4.
- Tests that compare the simulation with the experimental data.
- Test results and experimental data are stored in a database and easily available as web pages (web application) or programmatically (web service) → DoSSiER. Continuously run the tests: regression and validation testing.
- Should provide a one stop location covering all aspects of Geant4 (em, hadronic, medical, …..)

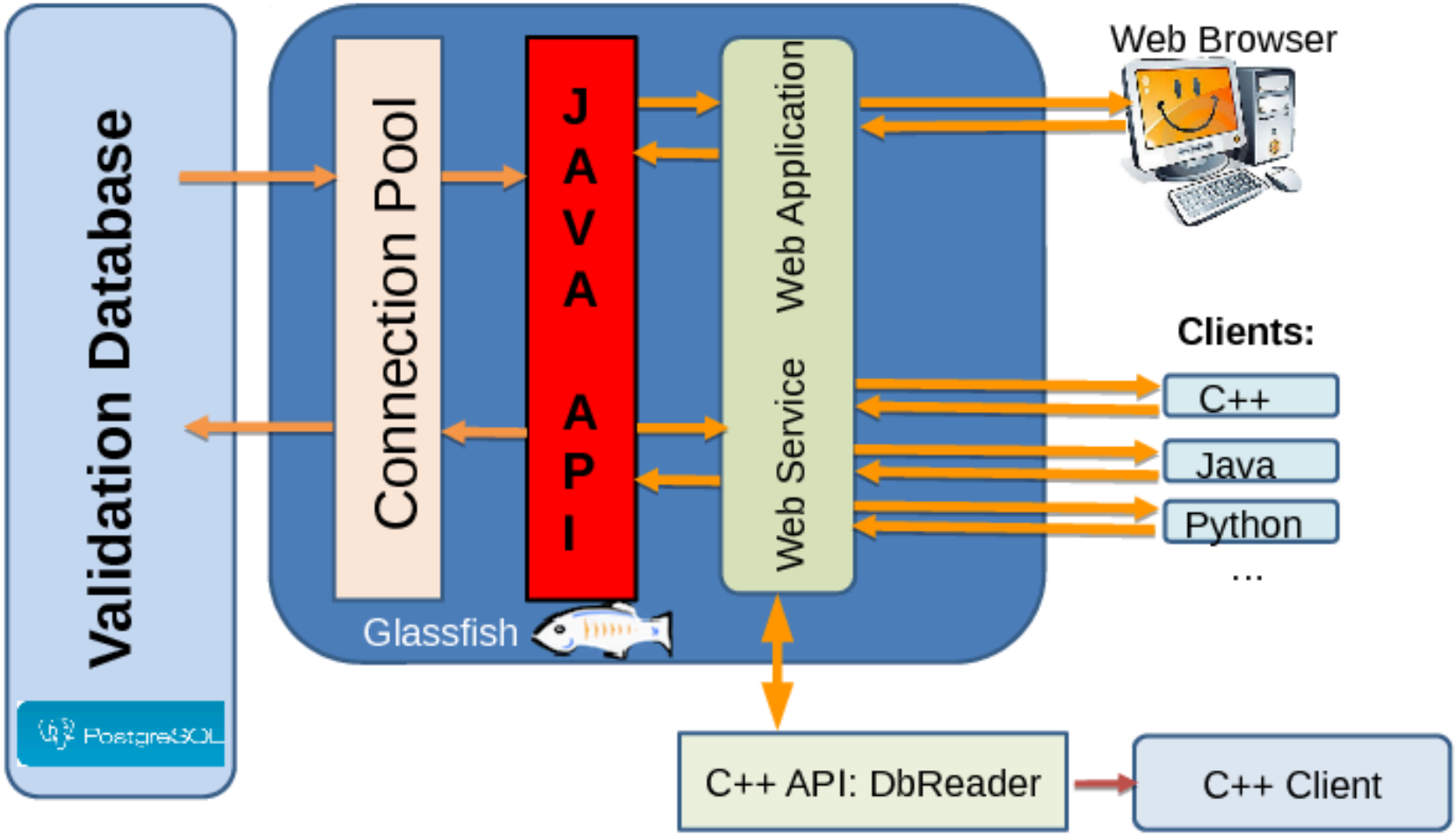
Many sources of experimental data

- Existing databases:
  - HEPData: https://hepdata.net/
  - ExFor: http://www.nndc.bnl.gov/exfor/exfor.htm
  - Inspire: https://inspirehep.net/ → cross link for references
  - Particle Data Group: http://pdg.lbl.gov/
- Directly from article/thesis: if only paper copy OCR, engauge digitizer….
- Both e.g. geant4 or Genie developers have data collections.
- Experimental web sites: https://spshadrons.web.cern.ch/spshadrons/
- Compilations: https://www.oecd-nea.org/dbdata/bara.html
- Labor intensive and error prone, biased. Currently ~ 3500 experimental data curves in the db. (compared to 128K simulated results)

Requirements

- Provide repository:
  - to store experimental validation data as raw data.
  - to store simulation results as raw data and as static plots.
- Provide display web-application which:
  - allows to select and overlay compatible tests,
  - allows to overlay experimental data,
  - allows automatic upload into repository,
  - allows to display static images,
  - provides search functions and easy navigation.
- Provide REST-ful Web service which:
  - allows programmatic access to the data
- Modern look, meaningful search, easy to navigate menus.
- Based on modern internet technology and industry standards.
- Secure!
- Must have defined deliveries/timeline.

Components



+ python ancillary tools: e.g. converter between various data formats: root, ascii, json to DoSSiER json format. Needs to be changed for new json format!

Choice of technologies

Technologies

Java	**Java Programming Language:** Java EE framework, JAX-RS RESTful web service, Maven software management tool.	GlassFish	**GlassFish:** Combined web application and web service server hosted on FermiCloud.
NetBeans	**Netbeans IDE:** Integrated Development Environment that works well with GitLab and GlassFish.	PostgreSQL	**PostgreSQL:** Open source relational database hosted by the Fermilab database group.
GitLab	**GitLab:** Web-based Git repository hosting service for managing collaborative revision.	PRIMEFACES	**PrimeFaces:** Java Server Faces based framework for creating clean and easy to navigate web pages.

Steps to create a new test

1. Create a new test description in db:
   - Required meta data:
     - Id, Name of test, Description, Responsible(s), working group, references to experimental data, keywords
   - Done manually by myself → not a very high frequency.
2. Provide your test data (e.g. as root files) and metadata → use python tool to obtain json files, that can be uploaded to the data base.
3. Authenticate and use web application to upload the data into Dossier.
4. Decide on default selection → manually done by myself.

5. For future uploads just repeat step 2 and 3. Usually very minor changes to meta data needed.

DoSSiER

DoSSiER **Database of Scientific Simulation and Experimental Results**

Home | Geant4 | Geant4 Collaborators | GeantV | GENIE | Fermilab | CERN | Search

Mon Sep 25 12:07:10 CDT 2017

Left

Main

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

ID	Name	Description	Responsible	Working group	Keywords
10000	Franz	Neutron-induced production of protons, deuterons and tritons by neutrons between 300-580 MeV	Hans Wenzel (Fermilab)	Geant4 hadronic working group	particle production differential cross section Thin Target neutron induced

References to experimental data used to validate this test

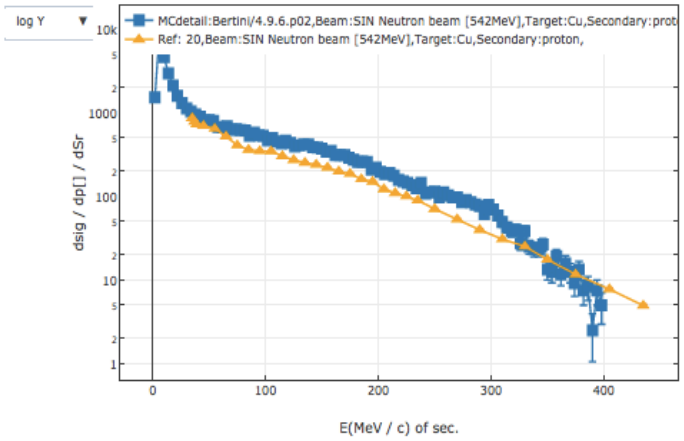
Reference ID	Title	Journal/URL	Authors	Link
20	Neutron Induced Production of Protons, Deuterons and Tritons on Copper and Bismuth	Nucl.Phys.**A510** (1990) , p. 774-802	Ero, J. et al.	link

Beam | MCdetail | Target | Secondary | ParValues | Submit

Table | Default

Print

RON-INDUCED PRODUCTION OF PROTONS, DEUTERONS AND TRITONS ON COPPER AND BIS



Python Upload Tool: usage

```
Usage:  plot_histofiles.py  [--comand|-c <cmd>] [--output|-o <ofile>] [--metadata|-m k[:type]=v] [--metadatafile <mdf>]
<files>
where:
  <files> are the files to read.
       File extension determines format. CSV is the text format
       from G4Analyais. For ROOT format, you need t specify the
       name of the file to be read in. Ex: file.root:h1
       pickle format is supported (file should be created with command "save")
  <cmd> is one of ("plot", "convert","save","genmd","list")
      "plot" (default) to plot the content of the file
      (requires matplotlib)
      "convert" creates an output file in JSON format suitable
      for FNALdb upload
      "save" saves histograms in internal format to pickle file "histos.pkl"
      "genmd" generates a metadata skeleton file as specified in <files>
      "list" shows content of ROOT File (TKey). Only for ROOT format.
  <ofile> is the output file name (default="output.json") for
      converted output for FNALdb
 <hn:k[:type]=v> is a key-value pair to add as metadata to FNALdb output
      hn is a regexp to assing the metadata to histogram based on names.
      k is the key of the metadata, type (default INT) can be INT if the value has to
      interpreted as integer value of STR if it must be interpreted as string
      or FLT if it is a floating point value
      e.g. -m .*A:INT=1 means add to all objects the integer metadata 1 with key A
 <mdf> is a json of pickle file containing the metadatada in a format of the type:
       { "regexpName" : { metadata } } where regexpName is a regular expression that
       matches a converted object name (the name being the ROOT TObject name or CSV full-filename)
```

e.g. Extracting TGraph from a root file and creating json file for upload to DoSSiER

Command used to add one record to the database
(cross sections for pion- on Aluminium)

```
python ./plot_histofiles.py --metadatafile meta_pion-_G4Al_totalxserr.json \
-c convert \
--output pion-_G4Al.json pion_xs.root:pion-_G4Al_totalxserr
```

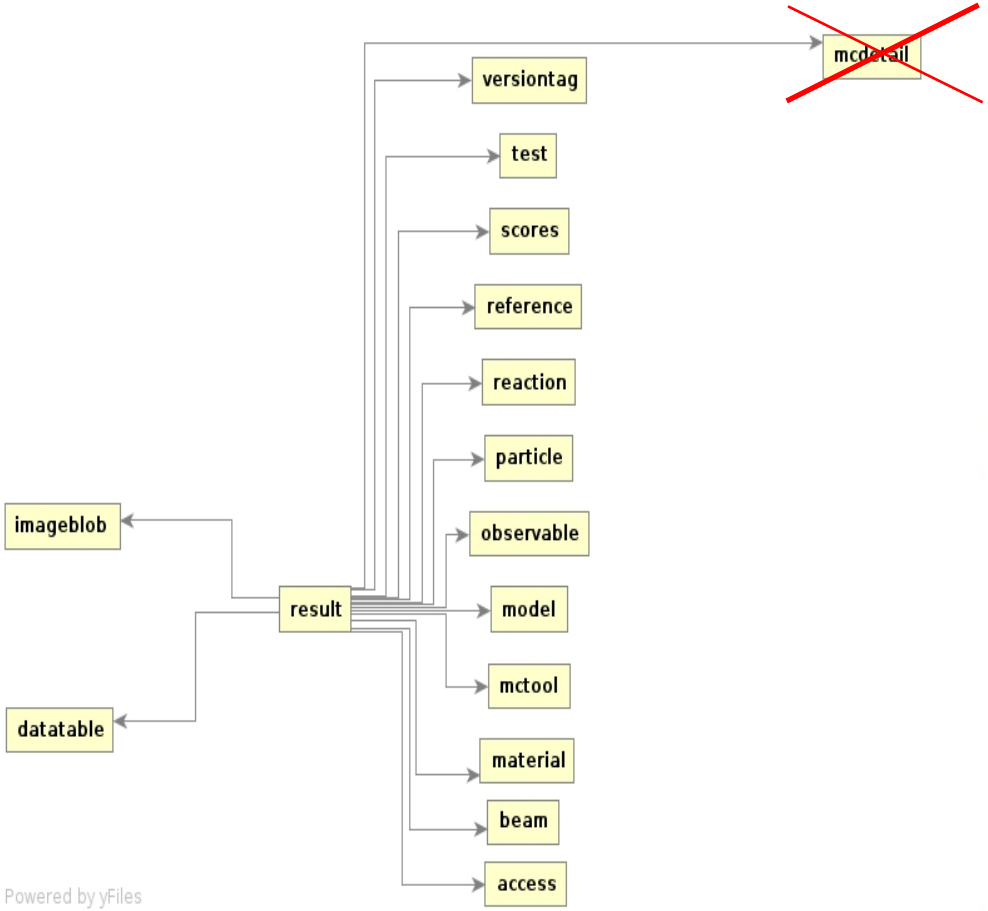
The Meta Data file: (in addition from the meta data provided by root)

Note: meta data format will change to more human readable form!

```
[wenzel@ironman uploader]$ more meta_pion-_G4Al_totalxserr.json
{
 ".*": {
   "trid": null,
  "testlnk":10002,
  "referencelnk": null,
  "targetlnk": 13,
  "reactionlnk": 4,
  "modtime": null,
  "secondarylnk": 0,
  "observablelnk": 10,
  "accesslnk": 1,
  "imageblobslnk": 0,
  "parnames": [],
  "beamlnk": 100,
  "parvalues": [],
  "scoreslnk": 1,
  "mcdetaillnk": 2
 }
}
```

Information about DoSSiER Dictionaries

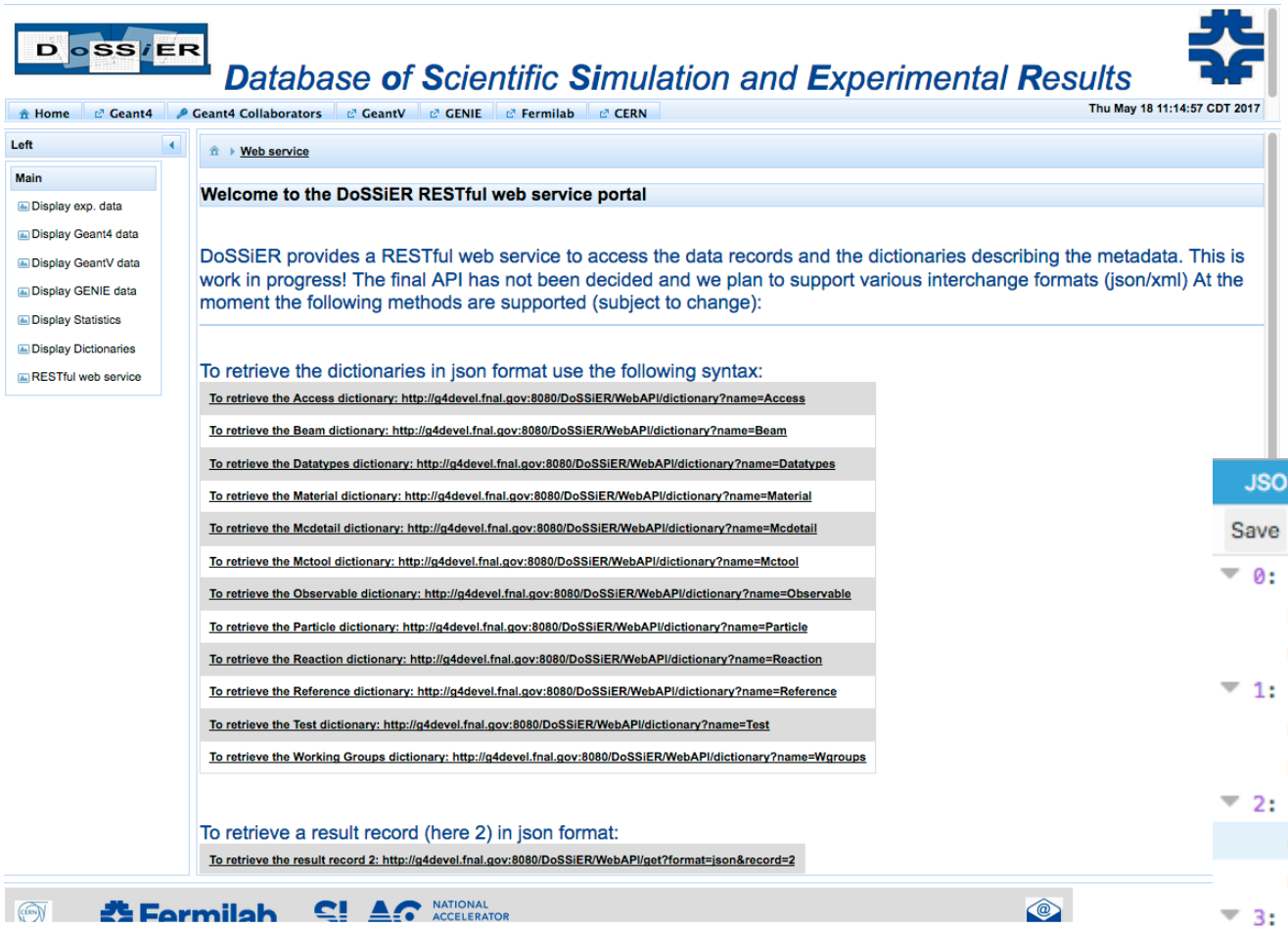
- DoSSiER uses “dictionaries” of different objects so that the metadata of results can be referenced by their ID keys or unique name, expandable, no duplication.



material

mid	SERIAL
z	INTEGER
mname	CHARACTER VARYING(50)
density	REAL
ion	REAL
ncomp	INTEGER
comp	INTEGER[]
frac	REAL[]

Retrieving dictionary data (reaction) via the web-service



Change to human readable json interchange format

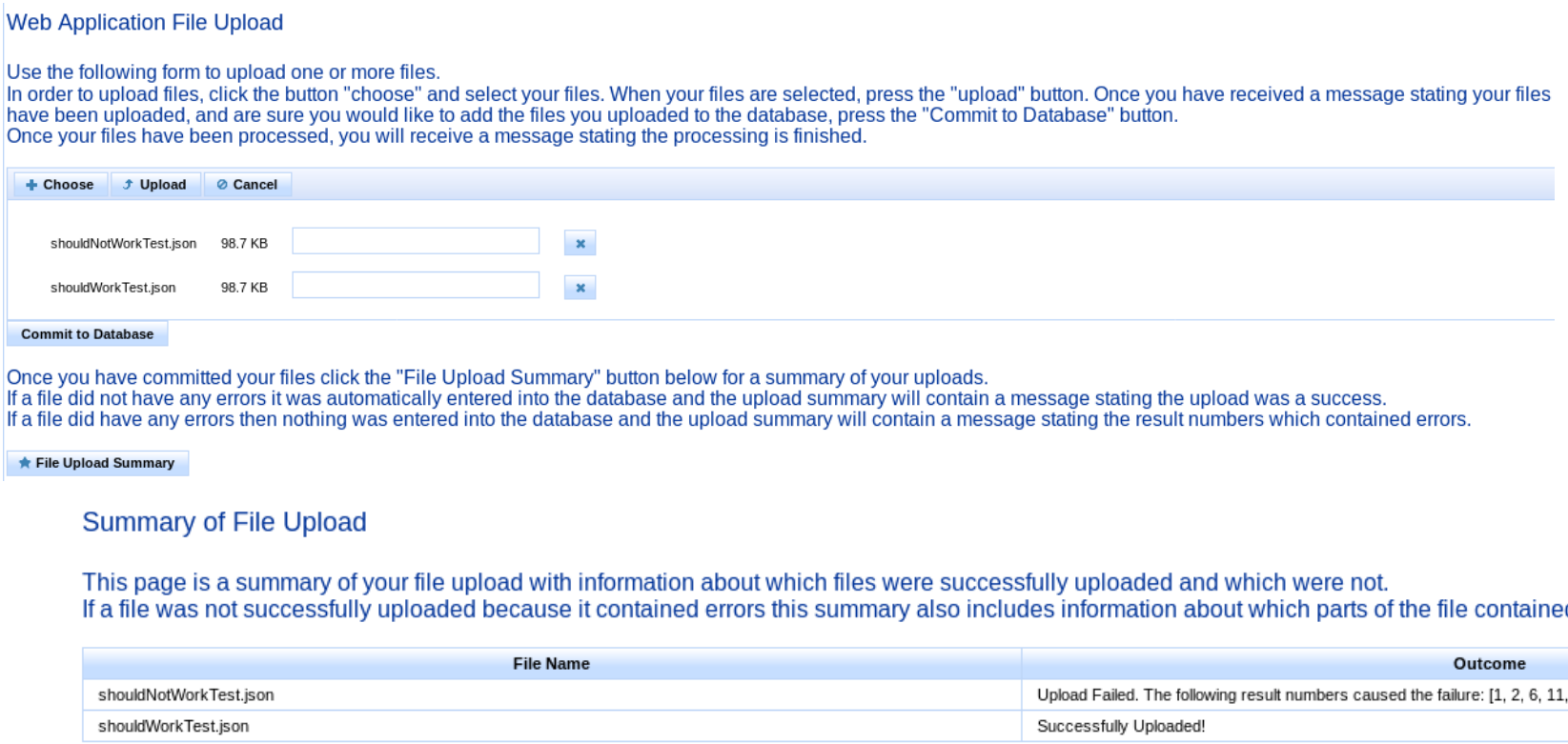

```
{
  "trid": 202,
  "testlnk": 0,
  "referencelnk": 4,
  "mcdetaillnk": 1,
  "beamlnk": 9,
  "targetlnk": 4,
  "observablelnk": 10,
  "secondarylnk": 0,
  "reactionlnk": 4,
  "datatable": {
    "dtid": 202,
    "datatypeslnk": 1000,
    "title": "Pion-Nucleus Total Cross-Sections from 80-MeV to 860-MeV",
    "npoints": 5,
    "nbins": [],
    "axisTitle": [
      "kinetic Energy T [MeV]",
      "crosssection [mb]"
    ],
    "val": [
      107.0,
      142.0,
      186.0,
      221.0,
      260.0,
      454.0,
      565.0,
      554.0,
      495.0,
      422.0
    ],
    "errStatPlus": [
      0.0,
      0.0,
      0.0,
      0.0,
      0.0,
      4.0,
      4.0,
      4.0,
      4.0,
      2.0
    ],
    "errStatMinus": [
      0.0,
      0.0,
      0.0,
      0.0,
      0.0,
      4.0,
      4.0,
      4.0,
      4.0,
      2.0
    ],
    "errSysPlus": [],
    "errSysMinus": [],
    "binMin": [],
    "binMax": []
  },
  "imageblobslnk": 0,
  "scoreslnk": 2,
  "accesslnk": 1,
  "parnames": [],
  "parvalues": [],
  "modtime": "Aug 2, 2016 1:52:36 PM"
}
```



```
{
  "referencekw": "A comparison of pi+ and pi- total cross-sections of light nuclei near the 3-3 resonance",
  "modelkw": "NA",
  "mctoolkw": "Experiment",
  "versiontagkw": "NA",
  "beamkw": "Nimrod Rutherford/sec.  (pi+)",
  "targetkw": "Be",
  "observablekw": "total  cross section",
  "secondarykw": "geantino",
  "reactionkw": "nuclear interaction",
  "scoreskw": "NA",
  "accesskw": "public",
  "datatable": {
    "title": "Pion-Nucleus Total Cross-Sections from 80-MeV to 860-MeV",
    "npoints": 5,
    "nbins": [],
    "axisTitle": [
      "kinetic Energy T [MeV]",
      "crosssection [mb]"
    ],
    "val": [
      107.0,
      142.0,
      186.0,
      221.0,
      260.0,
      454.0,
      565.0,
      554.0,
      495.0,
      422.0
    ],
    "errStatPlus": [
      0.0,
      0.0,
      0.0,
      0.0,
      0.0,
      4.0,
      4.0,
      4.0,
      4.0,
      2.0
    ],
    "errStatMinus": [
      0.0,
      0.0,
      0.0,
      0.0,
      0.0,
      4.0,
      4.0,
      4.0,
      4.0,
      2.0
    ],
    "errSysPlus": [],
    "errSysMinus": [],
    "binMin": [],
    "binMax": [],
    "datatypeskw": "1D Datapoint set"
  },
  "imageblobslnk": 0,
  "parnames": [],
  "parvalues": [],
  "modtime": "Aug 2, 2016 1:52:36 PM"
}
```


- Human readable
- Less error prone
- Makes current exchange format for simplified calo obsolete
- Deployed on Development sever

Images of the Web Page File Upload



Summary of File Upload

This page is a summary of your file upload with information about which files were successfully uploaded and which were not.
If a file was not successfully uploaded because it contained errors this summary also includes information about which parts of the file contained the errors.

File Name	Outcome
shouldNotWorkTest.json	Upload Failed. The following result numbers caused the failure: [1, 2, 6, 11, 40]
shouldWorkTest.json	Successfully Uploaded!

DoSSiER Summary and Conclusions

- DoSSiER: **D**atabase **o**f **S**cientific **Si**mulation and **E**xperimental **R**esults is actively being developed with participation by Geant4 groups at Fermilab and SLAC. https://g4devel.fnal.gov:8181/ (https protocol; for end to end encryption)
- We already identified features in Geant4 that needed fixing.
- Experimental data and results from simulation are stored in a relational database.
- Data can be imported and exported using json/xml formats. (python scripts are provided to extract data from root files or ASCII tables and convert to json/xml.)
- Web application:
  - allows to select and search.
  - allows to overlay experimental and simulated data.
  - authentication is necessary to have access to internal data and functions (e.g. upload, edit, delete).
- Web service: allows to programmatically access the repository (read only).

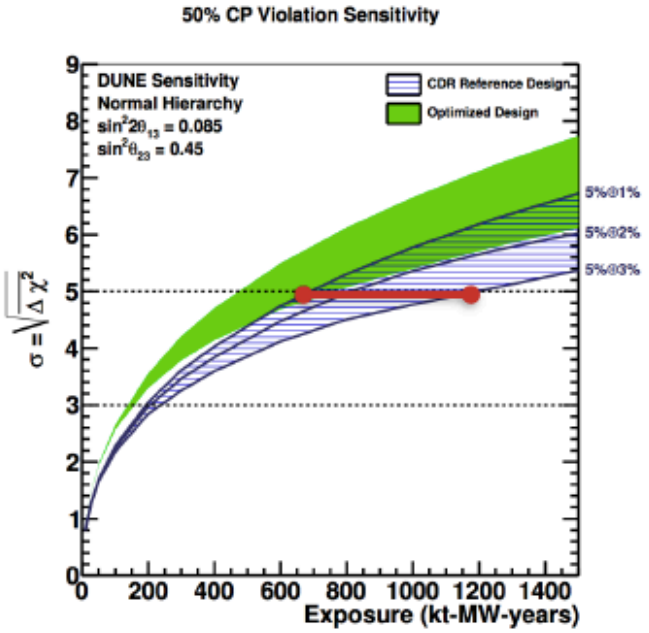
DoSSiER Summary and Conclusions

- Finding good experimental data and adding to DoSSiER is a continuing effort.
- So is developing new test to validate the Simulation vs. this data.
- We need more Geant 4 collaborators to upload their data, pretty easy with he ancillary python tools developed by Andrea, new uploads require only minor changes to the meta data→ will provide tutorial! Would like to have all geant4 tests in DoSSiER.
- Change to ‘human readable json format well under way.

- backup

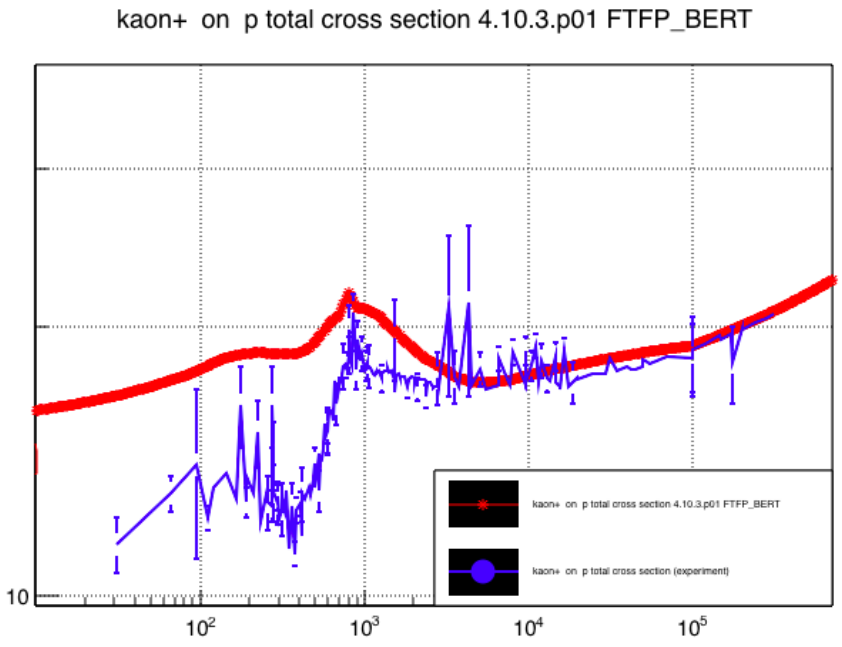
Motivation: e.g. Neutrino physics

- For DUNE, difference between 3% vs 1% relative signal normalization uncertainty **equivalent to nearly doubling exposure** time for some figures of merit
- We will **need unprecedented precision in models** of beams, physics, and detectors



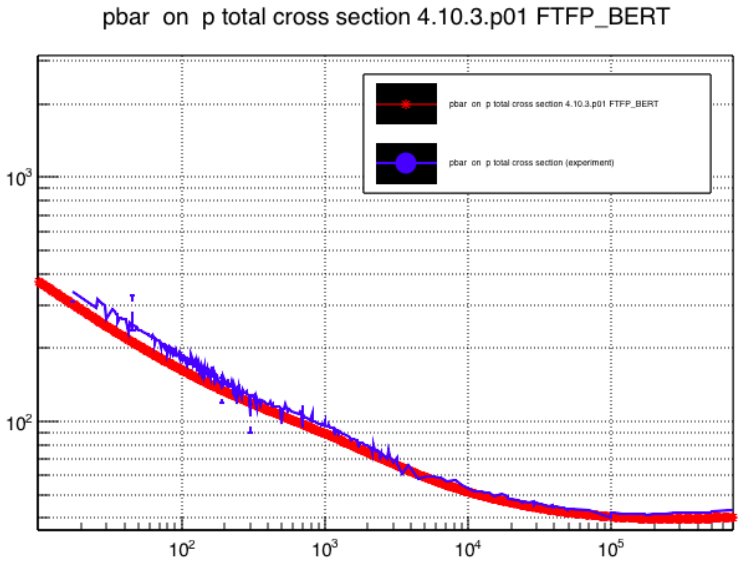
DUNE's physics reach will strongly depend on how low we are able to push systematic uncertainties, many of which will come from Detector/Beam modeling

Sometimes the result is surprising



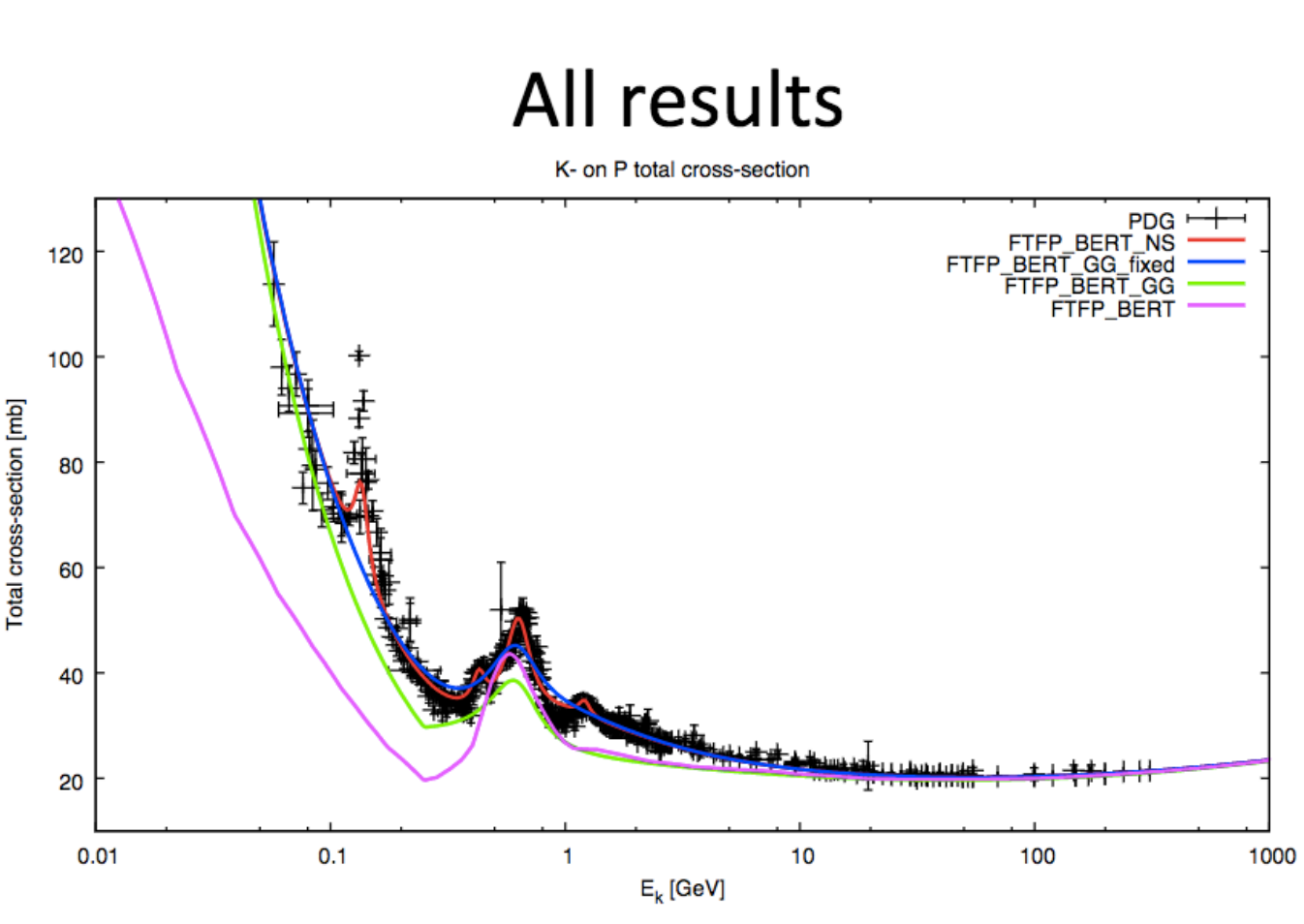
Kaon cross sections in reference physics list uses old Geisha tables→

Bug not caught because results were not compared to data and test used different method to access physics list

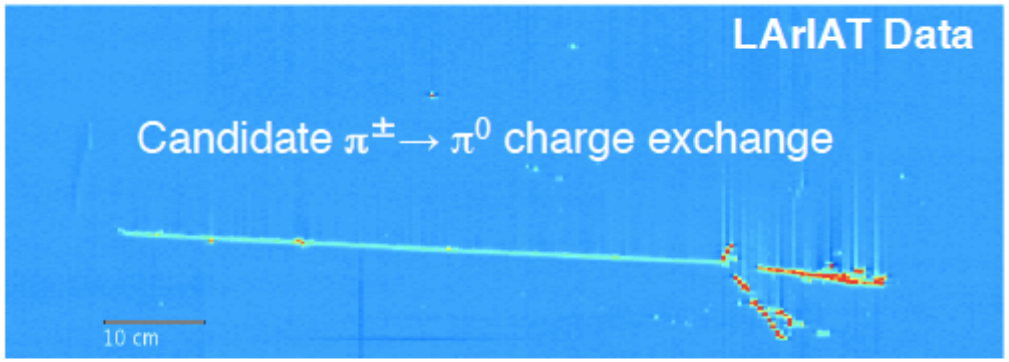


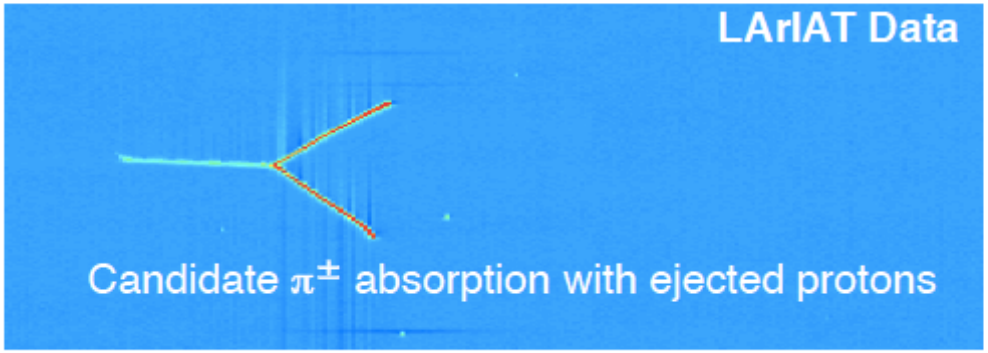
From Witek's presentation at Geant 4 hadronics meeting

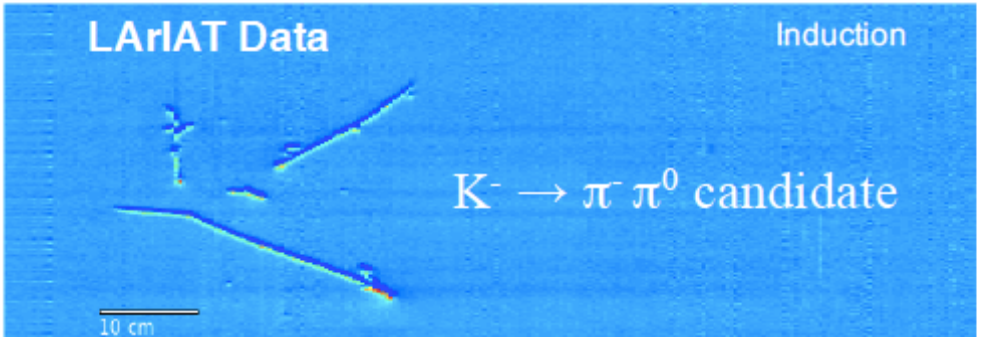
Good argument for continuously adding data and perform new tests. Sometimes new introduced bugs can worsen results that were previously validated or validated with different test. Sometimes the same thing is not really the same thing.



Detector Simulation





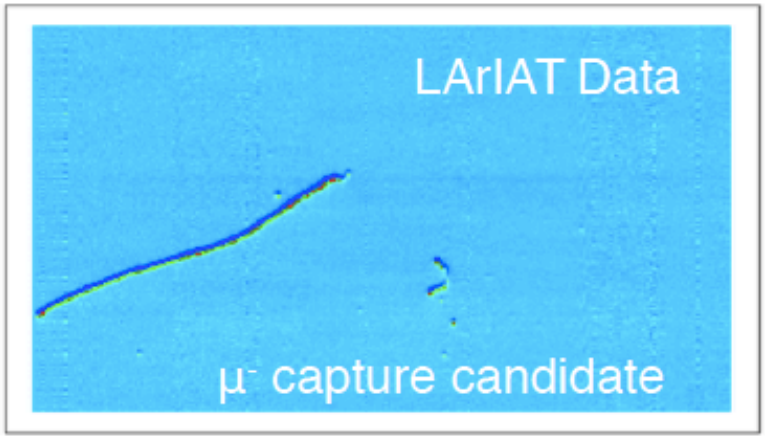


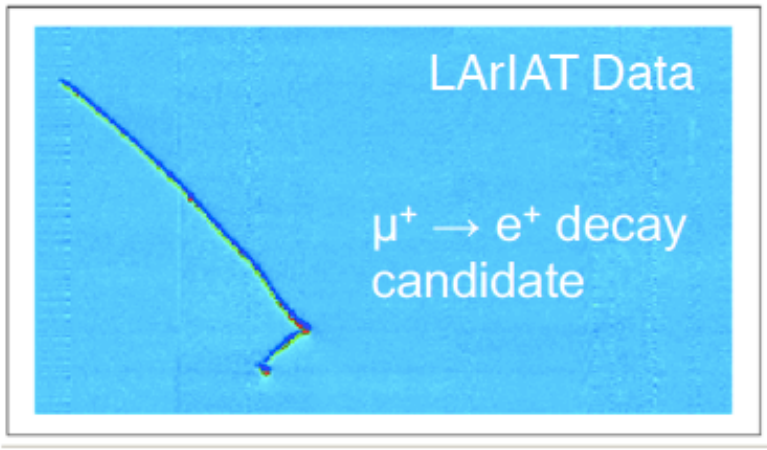
Modeling of liquid Argon Detectors is critical, as many new LAr detectors come online.

Important processes for oscillation physics: inelastic interactions/ response of < few GeV pions, protons, neutrons, photons, electrons

Kaons also important for nucleon decay analyses

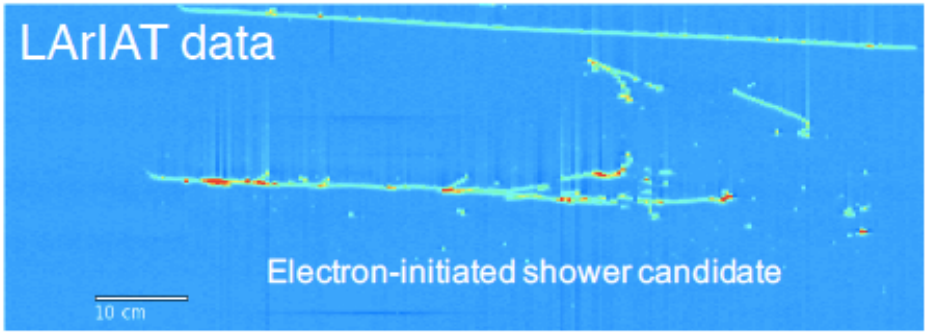
Detector Simulation

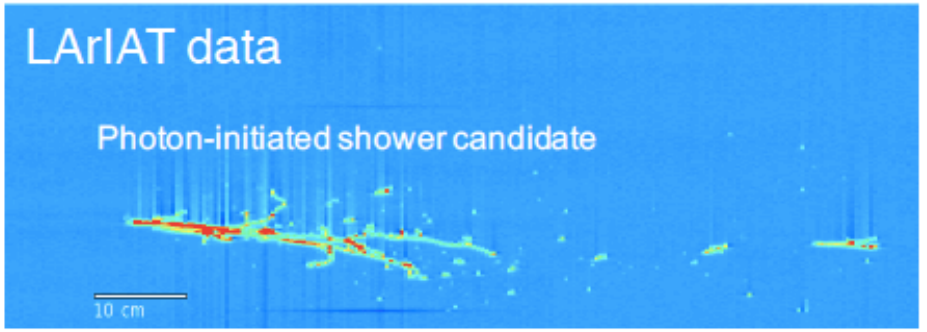




- Sign selection in detectors without a magnetic field is important
- Allows separation of neutrinos and antineutrinos (needed because observation of differences between neutrino and antineutrinos is a central physics goal)
- μ+ only decay, with e+ emission of known energy spectrum
- μ- capture on nuclei followed by γ/n emission (76%) or decay (24%)
- Capture rate higher in Argon than in lighter elements

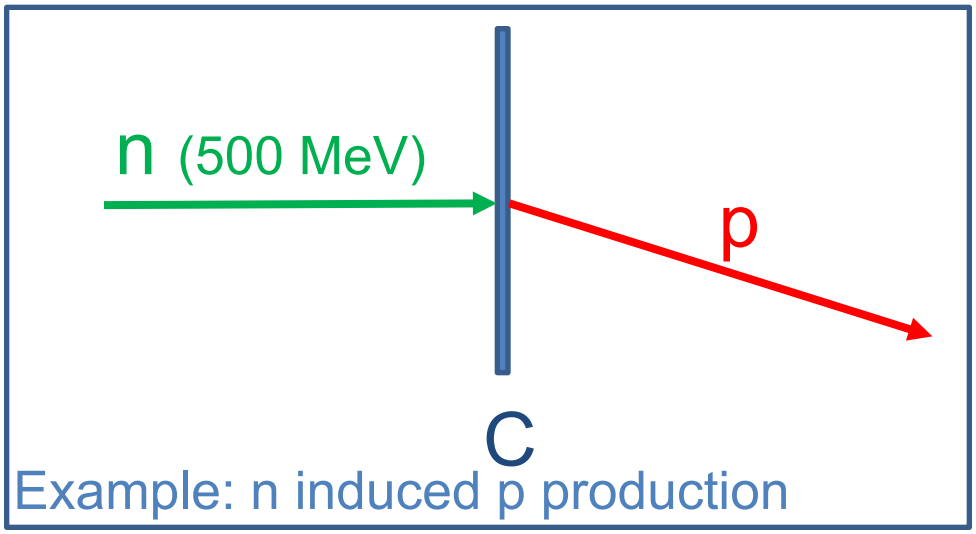
Detector Simulation





- e/γ separation also critical for oscillation measurements
- Separates electron neutrino appearance from backgrounds such as Neutral Current π^0 production

Meta data to classify a result



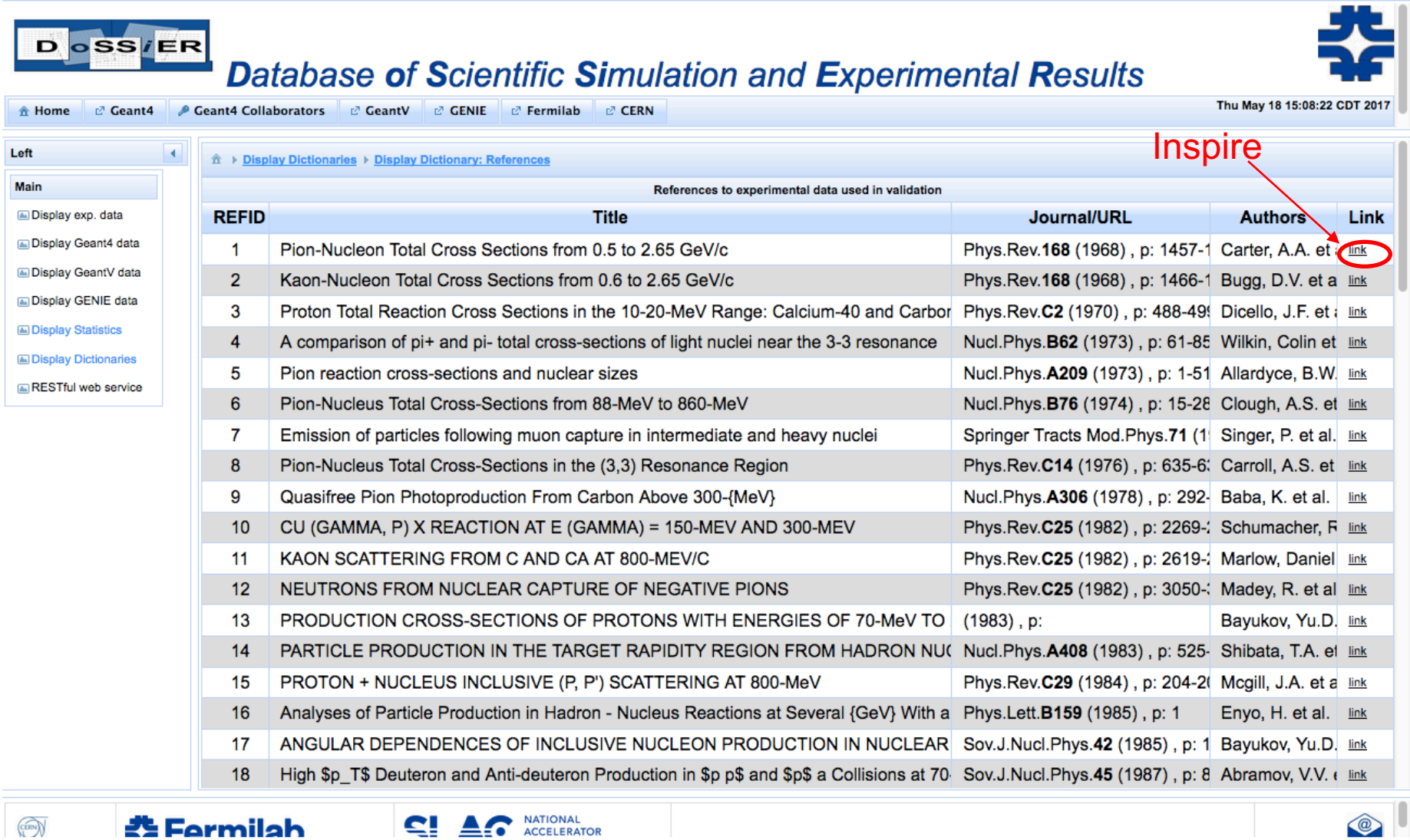
Example: n induced p production

Beam: mono energetic neutrons
Target: Carbon.
Secondary: protons.
Observable: differential cross section
Reaction: particle production
Model: e.g. BIC
Version: geant4.10.1.p02
In addition parnames/parvalue pairs can be added

Note:

- Values for metadata stored in dictionaries:
  (Beams, Materials, Particles, Observable, Reaction, Model, Version…).
- Meta data used to match experimental and simulated results.
- Complicated Beams (e.g. neutrino flux files, test beams consisting of many particles can be described by the schema).
- Ditto for Materials .
- Dictionaries can evolve as needed.

Experimental data



DoSSiER

Database of Scientific Simulation and Experimental Results

Home | Geant4 | Geant4 Collaborators | GeantV | GENIE | Fermilab | CERN

Thu May 18 15:08:22 CDT 2017

Left

Main

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

Display Dictionaries ▸ Display Dictionary: References

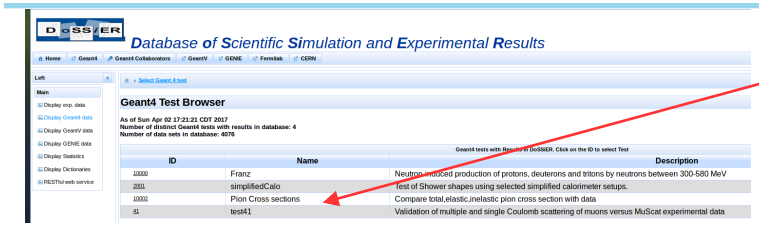
References to experimental data used in validation

REFID	Title	Journal/URL	Authors	Link
1	Pion-Nucleon Total Cross Sections from 0.5 to 2.65 GeV/c	Phys.Rev.**168** (1968) , p: 1457-1	Carter, A.A. et	link
2	Kaon-Nucleon Total Cross Sections from 0.6 to 2.65 GeV/c	Phys.Rev.**168** (1968) , p: 1466-1	Bugg, D.V. et a	link
3	Proton Total Reaction Cross Sections in the 10-20-MeV Range: Calcium-40 and Carbor	Phys.Rev.**C2** (1970) , p: 488-49	Dicello, J.F. et	link
4	A comparison of pi+ and pi- total cross-sections of light nuclei near the 3-3 resonance	Nucl.Phys.**B62** (1973) , p: 61-85	Wilkin, Colin et	link
5	Pion reaction cross-sections and nuclear sizes	Nucl.Phys.**A209** (1973) , p: 1-51	Allardyce, B.W.	link
6	Pion-Nucleus Total Cross-Sections from 88-MeV to 860-MeV	Nucl.Phys.**B76** (1974) , p: 15-28	Clough, A.S. et	link
7	Emission of particles following muon capture in intermediate and heavy nuclei	Springer Tracts Mod.Phys.**71** (1	Singer, P. et al.	link
8	Pion-Nucleus Total Cross-Sections in the (3,3) Resonance Region	Phys.Rev.**C14** (1976) , p: 635-63	Carroll, A.S. et	link
9	Quasifree Pion Photoproduction From Carbon Above 300-{MeV}	Nucl.Phys.**A306** (1978) , p: 292-	Baba, K. et al.	link
10	CU (GAMMA, P) X REACTION AT E (GAMMA) = 150-MEV AND 300-MEV	Phys.Rev.**C25** (1982) , p: 2269-2	Schumacher, R	link
11	KAON SCATTERING FROM C AND CA AT 800-MEV/C	Phys.Rev.**C25** (1982) , p: 2619-2	Marlow, Daniel	link
12	NEUTRONS FROM NUCLEAR CAPTURE OF NEGATIVE PIONS	Phys.Rev.**C25** (1982) , p: 3050-3	Madey, R. et al	link
13	PRODUCTION CROSS-SECTIONS OF PROTONS WITH ENERGIES OF 70-MeV TO	(1983) , p:	Bayukov, Yu.D.	link
14	PARTICLE PRODUCTION IN THE TARGET RAPIDITY REGION FROM HADRON NUC	Nucl.Phys.**A408** (1983) , p: 525-	Shibata, T.A. et	link
15	PROTON + NUCLEUS INCLUSIVE (P, P') SCATTERING AT 800-MeV	Phys.Rev.**C29** (1984) , p: 204-20	Mcgill, J.A. et a	link
16	Analyses of Particle Production in Hadron - Nucleus Reactions at Several {GeV} With a	Phys.Lett.**B159** (1985) , p: 1	Enyo, H. et al.	link
17	ANGULAR DEPENDENCES OF INCLUSIVE NUCLEON PRODUCTION IN NUCLEAR	Sov.J.Nucl.Phys.**42** (1985) , p: 1	Bayukov, Yu.D.	link
18	High p_T Deuteron and Anti-deuteron Production in $p p$ and p a Collisions at 70	Sov.J.Nucl.Phys.**45** (1987) , p: 8	Abramov, V.V.	link

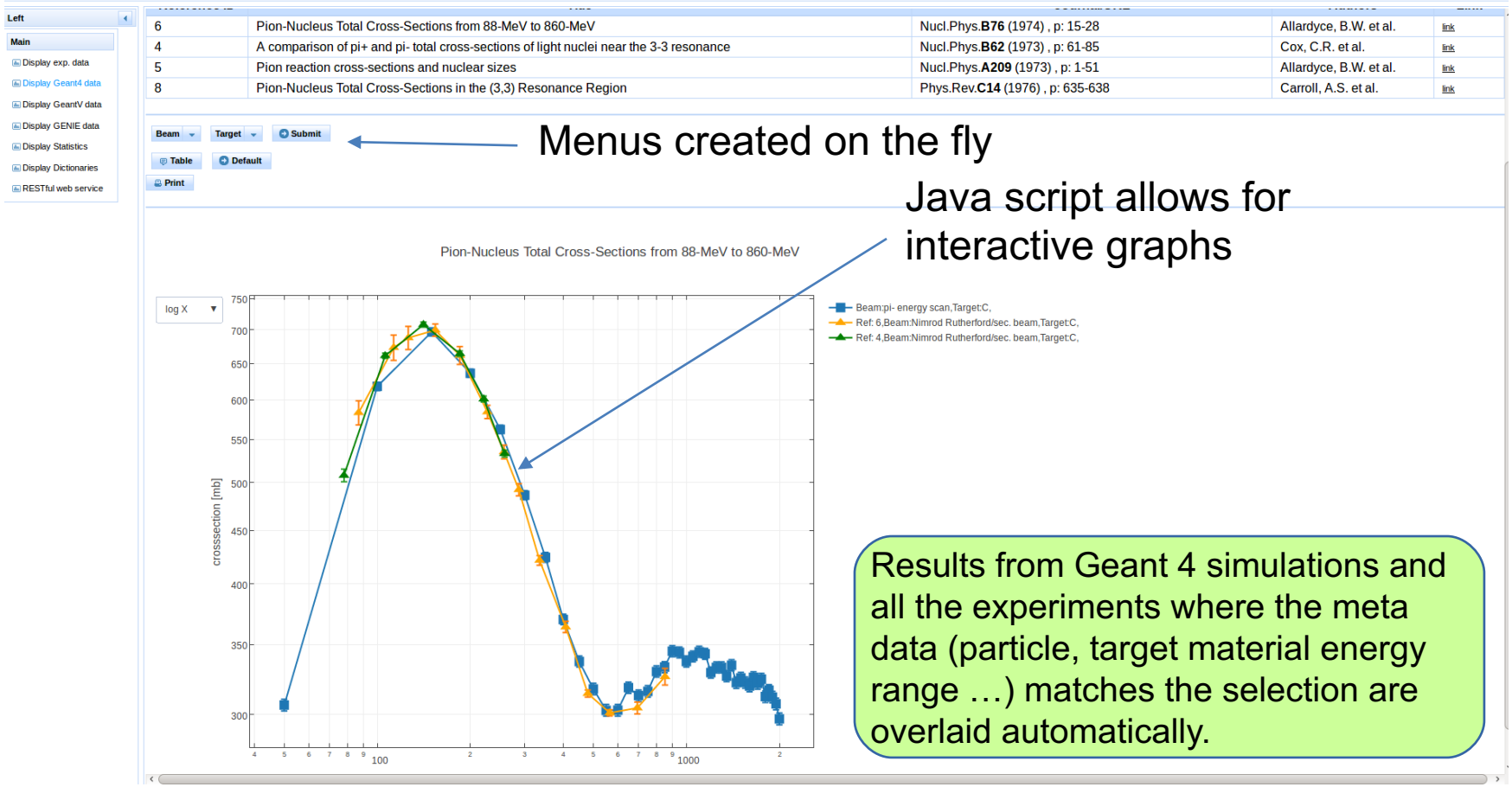
Inspire

Fermilab | NATIONAL ACCELERATOR

Overlay: simulation and matching experimental data



Select Pion Cross Sections in Geant 4 test browser

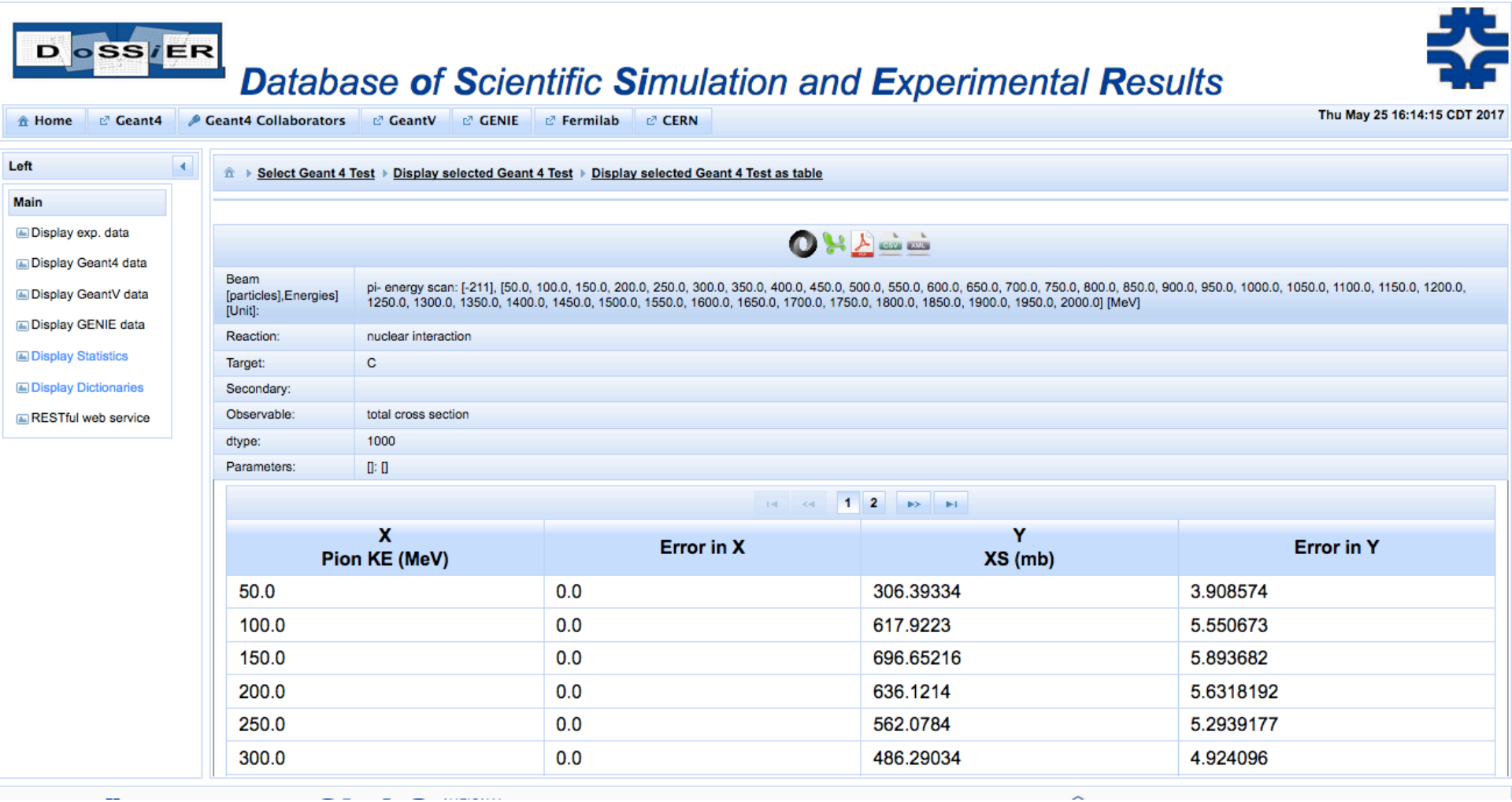


Menus created on the fly

Java script allows for interactive graphs

Results from Geant 4 simulations and all the experiments where the meta data (particle, target material energy range …) matches the selection are overlaid automatically.

Display as tables



Database of Scientific Simulation and Experimental Results

Home | Geant4 | Geant4 Collaborators | GeantV | GENIE | Fermilab | CERN

Thu May 25 16:14:15 CDT 2017

Left

Main

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

Select Geant 4 Test ▸ Display selected Geant 4 Test ▸ Display selected Geant 4 Test as table

Beam [particles],Energies] [Unit]:	pi- energy scan: [-211], [50.0, 100.0, 150.0, 200.0, 250.0, 300.0, 350.0, 400.0, 450.0, 500.0, 550.0, 600.0, 650.0, 700.0, 750.0, 800.0, 850.0, 900.0, 950.0, 1000.0, 1050.0, 1100.0, 1150.0, 1200.0, 1250.0, 1300.0, 1350.0, 1400.0, 1450.0, 1500.0, 1550.0, 1600.0, 1650.0, 1700.0, 1750.0, 1800.0, 1850.0, 1900.0, 1950.0, 2000.0] [MeV]
Reaction:	nuclear interaction
Target:	C
Secondary:	
Observable:	total cross section
dtype:	1000
Parameters:	[]: []

1 2

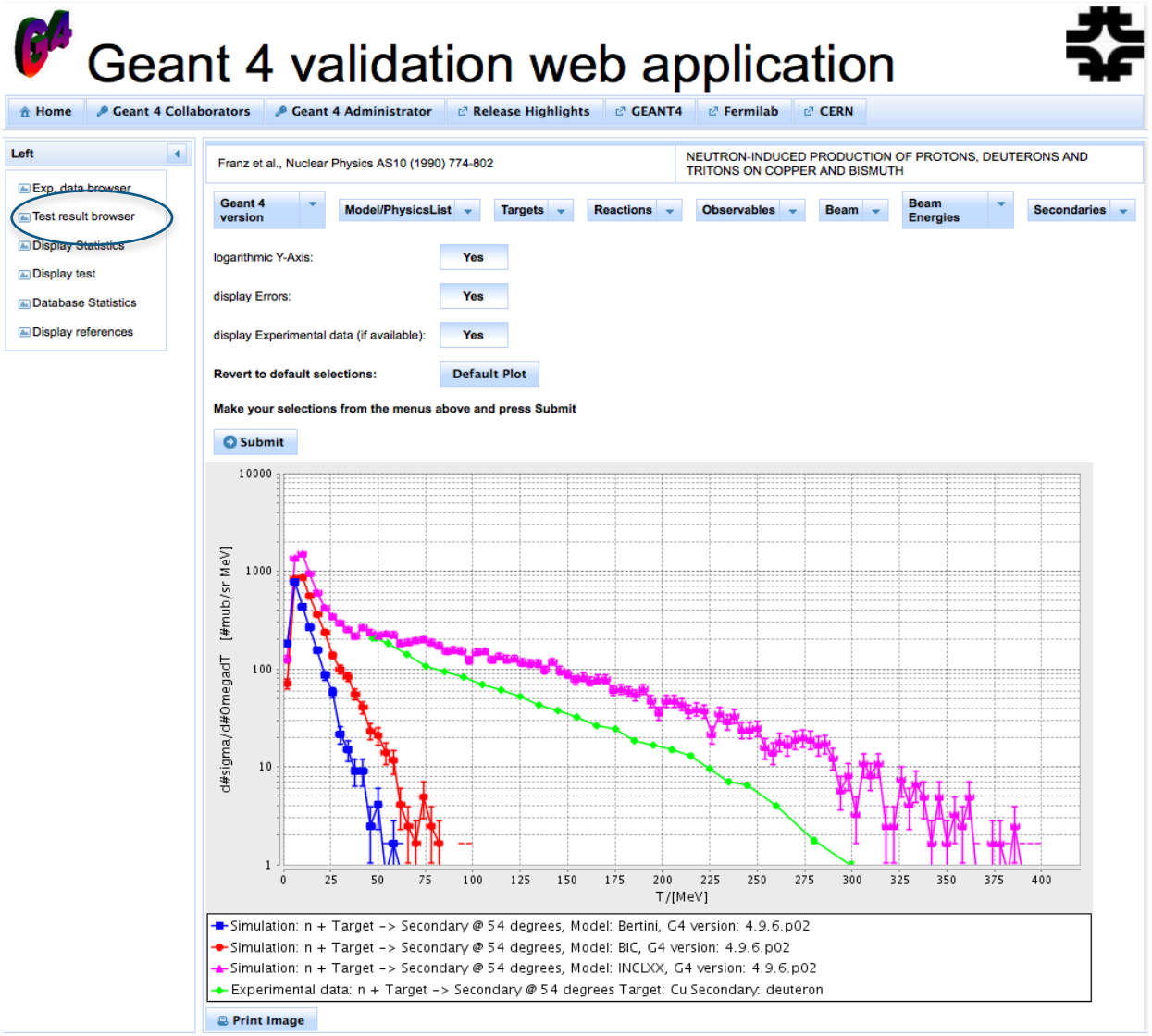
X Pion KE (MeV)	Error in X	Y XS (mb)	Error in Y
50.0	0.0	306.39334	3.908574
100.0	0.0	617.9223	5.550673
150.0	0.0	696.65216	5.893682
200.0	0.0	636.1214	5.6318192
250.0	0.0	562.0784	5.2939177
300.0	0.0	486.29034	4.924096

Test Result Browser

This allows to select Geant4 simulation results of interest, and to compare them to the experimental data as applicable. Shown on the right is neutron induced deuteron (default selection).

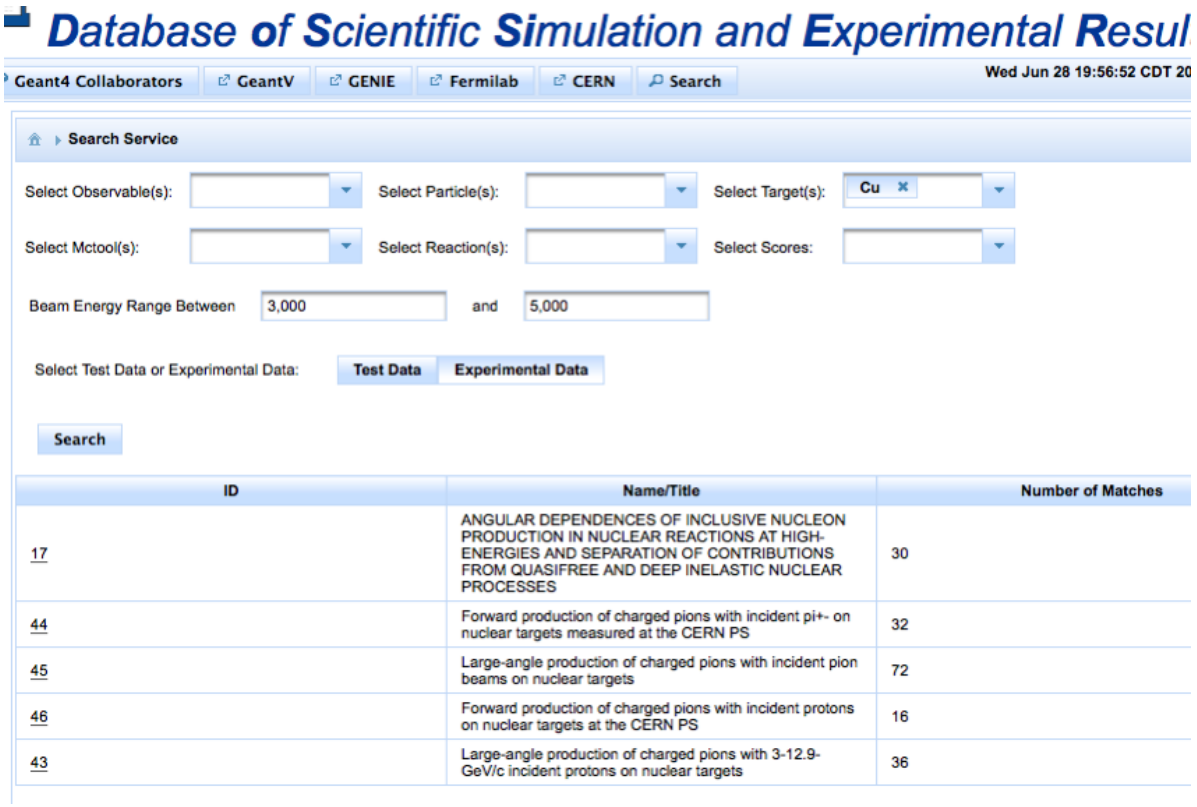
Different Models:

- BIC(blue),
- Bertini (red)
- INCL++(magenta)
- Experimental Data (green)



Web Page Search (Nicole Serpico)

- Options are selected from searchable drop down menus
- Beam energy is inputted or just left blank
- Test or experimental data is chosen
- The results include the ID of the matching experiment or test, which links to the display of the corresponding data



Web Page File Upload (Nicole Serpico)

- Accessed through the G4Expert secure web page
- Multiple file upload that can handle hundreds of small files (~100 results) as well as large files (~50,000 results or ~0.5 GB).
- Goes through each uploaded file twice: first checks to make sure it doesn't have any formatting errors, then after doing that, uploads the file to the database.
- Still need to elevate errors that happen when adding to the database.
- Links to a summary page that details which files were successfully uploaded and which had errors. If a file contains errors the summary page shows where those errors are so they can be easily corrected.

The Web service

- **Based on: Java API for REST - ful (**Representational State Transfer) **Services (JAX-RS)**
- Deployed on the development server: https://g4devel.fnal.gov:8181/
- Allows to programmatically retrieve results in json or xml format (with dictionaries expanded or not)→ these are the same formats used for uploads!
- C++ clients already used by selected Geant4 validation jobs!
- Dictionaries are retrieved by python upload tool to create look up tables so that dictionary can be accessed by keyword (not number)→ change to human readable json interchange format.
- Over the summer Niccole added:
  - Search functions like in INSPIRE/SPIRES.
  - Programmatic upload to database works but not enabled until we can secure the application.
  - Provide examples in python and java to deal with https.
  - Improved multiple file upload web application.