
Sumatra Documentation

Release 0.5.3

Andrew P. Davison

July 03, 2015

1	Table of Contents	3
1.1	Background	3
1.2	Installation	5
1.3	Getting started	6
1.4	smt command reference	9
1.5	Using the web interface	14
1.6	Graphical user interfaces	20
1.7	Parallel computations	20
1.8	Reproducible publications: including and linking to provenance information in documents	20
1.9	Using the Sumatra API within your own scripts	22
1.10	Parameter files	24
1.11	Upgrading your projects	26
1.12	Migrating records between record stores	27
1.13	Developers' guide	28
1.14	API reference	30
1.15	Frequently asked questions	43
1.16	Getting support	44
1.17	Release notes	44
1.18	Authors and contributors	47
	Python Module Index	49

Sumatra is a tool for managing and tracking projects based on numerical simulation and/or analysis, with the aim of supporting reproducible research. It can be thought of as an automated electronic lab notebook for computational projects.

It consists of:

- a command-line interface, *smt*, for launching simulations/analyses with automatic recording of information about the experiment, annotating these records, linking to data files, etc.
- a web interface with a built-in web-server, *smtweb*, for browsing and annotating simulation/analysis results.
- a LaTeX package and Sphinx extension for including Sumatra-tracked figures and links to provenance information in papers and other documents.
- a Python API, on which *smt* and *smtweb* are based, that can be used in your own scripts in place of using *smt*, or could be integrated into a GUI-based application.

Table of Contents

1.1 Background

1.1.1 Reproducibility, provenance and project management

Reproducibility of experiments is one of the foundation stones of science. A related concept is provenance, being able to track a given scientific result, such as a figure in an article, back through all the analysis steps (verifying the correctness of each) to the original raw data, and the experimental protocol used to obtain it.

In computational, simulation- or numerical analysis-based science, reproduction of previous experiments, and establishing the provenance of results, ought to be easy, given that computers are deterministic, not suffering from the problems of inter-subject and trial-to-trial variability that make reproduction of biological experiments more challenging.

In general, however, it is not easy, perhaps due to the complexity of our code and our computing environments, and the difficulty of capturing every essential piece of information needed to reproduce a computational experiment using existing tools such as spreadsheets, version control systems and paper notebooks.

1.1.2 What needs to be recorded?

To ensure reproducibility of a computational experiment we need to record:

- the code that was run
- any parameter files and command line options
- the platform on which the code was run

For an individual researcher trying to keep track of a research project with many hundreds or thousands of simulations and/or analyses, it is also useful to record the following:

- the reason for which the simulation/analysis was run
- a summary of the outcome of the simulation/analysis

Recording the code might mean storing a copy of the executable, or the source code (including that of any libraries used), the compiler used (including version) and the compilation procedure (e.g. the Makefile, etc.) For interpreted code, it might mean recording the version of the interpreter (and any options used in compiling it) as well as storing a copy of the main script, and of any external modules or packages that are included or imported into the script file.

For projects using version control, “storing a copy of the code” may be replaced with “recording the URL of the repository and the revision number”.

The platform includes the processor architecture(s), the operating system(s), the number of processors (for distributed simulations), etc.

1.1.3 Tools for recording provenance information

The traditional way of recording the information necessary to reproduce an experiment is by noting down all details in a paper notebook, together with copies or print-outs of any results. More modern approaches may replace or augment the paper notebook with a spreadsheet or other hand-rolled database, but still with the feature that all relevant information is entered by hand.

In other areas of science, particularly in applied science laboratories with high-throughput, highly-standardised procedures, electronic lab notebooks and laboratory information management systems (LIMS) are in widespread use, but none of these tools seem to be well suited for tracking simulation experiments.

1.1.4 Challenges for tracking computational experiments

In developing a tool for tracking simulation experiments, something like an electronic lab notebook for computational science, there are a number of challenges:

- different researchers have very different ways of working and different workflows: command line, GUI, batch-jobs (e.g. in supercomputer environments), or any combination of these for different components (simulation, analysis, graphing, etc.) and phases of a project.
- some projects are essentially solo endeavours, others collaborative projects, possibly distributed geographically.
- as much as possible should be recorded automatically. If it is left to the researcher to record critical details there is a risk that some details will be missed or left out, particularly under pressure of deadlines.

The solution we propose is to develop a core library, implemented as a Python package, `sumatra`, and then to develop a series of interfaces that build on top of this: a command-line interface, a web interface, a graphical interface. Each of these interfaces will enable:

- launching simulations/analyses with automated recording of provenance information;
- managing a computational project: browsing, viewing, deleting simulations/analyses.

Alternatively, modellers can use the `sumatra` package directly in their own code, to enable provenance recording, then simply launch experiments in their usual way.

The core `sumatra` package needs to:

- interact with version control systems, such as [Subversion](#), [Git](#), [Mercurial](#), or [Bazaar](#);
- support launching serial, distributed (via [MPI](#)) or batch computations;
- link to data generated by the computation, whether stored in files or databases;
- support all and any command-line drivable simulation or analysis programs;
- support both local and networked storage of information;
- be extensible, so that components can easily be added for new version control systems, etc.
- be very easy to use, otherwise it will only be used by the very conscientious.

1.1.5 Further resources

For further background, see the following article:

Davison A.P. (2012) Automated capture of experiment context for easier reproducibility in computational research. *Computing in Science and Engineering* **14**: 48-56 [[preprint](#)]

You may also be interested in exploring this interactive poster about Sumatra:



or in watching a talk given at a [workshop](#) on “*Reproducible Research-Tools and Strategies for Scientific Computing*” in Vancouver, Canada in July 2011. [[video with slides](#) (Silverlight required)] [[video only](#) (YouTube)] [[slides](#)].

1.2 Installation

To run Sumatra you will need Python installed on your machine. If you are running Linux or OS X, you almost certainly already have it. If you don't have Python, you can install it from [python.org](#), or install one of the “value-added” distributions aimed at scientific users of Python: [Enthought](#) or [Python\(x,y\)](#).

The easiest way to install Sumatra is directly from the [Python Package Index \(PyPI\)](#):

```
$ pip install sumatra
```

or:

```
$ easy_install sumatra
```

Alternatively, you can download the Sumatra package from either PyPI or the [INCF Software Centre](#) and install it as follows:

```
$ tar xzf Sumatra-0.5.3.tar.gz
$ cd Sumatra-0.5.3
# python setup.py install
```

The last step may need to be run as root, or using `sudo`.

1.2.1 Installing Django

If you wish to use the web interface, you will also need to install [Django](#). On Linux, you may be able to do this via your package management system: see <https://code.djangoproject.com/wiki/Distributions>.

Otherwise, it is very easy to install manually: see <https://docs.djangoproject.com/en/dev/topics/install/#installing-official-release>

You will also need to install the [django-tagging](#) and [docutils](#) packages, which may be in your package management system, otherwise they can be installed from PyPI:

```
$ pip install django-tagging
$ pip install docutils
```

1.2.2 Installing Python bindings for your version control system

Sumatra currently supports [Mercurial](#), [Subversion](#), [Git](#) and [Bazaar](#). If you are using Subversion, you will need to install the [pysvn bindings](#). Since Mercurial and Bazaar are mostly written in Python, just installing the main Mercurial/Bazaar packages is sufficient. For Git, you need to install the [GitPython](#) package.

1.3 Getting started

Let us assume that you already have a project based on numerical simulation, which you wish to start managing using Sumatra, and that the code for this project is under version control. Note that the following is equally valid if your project is based on data analysis rather than, or as well as, simulation: just mentally replace “simulation” with “analysis” in the following.

Change to the working directory for your project, and then create a new Sumatra project in this directory using the `smt init` command:

```
$ cd myproject
$ smt init MyProject
```

where `MyProject` is the project name. This creates a sub-directory named `.smt`.

Sumatra tracks data files created by your simulation by searching for newly created files within a given directory tree. By default, it assumes that your simulation will create files in a sub-directory `Data` of your working directory. (You can change this by providing the `--datapath` option to `smt init` or `smt configure`.)

Now let’s run a simulation. We will assume that your simulation code is written in [Python](#), and that you run the simulation by executing a file called `main.py`, passing it the name of a parameter file on the command line, i.e., you would normally run a simulation using:

```
$ python main.py default.param
```

To run it using Sumatra, you would use:

```
$ smt run --executable=python --main=main.py default.param
```

Now we can see a list of the simulations we have run:

```
$ smt list
20121017-114820
```

This shows the label for each simulation we have run. Since we did not specify a label, one was automatically generated from the timestamp. To see more detail, use the `--long` option:

```
$ smt list --long
-----
Label           : 20121017-114820
Timestamp      : 2012-10-17 11:48:20.421631
Reason         :
Outcome        :
Duration       : 0.119576931
Repository     : GitRepository at /path/to/myproject
Main_File      : main.py
Version        : 560a4afae1565799a29ca259b6a400aa389e59dd
Script_Arguments : <parameters>
Executable     : Python (version: 2.7.1) at /usr/local/bin/python
```

```
Parameters      : seed = 45245
                 : distr = "uniform"
                 : n = 100
                 : tau_m = 20.0
Input_Data      : []
Launch_Mode     : serial
Output_Data     : [output.dat (dcb86788c2c793804a04e683fae99ad0bac8fb99)]
Tags           :
```

(most options also have a short form, `-l` in this case.)

It is a bit tedious to have to tell Sumatra which simulator and which file to run every time. Presumably, the name of the main file changes infrequently and the simulator almost never. Therefore, these can be set as defaults for a given project:

```
$ smt configure --executable=python --main=main.py
```

(you could also have given these options to `smt init`. `init` is used to create a project and `configure` to change its configuration later, but they mostly accept the same arguments).

Now you can run a simulation with a much shorter command line:

```
$ smt run default.param
```

To see the current configuration of your project, use the `info` command:

```
$ smt info
Project name      : MyProject
Default executable : Python (version: 2.7.1) at /usr/local/bin/python
Default repository : GitRepository at /path/to/my/project
Default main file  : main.py
Default launch mode : serial
Data store (output) : ./Data
.                 (input) : /
Record store      : Relational database record store using the Django ORM (database file=/path/
Code change policy : error
Append label to   : None
Label generator   : timestamp
```

Sumatra automatically records the identity and versions of the simulation files and the simulator executable, stores links to any files created by the simulation, records any error messages, the date and time at which the simulation was run, and its duration. You may also add your own annotations, in several different ways. On running the simulation, you can specify a unique label, and the reason for which you are running the simulation:

```
$ smt run --label=hagglng --reason="determine whether the gourd is worth 3 or 4 shekels" romans.
```

After the simulation is complete, you can add a description of the outcome:

```
$ smt comment "apparently, it is worth NaN shekels."
```

This adds the comment to the most recent simulation. You may also describe the outcome of an earlier simulation, by specifying its label:

```
$ smt comment 20121017-114820 "Eureka! Nobel prize here we come."
```

You can also tag a simulation record with one or more short keywords:

```
$ smt tag foobar
$ smt tag barfoo
```

and remove tags:

```
$ smt tag --remove barfoo
```

The parameter file may be in any format - it is your script which is responsible for reading it. However, if it is in one of the [formats that Sumatra understands](#) then it is possible to modify parameter values on the command line. Suppose `default.param` contains a parameter `tau_m = 20.0`, as well as a number of other parameters, then:

```
$ smt run --reason="test effect of a smaller time constant" default.param tau_m=10.0
```

will generate a new parameter file identical to `default.param` but with `tau_m` equal to 10.0, and then will pass this new parameter file to your script. This can be very convenient when you wish to study the effects of changing one or two parameters, without having to edit your parameter file each time.

One of the main aims of Sumatra is to ensure the reproducibility of simulation results. The `repeat` command re-runs a previous simulation, and checks that the output is identical to that of the original run:

```
$ smt repeat haggling
The new record exactly matches the original.
```

Although it is better not to delete simulation records (so as to preserve a full record of the project, false starts and all), it is possible:

```
$ smt delete 20121017-123706
```

It is also possible to delete all simulations with a given tag:

```
$ smt delete --tag foobar
```

Most of the commands described here have further options that we have not described. A full description of the options for each command is given in the [command reference](#). The full list of commands is available by running `smt` by itself:

```
$ smt
Usage: smt <subcommand> [options] [args]

Simulation/analysis management tool version 0.4.0.dev

Available subcommands:
  init
  configure
  info
  run
  list
  delete
  comment
  tag
  repeat
  diff
  export
  upgrade
  sync
```

and help on a given command is available by running the command with the `--help` option, e.g.:

```
$ smt comment --help
Usage: smt comment [options] [LABEL] [COMMENT]

This command is used to describe the outcome of the simulation/analysis. If LABEL is omitted, the comment will be added to the most recent experiment (any existing comment will be overwritten). If the '-f/--file' option is set, COMMENT should be the name of a file containing the comment, otherwise it should be a string of text.

Options:
  -h, --help      show this help message and exit
  -r, --replace   if this flag is set, any existing comment will be
```

```

overwritten, otherwise, the new comment will be appended to
the end, starting on a new line
-f, --file    interpret COMMENT as the path to a file containing the
              comment

```

or `smt help CMD`, where `CMD` is the name of the command.

This tutorial has covered using `smt` for serial simulations/analyses. A further tutorial covers [using smt for parallel computations \(using MPI\)](#).

Also see [smtweb](#), which provides a more graphical interface to viewing lists of records than `smt list`.

1.4 smt command reference

1.4.1 comment

```
Usage: smt comment [options] [LABEL] [COMMENT]
```

This command is used to describe the outcome of the simulation/analysis. If `LABEL` is omitted, the comment will be added to the most recent experiment (any existing comment will be overwritten). If the `'-f/--file'` option is set, `COMMENT` should be the name of a file containing the comment, otherwise it should be a string of text.

Options:

```

-h, --help    show this help message and exit
-r, --replace if this flag is set, any existing comment will be
              overwritten, otherwise, the new comment will be appended to
              the end, starting on a new line
-f, --file    interpret COMMENT as the path to a file containing the
              comment

```

1.4.2 configure

```
Usage: smt configure [options]
```

Commands provided by the `smt` tool. Each command corresponds to a function in this module.

Options:

```

-h, --help          show this help message and exit
-d PATH, --datapath=PATH
                    set the path to the directory in which smt will search
                    for datafiles generated by the simulation or analysis.
                    Defaults to none
-i PATH, --input=PATH
                    set the path to the directory in which smt will search
                    for input datafiles.
-l OPTION, --adddlabel=OPTION
                    If this option is set, smt will append the record
                    label either to the command line (option 'cmdline') or
                    to the parameter file (option 'parameters'), and will
                    add the label to the datapath when searching for
                    datafiles. It is up to the user to make use of this
                    label inside their program to ensure files are created
                    in the appropriate location.
-e PATH, --executable=PATH
                    set the path to the executable.
-r REPOSITORY, --repository=REPOSITORY

```

```
the URL of a Subversion or Mercurial repository
containing the code. This will be checked out/cloned
into the current directory.
-m MAIN, --main=MAIN the name of the script that would be supplied on the
command line if running the simulator normally, e.g.
init.hoc.
-c ON_CHANGED, --on-changed=ON_CHANGED
may be 'store-diff' or 'error': the action to take if
the code in the repository or any of the dependencies
has changed. Defaults to 'error'
-A PATH, --archive=PATH
specify a directory in which to archive output
datafiles. If not specified, or if 'false', datafiles
are not archived.
-g OPTION, --labelgenerator=OPTION
specify which method Sumatra should use to generate
labels (options: timestamp, uuid)
-t TIMESTAMP_FORMAT, --timestamp_format=TIMESTAMP_FORMAT
the timestamp format given to strftime
```

1.4.3 delete

Usage: smt delete [options] LIST

LIST should be a space-separated list of labels for individual records or of tags. If it contains tags, you must set the --tag/-t option (see below). The special value "last" allows you to delete the most recent simulation/analysis. If you want to delete all records, just delete the .smt directory and use smt init to create a new, empty project.

Options:

```
-h, --help show this help message and exit
-t, --tag interpret LIST as containing tags. Records with any of these
tags will be deleted.
-d, --data also delete any data associated with the record(s).
```

1.4.4 diff

Usage: smt diff [options] LABEL1 LABEL2

Show the differences, if any, between two records.

Options:

```
-h, --help show this help message and exit
-i IGNORE, --ignore=IGNORE
a regular expression pattern for filenames to ignore
when evaluating differences in output data. To supply
multiple patterns, use the -i option multiple times.
-l, --long prints full information for each record
```

1.4.5 export

Usage: smt export

Export a Sumatra project and its records to JSON. This is needed before running upgrade.

Options:

-h, --help show this help message and exit

1.4.6 help

Usage: smt help [CMD]

Get help on an smt command.

Options:

-h, --help show this help message and exit

1.4.7 info

Usage: smt info

Print information about the current project.

Options:

-h, --help show this help message and exit

1.4.8 init

Usage: smt init [options] NAME

Create a new project called NAME in the current directory.

Options:

-h, --help show this help message and exit

-d PATH, --datapath=PATH set the path to the directory in which smt will search for output datafiles generated by the simulation/analysis. Defaults to ./Data

-i PATH, --input=PATH set the path to the directory relative to which input datafile paths will be given. Defaults to the filesystem root.

-l OPTION, --addlabel=OPTION If this option is set, smt will append the record label either to the command line (option 'cmdline') or to the parameter file (option 'parameters'), and will add the label to the datapath when searching for datafiles. It is up to the user to make use of this label inside their program to ensure files are created in the appropriate location.

-e PATH, --executable=PATH set the path to the executable. If this is not set, smt will try to infer the executable from the value of the --main option, if supplied, and will try to find the executable from the PATH environment variable, then by searching various likely locations on the filesystem.

-r REPOSITORY, --repository=REPOSITORY the URL of a Subversion or Mercurial repository containing the code. This will be checked out/cloned into the current directory.

-m MAIN, --main=MAIN the name of the script that would be supplied on the command line if running the simulation or analysis

```
normally, e.g. init.hoc.
-c ON_CHANGED, --on-changed=ON_CHANGED
    the action to take if the code in the repository or
    any of the dependencies has changed. Defaults to
    error
-s STORE, --store=STORE
    specify the path to the record store, either an
    existing one or one to be created.
-A PATH, --archive=PATH
    specify a directory in which to archive output
    datafiles. If not specified, datafiles are not
    archived.
-g OPTION, --labelgenerator=OPTION
    specify which method Sumatra should use to generate
    labels (options: timestamp, uuid)
-t TIMESTAMP_FORMAT, --timestamp_format=TIMESTAMP_FORMAT
    the timestamp format given to strftime
-M URL, --mirror=URL
    specify a URL at which your datafiles will be
    mirrored.
```

1.4.9 list

Usage: smt list [options] [TAGS]

If TAGS (optional) is specified, then only records with a tag in TAGS will be listed.

Options:

```
-h, --help          show this help message and exit
-l, --long          prints full information for each record
-T, --table        prints information in tab-separated columns
-f FMT, --format=FMT
    FMT can be 'text' (default) or 'html'.
```

1.4.10 repeat

Usage: smt repeat LABEL

Re-run a previous simulation/analysis under (in theory) identical conditions, and check that the results are unchanged.

Options:

```
-h, --help  show this help message and exit
```

1.4.11 run

Usage: smt run [options] [arg1, ...] [param=value, ...]

The list of arguments will be passed on to the simulation/analysis script. It should normally contain at least the name of a parameter file, but can also contain input files, flags, etc. If the parameter file should be in a format that Sumatra understands (see documentation), then the parameters will be stored to allow future searching, comparison, etc. of records. For convenience, it is possible to specify a file with default parameters and then specify those parameters that are different from the default values on the command line with any number of param=value pairs (note no space around the equals sign).

Options:

```

-h, --help          show this help message and exit
-v REV, --version=REV
                    use version REV of the code (if this is not the same
                    as the working copy, it will be checked out of the
                    repository). If this option is not specified, the most
                    recent version in the repository will be used. If
                    there are changes in the working copy, the user will
                    be prompted to commit them first
-l LABEL, --label=LABEL
                    specify a label for the experiment. If no label is
                    specified, the label will be based on PARAMFILE and
                    the timestamp.
-r REASON, --reason=REASON
                    explain the reason for running this
                    simulation/analysis.
-e PATH, --executable=PATH
                    Use this executable for this run. If not specified,
                    the project's default executable will be used.
-m MAIN, --main=MAIN
                    the name of the script that would be supplied on the
                    command line if running the simulation/analysis
                    normally, e.g. init.hoc. If not specified, the
                    project's default will be used.
-n N, --num_processes=N
                    run a distributed computation on N processes using
                    MPI. If this option is not used, or if N=0, a normal,
                    serial simulation/analysis is run.
-t TAG, --tag=TAG   tag you want to add to the project
-D, --debug         print debugging information.

```

1.4.12 sync

```
Usage: smt sync PATH1 [PATH2]
```

Synchronize two record stores. If both PATH1 and PATH2 are given, the record stores at those locations will be synchronized. If only PATH1 is given, and the command is run in a directory containing a Sumatra project, only that project's records be synchronized with the store at PATH1. Note that PATH1 and PATH2 may be either filesystem paths or URLs.

Options:

```
-h, --help  show this help message and exit
```

1.4.13 tag

```
Usage: smt tag [options] TAG [LIST]
```

If TAG contains spaces, it must be enclosed in quotes. LIST should be a space-separated list of labels for individual records. If it is omitted, only the most recent record will be tagged. If the '-d/--delete' option is set, the tag will be removed from the records.

Options:

```
-h, --help  show this help message and exit
-r, --remove  remove the tag from the record(s), rather than adding it.
```

1.4.14 upgrade

```
Usage: smt upgrade
```

```
Upgrade an existing Sumatra project. You must have previously run "smt export" or the standalone 'export.py' script.
```

```
Options:
```

```
-h, --help show this help message and exit
```

1.5 Using the web interface

The web interface is built using the [Django](#) web framework, and requires that Django be installed (see [Installation](#)).

1.5.1 Starting the web interface

Before using the web interface, you must have created a Sumatra project using `smt init`.

To launch the web interface, in your project directory run:

```
$ smtweb &
```

This will launch a simple web server that listens on port 8000, and will automatically open a new tab in your browser at <http://127.0.0.1:8000/>. You can specify the `-n` option which will disable automatic opening of the new tab:

```
$ smtweb -n
```

If port 8000 is already in use, you can specify a different port with the `-p` option to `smtweb`, e.g.:

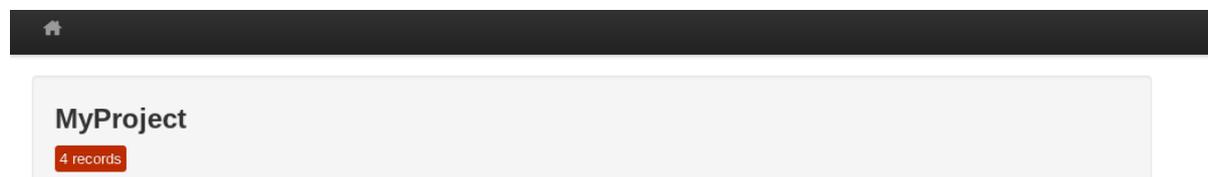
```
$ smtweb -p 8001
```

If you are using a single record store for multiple projects, you can run `smtweb` from anywhere and specify the location of the record store on the command line, e.g.:

```
$ smtweb ~/sumatra.db
```

1.5.2 List of projects

When you first start `smtweb`, the first page you see is a list of your projects.



Click on the project name to see the records of your simulations/analyses in that project.

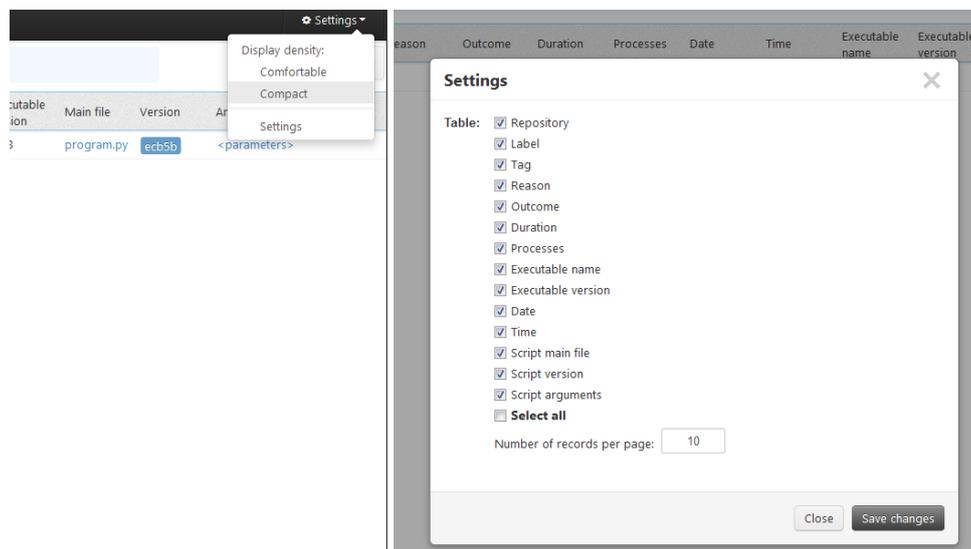
1.5.3 List of records

The list of records page contains a table with the following columns:

- version control repository

- label
- tags
- reason
- outcome
- duration
- number of processes
- date
- time
- executable name
- executable version
- main file
- version
- command line arguments

You can change which columns to display by clicking on **Settings**.



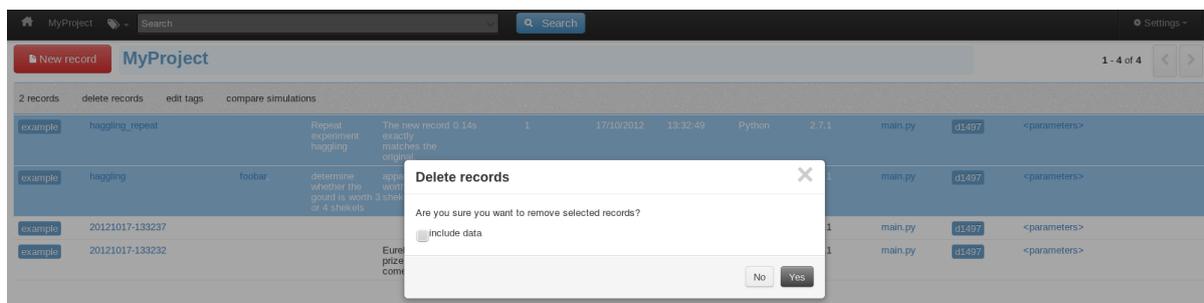
Selecting records

Each record is represented as one row in the table. The rows can be selected by dragging the mouse over them. As soon as you start doing that, the header of the table will be changed to contain actions you can perform with selected records. Actions you can perform are:

- set tags for the selected records
- delete records
- compare records

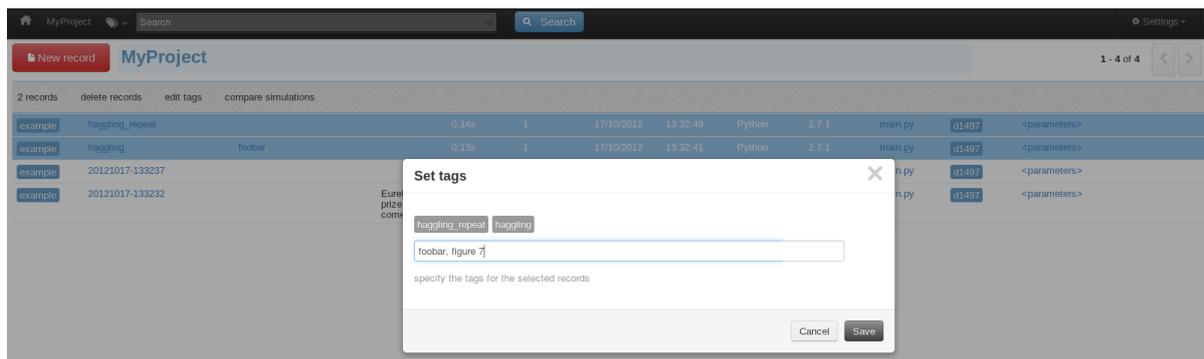
Deleting records

When deleting records, you have the option of also deleting any data generated by that simulation or analysis.



Editing tags

For the selected records you can specify additional tags, edit or remove them.



Comparing records

From the set of selected records you can choose any two to compare them.

Reviewing your code

You can see the contents of your main script file by clicking the corresponding link in the table. It will be shown in the modal window which can be dragged around.

Important: for now this works only for Git and Mercurial repositories.

Filtering the records

You can filter the records by clicking on the 'tag' button or by using the search form. This form contains the following items:

- label
- tags

Comparison of simulations

Click the records you would like to compare:

[hagglng](#) [20121017-133237](#)

20121017-133237			hagglng																																
Reason:																																			
Outcome:																																			
Timestamp:	17/10/2012 13:32:37		17/10/2012 13:32:41																																
Duration:	0.12s		0.13s																																
Executable:	Python version 2.7.1 (/home/andrew/erw/sumatra_py27_django13/bin/python)		Python version 2.7.1 (/home/andrew/erw/sumatra_py27_django13/bin/python)																																
Launch mode:	serial		serial																																
Repository:	/home/andrew/tmp/example		/home/andrew/tmp/example																																
Main file:	main.py		main.py																																
Version:	d1497c334b4df1f06a9c5e55a894bcc55b58e874		d1497c334b4df1f06a9c5e55a894bcc55b58e874																																
Arguments:	seed = 45245 distr = "uniform" n = 100		seed = 348570 distr = "uniform" n = 100																																
Tags:			foobar, figure 7																																
Input files:	no input files		no input files																																
Output files:	/Data output.dat dcb86788c2c793804a04e683fae99ad0bac8tb99 2.4 KB		/Data output.dat b8c62d32b6175b8d4cce899cd7d3f598fa82a94 2.4 KB																																
Dependencies:	<table border="1"> <thead> <tr> <th>Name</th> <th>Path</th> <th>Version</th> </tr> </thead> <tbody> <tr> <td>distutils</td> <td>/home/andrew/erw/sumatra_py27_django13/lib/python2.7/distutils</td> <td>2.7.1</td> </tr> <tr> <td>encodings</td> <td>/home/andrew/erw/sumatra_py27_django13/lib/python2.7/encodings</td> <td>unknown</td> </tr> <tr> <td>nose</td> <td>/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/nose</td> <td>1.1.2</td> </tr> <tr> <td>numpy</td> <td>/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy</td> <td>1.6.2</td> </tr> </tbody> </table>	Name	Path	Version	distutils	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/distutils	2.7.1	encodings	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/encodings	unknown	nose	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/nose	1.1.2	numpy	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy	1.6.2		<table border="1"> <thead> <tr> <th>Name</th> <th>Path</th> <th>Version</th> </tr> </thead> <tbody> <tr> <td>distutils</td> <td>/home/andrew/erw/sumatra_py27_django13/lib/python2.7/distutils</td> <td>2.7.1</td> </tr> <tr> <td>encodings</td> <td>/home/andrew/erw/sumatra_py27_django13/lib/python2.7/encodings</td> <td>unknown</td> </tr> <tr> <td>nose</td> <td>/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/nose</td> <td>1.1.2</td> </tr> <tr> <td>numpy</td> <td>/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy</td> <td>1.6.2</td> </tr> </tbody> </table>	Name	Path	Version	distutils	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/distutils	2.7.1	encodings	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/encodings	unknown	nose	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/nose	1.1.2	numpy	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy	1.6.2		
Name	Path	Version																																	
distutils	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/distutils	2.7.1																																	
encodings	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/encodings	unknown																																	
nose	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/nose	1.1.2																																	
numpy	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy	1.6.2																																	
Name	Path	Version																																	
distutils	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/distutils	2.7.1																																	
encodings	/home/andrew/erw/sumatra_py27_django13/lib/python2.7/encodings	unknown																																	
nose	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/nose	1.1.2																																	
numpy	/home/andrew/erw/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy	1.6.2																																	
Platform information:	<ol style="list-style-type: none"> Name: retina IP address: 127.0.1.1 Processor: x86_64 x86_64 Architecture: 64bit 		<ol style="list-style-type: none"> Name: retina IP address: 127.0.1.1 Processor: x86_64 x86_64 Architecture: 64bit 																																

MyProject

New record

1 records delete records edit tags compare simulations

example hagglng_repeat figure 7, foobar,

example hagglng figure 7 foobar

example 20121017-133237

example 20121017-133232

Eureka! Nobel prize here we come.

main.py: d1497c334b4df1f06a9c5e55a894bcc55b58e874

```

1 import numpy
2 import sys
3
4 def main(parameters):
5     numpy.random.seed(parameters["seed"])
6     distr = getattr(numpy.random, parameters["distr"])
7     data = distr(size=parameters["n"])
8     output_file = "Data/output.dat"
9     numpy.savetxt(output_file, data)
10
11 parameter_file = sys.argv[1]
12 parameters = {}
13 execfile(parameter_file, parameters)
14
15 main(parameters)
16
17

```

main.py d1497 <parameters>

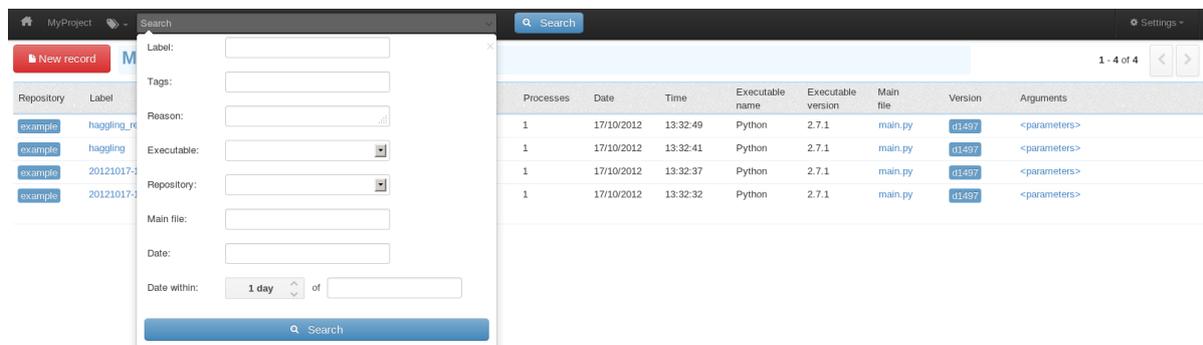
main.py d1497 <parameters>

main.py d1497 <parameters>

main.py d1497 <parameters>

1 - 4 of 4

- reason
- executable
- repository
- main file
- date
- interval of dates



Search by **variable name** allows filtering the records using the name of parameter. If for each new simulation you have different parameter set, this feature can be useful for narrowing the set of possible records.

1.5.4 Accessing record details

You can access the record detail by clicking the corresponding label name in the main table. The record detail page contains the following sections:

- general info
- input files
- output files
- parameters
- dependencies
- platform information
- stdout & stderr

1.5.5 Finishing up

Don't forget to kill the webserver process (e.g. with `fg, Ctrl-C`) when you are finished with it.

1.5.6 Launching computations from the web interface

It is possible to run simulations/analyses from within the web interface. Clicking on the “New record” button will bring up the following dialog:

You can specify **label**, **reason**, **tag**, **main file**, **arguments**, and **executable** there. **Main file** and **arguments** are drop-down lists with the names of the files from the folder of the Sumatra project. As soon as the file is picked from the list, you can see its content. Moreover, the **argument** file is editable and any changes you made can be saved from this web page. You can hide and open the content of the files by clicking the corresponding links below the drop-down list.

On clicking run, the computation is launched and a progress bar appears. On successful completion, a new record is added to the **list of records** page.

MyProject 20121017-133232

General info

Label: 20121017-133232

Reason:

Outcome:

Timestamp: 17/10/2012 13:32:32

Duration: 0.12s

Executable: Python version 2.7.1 (/home/andrew/sumatra_py27_django13/bin/python)

Launch mode: serial

Repository: /home/andrew/tmp/example

Main file: main.py

Version: d1497c334b4df1f06a0c5e55a894bcc55b56e874

Arguments: <parameters>

Tags:

Parameters

seed
45245

dist
uniForm

n
100

Dependencies

Name	Path	Version
distutils	/home/andrew/sumatra_py27_django13/lib/python2.7/distutils	2.7.1
encodings	/home/andrew/sumatra_py27_django13/lib/python2.7/encodings	unknown
nose	/home/andrew/sumatra_py27_django13/local/lib/python2.7/site-packages/nose	1.1.2
numpy	/home/andrew/sumatra_py27_django13/local/lib/python2.7/site-packages/numpy	1.6.2

Platform information

Name	IP address	Processor	Architecture	System type	Release	Version
retina	127.0.1.1	x86_64 x86_64	64bit ELF	Linux	2.6.38-11-generic	#50-Ubuntu SMP Mon Sep 12 21:17:25 UTC 2011

Stdout & Stderr

No output.

Input files

no input files

Output files

JData
[output.dat](#) dcb86788c2c793804a04e683fae99e02bac8f899 2.4 KB

New computation

Label:

Reason:

Tag:

Executable path:

Main file: [show content](#)

Arguments: [show content](#)

MyProject Search Settings

New computation << >> romans.param

Label:

Reason:

Tag:

Executable path:

Main file: [show content](#)

Arguments: [hide content](#)

```
seed = 348571
dist = 'uniForm'
n = 1000
```

Customizing the web interface

You can customize the web interface on a per-project basis by placing your own [Django templates](#) in a “templates” subdirectory of the Sumatra “.smt” directory. The templates you can customize are called “base.html”, “record_list.html”, “show_file.html”, “project_detail.html”, “show_csv.html”, “show_image.html”, “record_detail.html”, “show_diff.html”, “tag_list.html”. The best way to proceed is to copy the default template from “/path/to/sumatra/web/templates” to “/path/to/myproject/.smt/templates” and then modify it.

1.6 Graphical user interfaces

Sumatra has a command line interface and a web interface, but does not currently have a graphical desktop client.

We will probably write one at some point, but we hope that other people will also write their own, building on top of the tools and functionality provided in the Sumatra package.

The reason we hope for multiple desktop clients is that everyone has their own preferred workflow, and it seems unlikely that one graphical interface will work equally well for everyone.

1.7 Parallel computations

As well as launching computations on your local machine, Sumatra can launch distributed, MPI-based computations on a cluster, at least for simple use-cases. We assume you already have your hosts files, etc. set up. Then, to run your computation on 17 nodes, run:

```
$ smt run -n 17 default.param
```

(assuming you have already configured your default executable and main script file). This will call `mpiexec` for you with the appropriate arguments.

If this is insufficiently configurable for you, please take a look at the `DistributedLaunchMode` class in `launch.py` within the source distribution, and get in touch with the Sumatra developers, for example by [creating a ticket](#) or asking a question on the [mailing list](#).

1.8 Reproducible publications: including and linking to provenance information in documents

Sumatra provides tools to include figures and other results generated by Sumatra-tracked computations in documents, with links to full provenance information: i.e. the full details of the code, input data and computational environment used to generate the figure/result.

1.8.1 LaTeX

To include figures from a Sumatra project in a LaTeX document, copy the file `sumatra.sty` from the `src/publishing/latex` directory in the Sumatra source distribution into your working directory, then add

```
\usepackage{sumatra}
```

in the preamble of your LaTeX document. If your LaTeX working directory is not the same as your Sumatra project directory, you can specify the location of the record store and the project name as package options:

```
\usepackage[recordstore=/path/to/db, project=MyProject]{sumatra}
```

You can then use the `\smtincludegraphics` command in place of the usual `\includegraphics`:

```
\smtincludegraphics{20120907-153528:my_figure.png}
```

Here the argument is the label of a Sumatra record. This command will look up the record in your Sumatra project, find the location of the image file (whether a local file path or a URL) and include the figure in the document. If you are using the `HttpRecordStore`, the figure will also be a hyperlink to the corresponding Sumatra record, so you can easily check the full provenance of the figure.

If `my_figure.png` is the only image file produced by that Sumatra run, then you can use just:

```
\smtincludegraphics{20120907-153528}
```

without the path part.

All the usual `\includegraphics` options are supported, e.g.:

```
\smtincludegraphics[width=\textwidth]{20120907-153528:my_figure.png}
```

You can also include as an option the SHA1 hash of the image file contents, as captured by Sumatra, to ensure that the image you include is the correct one, and that it hasn't been accidentally overwritten or replaced by another:

```
\smtincludegraphics[width=\textwidth, digest=e2d1054c2893f19f50c43ddd5a344b59383df648]{20120907-153528:my_figure.png}
```

Note: you will have to run `latex/pdflatex` with the `-shell-escape` option.

1.8.2 Sphinx

To include figures from a Sumatra project, or links to Sumatra records of simulations or analyses, in a [Sphinx](#) document, you should first add `'sumatra.publishing.sphinxext'` to the extensions list in your `conf.py` and then set the following options:

```
sumatra_record_store = "/path/to/record/store"
sumatra_project_name = "MyProject"
sumatra_link_icon = "icon_info.png"
```

The `sumatra.publishing.sphinxext` extension provides a `reStructuredText` directive `smtimage` and a `reStructuredText` role `smtlink`.

The `smtimage` directive

`smtimage` includes an image file retrieved from a Sumatra project in your document. If you are using the `HttpRecordStore`, the image will also be a hyperlink to the corresponding Sumatra record.

```
.. smtimage:: 20120907-153528:my_figure.png
```

All the usual options to the normal `image` directive can be used (`'width'`, `'height'`, `'align'`, etc.), as well as the Sumatra-specific options `"digest"` (which is used to check the identity of the included image, as detailed in the section on LaTeX above) `"record_store"` and `"project"`, which can be used to over-ride the global options specified in `conf.py`, and so mix results from multiple projects in a single document. Here is a rather complete example.

```
.. smtimage:: 20120907-153528:my_figure.png
   :record_store: http://sumatra.example.com/
   :project: MyOtherProject
   :digest: e2d1054c2893f19f50c43ddd5a344b59383df648
   :width: 800 px
   :align: center
   :class: some-css-class
```

The smtlink role

If using a `HttpRecordStore`, the `smtlink` role inserts an icon in the document, which is a hyperlink to a given record in the store.

To use this, you must set the `sumatra_link_icon` option, as discussed above.

1.9 Using the Sumatra API within your own scripts

Using the `smt run` command is quick and convenient, but it does require you to change the way you launch your simulations/analyses.

One way to avoid this, if you use Python, is to use the `sumatra` package within your own scripts to perform the record-keeping tasks performed by `smt run`.

You may also wish to write your own custom script for creating a Sumatra project, instead of using `smt init`, but we do not cover this scenario here.

We will start with a simple example script, a dummy simulation, that reads a parameter file, generates some random numbers, and writes some data to file:

```
import numpy
import sys

def main(parameters):
    numpy.random.seed(parameters["seed"])
    distr = getattr(numpy.random, parameters["distr"])
    data = distr(size=parameters["n"])
    output_file = "example.dat"
    numpy.savetxt(output_file, data)

parameter_file = sys.argv[1]
parameters = {}
execfile(parameter_file, parameters) # this way of reading parameters
                                     # is not necessarily recommended
main(parameters)
```

Let's suppose this script is in a file named `myscript.py`, and that we have a parameter file named `defaults.param`, which contains:

```
seed = 65784
distr = "uniform"
n = 100
```

Without Sumatra, we would normally run this script using something like:

```
$ python myscript.py defaults.param
```

To run the script using the `smt` command line tool, we would use:

```
$ smt run --reason="reason for running this simulation" defaults.param
```

(This assumes we have previously used `smt init` or `smt configure` to specify that our executable is `python` and our main file is `myscript.py`.)

To benefit from the functionality of Sumatra without having to use `smt run`, we have to integrate the steps performed by `smt run` into our script.

First, we have to load the Sumatra project:

```
from sumatra.projects import load_project
project = load_project()
```

We're going to want to record the simulation duration, so we import the standard Python `time` module and record the start time:

```
import time
start_time = time.time()
```

We need to slightly modify the procedure for reading parameters. Sumatra stores the parameters for later use in searching and comparison, so they need to be transformed into a form Sumatra can use. This is very simple, we just replace the `execfile()` call with a `build_parameters()` call:

```
from sumatra.parameters import build_parameters
parameters = build_parameters(parameter_file)
```

Now we create a new `Record` object, telling it that the script is the current file; this automatically registers information about the simulation environment:

```
record = project.new_record(parameters=parameters,
                            main_file=__file__,
                            reason="reason for running this simulation")
```

Now comes the main body of the simulation, which is unchanged except that we take the opportunity to give the output data file a more informative name by adding the record label to the parameter file:

```
output_file = "%s.dat" % parameters["sumatra_label"]
```

At the end of the simulation, we calculate the simulation duration and search for newly created files:

```
record.duration = time.time() - start_time
record.output_data = record.datastore.find_new_data(record.timestamp)
```

Now we add this simulation record to the project, and save the project:

```
project.add_record(record)
project.save()
```

Putting this all together:

```
import numpy
import sys
import time
from sumatra.projects import load_project
from sumatra.parameters import build_parameters

def main(parameters):
    numpy.random.seed(parameters["seed"])
    distr = getattr(numpy.random, parameters["distr"])
    data = distr(size=parameters["n"])
    output_file = "%s.dat" % parameters["sumatra_label"]
    numpy.savetxt(output_file, data)

parameter_file = sys.argv[1]
parameters = build_parameters(parameter_file)

project = load_project()
record = project.new_record(parameters=parameters,
                            main_file=__file__,
                            reason="reason for running this simulation")
parameters.update({"sumatra_label": record.label})
start_time = time.time()

main(parameters)

record.duration = time.time() - start_time
record.output_data = record.datastore.find_new_data(record.timestamp)
project.add_record(record)
```

```
project.save()
```

Now you can run the simulation in the original way:

```
python myscript.py defaults.param
```

and still have the simulation recorded in your Sumatra project. For such a simple script and simple run environment there is no advantage to doing it this way: `smt run` is much simpler. However, if you already have a fairly complex run environment, this provides a straightforward way to integrate Sumatra's functionality into your existing system.

You will have noticed that much of the Sumatra code you have to add is effectively boilerplate, which will be the same for all your scripts. To save time, and typing therefore, Sumatra provides a `@capture` decorator for your `main()` function:

```
import numpy
import sys
from sumatra.parameters import build_parameters
from sumatra.decorators import capture

@capture
def main(parameters):
    numpy.random.seed(parameters["seed"])
    distr = getattr(numpy.random, parameters["distr"])
    data = distr(size=parameters["n"])
    numpy.savetxt("%s.dat" % parameters["sumatra_label"], data)

parameter_file = sys.argv[1]
parameters = build_parameters(parameter_file)
main(parameters)
```

This is now hardly any longer than the original script.

1.10 Parameter files

There is no requirement to put parameters in a separate file to use Sumatra, nor is it required to use a particular parameter file format. However, if you do use one of the formats Sumatra supports then you will gain extra functionality: currently, the ability to modify/add parameters on the command line and to have Sumatra automatically add the record label to the parameter file; in future versions, the ability to search and filter your records based on parameters.

1.10.1 Supported formats

Simple

Each parameter on a separate line, in “name = value” format. Values may be numbers, strings or lists (denoted with square brackets, comma-separated). Comments are indicated by a leading “#”. Example:

```
# Example parameter file
nx = 10 # } grid
ny = 12 # } size
inputs = [1e-3, 2e-3, 5e-3, 1e-2]
label = "default"
```

Config/ini-style

Traditional config file format, as parsed by the standard Python `ConfigParser` module. Note that this format does not distinguish numbers from string representations of those numbers, so all parameter values are treated as

strings. This format allows one level of nesting: you can create sections within which you can define parameters. Comments are indicated by a leading "#". Example:

```
[sectionA]
  a: 2
  b: 3

[sectionB]
  c: hello
  d: world
```

See the `ConfigParser` docs for more details.

JSON

See <http://www.json.org/>.

YAML

See <http://www.yaml.org>

NeuroTools format

`NeuroTools` parameter format is essentially the same as JSON, but with the addition of a `url()` function which allows sub-parameter sets to be included from other files. Example:

```
{
  "network": {
    "excitatory_cells": url("https://neuralensemble.org/svn/NeuroTools/trunk/doc/example.param"),
    "inhibitory_cells": {
      "tau_m": 15.0,
      "cm": 0.75,
    },
  },
  "sim": {
    "tstop": 1000.0,
    "dt": 0.11,
  }
}
```

1.10.2 Adding new formats

If your parameter file format is not supported by Sumatra, there is no problem: Sumatra will treat your parameter file as any other input data file and pass it directly through to your simulation/analysis script.

However, it is fairly straightforward to add support for a new format. You will need to write a Python class according to the following skeleton:

```
class MyParameterSet(object):

    def __init__(self, initialiser):
        # initialiser could be either a filesystem path or a string containing
        # the contents of your parameter file, and should raise a SyntaxError
        # if it cannot make sense of the contents.

    def __getitem__(self, name):
        # return the parameter or sub-parameter set with the given name

    def __eq__(self, other):
```

```
# must be implemented

def __ne__(self, other):
    # must be implemented

def as_dict(self):
    # return the parameter set as a Python dict containing only numerical
    # types, lists, or other dicts.

def save(self, filename):
    # self-explanatory

def pop(self, k, d=None):
    # same behaviour as Python dict

def update(self, E, **F):
    # same behaviour as Python dict

def pretty(self, expand_urls=False):
    # Return a string representation of the parameter set, suitable for
    # creating a new, identical parameter set.
    # expand_urls is present for compatibility with NTPParameterSet, and need
    # not be used.
```

For this version of Sumatra, you will have to include this class within the file `parameters.py` of your Sumatra installation, or send it to the developers to include in the Sumatra repository (see [Developers' guide](#)), as well as editing the `build_parameters()` function within `parameters.py` so that it tries to use your class. In the next version of Sumatra, we plan to include a plugin system which will greatly simplify adding your own customizations.

1.11 Upgrading your projects

Since new versions of Sumatra extend its capabilities, and may change the way records are stored, when you install a new version of Sumatra you will need to upgrade your existing Sumatra projects to work with the new version.

In future, this will probably be done automatically, but while Sumatra is still rapidly evolving we are keeping it as a simple manual process to minimize the risk of data loss.

1.11.1 Export using the old version

Before installing the new version of Sumatra, you must export your project to a file.

For Sumatra 0.1-0.3

First, download `export.py` to your project directory, then run:

```
$ python export.py
```

This will export your project in JSON format to two files in the `.smt` directory: `records_export.json` and `project_export.json`.

You can now delete `export.py`

For Sumatra 0.4 and later

Run:

```
$ smt export
```

This will export your project in JSON format to two files in the `.smt` directory: `records_export.json` and `project_export.json`.

1.11.2 Install the new version and upgrade

Now you can install the new version, e.g. with:

```
$ pip install --upgrade sumatra
```

or:

```
$ easy_install -U sumatra
```

(or you can install from source, as explained in *doc:installation*).

Then run:

```
$ smt upgrade
```

The original `.smt` directory will be copied to a time-stamped directory, e.g. `.smt_backup_20110209132422`.

1.12 Migrating records between record stores

Sumatra supports multiple back-ends for storing records, `ShelveRecordStore`, based on the `shelve` module from the Python standard library, `DjangoRecordStore`, based on a relational database accessed via the [Django ORM](#), and `HttpRecordStore`, which stores records on a remote database with communication based on JSON over HTTP.

Suppose you created a Sumatra project using the `ShelveRecordStore` since you didn't want to install Django, then decided you'd like to change to `DjangoRecordStore`. This is what the project looks like at the beginning:

```
$ smt info
Sumatra project
-----
Name           : MyProject
Default executable : Python (version: 2.5.2) at /usr/local/bin/python
Default repository : MercurialRepository at /path/to/working/directory
Default main file  : main.py
Default launch mode : serial
Data store       : /path/to/data
Record store     : Record store using the shelve package (database file=.smt/records)
Code change policy : store-diff
Append label to   : cmdline
$ smt list
20110309-141853
20110309-141849
```

First, rename the `.smt` directory, and then create a new project:

```
$ mv .smt .smt_orig
$ smt init MyProject --addlabel cmdline --executable=python --on-changed=store-diff --main=main.p
```

Now we synchronize the old and new databases:

```
$ smt sync .smt_orig/records
```

and just to check it worked:

```
$ smt info
Sumatra project
-----
Name           : MyProject
Default executable : Python (version: 2.5.2) at /usr/local/bin/python
Default repository : MercurialRepository at /path/to/working/directory
Default main file  : main.py
Default launch mode : serial
Data store       : /path/to/data
Record store     : Relational database record store using the Django ORM (database file=.smt/r
Code change policy : store-diff
Append label to   : cmdline
$ smt list
20110309-141853
20110309-141849
```

1.13 Developers' guide

These instructions are for developing on a Unix-like platform, e.g. Linux or Mac OS X, with the bash shell.

1.13.1 Requirements

- Python 2.6 and/or 2.7
- Django >= 1.2
- django-tagging >= 0.3
- nose >= 0.11.4
- if using Python 2.6, unittest2 >= 0.5.1
- Distribute >= 0.6.14
- (optional) mpi4py >= 1.2.2
- (optional) twill >= 0.9 (for testing web interface)
- (optional) coverage >= 3.3.1 (for measuring test coverage)

We strongly recommend developing within a `virtualenv`.

1.13.2 Getting the source code

We use the Mercurial version control system. To get a copy of the code:

```
$ cd /some/directory
$ hg clone https://neuralensemble.org/hg/sumatra sumatra_src
```

If you get an `SSL3_GET_SERVER_CERTIFICATE` error, you can either use the `--insecure` option to `hg clone`, or add the following lines to your `~/.hgrc` file:

```
[hostfingerprints]
neuralensemble.org = 7c:a8:72:19:79:b2:ec:d8:27:59:34:dc:a3:13:d9:5e:c3:aa:53:0e
```

Note: an alternative is to fork the [Sumatra mirror on Bitbucket](#), then clone your own fork.

Now you need to make sure that the `sumatra` package is on your `PYTHONPATH` and that the `smt` and `smtweb` scripts are on your `PATH`. You can do this either by installing Sumatra:

```
$ cd sumatra_src
$ python setup.py install
```

(if you do this, you will have to re-run `setup.py install` any time you make changes to the code) *or* by creating symbolic links from somewhere on your `PATH` and `PYTHONPATH`, for example:

```
$ cd /some/directory
$ ln -s sumatra_src/src sumatra
$ export PYTHONPATH=/some/directory:${PYTHONPATH}
$ export PATH=/some/directory/sumatra_src/bin:${PATH}
```

To update to the latest version from the repository:

```
$ hg pull -u
```

1.13.3 Running the test suite

Before you make any changes, run the test suite to make sure all the tests pass on your system:

```
$ cd sumatra_src/test/unittests
$ nosetests
```

You will see some error messages, but don't worry - these are just tests of Sumatra's error handling. At the end, if you see "OK", then all the tests passed, otherwise it will report how many tests failed or produced errors.

If any of the tests fail, check out the [continuous integration server](#) to see if these are "known" failures, otherwise please [open a bug report](#).

(many thanks to the [NEST Initiative](#) for hosting the CI server).

1.13.4 Writing tests

You should try to write automated tests for any new code that you add. If you have found a bug and want to fix it, first write a test that isolates the bug (and that therefore fails with the existing codebase). Then apply your fix and check that the test now passes.

To see how well the tests cover the code base, run:

```
$ nosetests --coverage --cover-package=sumatra --cover-erase
```

1.13.5 Committing your changes

Once you are happy with your changes, you can commit them to your local copy of the repository:

```
$ hg commit -m 'informative commit message'
```

If you have a NeuralEnsemble account, you can now push your changes to the central repository:

```
$ hg push https://neuralensemble.org/hg/sumatra
```

If you are using Bitbucket, you can open a pull request.

Otherwise, you can create a patch:

```
$ hg export tip > descriptive_name.patch
```

and attach it to a ticket in the [issue tracker](#). If you have made more than one commit, determine the revision number of when you cloned or last updated from the central repository (using `hg log`), and then give a range of revisions to include in the patch:

```
$ hg export start-revision:tip > descriptive_name.patch
```

Before either pushing or creating a patch, run the test suite again to check that you have not introduced any new bugs.

1.13.6 Coding standards and style

All code should conform as much as possible to [PEP 8](#), and should run with Python 2.6 and 2.7.

1.14 API reference

1.14.1 Input and output data

The `datastore` module provides an abstraction layer around data storage, allowing different methods of storing simulation/analysis results (local filesystem, remote filesystem, database, etc.) to provide a common interface.

The interface is built around three types of object: a `DataStore` may contain many `DataItems`, each of which is identified by a `DataKey`.

There is a single `DataKey` class. `DataStore` and `DataItem` are abstract base classes, and must be subclassed to provide different functionality.

Base classes

class `sumatra.datastore.DataKey` (*path, digest, **metadata*)

Identifies a `DataItem`, and may be used to retrieve a `DataItem` from a `DataStore`.

May also be used to store metadata (e.g. file size, mimetype) and be used as a proxy for the `DataItem` on a system where the actual data is not available.

path

a token used to retrieve a `DataItem`. For filesystem-based `DataStores`, this will be a relative path. For database-backed stores (none of which have been implemented yet :-)) it could be a primary key or an object encapsulating a query.

digest

the SHA1 digest of the contents of the associated `DataItem`. This attribute is calculated on creation of the `DataKey`.

metadata

a dict containing metadata, such as file size and mimetype.

class `sumatra.datastore.base.DataItem`

Base class for data item classes, that may represent files or database records.

digest

docstring

generate_key ()

Generate a `DataKey` uniquely identifying this data item.

get_content (*max_length=None*)

Return the contents of the data item as a string.

If *max_length* is specified, return that number of bytes, otherwise return the entire content.

save_copy (*path*)

Save a copy of the data to a local file.

If *path* is an existing directory, the data item path will be appended to it, otherwise *path* is treated as a full path including filename, either absolute or relative to the working directory.

Return the full path of the final file.

sorted_content ()

Return the contents of the data item, sorted by line.

class `sumatra.datastore.base.DataStore`

Base class for data storage abstractions.

contains_path (*path*)

Does the store contain a data item with the given path?

copy ()

delete (**keys*)

Delete the files corresponding to the given keys.

find_new_data (*timestamp*)

Finds newly created/changed data items

generate_keys (**paths*)

Given a number of “paths”, return a list of keys enabling the data at those paths to be retrieved from this store later.

get_content (*key, max_length=None*)

Return the contents of a file identified by a key.

If *max_length* is given, the return value will be truncated.

get_data_item (*key*)

Return the file that matches the given key.

Storing data on the local filesystem

class `sumatra.datastore.FileSystemDataStore` (*root*)

Bases: `sumatra.datastore.base.DataStore`

Represents a locally-mounted filesystem. The root of the data store will generally be a subdirectory of the real filesystem.

root

The absolute path on the underlying file system to the root directory of the data store.

class `sumatra.datastore.filesystem.DataFile` (*path, store*)

Bases: `sumatra.datastore.base.DataItem`

A file-like object, that represents a file in a local filesystem.

path

path relative to the `FileSystemDataStore` root

full_path

absolute path relative to the underlying filesystem.

size

file size in bytes

name

file name

extension

file extension

mimetype

if the mimetype cannot be guessed, this will be None

Automatic archiving of data written to the local filesystem

class `sumatra.datastore.ArchivingFileSystemDataStore` (*root*, *archive*='.smt/archive')

Bases: `sumatra.datastore.filesystem.FileSystemDataStore`

Represents a locally-mounted filesystem that archives any new files created in it. The root of the data store will generally be a subdirectory of the real filesystem.

archive_store

Directory within which data will be archived.

class `sumatra.datastore.archivingfs.ArchivedDataFile` (*path*, *store*)

Bases: `sumatra.datastore.base.DataItem`

A file-like object, that represents a file inside a tar archive

Mirroring data to a remote webserver

class `sumatra.datastore.MirroredFileSystemDataStore` (*root*, *mirror_base_url*)

Bases: `sumatra.datastore.filesystem.FileSystemDataStore`

Represents a locally-mounted filesystem whose contents are mirrored on a webserver, so that the files can be accessed via an HTTP URL.

mirror_base_url

URL to which the file path will be appended to obtain the final URL of a file

class `sumatra.datastore.mirroredfs.MirroredDataFile` (*path*, *store*)

Bases: `sumatra.datastore.base.DataItem`

A file-like object, that represents a file existing both on a local file system and on a webserver.

1.14.2 Finding dependencies

The `dependency_finder` sub-package attempts to determine all the dependencies of a given script, including the version of each dependency.

For each executable that is supported there is a sub-module containing a `find_dependencies()` function, and a series of heuristics for finding version information. There is also a sub-module `core`, which contains heuristics that are independent of the language, e.g. where the dependencies are under version control.

For users of the API, the principal function of interest is the following.

`sumatra.dependency_finder.find_dependencies` (*filename*, *executable*)

Return a list of dependencies for a given script and programming language.

filename: the path to the script whose dependencies should be found.

executable: an instance of `Executable` or one of its subclasses.

This function returns a list of `Dependency` objects. There is a different `Dependency` subclass for each programming language, but all have the following attributes:

class `sumatra.dependency_finder.core.BaseDependency` (*name*, *path*=None, *version*='unknown', *diff*='', *source*=None)

Contains information about a program component, and tries to determine version information.

name: an identifying name, e.g. the module name in Python

path: the location of the dependency file in the local filesystem

version: the version of the dependency, if that can be determined, otherwise 'unknown'. Always a string, even if the version can also be represented as a number.

diff: if the dependency is under version control and has been modified, the diff between the actual version and the last-committed version.

source: an identifier for where the dependency came from, if known, e.g. the url of a version control repository or the name of a Linux package.

If you are interested in improving the dependency finder for an existing program/language, or in adding a dependency finder for a new program or language, you may be interested in the following.

Language-independent heuristics and utilities

`sumatra.dependency_finder.core.find_versions` (*dependencies, heuristics*)

Try to find version information by calling a series of functions in turn.

dependencies: a list of Dependency objects.

heuristics: a list of functions that accept a component as the single argument and return a version number or 'unknown'.

Returns a possibly modified list of dependencies

`sumatra.dependency_finder.core.find_versions_from_versioncontrol` (*dependencies*)

Determine whether a file is under version control, and if so, obtain version information from this.

`sumatra.dependency_finder.core.find_file` (*path, current_directory, search_dirs*)

Look for *path* as an absolute path then relative to the current directory, then relative to *search_dirs*. Return the absolute path.

Python

`sumatra.dependency_finder.python.find_versions_by_attribute` (*dependencies, executable*)

Try to find version information from the attributes of a Python module.

`sumatra.dependency_finder.python.find_versions_from_egg` (*dependencies*)

Determine whether a Python module is provided as an egg, and if so, obtain version information from this.

`sumatra.dependency_finder.python.find_imported_packages` (*filename, executable_path, debug=0, exclude_stdlib=True*)

Find all imported top-level packages for a given Python file.

We cannot assume that the version of Python being used to run Sumatra is the same as that used to run the simulation/analysis. Therefore we need to run all the dependency finding and version checking in a subprocess with the correct version of Python.

`sumatra.dependency_finder.python.find_dependencies` (*filename, executable*)

Return a list of Dependency objects representing all the top-level modules or packages imported (directly or indirectly) by a given Python file.

Matlab

`sumatra.dependency_finder.matlab.find_dependencies` (*filename, executable*)

`sumatra.dependency_finder.matlab.save_dependencies` (*cmd, filename*)

save all dependencies to the file in the current folder

NEURON

`sumatra.dependency_finder.neuron.find_xopened_files (file_path)`

Find all files that are xopened, whether directly or indirectly, by a given Hoc file. Note that this only handles cases whether the path is given directly, not where it has been previously assigned to a `strdef`.

`sumatra.dependency_finder.neuron.find_loaded_files (file_path, executable_path)`

Find all files that are loaded with `load_file()`, whether directly or indirectly, by a given Hoc file. Note that this only handles cases whether the path is given directly, not where it has been previously assigned to a `strdef`. Also note that this is more complicated than `xopen()`, since NEURON also looks in any directories in `$HOC_LIBRARY_PATH` and `$NEURONHOME/lib/hoc`.

`sumatra.dependency_finder.neuron.find_dependencies (filename, executable)`

Return a list of Dependency objects representing all Hoc files imported (directly or indirectly) by a given Hoc file.

GENESIS

`sumatra.dependency_finder.genesis.find_included_files (file_path)`

Find all files that are included, whether directly or indirectly, by a given .g file.

`sumatra.dependency_finder.genesis.find_dependencies (filename, executable)`

Return a list of Dependency objects representing all files included, whether directly or indirectly, by a given .g file.

1.14.3 Storing provenance information

The `recordstore` module provides an abstraction layer around storage of simulation/analysis records, providing a common interface to different storage methods (simple serialisation, relational database, etc.)

Base class

All record store classes have the following methods. Some stores have additional methods (see below).

class `sumatra.recordstore.base.RecordStore`

Base class for record store implementations.

delete (*project_name, label*)

Delete the record with the given label from the given project.

delete_all ()

Delete all records from the store.

delete_by_tag (*project_name, tag*)

Delete all records from the given project that have been tagged with the given tag.

export (*project_name, indent=2*)

Export store contents as JSON.

get (*project_name, label*)

Retrieve the record with the given label from the given project.

has_project (*project_name*)

Does the store contain any records for the given project?

import_ (*project_name, content*)

Import records in JSON format.

labels (*project_name*)

Return the labels of all records in the given project.

list (*project_name*, *tags=None*)

Return a list of records for the given project.

If *tags* is not provided, list all records, otherwise list only records that have been tagged with one or more of the tags.

list_projects (*project_name*)

Return the names of all projects that have records in the store.

most_recent (*project_name*)

Return the most recent record from the given project.

save (*project_name*, *record*)

Store the given record under the given project.

sync (*other*, *project_name*)

Synchronize two record stores so that they contain the same records for a given project.

Where the two stores have the same label (within a project) for different records, those records will not be synced. The method returns a list of non-synchronizable records (empty if the sync worked perfectly).

sync_all (*other*)

Synchronize all records from all projects between two record stores.

Minimal record store

class `sumatra.recordstore.ShelveRecordStore` (*shelf_name='smt/records'*)

Bases: `sumatra.recordstore.base.RecordStore`

Handles storage of simulation/analysis records based on the Python standard `shelve` module.

The advantage of this record store is that it has no dependencies. The disadvantages are that it allows only local access and does not support the *smtweb* interface.

Django-based record store

class `sumatra.recordstore.DjangoRecordStore` (*db_file='smt/records'*)

Bases: `sumatra.recordstore.base.RecordStore`

Handles storage of simulation/analysis records in a relational database, via the Django object-relational mapper (ORM), which means that any database supported by Django could in principle be used, although for now we assume SQLite.

This record store is needed for the *smtweb* interface.

Client for remote record store

class `sumatra.recordstore.HttpRecordStore` (*server_url*, *username=None*, *password=None*, *disable_ssl_certificate_validation=True*)

Bases: `sumatra.recordstore.base.RecordStore`

Handles storage of simulation/analysis records on a remote server using HTTP.

The server should support the following URL structure and HTTP methods:

/	GET
/<project_name>/[?tags=<tag1>,<tag2>,...]	GET
/<project_name>/tag/<tag>/	GET, DELETE
/<project_name>/<record_label>/	GET, PUT, DELETE

and should both accept and return JSON-encoded data when the Accept header is “application/json”.

The required JSON structure can be seen in `recordstore.serialization`.

create_project (*project_name*, *long_name*='', *description*='')

Create an empty project in the record store.

project_info (*project_name*)

Return a project's long name and description.

update_project_info (*project_name*, *long_name*='', *description*='')

Update a project's long name and description.

Module functions

`sumatra.recordstore.get_record_store` (*uri*)

Return the RecordStore object found at the given URI (which may be a URL or filesystem path).

Transferring provenance information

Handles serialization/deserialization of record store contents to/from JSON.

`sumatra.recordstore.serialization.encode_record` (*record*, *indent=None*)

Encode a Sumatra record as JSON.

`sumatra.recordstore.serialization.encode_project_info` (*long_name*, *description*)

Encode a Sumatra project as JSON

`sumatra.recordstore.serialization.build_record` (*data*)

Create a Sumatra record from a nested dictionary.

`sumatra.recordstore.serialization.decode_record` (*content*)

Create a Sumatra record from a JSON string.

`sumatra.recordstore.serialization.decode_records` (*content*)

Create multiple Sumatra records from a JSON string.

1.14.4 Version control

The `versioncontrol` sub-package provides an abstraction layer around different revision/version control systems (VCSs). Only the functionality required for recording version numbers and switching the working copy between different versions is wrapped - for more complex tasks such as merging, branching, etc., the version control tool should be used directly.

Repository objects

A `Repository` object represents a version control repository. Its main roles in Sumatra are

1. to contain the information necessary to identify the repository for reproducibility purposes (i.e. its URL)
2. to provide a uniform interface for obtaining a working copy (“checkout” in Subversion parlance, “clone” for Git/Mercurial)

There are four subclasses of the abstract base `Repository` class:

- `SubversionRepository`,
- `MercurialRepository`,
- `GitRepository`,
- `BazaarRepository`.

These subclasses are only available if the appropriate Python bindings for the underlying VCS are installed - see [Installation](#). Each of the subclasses implements the following interface:

class `sumatra.versioncontrol.base.Repository` (*url*, *upstream=None*)
 Represents, and enables limited interaction with, the version control system repository located at *url*.
 If *upstream* is not provided, this information will be obtained, if possible, from the version control system.

url
 The repository URL, generally a local file system path for distributed VCSs.

upstream
 For distributed VCSs, the repository from which the local repository was cloned.

checkout (*path='.'*)
 Clone a repository (“checkout” in Subversion) from *self.url* to the local filesystem at *path*.

exists
 Does the repository represented by this object actually exist?

get_working_copy (*path=None*)
 Return a `WorkingCopy` object corresponding to a checkout of this repository.

Working copy objects

`WorkingCopy` objects provide functionality for inspecting the status of a version control working copy (which files have been modified, what version is currently checked out) and to change the version in use (to repeat previous computations, etc.)

There are four subclasses of the abstract base `WorkingCopy` class:

- `SubversionWorkingCopy`,
- `MercurialWorkingCopy`,
- `GitWorkingCopy`,
- `BazaarWorkingCopy`.

Each of these subclasses implements the following interface:

class `sumatra.versioncontrol.base.WorkingCopy` (*path=None*)
 Represents, and enables limited interaction with, the version control system working copy located in the *path* directory.

If *path* is not specified, the current working directory is assumed.

For each version control system supported by Sumatra, there is a specific subclass of the abstract `WorkingCopy` base class.

contains (*path*)
 Does the repository contain the file with the given path?
 where *path* is relative to the working copy root.

current_version ()
 Return the version of the current state of the working copy.

diff ()
 Return the difference between working copy and repository.

get_username ()
 Return the username and e-mail of the current user, as understood by the version control system, in the format ‘username <e-mail>’.

has_changed ()
 Are there any uncommitted changes to the working copy?

status ()
 Return a dict containing the sets of files that have been modified, added, removed, are missing, not under version control (‘unknown’), are being ignored, or are unchanged (‘clean’).

use_latest_version ()

Switch the working copy to the most recent version.

Any uncommitted changes are retained/merged in.

use_version (version)

Switch the working copy to *version*.

If the working copy has uncommitted changes, raises an *UncommittedModificationsError*.

Functions

It is seldom necessary to create a `Repository` or `WorkingCopy` object directly, or even to know which version control system is in use. Instead, the following functions will return the correct object, given a URL or a filesystem path.

`sumatra.versioncontrol.get_repository (url)`

Return a `Repository` object which represents, and enables limited interaction with, the version control repository at *url*.

If no repository is found at *url*, raises a *VersionControlError*.

`sumatra.versioncontrol.get_working_copy (path=None)`

Return a `WorkingCopy` object which represents, and enables limited interaction with, the version control working copy at *path*.

If *path* is not specified, the current working directory is used. If no working copy is found at *path*, raises a *VersionControlError*.

Exceptions

class `sumatra.versioncontrol.VersionControlError`

class `sumatra.versioncontrol.UncommittedModificationsError`

1.14.5 Formatting output

The formatting module provides classes for formatting simulation/analysis records in different ways: summary, list or table; and in different mark-up formats: currently text or HTML.

Formatting Sumatra records

class `sumatra.formatting.TextFormatter (records)`

Bases: `sumatra.formatting.Formatter`

Format the information from a list of Sumatra records as text.

format (mode='short')

Format a record according to the given mode. mode may be 'short', 'long' or 'table'.

long (text_width=80, left_column_width=17)

Return detailed information about a list of records, as text with a limited column width. Lines that are too long will be wrapped round.

short ()

Return a list of record labels, one per line.

table ()

Return information about a list of records as text, in a simple tabular format.

class `sumatra.formatting.HTMLFormatter` (*records*)

Bases: `sumatra.formatting.Formatter`

Format information about a group of Sumatra records as HTML fragments, to be included in a larger document.

format (*mode='short'*)

Format a record according to the given mode. mode may be 'short', 'long' or 'table'.

long ()

Return detailed information about a list of records as an HTML description list.

short ()

Return a list of record labels as an HTML unordered list.

table ()

Return detailed information about a list of records as an HTML table.

`sumatra.formatting.get_formatter` (*format*)

Return a `Formatter` object of the appropriate type. *format* may be 'text', 'html' or 'textdiff'

Formatting the difference between two records

class `sumatra.formatting.TextDiffFormatter` (*diff*)

Bases: `sumatra.formatting.Formatter`

Format information about the differences between two Sumatra records in text format.

format (*mode='short'*)

Format a record according to the given mode. mode may be 'short', 'long' or 'table'.

long ()

Return a detailed description of the differences between two records.

short ()

Return a summary of the differences between two records.

`sumatra.formatting.get_diff_formatter` ()

Return a `DiffFormatter` object of the appropriate type. Only text format is currently available.

1.14.6 Launching programs

The launch module handles launching of simulations/analyses as sub-processes, and obtaining information about the platform(s) on which the simulations are run.

class `sumatra.launch.SerialLaunchMode` (*working_directory=None*)

Enable running serial computations.

generate_command (*executable, main_file, arguments*)

Return a string containing the command to be launched.

get_platform_information ()

Return a list of `PlatformInformation` objects, containing information about the machine(s) and environment(s) the computations are being performed on/in.

pre_run (*executable*)

Run tasks before the simulation/analysis proper.

run (*executable, main_file, arguments, append_label=None*)

Run a computation in a shell, with the given executable, script and arguments. If *append_label* is provided, it is appended to the command line. Return True if the computation finishes successfully, False otherwise.

```
class sumatra.launch.DistributedLaunchMode (n,          mpirun='mpiexec',      hosts=[],
                                           pfi_path='/usr/local/bin/pfi.py',    work-
                                           ing_directory=None)
```

Enable running distributed computations using MPI.

The current implementation is specific to MPICH2, but this will be generalised in future releases.

```
generate_command (executable, main_file, arguments)
    Return a string containing the command to be launched.
```

```
get_platform_information ()
    Return a list of PlatformInformation objects, containing information about the machine(s) and envi-
    ronment(s) the computations are being performed on/in.
```

Requires the script `pfi.py` to be placed on the user's path on each node of the machine.

This is currently not useful, as I don't think there is any guarantee that we get the same *n* nodes that the command is run on. Need to look more into this.

```
pre_run (executable)
    Run tasks before the simulation/analysis proper.
```

```
run (executable, main_file, arguments, append_label=None)
    Run a computation in a shell, with the given executable, script and arguments. If append_label is
    provided, it is appended to the command line. Return True if the computation finishes successfully,
    False otherwise.
```

```
class sumatra.launch.PlatformInformation (**kwargs)
    A simple container for information about the machine and environment the computations are being per-
    formed on/in.
```

1.14.7 Parameter file formats

The parameters module handles different parameter file formats.

The original idea was that all parameter files will be converted to a single internal parameter format, the NeuroTools ParameterSet class. This will allow fancy searching/comparisons based on parameters. However, we don't do this at the moment, the only methods that are used are *update()* and *save()*

Classes

NTPParameterSet: handles parameter files in the NeuroTools parameter set format, based on nested dictionaries.

SimpleParameterSet: handles parameter files in a simple "name = value" format, with no nesting or grouping.

ConfigParserParameterSet handles parameter files in traditional config file format, as parsed by the standard Python `ConfigParser` module.

JSONParameterSet handles parameter files in JSON format

YAMLParameterSet handles parameter files in YAML format

1.14.8 Providing information about programs

The programs module handles simulator and analysis programs, i.e. executable files, to support the ability to customize Sumatra's behaviour for specific tools.

Classes

Executable represents a generic executable, about which nothing is known except its name. The base class for specific simulator/analysis tool classes.

PythonExecutable represents the Python interpreter executable.

MatlabExecutable represents the Matlab interpreter executable.

NESTSimulator represents the NEST neuroscience simulator.

NEURONSimulator represents the NEURON neuroscience simulator.

GENESSimulator represents the GENESIS neuroscience simulator.

Functions

get_executable() Return an appropriate subclass of Executable, given either the path to an executable file or a script file that can be run with a given tool.

register_executable() Register new subclasses of Executable that can be returned by get_executable().

class `sumatra.programs.Executable` (*path, version=None, options='', name=None*)

Bases: `object`

name = `None`

requires_script = `False`

static write_parameters (*parameters, filebasename*)

`sumatra.programs.get_executable` (*path=None, script_file=None*)

Given the path to an executable, determine what program it is, if possible. Given the name of a script file, try to infer the program that runs that script. Return an appropriate subclass of Executable

`sumatra.programs.register_executable` (*cls, name, executables, extensions*)

Register a new subclass of Executable that can be returned by get_executable().

1.14.9 Managing projects

The projects module defines the `Project` class, which stores information about a computation-based project and contains a number of methods for managing and running computational experiments, whether simulations, analyses or whatever. This is the main class that is used directly when using Sumatra within your own scripts.

class `sumatra.projects.Project` (*name, default_executable=None, default_repository=None, default_main_file=None, default_launch_mode=None, data_store='default', record_store='default', on_changed='error', description='', data_label=None, input_datastore=None, label_generator='timestamp', timestamp_format='%Y%m%d-%H%M%S'*)

Bases: `object`

add_comment (*label, comment*)

add_record (*record*)

Add a simulation or analysis record.

add_tag (*label, tag*)

compare (*label1, label2, ignore_mimetypes=[], ignore_filenames=[]*)

data_label

delete_by_tag (*tag, delete_data=False*)

Delete all records with a given tag. Return the number of records deleted.

delete_record (*label, delete_data=False*)

Delete a record. Return 1 if the record is found. Otherwise return 0.

export ()

format_records (*format='text', mode='short', tags=None*)

get_record (*label*)

Search for a record with the supplied label and return it if found. Otherwise return None.

info ()

Show some basic information about the project.

launch (*parameters*={}, *input_data*=[], *script_args*='', *executable*='default', *repository*='default', *main_file*='default', *version*='latest', *launch_mode*='default', *label*=None, *reason*=None, *timestamp_format*='default')

Launch a new simulation or analysis.

most_recent ()

new_record (*parameters*={}, *input_data*=[], *script_args*='', *executable*='default', *repository*='default', *main_file*='default', *version*='latest', *launch_mode*='default', *label*=None, *reason*=None, *timestamp_format*='default')

remove_tag (*label*, *tag*)

save ()

Save state to some form of persistent storage. (file, database).

show_diff (*label1*, *label2*, *mode*='short', *ignore_mimetypes*=[], *ignore_filenames*=[])

update_code (*working_copy*, *version*='latest')

Check if the working copy has modifications and prompt to commit or revert them.

`sumatra.projects.load_project` (*path*=None)

Read project from directory passed as the argument and return Project object. If no argument is given, the project is read from the current directory.

1.14.10 Handling provenance information

The records module defines the *Record* class, which gathers and stores information about an individual simulation or analysis run.

class `sumatra.records.Record` (*executable*, *repository*, *main_file*, *version*, *launch_mode*, *datastore*, *parameters*={}, *input_data*=[], *script_arguments*='', *label*=None, *reason*='', *diff*='', *user*='', *on_changed*='error', *input_datastore*=None, *stdout_stderr*='Not launched.', *timestamp*=None, *timestamp_format*='%Y%m%d-%H%M%S')

Bases: `object`

The *Record* class has two main roles: capturing information about the context of a computation, and storing this information for later retrieval.

command_line

Return the command-line string for the computation captured by this record.

delete_data ()

Delete any data files associated with this record.

describe (*format*='text', *mode*='long')

Return a description of the record.

mode: either 'long' or 'short'

format either 'text' or 'html'

difference (*other_record*, *ignore_mimetypes*=[], *ignore_filenames*=[])

Determine the difference between this computational experiment and another (code, platform, results, etc.).

Return a RecordDifference object.

register (*working_copy*)

Record information about the environment.

run (*with_label=False*)
Launch the simulation or analysis.

with_label adds the record label either to the parameter file (*with_label="parameters"*) or to the end of the command line (*with_label="cmdline"*), and appends the label to the datastore root. This allows the program being run to create files in a directory specific to this run.

class `sumatra.records.RecordDifference` (*recordA*, *recordB*, *ignore_mimetypes=[]*, *ignore_filenames=[]*)

Bases: `object`

Represents the difference between two Record objects.

code_differs

dependencies_differ

dependency_differences

ignore_filenames = ['\\.log', '^log']

ignore_mimetypes = []

input_data_differ

input_data_differences

launch_mode_differences

output_data_differ

output_data_differences

1.15 Frequently asked questions

1.15.1 Where does the name “Sumatra” come from?

It is based on the initial letters of “Simulation Management Tool”. (Sumatra was originally conceived for tracking simulations, it was only later that I realized it could equally well be used for any command-line driven computation). Despite a certain [geographical proximity](#), it has nothing to do with Java :-)

1.15.2 When I run more than one simulation at once, Sumatra mixes up the output files. How can I make it associate the right files with the right simulation?

When you run a simulation/analysis, Sumatra looks for any new files within your datastore root directory, and associates them with your computation. This means that if you launch a second computation before the first one has finished Sumatra can’t distinguish which files were produced by which computation. The solution is to save the results for a given computation in a subdirectory whose name is a unique id, and for Sumatra to look only in this subdirectory for output files.

The easiest way to do this is to use the record label. First run:

```
$ smt configure --addlabel=cmdline
```

or:

```
$ smt configure --addlabel=parameters
```

Then Sumatra will add the record label (which is generated from the timestamp unless you use the ‘`--label`’ option to `smt run`) to either the command line or the parameter file for your script. It is then up to your script to read this value and use it to name your output files accordingly. Here is an example for a Python script, using the `cmdline` option and `"/Data"` set as the datastore root:

```
import sys
import os.path
options = sys.argv[1:]
label = options[-1] # label is added to the end of the command line

# computations happen here, results stored in `output_data`

output_dir = os.path.join("Data", label)
with open(os.path.join(output_dir, "mydata.txt"), 'w') as fp:
    fp.write(output_data)
```

1.16 Getting support

If you have a question about, or problems with using Sumatra, please post a message on either the [sumatra-users](#) or [neuralensemble](#) Google Groups.

1.17 Release notes

1.17.1 Sumatra 0.5.3 release notes

April 9th 2014

Sumatra 0.5.3 is a minor release that adds support for the ISO datetime format and fixes a problem when trying to move a project to another machine.

1.17.2 Sumatra 0.5.2 release notes

April 2nd 2013

Sumatra 0.5.2 is a bug-fix release. It adds support for Django 1.5, and fixes the following bugs:

- ticket:162 “Can’t use main files not in current directory when using git”
- ticket:163 “If main file is set but executable is unrecognized, main file isn’t run”
- ticket:164 “Command line JSON parameter files which are not intended as ‘parameter files’ get broken”
- uncaught exception in *have_internet_connection()*

1.17.3 Sumatra 0.5.0 release notes

February 18th 2013

Welcome to Sumatra 0.5.0!

Overview

Sumatra 0.5 development has mostly been devoted to polishing. There were a bunch of small improvements, with contributions from several new contributors. The [Bitbucket](#) pull request workflow seemed to work well for this. The main changes are:

- working directory now captured (as a parameter of `LaunchMode`);
- data differences are now based on content, not name, i.e. henceforth two files with identical content but different names (e.g. because the name contains a timestamp) will evaluate as being the same;
- improved error messages when a required version control wrapper is not installed;

- dependencies now capture the source from which the version was obtained (e.g. repository url);
- YAML-format parameter files are now supported (thanks to Tristan Webb);
- added “upstream” attribute to the `Repository` class, which may contain the URL of the repository from which your local repository was cloned;
- added `MirroredFileSystemDataStore`, which supports the case where files exist both on the local filesystem and on some web server (e.g. DropBox);
- the name/e-mail of the user who launched the computation is now captured (first trying `~/ .smt.rc`, then the version control system);
- there is now a choice of methods for auto-generating labels when they are not supplied by the user: timestamp-based (the default and previously the only option) and uuid-based. Use the “-g” option to **smt configure**;
- you can also specify the timestamp format to use (thanks to Yoav Ram);
- improved API reference documentation.

Interfaces to documentation systems

The one big addition to Sumatra is a set of tools to include figures and other results generated by Sumatra-tracked computations in documents, with links to full provenance information: i.e. the full details of the code, input data and computational environment used to generate the figure/result.

The following tools are available:

- for reStructuredText/Sphinx: an “smtlink” role and “smtimage” directive.
- for LaTeX, a “sumatra” package, which provides the “\smtincludegraphics” command.

see [Reproducible publications: including and linking to provenance information in documents](#) for more details.

Bug fixes

A [handful of bugs](#) have been fixed.

1.17.4 Sumatra 0.4.0 release notes

October 18th 2012

Welcome to Sumatra 0.4.0!

Overview

The biggest change in Sumatra 0.4 is the redesign of the browser-based interface, launched with **smtweb**. Thanks to the [Google Summer of Code](#) program, Dmitry Samarkanov was able to spend his summer working on improving Sumatra, with the results being a much improved web interface, better support for running Sumatra on Windows, and better support for running Matlab scripts with Sumatra. Many thanks to Google and to the [INCF](#) as mentoring organisation. In addition to Dmitry’s improvements, handling of input and output data files is much improved, and Sumatra now captures and stores standard output (stdout) and standard error (stderr) streams. More details on all of these, plus a bunch of minor improvements and bug fixes, is given below. Finally, Sumatra no longer supports Python 2.5 - the minimum requirement is Python 2.6.

Web interface

The Sumatra browser-based interface runs a local webserver on your computer, and allows you to browse the information that Sumatra captures about your analyses, simulations or other computations, including code versions, input and output data files, parameter/configuration files, the operating system and processor architecture.

The interface has been completely redesigned for Sumatra 0.4, and includes dozens of large and small improvements, including:

- a more modern, attractive design
- the ability to select which columns to display in the record list view
- the ability to search all of your records based on date, tags or full-text
- side-by-side comparison of records
- sorting of records based on any column
- selection of multiple records by clicking or dragging for deletion, comparison and tagging

Furthermore, it is now possible to launch computations from the browser interface.

For more information, see [Using the web interface](#) and [this blog post](#) from Dmitry Samarkanov.

Data file handling

In earlier versions of Sumatra, the filename (or rather, the file path relative to a user-defined root) was used as the identifier for input and output data files. The problem with this, of course, is that it is possible to overwrite a given file with new data. For this reason, Sumatra 0.4 now calculates and stores the SHA1 hash of the file contents. If the file contents change, the hash will also change, so that Sumatra can alert you if a file is accidentally overwritten, for example.

Sumatra 0.4 also adds a new data store which automatically archives a copy of the output data from your computations in a user-selected location. This data store is accessible through the API as the `ArchivingFileSystemDataStore` class, or through the `smt` command-line interface with the “`-archive`” option to the “`init`” and “`configure`” commands.

Finally, Sumatra now allows the user the choice of whether to use an absolute or relative path for the data store root directory. Using a relative path makes projects easier to move and easier to access from other locations (e.g. with symbolic links or NFS).

Matlab support

Sumatra can capture certain information for *any* command-line tool: input and output data, version of the main codebase, operating system and processor architecture, etc. For dependency information, however (i.e. which libraries, modules or packages are imported/included by your main script), a separate plugin is required for each language. Sumatra already has a dependency tracking plugin for Python and for two computational neuroscience simulation environments, NEURON and GENESIS. Sumatra 0.4 adds dependency tracking for Matlab scripts.

Recording of stdout and stderr

Sumatra 0.4 now supports recording and storage of the standard output and standard error streams from your scripts.

Other new features

- added support for JSON-format parameter files;
- added `smt export` command, which allows the contents of a Sumatra record store to be exported in JSON format;

- more information is now printed by `smt list --long`;
- the Python dependency finder now supports scripts run with Python 3 (although Sumatra itself still needs Python 2);
- can now specify `HttpRecordStore` username and password as part of the URL passed to `smt init`;
- added support for markup using `reStructuredText` in the project description
- it is no longer required to have a script file, which makes it possible to use Sumatra with your own compiled executables. Further support for compiled languages is planned for the next release.

Bug fixes

A whole bunch of bugs were fixed in Sumatra 0.4.

1.18 Authors and contributors

The following people have contributed code to Sumatra. The institutional affiliations are those at the time of the contribution, and may not be the current affiliation of a contributor.

- Andrew Davison [1]
- Dmitry Samarkanov [2]
- Bartosz Telenczuk [1, 3]
- Michele Mattioni [4]
- Eilif Muller [5]
- Konrad Hinsén [6]
- Stephan Gabler [7]
- Takafumi Arakaki [8]
- Yoav Ram [9]
- Tristan Webb [10]
- Maximilian Albert [11]

1. Unité de Neurosciences, Information et Complexité, CNRS UPR 3293, Gif-sur-Yvette, France

2. Ecole Centrale de Lille, Lille, France

3. Institute for Theoretical Biology, Humboldt University zu Berlin, Berlin, Germany

4. European Bioinformatics Institute, Hinxton, UK

5. Blue Brain Project, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

6. Centre de biophysique moléculaire, CNRS UPR 4301, Orléans, France

7. Max Planck Institute for Human Development, Berlin, Germany

8. Laboratoire de Neurophysique et Physiologie, CNRS UMR 8119, Université Paris Descartes, Paris, France

9. ?

10. Warwick University, UK

11. ?

If I've somehow missed you off the list I'm very sorry - please let us know.

Many thanks also go to everyone who has reported bugs on the issue tracker.

1.18.1 Licence

Sumatra is freely available under the CeCILL v2 license, which is equivalent to, and compatible with, the GNU GPL license, but conforms to French law (and is also perfectly suited to international projects) - see <http://www.cecill.info/index.en.html> for more information.

If you are interested in using Sumatra, but the choice of licence is a problem for you, please contact us - we are open to persuasion.

S

`sumatra.dependency_finder`, 32
`sumatra.formatting`, 38
`sumatra.launch`, 39
`sumatra.parameters`, 40
`sumatra.programs`, 40
`sumatra.recordstore.serialization`, 36
`sumatra.versioncontrol`, 36

A

add_comment() (sumatra.projects.Project method), 41

add_record() (sumatra.projects.Project method), 41

add_tag() (sumatra.projects.Project method), 41

archive_store (sumatra.datastore.ArchivingFileSystemDataStore attribute), 32

ArchivedDataFile (class in sumatra.datastore.archivingfs), 32

ArchivingFileSystemDataStore (class in sumatra.datastore), 32

B

BaseDependency (class in sumatra.dependency_finder.core), 32

build_record() (in module sumatra.recordstore.serialization), 36

C

checkout() (sumatra.versioncontrol.base.Repository method), 37

code_differs (sumatra.records.RecordDifference attribute), 43

command_line (sumatra.records.Record attribute), 42

compare() (sumatra.projects.Project method), 41

contains() (sumatra.versioncontrol.base.WorkingCopy method), 37

contains_path() (sumatra.datastore.base.DataStore method), 31

copy() (sumatra.datastore.base.DataStore method), 31

create_project() (sumatra.recordstore.HttpRecordStore method), 35

current_version() (sumatra.versioncontrol.base.WorkingCopy method), 37

D

data_label (sumatra.projects.Project attribute), 41

DataFile (class in sumatra.datastore.filesystem), 31

DataItem (class in sumatra.datastore.base), 30

DataKey (class in sumatra.datastore), 30

DataStore (class in sumatra.datastore.base), 31

decode_record() (in module sumatra.recordstore.serialization), 36

decode_records() (in module sumatra.recordstore.serialization), 36

delete() (sumatra.datastore.base.DataStore method), 31

delete() (sumatra.recordstore.base.RecordStore method), 34

delete_all() (sumatra.recordstore.base.RecordStore method), 34

delete_by_tag() (sumatra.projects.Project method), 41

delete_by_tag() (sumatra.recordstore.base.RecordStore method), 34

delete_data() (sumatra.records.Record method), 42

delete_record() (sumatra.projects.Project method), 41

dependencies_differ (sumatra.records.RecordDifference attribute), 43

dependency_differences (sumatra.records.RecordDifference attribute), 43

describe() (sumatra.records.Record method), 42

diff() (sumatra.versioncontrol.base.WorkingCopy method), 37

difference() (sumatra.records.Record method), 42

digest (sumatra.datastore.base.DataItem attribute), 30

digest (sumatra.datastore.DataKey attribute), 30

DistributedLaunchMode (class in sumatra.launch), 39

DjangoRecordStore (class in sumatra.recordstore), 35

E

encode_project_info() (in module sumatra.recordstore.serialization), 36

encode_record() (in module sumatra.recordstore.serialization), 36

Executable (class in sumatra.programs), 41

exists (sumatra.versioncontrol.base.Repository attribute), 37

export() (sumatra.projects.Project method), 41

export() (sumatra.recordstore.base.RecordStore method), 34

extension (sumatra.datastore.DataFile attribute), 31

F

FileSystemDataStore (class in sumatra.datastore), 31

find_dependencies() (in module sumatra.dependency_finder), 32

find_dependencies() (in module sumatra.dependency_finder.genesis), 34

- [find_dependencies\(\)](#) (in module `sumatra.dependency_finder.matlab`), 33
[find_dependencies\(\)](#) (in module `sumatra.dependency_finder.neuron`), 34
[find_dependencies\(\)](#) (in module `sumatra.dependency_finder.python`), 33
[find_file\(\)](#) (in module `sumatra.dependency_finder.core`), 33
[find_imported_packages\(\)](#) (in module `sumatra.dependency_finder.python`), 33
[find_included_files\(\)](#) (in module `sumatra.dependency_finder.genesis`), 34
[find_loaded_files\(\)](#) (in module `sumatra.dependency_finder.neuron`), 34
[find_new_data\(\)](#) (`sumatra.datastore.base.DataStore` method), 31
[find_versions\(\)](#) (in module `sumatra.dependency_finder.core`), 33
[find_versions_by_attribute\(\)](#) (in module `sumatra.dependency_finder.python`), 33
[find_versions_from_egg\(\)](#) (in module `sumatra.dependency_finder.python`), 33
[find_versions_from_versioncontrol\(\)](#) (in module `sumatra.dependency_finder.core`), 33
[find_xopened_files\(\)](#) (in module `sumatra.dependency_finder.neuron`), 34
[format\(\)](#) (`sumatra.formatting.HTMLFormatter` method), 39
[format\(\)](#) (`sumatra.formatting.TextDiffFormatter` method), 39
[format\(\)](#) (`sumatra.formatting.TextFormatter` method), 38
[format_records\(\)](#) (`sumatra.projects.Project` method), 41
[full_path](#) (`sumatra.datastore.DataFile` attribute), 31
- ## G
- [generate_command\(\)](#) (`sumatra.launch.DistributedLaunchMode` method), 40
[generate_command\(\)](#) (`sumatra.launch.SerialLaunchMode` method), 39
[generate_key\(\)](#) (`sumatra.datastore.base.DataItem` method), 30
[generate_keys\(\)](#) (`sumatra.datastore.base.DataStore` method), 31
[get\(\)](#) (`sumatra.recordstore.base.RecordStore` method), 34
[get_content\(\)](#) (`sumatra.datastore.base.DataItem` method), 30
[get_content\(\)](#) (`sumatra.datastore.base.DataStore` method), 31
[get_data_item\(\)](#) (`sumatra.datastore.base.DataStore` method), 31
[get_diff_formatter\(\)](#) (in module `sumatra.formatting`), 39
[get_executable\(\)](#) (in module `sumatra.programs`), 41
[get_formatter\(\)](#) (in module `sumatra.formatting`), 39
- [get_platform_information\(\)](#) (`sumatra.launch.DistributedLaunchMode` method), 40
[get_platform_information\(\)](#) (`sumatra.launch.SerialLaunchMode` method), 39
[get_record\(\)](#) (`sumatra.projects.Project` method), 41
[get_record_store\(\)](#) (in module `sumatra.recordstore`), 36
[get_repository\(\)](#) (in module `sumatra.versioncontrol`), 38
[get_username\(\)](#) (`sumatra.versioncontrol.base.WorkingCopy` method), 37
[get_working_copy\(\)](#) (in module `sumatra.versioncontrol`), 38
[get_working_copy\(\)](#) (`sumatra.versioncontrol.base.Repository` method), 37
- ## H
- [has_changed\(\)](#) (`sumatra.versioncontrol.base.WorkingCopy` method), 37
[has_project\(\)](#) (`sumatra.recordstore.base.RecordStore` method), 34
[HTMLFormatter](#) (class in `sumatra.formatting`), 38
[HttpRecordStore](#) (class in `sumatra.recordstore`), 35
- ## I
- [ignore_filenames](#) (`sumatra.records.RecordDifference` attribute), 43
[ignore_mimetypes](#) (`sumatra.records.RecordDifference` attribute), 43
[import_\(\)](#) (`sumatra.recordstore.base.RecordStore` method), 34
[info\(\)](#) (`sumatra.projects.Project` method), 42
[input_data_differ](#) (`sumatra.records.RecordDifference` attribute), 43
[input_data_differences](#) (`sumatra.records.RecordDifference` attribute), 43
- ## L
- [labels\(\)](#) (`sumatra.recordstore.base.RecordStore` method), 34
[launch\(\)](#) (`sumatra.projects.Project` method), 42
[launch_mode_differences](#) (`sumatra.records.RecordDifference` attribute), 43
[list\(\)](#) (`sumatra.recordstore.base.RecordStore` method), 34
[list_projects\(\)](#) (`sumatra.recordstore.base.RecordStore` method), 35
[load_project\(\)](#) (in module `sumatra.projects`), 42
[long\(\)](#) (`sumatra.formatting.HTMLFormatter` method), 39
[long\(\)](#) (`sumatra.formatting.TextDiffFormatter` method), 39
[long\(\)](#) (`sumatra.formatting.TextFormatter` method), 38

M

metadata (sumatra.datastore.DataKey attribute), 30
 mimetype (sumatra.datastore.DataFile attribute), 31
 mirror_base_url (sumatra.datastore.MirroredFileSystemDataStore attribute), 32
 MirroredDataFile (class in sumatra.datastore.mirroredfs), 32
 MirroredFileSystemDataStore (class in sumatra.datastore), 32
 most_recent() (sumatra.projects.Project method), 42
 most_recent() (sumatra.recordstore.base.RecordStore method), 35

N

name (sumatra.datastore.DataFile attribute), 31
 name (sumatra.programs.Executable attribute), 41
 new_record() (sumatra.projects.Project method), 42

O

output_data_differ (sumatra.records.RecordDifference attribute), 43
 output_data_differences (sumatra.records.RecordDifference attribute), 43

P

path (sumatra.datastore.DataFile attribute), 31
 path (sumatra.datastore.DataKey attribute), 30
 PlatformInformation (class in sumatra.launch), 40
 pre_run() (sumatra.launch.DistributedLaunchMode method), 40
 pre_run() (sumatra.launch.SerialLaunchMode method), 39
 Project (class in sumatra.projects), 41
 project_info() (sumatra.recordstore.HttpRecordStore method), 36

R

Record (class in sumatra.records), 42
 RecordDifference (class in sumatra.records), 43
 RecordStore (class in sumatra.recordstore.base), 34
 register() (sumatra.records.Record method), 42
 register_executable() (in module sumatra.programs), 41
 remove_tag() (sumatra.projects.Project method), 42
 Repository (class in sumatra.versioncontrol.base), 36
 requires_script (sumatra.programs.Executable attribute), 41
 root (sumatra.datastore.FileSystemDataStore attribute), 31
 run() (sumatra.launch.DistributedLaunchMode method), 40
 run() (sumatra.launch.SerialLaunchMode method), 39
 run() (sumatra.records.Record method), 42

S

save() (sumatra.projects.Project method), 42

save() (sumatra.recordstore.base.RecordStore method), 35
 save_copy() (sumatra.datastore.base.DataItem method), 30
 save_dependencies() (in module sumatra.dependency_finder.matlab), 33
 SerialLaunchMode (class in sumatra.launch), 39
 ShelveRecordStore (class in sumatra.recordstore), 35
 short() (sumatra.formatting.HTMLFormatter method), 39
 short() (sumatra.formatting.TextDiffFormatter method), 39
 short() (sumatra.formatting.TextFormatter method), 38
 show_diff() (sumatra.projects.Project method), 42
 size (sumatra.datastore.DataFile attribute), 31
 sorted_content() (sumatra.datastore.base.DataItem method), 31
 status() (sumatra.versioncontrol.base.WorkingCopy method), 37
 sumatra.dependency_finder (module), 32
 sumatra.formatting (module), 38
 sumatra.launch (module), 39
 sumatra.parameters (module), 40
 sumatra.programs (module), 40
 sumatra.recordstore.serialization (module), 36
 sumatra.versioncontrol (module), 36
 sync() (sumatra.recordstore.base.RecordStore method), 35
 sync_all() (sumatra.recordstore.base.RecordStore method), 35

T

table() (sumatra.formatting.HTMLFormatter method), 39
 table() (sumatra.formatting.TextFormatter method), 38
 TextDiffFormatter (class in sumatra.formatting), 39
 TextFormatter (class in sumatra.formatting), 38

U

UncommittedModificationsError (class in sumatra.versioncontrol), 38
 update_code() (sumatra.projects.Project method), 42
 update_project_info() (sumatra.recordstore.HttpRecordStore method), 36
 upstream (sumatra.versioncontrol.Repository attribute), 37
 url (sumatra.versioncontrol.Repository attribute), 37
 use_latest_version() (sumatra.versioncontrol.base.WorkingCopy method), 37
 use_version() (sumatra.versioncontrol.base.WorkingCopy method), 38

V

VersionControlError (class in sumatra.versioncontrol), 38

W

WorkingCopy (class in `sumatra.versioncontrol.base`),
37

`write_parameters()` (`sumatra.programs.Executable`
static method), 41