

# AUTOSUBMIT: A VERSATILE TOOL TO MANAGE WEATHER AND CLIMATE EXPERIMENTS IN DIVERSE SUPERCOMPUTING ENVIRONMENTS

D. Manubens<sup>(1)</sup>, J. Vegas<sup>(1)</sup>, F. Doblas-Reyes<sup>(1,2)</sup>

<sup>(1)</sup> Barcelona Supercomputing Center (BSC) <sup>(2)</sup> Institutió Catalana de Recerca i Estudis Avançats (ICREA)

## 1. What is Autosubmit ?

Autosubmit is a python-based tool for submitting, managing and monitoring weather and climate simulations. It was developed at BSC to fill an uncovered need at the moment: allow scientists to run experiments easily on different supercomputers and to reduce the time they spend monitoring model runs.

Autosubmit is currently used at BSC to run EC-Earth and NEMO climate models and to run WRF weather model and NMMB/BSC-CTM air quality model.

Autosubmit has been used to manage models running at supercomputers in IC3, BSC, CERFACS, ECMWF, EPCC, PDC, SMHI and OLCF.

Autosubmit 3.7 version is now available via PyPi package under the terms of GNU General Public License.

## 2. Why is Autosubmit needed ?

Autosubmit is the only existing tool that satisfies the following requirements from the weather and climate community:

- **Automatisation:** Job submission to machines and dependencies between jobs are managed by Autosubmit. No user intervention is needed.
- **Data provenance:** Assigns unique identifiers for each experiment and stores information about model version, experiment configuration and computing facilities used in the whole process.
- **Failure tolerance:** Automatic retrials and ability to rerun chunks in case of corrupted or missing data.
- **Resource management:** Autosubmit manages supercomputer particularities, allowing users to run their experiments in the available machine without having to adapt the code. Autosubmit also allows to submit tasks from the same experiment to different platforms.

## 3. How does Autosubmit work ?

### Experiment creation

To create a new experiment, run the command "autosubmit expid -h HPCname -d Description". HPCname is the name of the main HPC platform for the experiment: it will be the default platform for the tasks. Description is a brief experiment description.

This command assigns a unique four character identifier (xxxx) to the experiment and creates a new folder in experiments repository with structure shown in Figure 1.

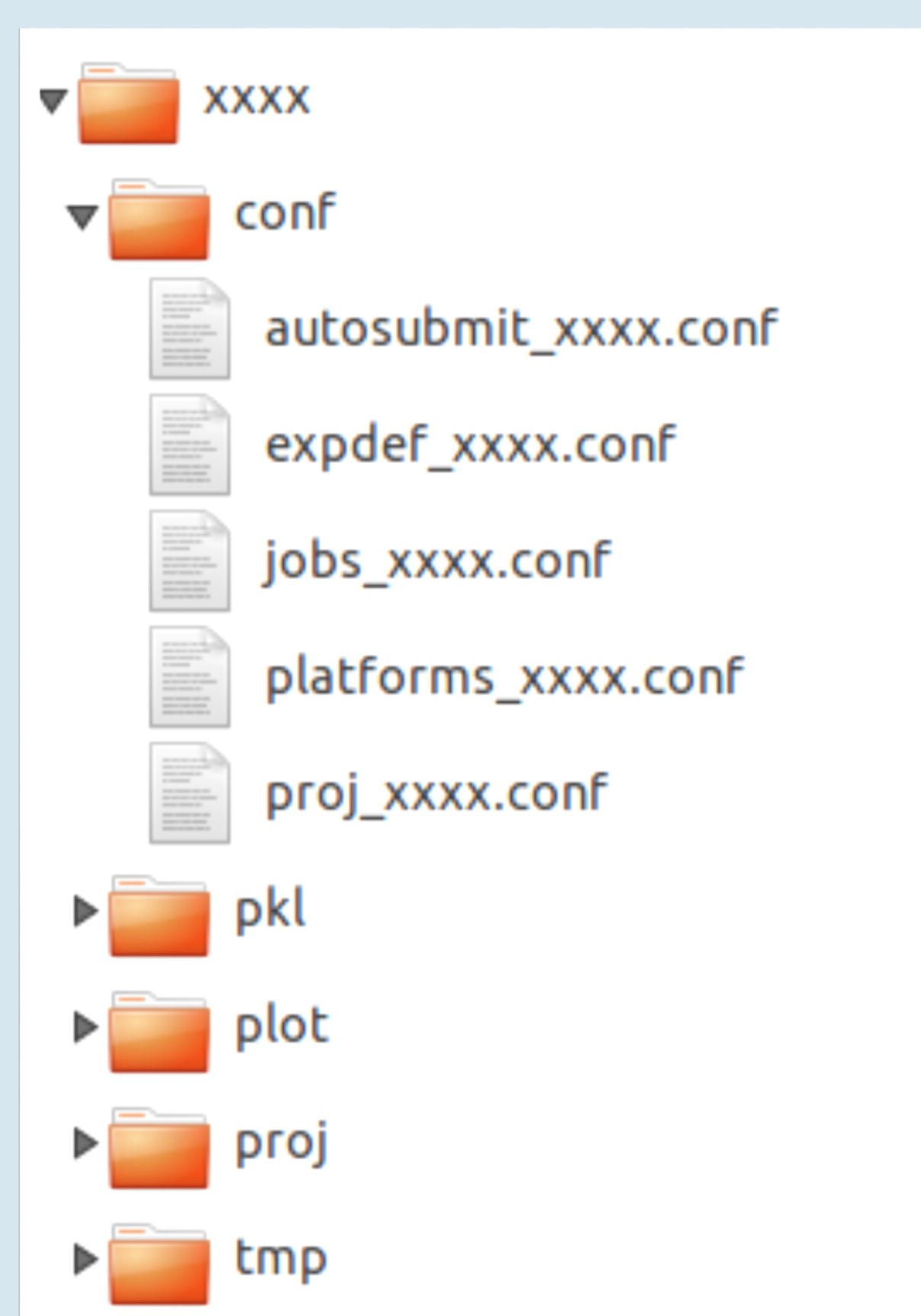
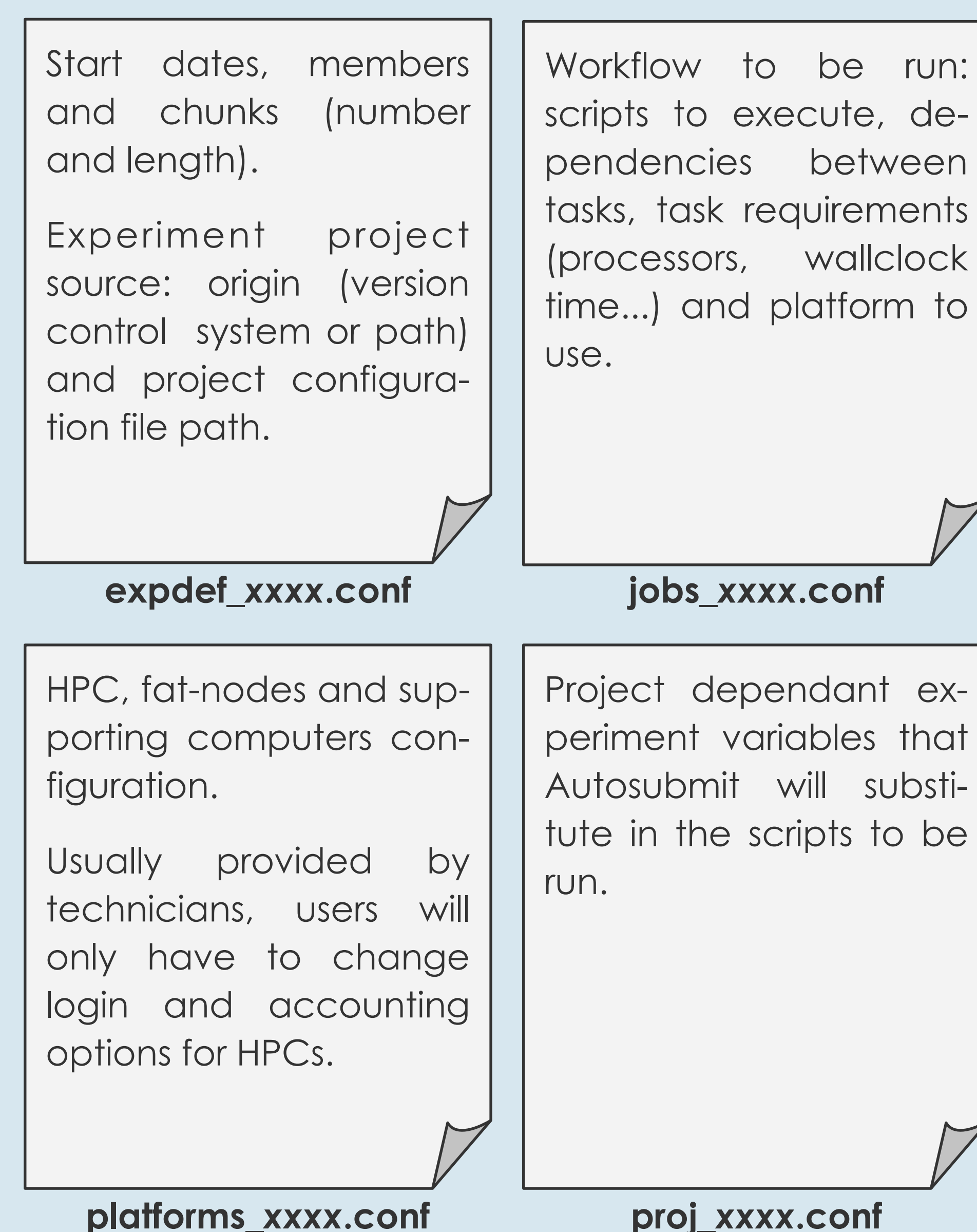


Figure 1: Example of an experiment directory tree.

### Experiment configuration

To configure the experiment, edit expdef\_xxxx.conf, jobs\_xxxx.conf and platforms\_xxxx.conf in the conf folder of the experiment.

After that, the user is expected to run "autosubmit create xxxx". This command creates the experiment project in the proj folder. The experiment project contains the scripts specified in jobs\_xxxx.conf and a copy of model source code and data specified in expdef\_xxxx.conf.



### Experiment run

To run the experiment, just execute the command "autosubmit run xxxx". Autosubmit will start submitting jobs to the relevant platforms (both HPC and supporting computers) by using the scripts specified in jobs\_xxxx.conf. Autosubmit will substitute variables present on scripts where handlers appear in % variable\_name% format. Autosubmit provides variables for current chunk, start date, member, computer configuration and more., and also will replace variables from proj\_xxxx.conf.

To monitor the status of the experiment, the command "autosubmit monitor xxxx" is available. This will plot the workflow of the experiment and the current status.

## 4. Autosubmit and EC-Earth

Autosubmit is used at BSC to run EC-Earth. To do that, a git repository has been created that contains the model source code and the scripts used to run the tasks. The workflow is defined using seven job types, as shown in figure 2. These job types are:

- **Local\_setup:** prepares a patch for model changes and copies it to HPC
- **Remote\_setup:** creates a model copy and applies the patch to it
- **Ini:** prepares model to start the simulation of one member

- **Sim:** runs a simulation chunk (usually 1 to 3 months)
- **Post:** post-process and CMORize outputs for one simulation chunk
- **Clean:** removes unnecessary outputs from the simulated chunk
- **Transfer:** transfers post-processed outputs to definitive storage

## 5. Future work

Plans for further development of Autosubmit include:

- Expand metadata stored for CMIP6 and use pyesdoc (<http://es-doc.org>).
- Create a wrapper for jobs that allows to send various independent tasks as one big job to reduce queueing time and to allow participation in tier-0 projects.
- Evaluate the suitability of using ecFlow (<http://software.ecmwf.int/wiki/display/ECFLOW>) for workflow configuration and as Graphical User Interface.

## Acknowledgements

The work leading to these results has received funding from the EU Seventh Framework Programme FP7(2007-2013) under grant agreements n° 308378 (SPECS) and n° 312979 (IS-ENES2), and the Catalan Government.