



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

Autosubmit: Investigando el clima con Python

Javier Vegas – Domingo Manubens



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

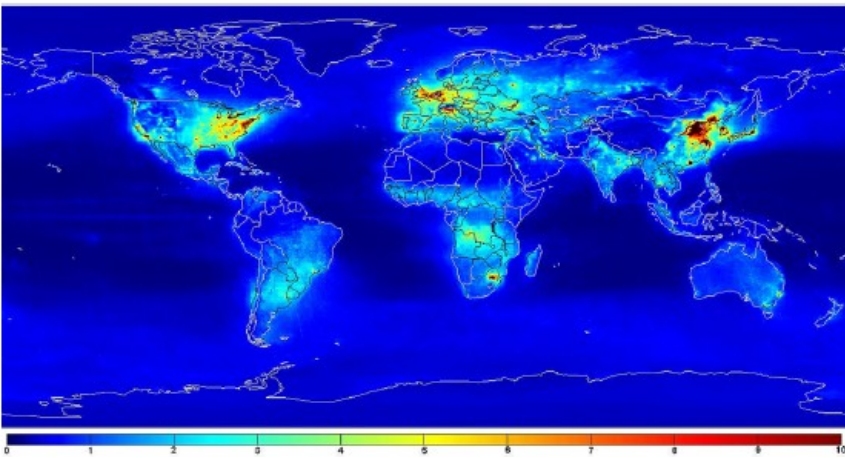
BSC - Earth Sciences Department

- Created in 2005; 350 employees.
- Research, develop and manage information technology.
- Facilitate scientific progress and its application in society.

Earth Sciences Department

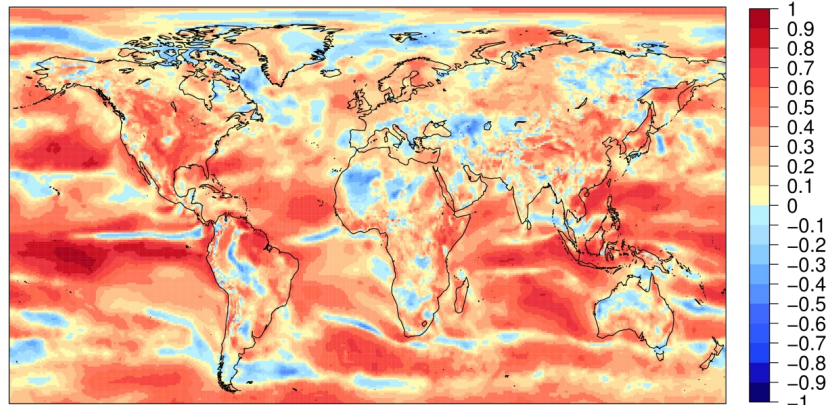
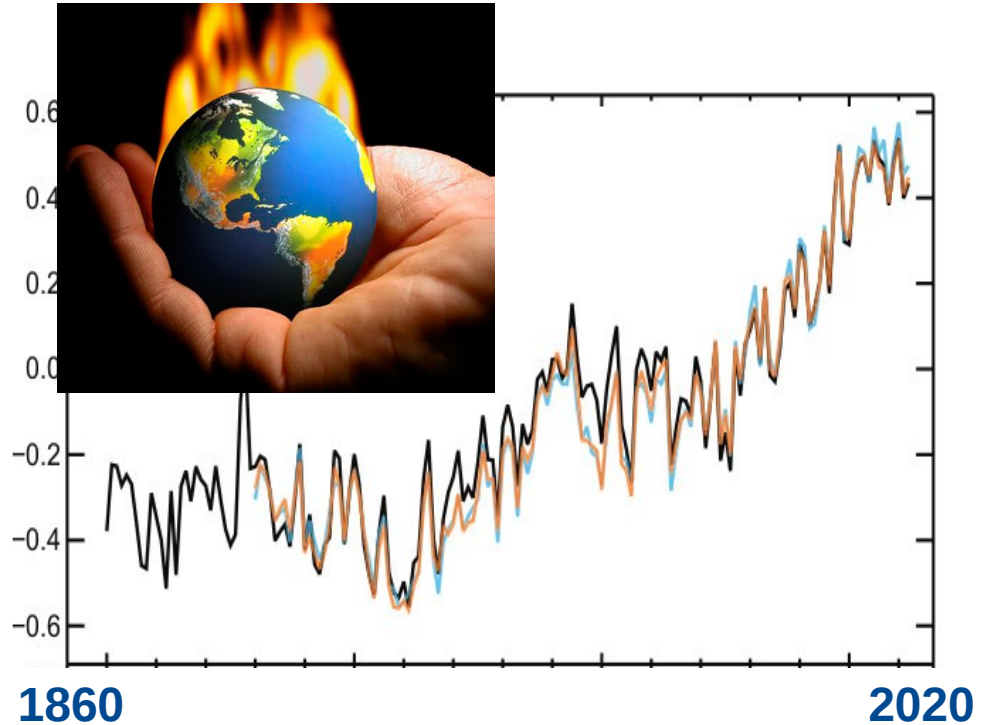
- Atmospheric composition modelling.
- Climate prediction modelling.
- Computational Earth Sciences.**
- Earth Sciences Services.





Calidad del aire

Simulación del cambio climático



Predicción estacional del viento





**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Introducción

¿Cómo investigar el clima?

- « Es inviable experimentar con el sistema climático.
- « Hay que utilizar modelos.

Nuestro modelo de predicción climática:



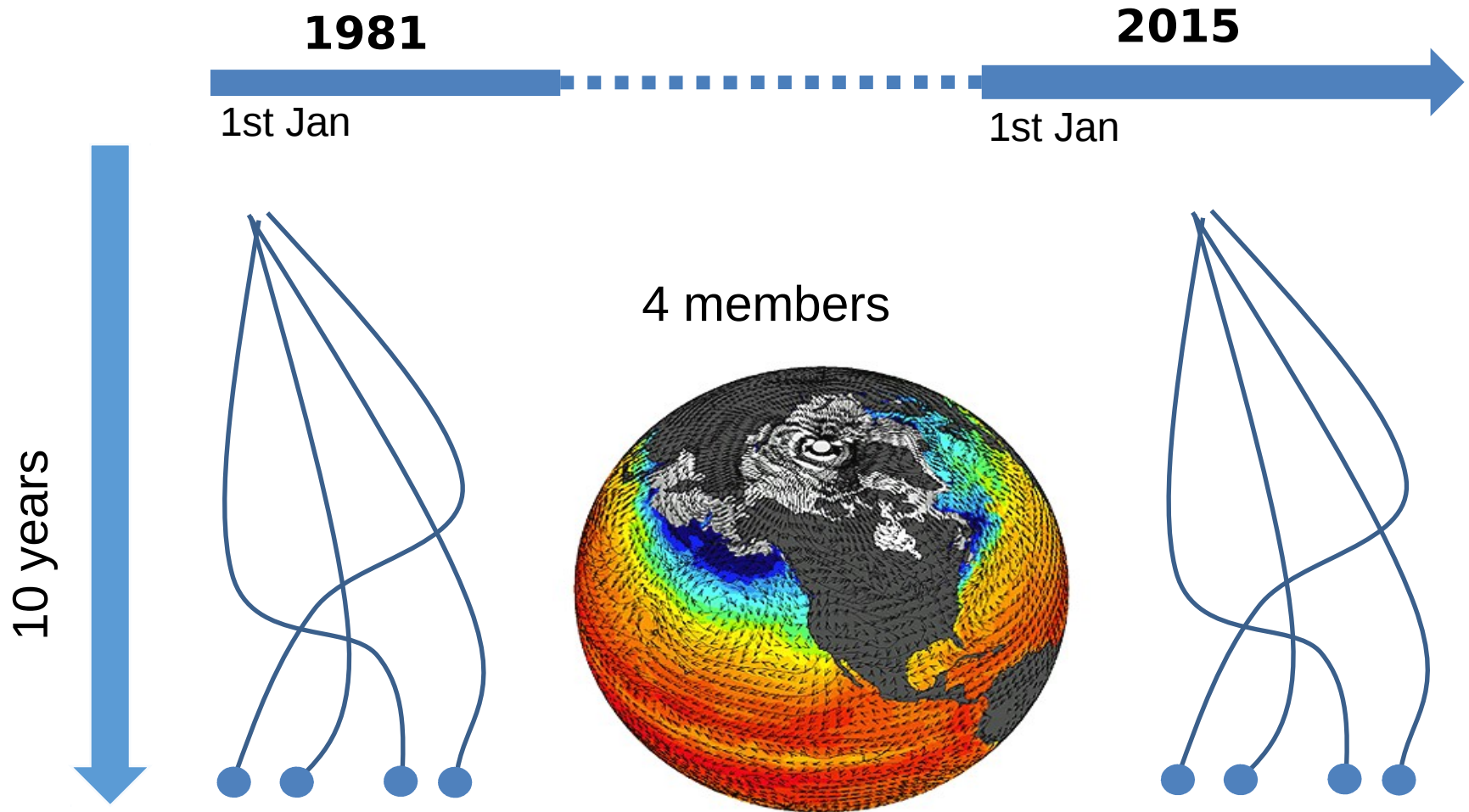
- ⌋ Entre 75 y 2256 cores por ejecución.
- ⌋ Entre 3 y 18 horas para simular un año.
- ⌋ Entre 10 y 100 GB generados por año de simulación

¿Cual es el problema? (1/3)

Un experimento típico no es una simulación lineal de X años desde una fecha Y:

- Se realizan simulaciones desde varias fechas de inicio (startdates).
- Para cada fecha de inicio se realizan varias simulaciones (miembros).
- Para una previsión decadal, cada simulacion debe alcanzar al menos 10 años.

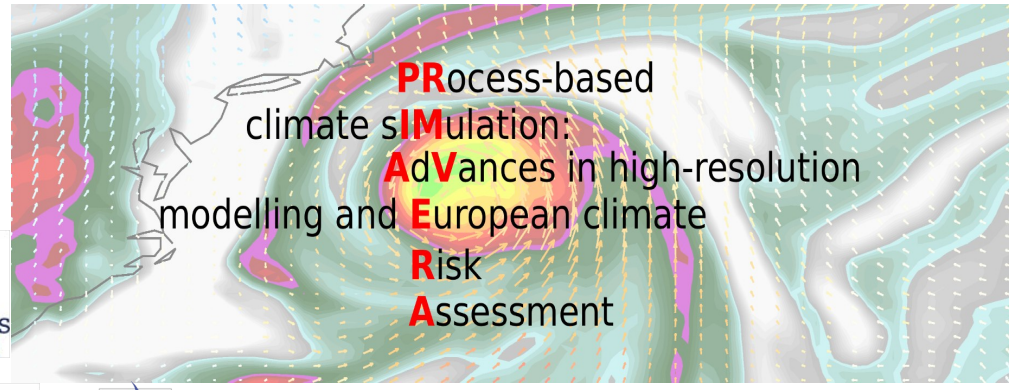
¿Cuál es el problema? (2/3)



¿Cuál es el problema? (3/3)



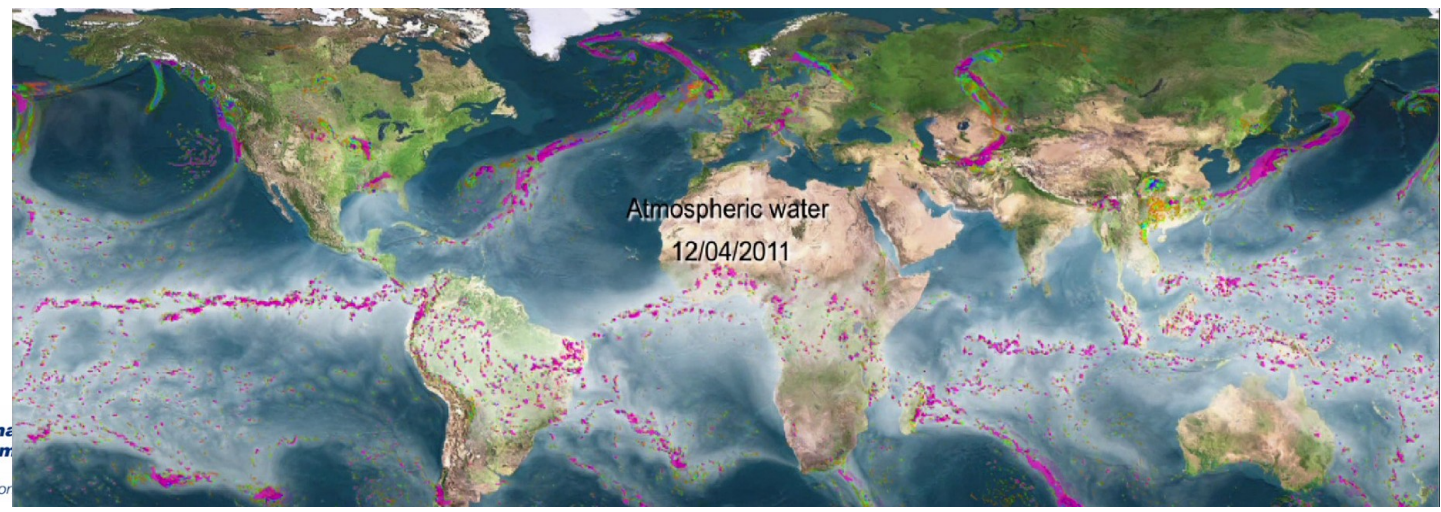
UNIVERSITY OF LEEDS



Max-Planck-Institut
f r Meteorologie



“To develop a new generation of advanced and well-evaluated high-resolution global climate models, capable of simulating and predicting regional climate with unprecedented fidelity, for the benefit of governments,



¿Cual es la solución?

« Correr en varias máquinas:

- Conseguir horas de cálculo con: RES, PRACE, INCITE.



Copyright 2013. Barcelona Supercomputing Center - BSC



Otras máquinas, otros mundos:

« Diferentes schedulers.

« Diferentes políticas de acceso.

- Imposibilidad de conectarse al exterior.
- Imposibilidad de instalar software.

« Falta de almacenamiento.



**Barcelona
Supercomputing
Center**

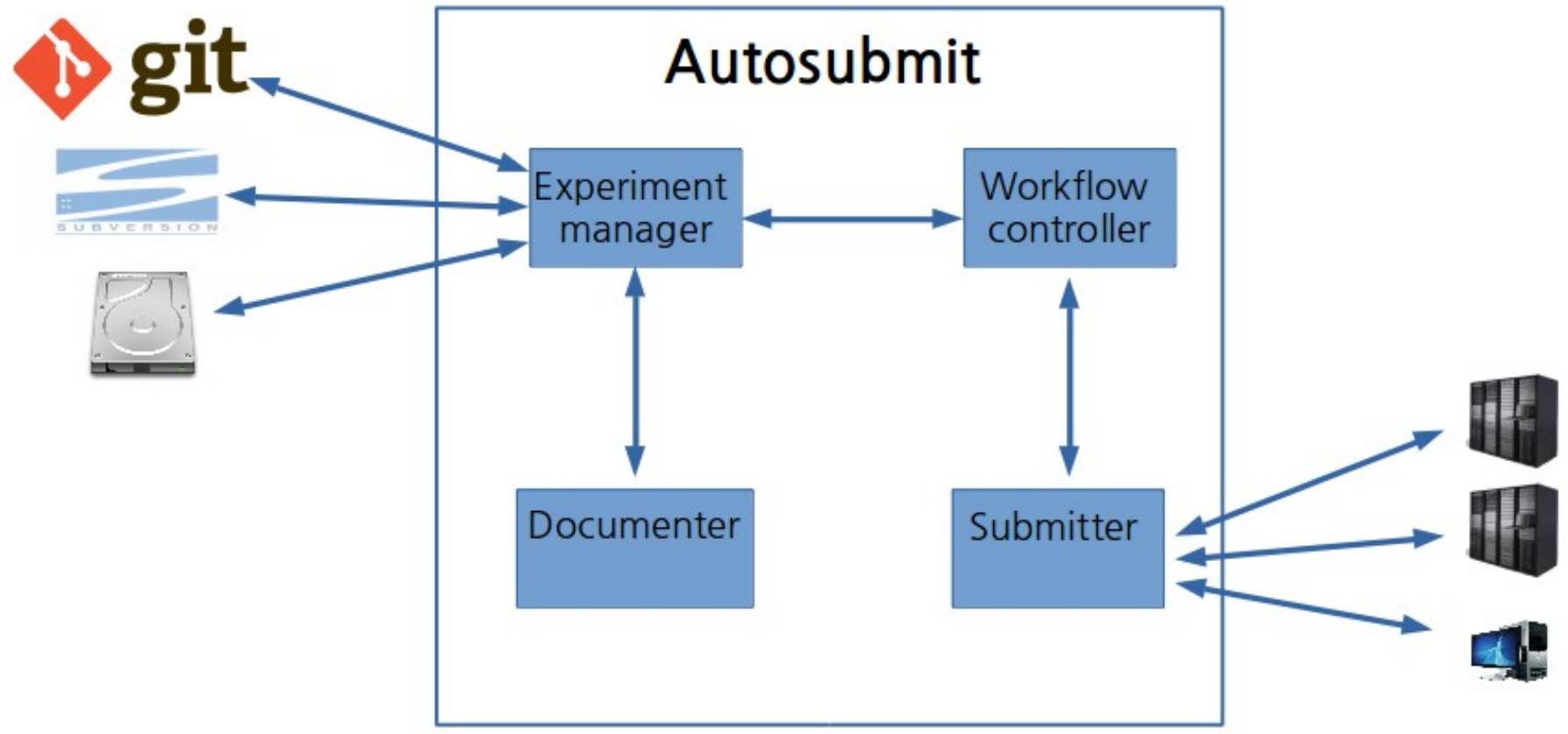
Centro Nacional de Supercomputación

Autosubmit

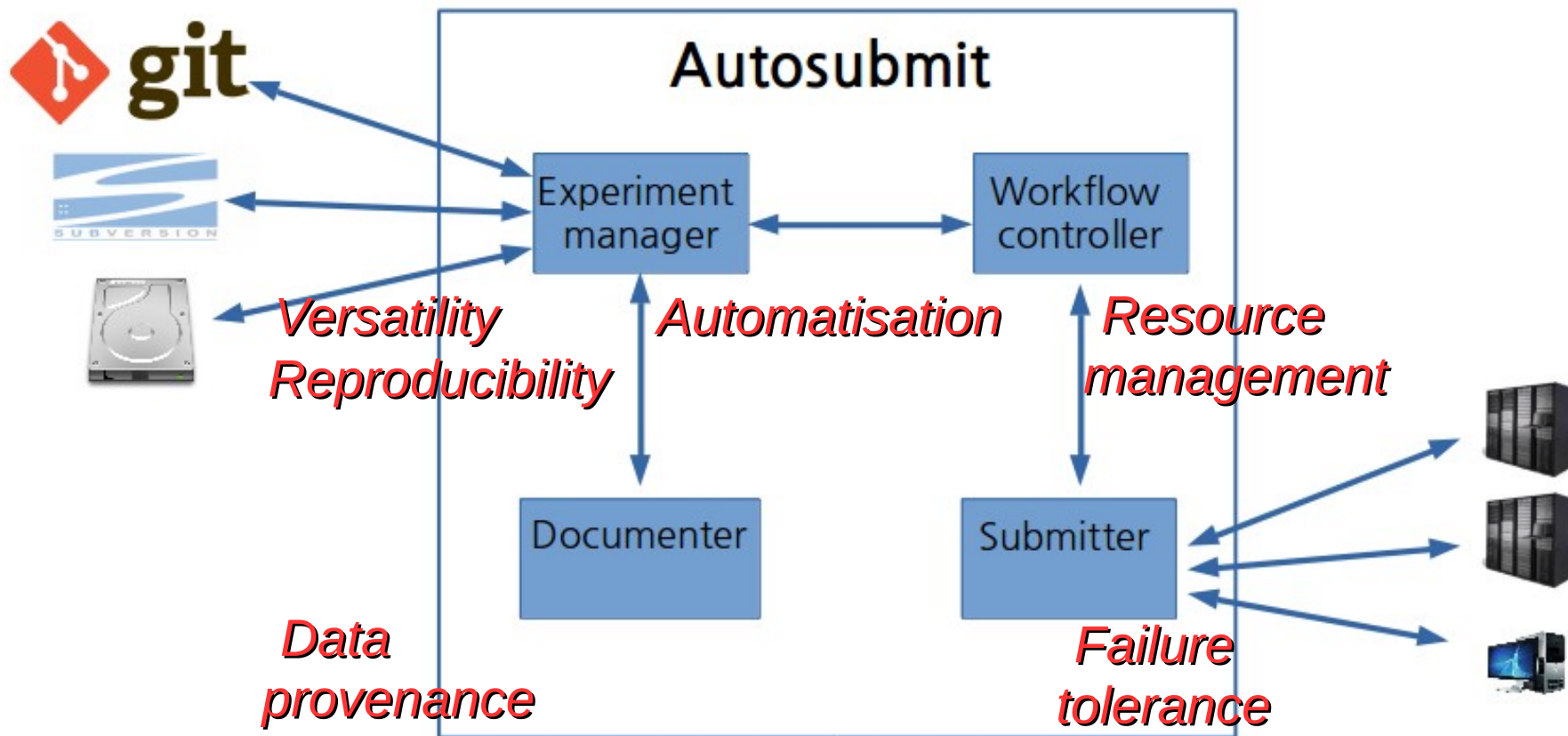
¿Qué necesitamos de Autosubmit?

- « Correr varias simulaciones como una sola.
- « Gestión transparente de máquinas y schedulers.
- « Postprocesar y transferir automáticamente los resultados.
- « Reproducibilidad.
- « Tolerancia a fallos.
- « Estadísticas.

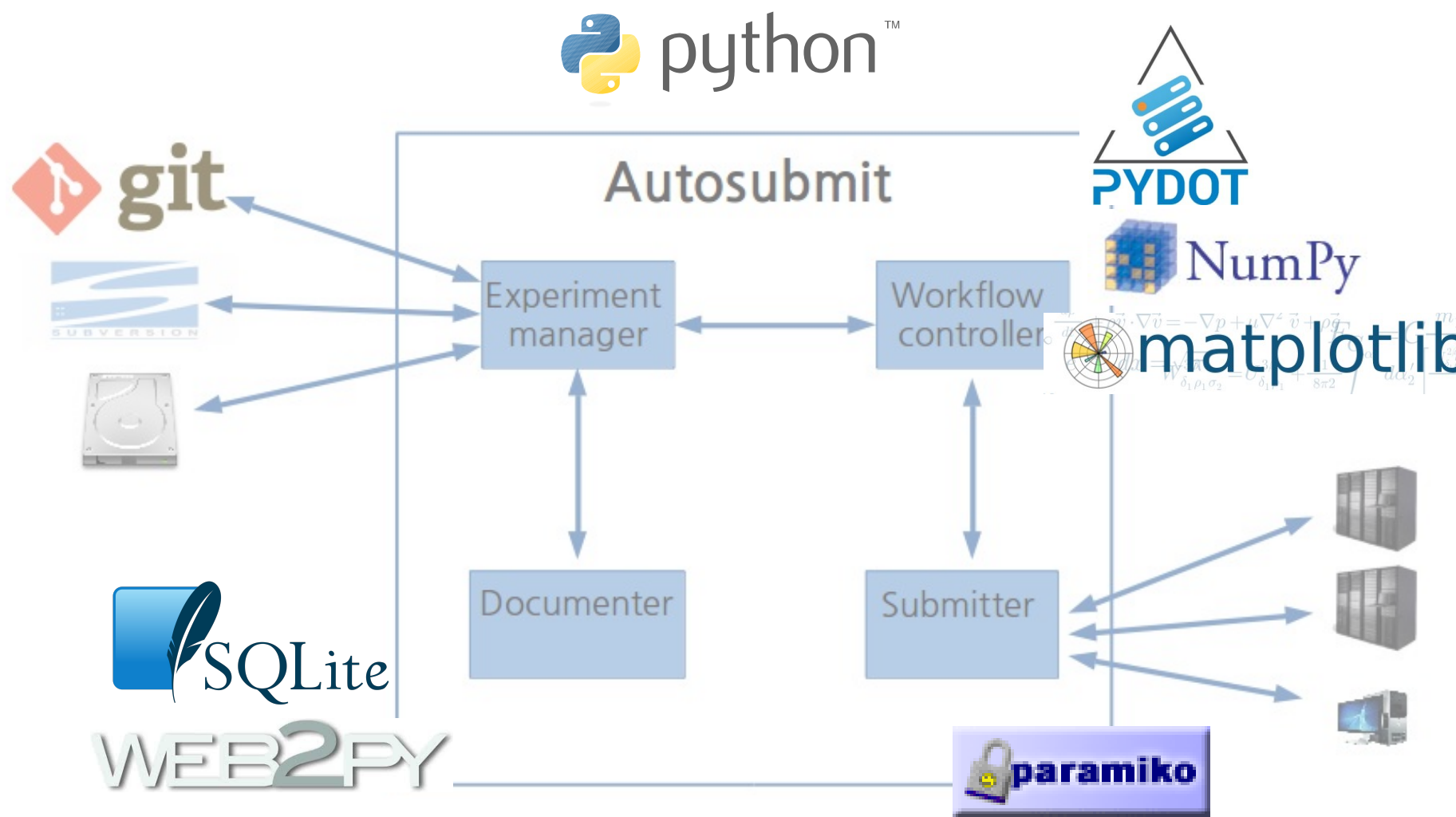
¿Qué necesitamos de Autosubmit?



¿Qué necesitamos de Autosubmit?



¿Qué necesitamos de Autosubmit?



Gestión de máquinas (1/2)

- « Conecta a máquinas remotas por SSH (mediante paramiko) u otros métodos (ecaccess).

- « Genera los scripts a ejecutar.
 - Parte de templates generados por el usuario.

 - Añade dos fragmentos de código usados por autosubmit: antes y después del código del template.

 - Añade la cabecera adecuada para el scheduler.

 - Sustituye variables dependientes de la configuración.

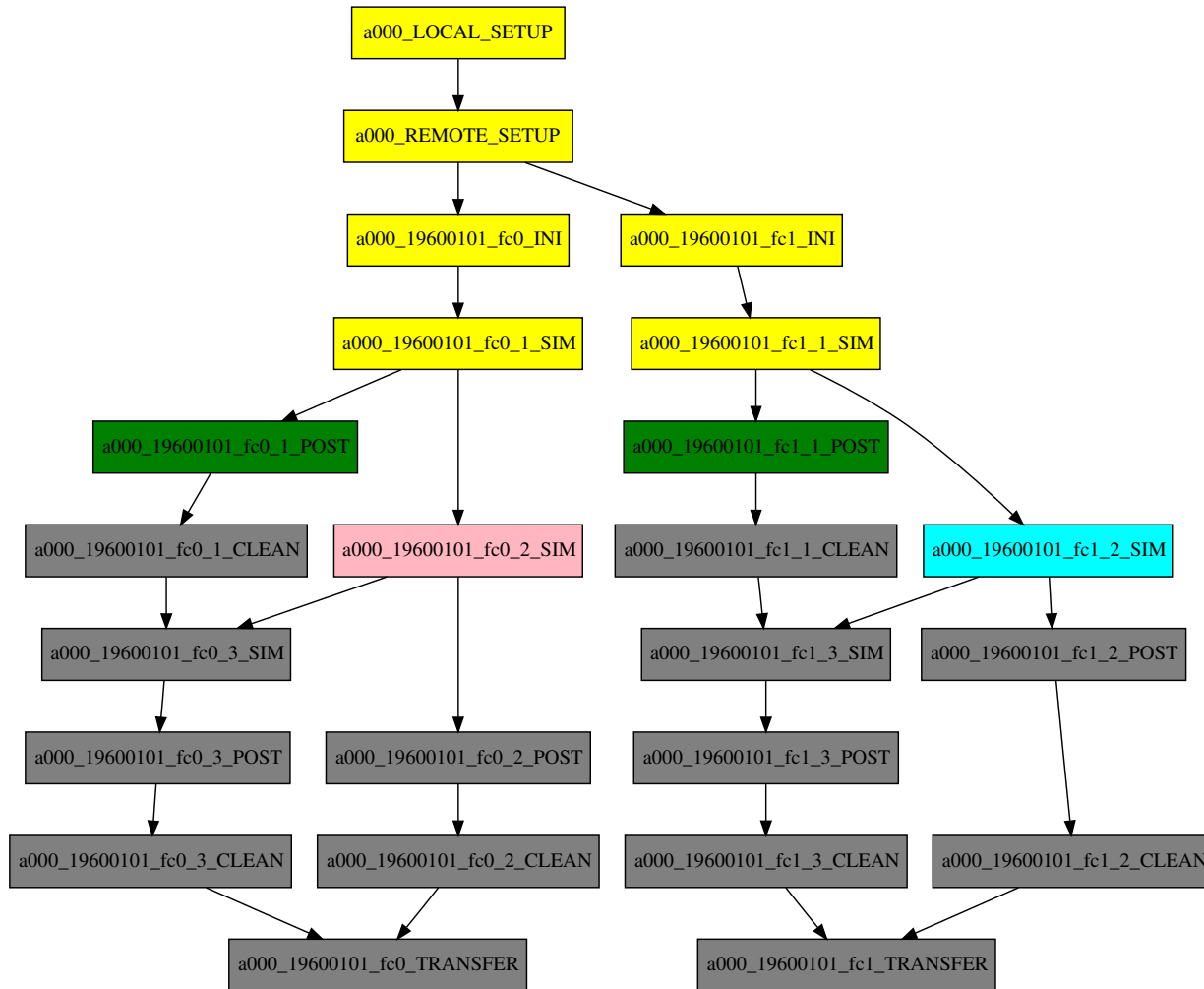
Gestión de máquinas (2/2)

- « Programa la ejecución en el scheduler y comprueba el estado de ejecución.
 - No nos fiamos por completo del estado que nos da el scheduler: generamos un fichero COMPLETED al acabar el job
- « Migrando a SAGA: librería para gestionar las conexiones con las máquinas y los schedulers.

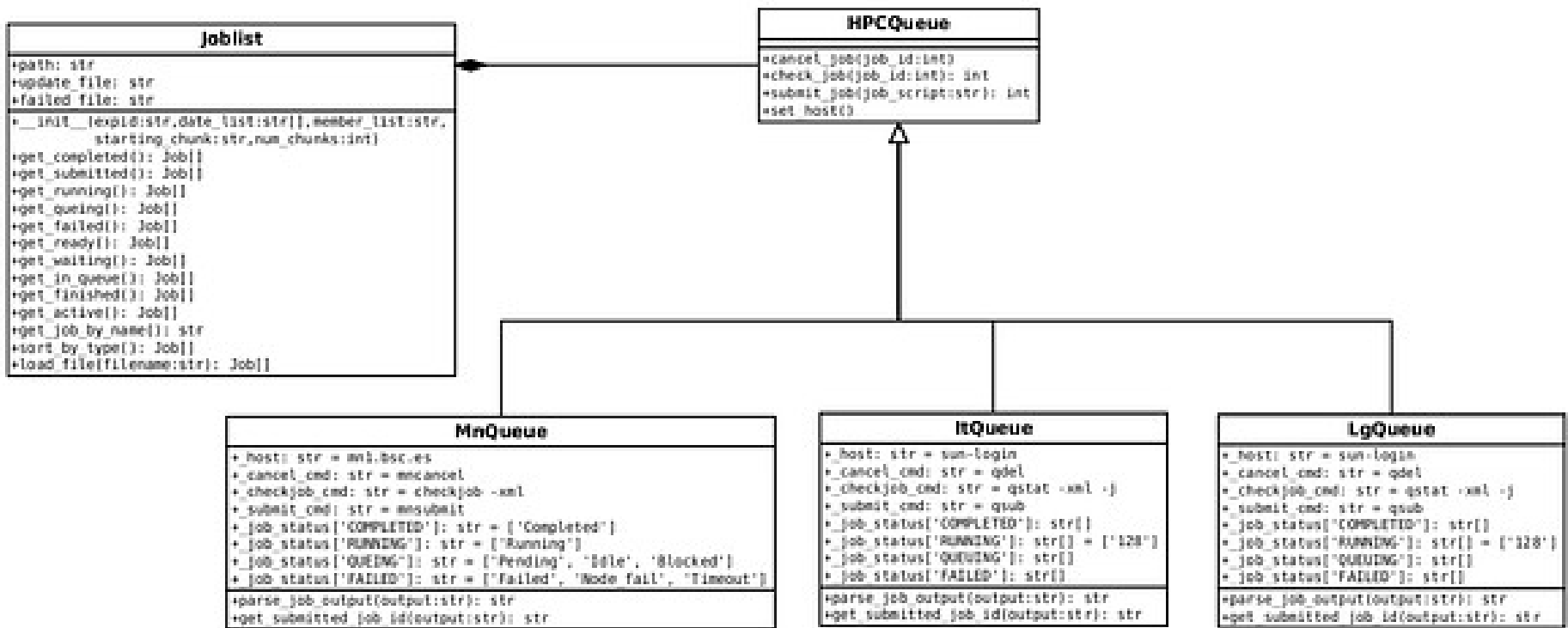
Gestor de workflows (1/3)

- « Reintenta automáticamente los jobs que fallan.
- « Los jobs pueden ejecutarse en máquinas distintas.
- « Permite volver a correr fragmentos de simulación en cualquier momento.

Gestor de workflows (2/3)



Gestor de workflows (3/3)

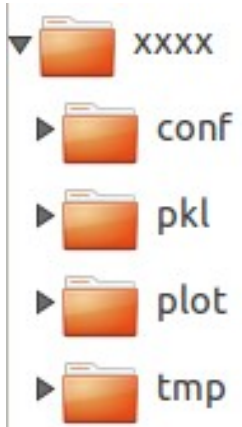


Gestor de configuraciones

- « La configuración básica del experimento (miembros, startdates, ...) se almacena en ficheros.
- « Cada proyecto puede añadir un fichero de configuración propio para las variables que necesite en sus templates.
- « Guarda el commit exacto que utiliza un experimento.

Experiment creation

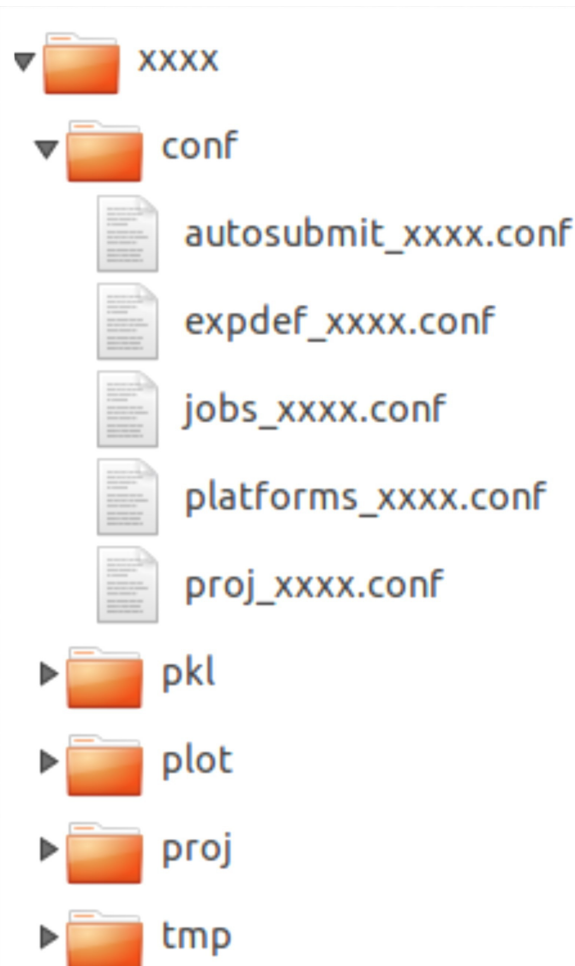
```
autosubmit expid -H HPCname
```



Experiment creation Experiment configuration

```
autosubmit expid -H HPCname
```

```
autosubmit create xxxx
```



Start dates, members and chunks (number and length).

Experiment project source: origin (version control system or path) and project configuration file path.

expdef_xxxx.conf

Workflow to be run: scripts to execute, dependencies between tasks, task requirements (processors, wallclock time...) and platform to use.

jobs_xxxx.conf

HPC, fat-nodes and supporting computers configuration.

Usually provided by technicians, users will only have to change login and accounting options for HPCs.

platforms_xxxx.conf

Project dependant experiment variables that Autosubmit will substitute in the scripts to be run.

proj_xxxx.conf

Experiment creation

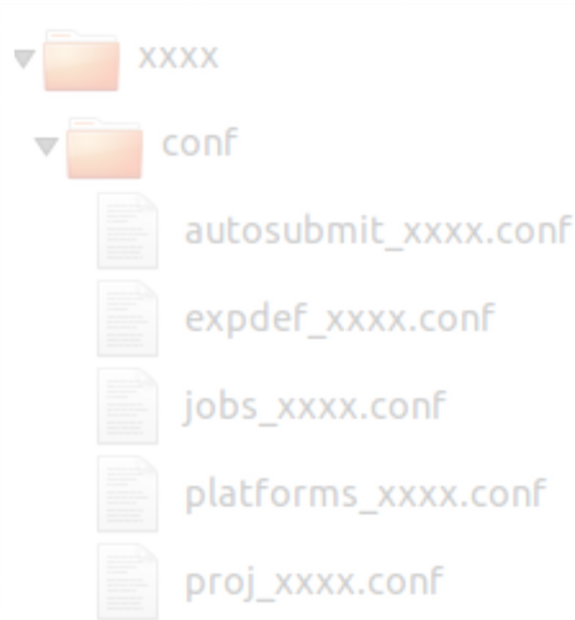
Experiment configuration

Experiment run

```
autosubmit expid -H HPCname
```

```
autosubmit create xxxx
```

```
autosubmit run xxxx
```



Start dates, members and chunks (number and length).

Experiment project source: origin (version control system or path) and project configuration file path.

expdef_xxxx.conf

Workflow to be run: scripts to execute, dependencies between tasks, task requirements (processors, wallclock time...) and platform to use.

jobs_xxxx.conf

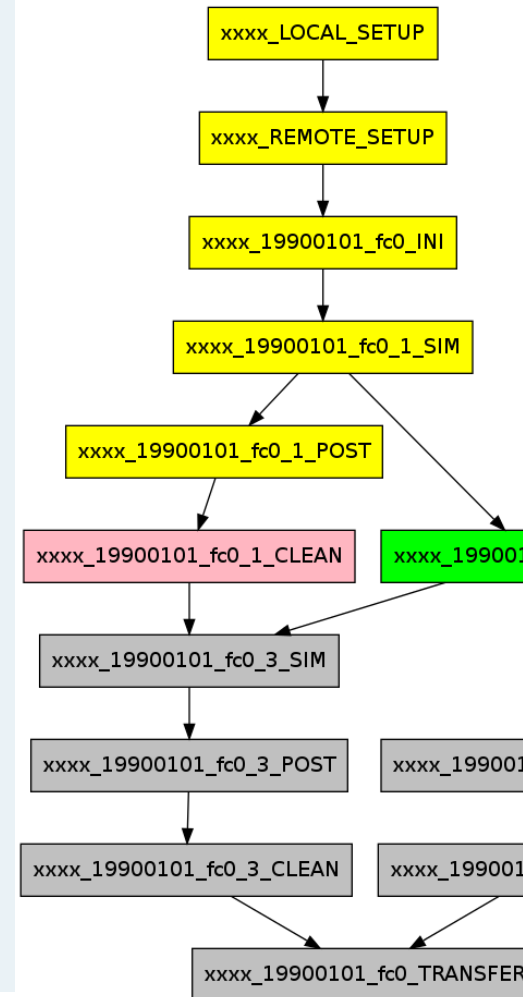
HPC, fat-nodes and supporting computers configuration.

Usually provided by technicians, users will only have to change login and accounting options for HPCs.

platforms_xxxx.conf

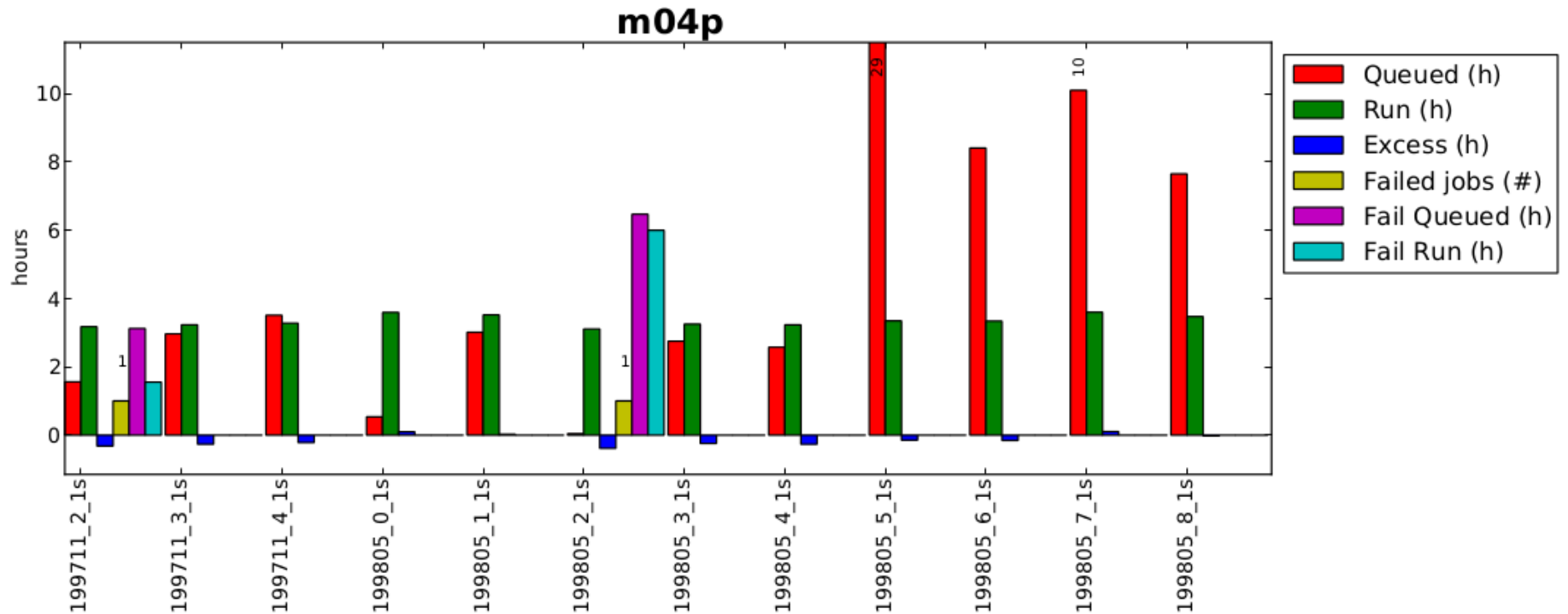
Project dependant experiment variables that Autosubmit will substitute in the scripts to be run.

proj_xxxx.conf



Estadísticas automáticas:

autosubmit stats cxxx



Para usar autosubmit:

« Instala el paquete python

- pip install autosubmit

« Sigue las instrucciones de la documentación

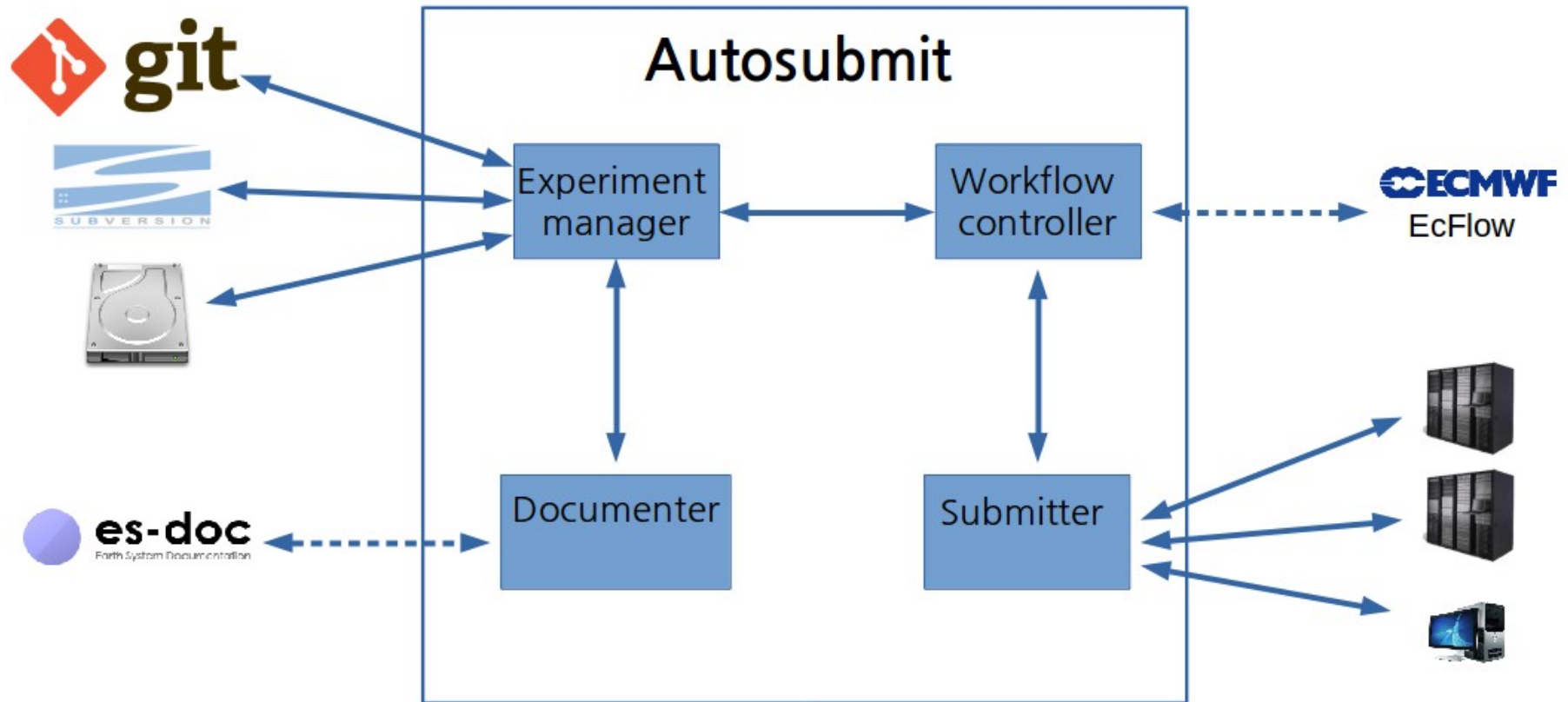
- <http://www.bsc.es/projects/earthscience/autosubmit/>



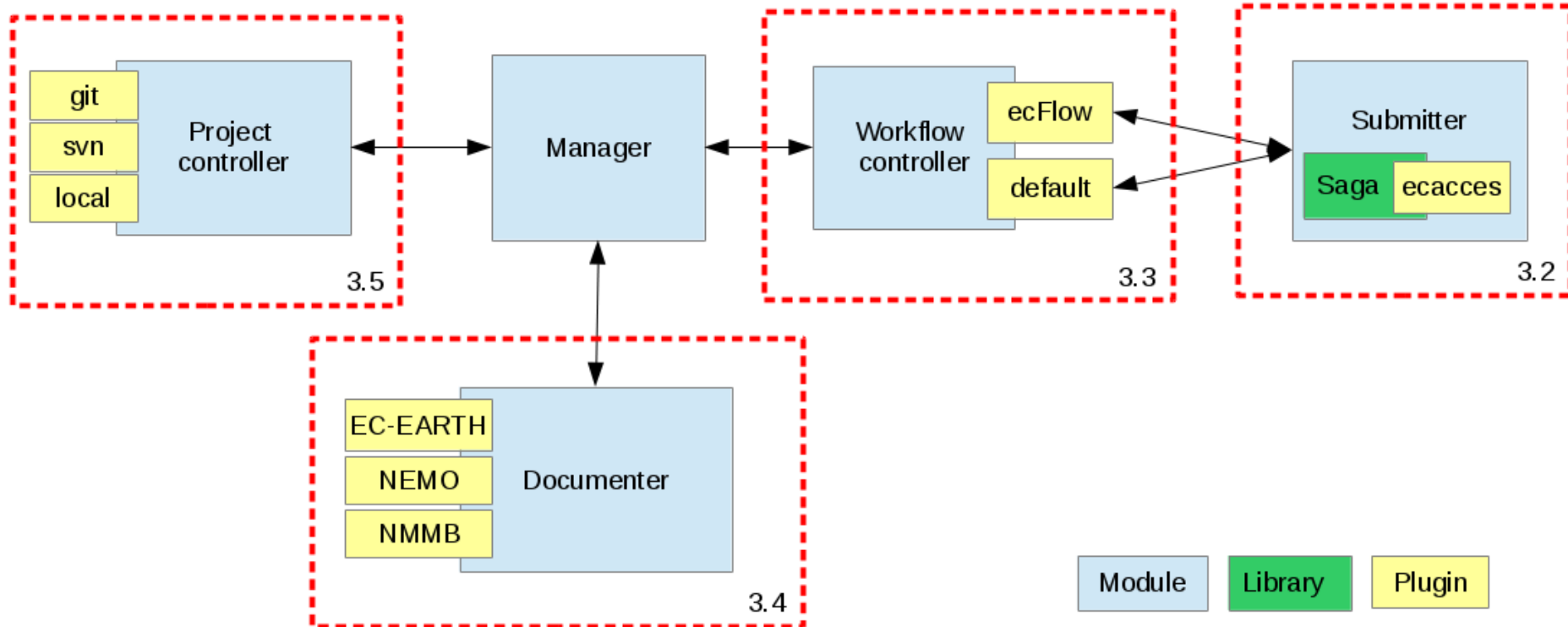
**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

Future work

Future work:



Roadmap:



Para desarrollar autosubmit:

☞ Código fuente disponible en gitlab:

- <https://earth.bsc.es/gitlab/es/autosubmit/>

☞ Plaza abierta para desarrollador junior:

- <https://www.bsc.es/about-bsc/employment/vacancies/phytondev>



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Gracias !

Consúltanos tus dudas y envíanos tus sugerencias:

javier.vegas@bsc.es

domingo.manubens@bsc.es