A WEB APPLICATION TO PUBLISH R SCRIPTS AS-A-SERVICE ON A CLOUD COMPUTING PLATFORM

Gianpaolo Coro, Giancarlo Panichi, Pasquale Pagano

Istituto di Scienza e Tecnologie dell'Informazione A. Faedo - CNR, Pisa, Italy

CONTEXT

Prototype scripting is the base of most models in computational biology and environmental sciences.

Scientists making prototype scripts (e.g. using R and Matlab) often need to **share results** and make their models **reusable** by other scientists on **new data**. To this aim, one approach is to **publish scripts as-a-Service**, possibly under a recognized standard.

But there are issue with prototype scripts:

- Generally **not meant to be transformed into Web services**.
- Do not manage **multi-tenancy, concurrency** etc.
- Porting to **more efficient programming** languages is **not affordable** (demands time, competencies and money).

SOLUTION

We present the **Statistical Algorithms Importer (SAI)**, a tool that allows scientists to:

- Easily and quickly **import R scripts** onto a distributed **e-Infrastructure**,
- **Publish** prototype scripts **as-a-Service** on a Cloud computing platform,
- Automatically obtain a **Web user interface** for the script,
- Manage **multi-tenancy** and **concurrency**,
- **Update** scripts without following long software re-deployment procedures,
- Manage **different versions** of the **R** interpreter.

SAI uses the **D4Science e-Infrastructure** (**www.d4science.org**), a distributed computer system supporting large-scale resource sharing and **Cloud computing**, via the definition of **Virtual Research Environments (VREs)**.

VREs define groups of scientists working together in the same domain and are endowed with social networking and collaborative facilities.
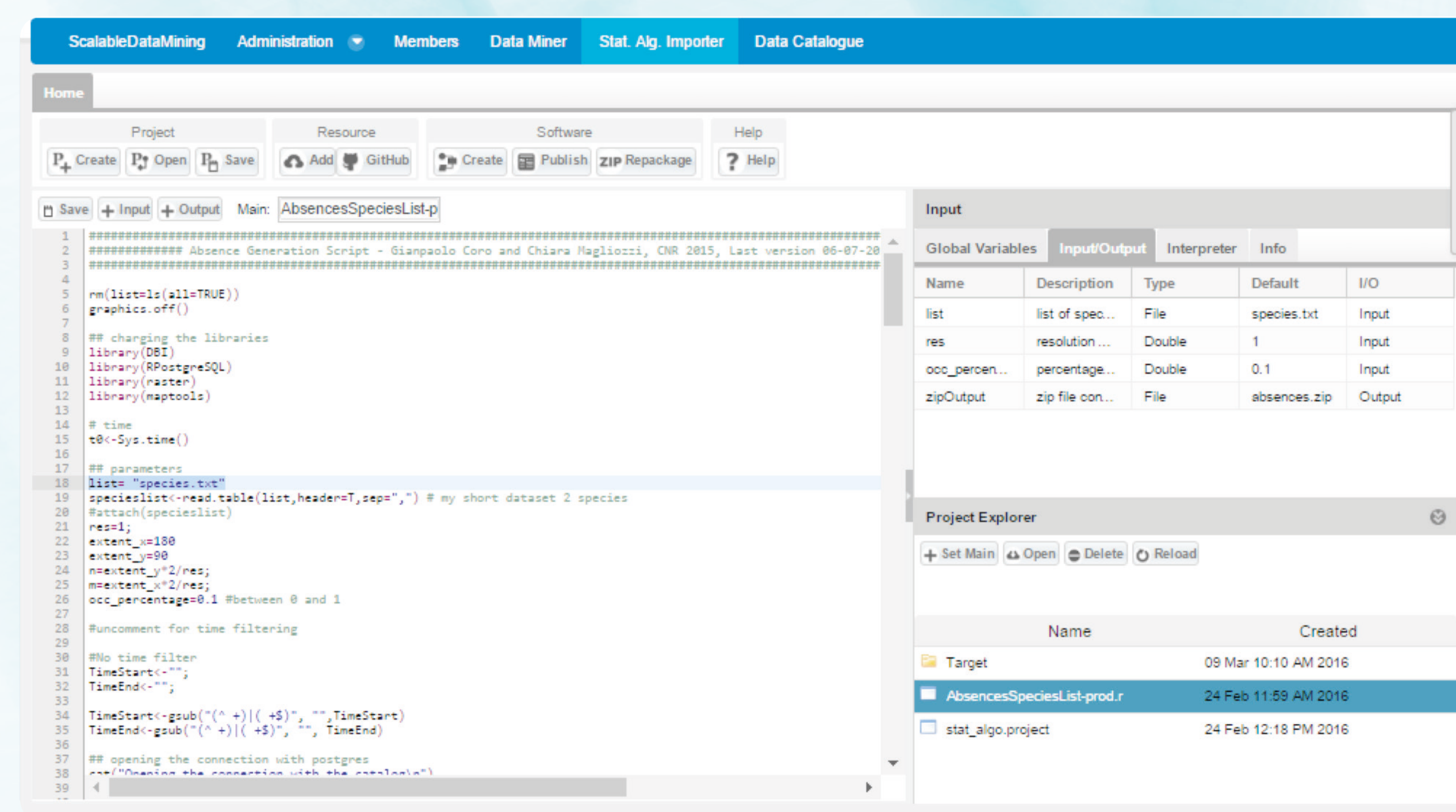
HOW SAI WORKS

A SAI user **uploads** (with drag-n-drop) **scripts to a Workspace area** (lower-right panel) and indicates the starting (main) script to be executed by the computational platform.
Script **inputs and outputs** are selected using the +Input and +Output buttons and are later transformed into **Web interface elements**. Alternatively, SAI can read 52North WPS4R annotations.

The other tabs allow setting **global variables** and adding **metadata** to the process (e.g. a name and description, the R interpreter version, the required packages etc.).

The functional buttons operate the transformation into a service:

Create: generates an as-a-Service version of the script,
Publish: prepares the Cloud computing system to execute the script,
Repackage: notifies the e-Infrastructure that the R code has been updated.

POLICIES MANAGEMENT

SAI manages privacy for both the script provider and the script user(s):

- The script is saved in an area that is **accessible by the provider and by a Java-compiled program only**,
- The computing machines **download, execute and then delete the script**(s),
- The script provider **indicates the VREs** where the script will be usable,
- These **VREs may have public or moderated access**, in order to monitor or filter users,
- Input data and **results remain private** to the user of the script/service,
- The **results can be shared with selected people**, using the D4Science native sharing facilities (Workspace).
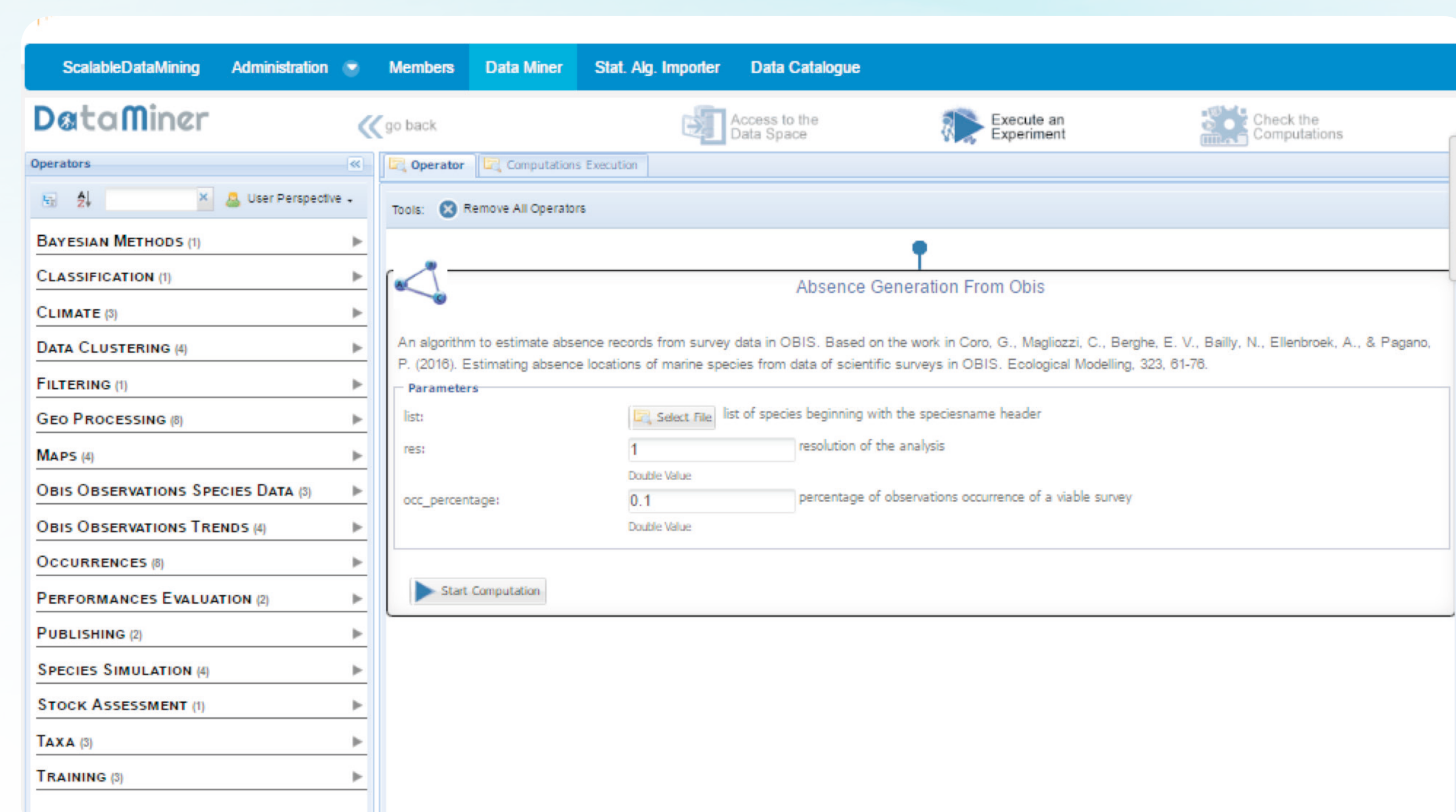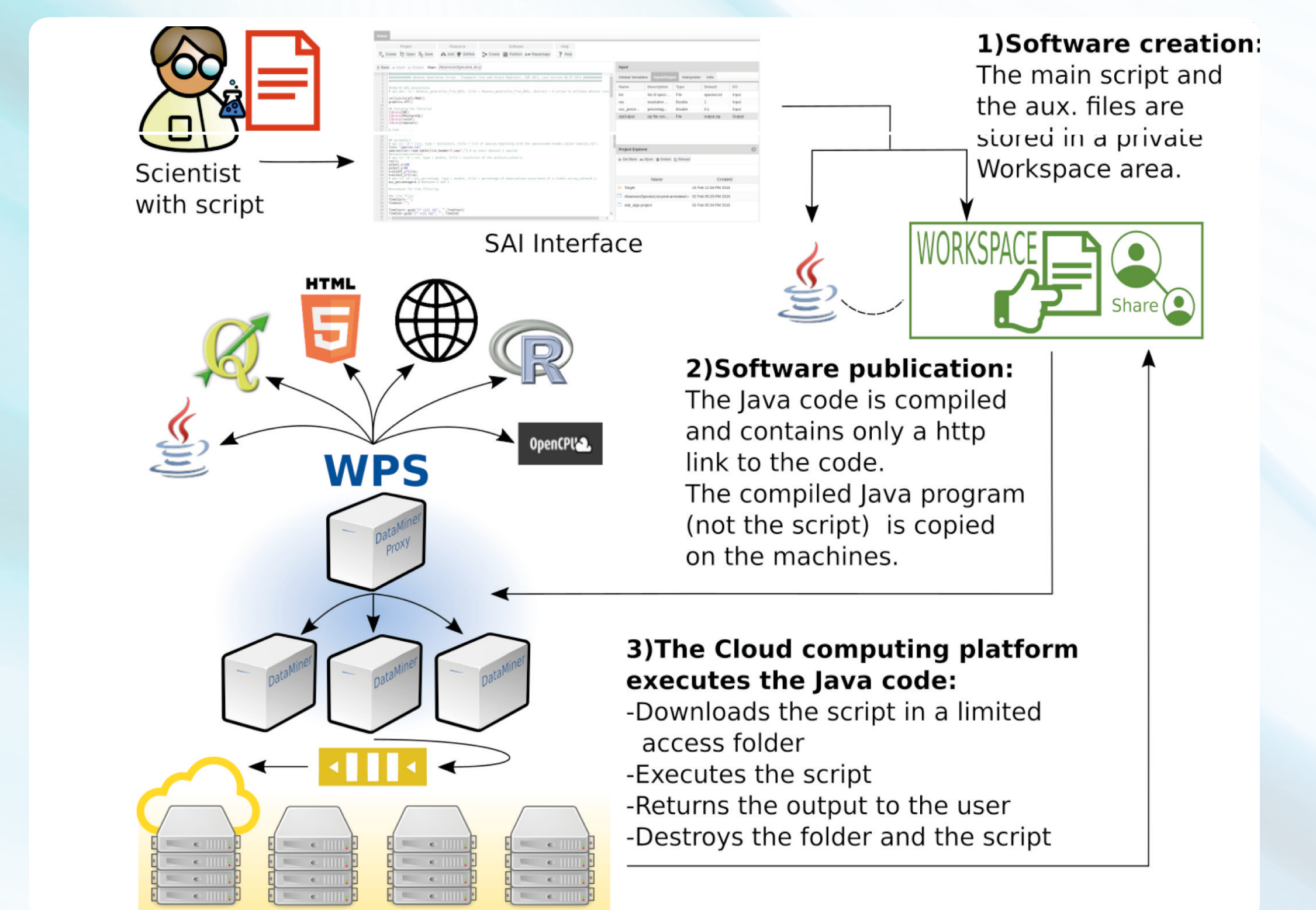
USING SAI

1. Register to the D4Science Web portal: https://services.d4science.org and **subscribe** to a public-access VRE (e.g. the ScalableDataMining VRE or the BiodiversityLab VRE).

2. Follow the **SAI user's guide to integrate your process**:
https://wiki.gcube-system.org/gcube/Statistical_Algorithms_Importer

3. Test your process on the D4Science computational platform (DataMiner), following its user's guide: https://wiki.gcube-system.org/gcube/DataMiner_Manager