

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.

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**,

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).

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.

_												
Scala	bleDataMining Adm	inistration 🕤 Men	nbers Da	ta Miner	Stat. Alg. Impo	ter Data Cat	alogue					
Home												
	Project	Resource		Software	E.	Help						
P+ Creat	te P_{\bullet} Open P_{\Box} Save	🔥 Add 🏺 GitHub	Create	Publish	ZIP Repackage	? Help						
🗂 Save -	+ Input + Output Main:	AbsencesSpeciesList	i-p					Input				
1 ###	######################################	eration Script - Giang	paolo Coro an	nd Chiara Ha	gliozzi, CNR 201	5, Last version	06-07-20	Global Variab	les Input/Out	put Interpr	eter Info	
4	(list=ls(sll=TRUF))	******************	**********	*****	*******	**********		Name	Description	Туре	Default	I/O
6 gr:	phics.off()							list	list of spec	File	species.txt	Input
8 ##	8 ## charging the libraries 9 library(DBT)					res	resolution	Double	1	Input		
10 lit	10 library(RPostgreSQL)					occ_percen	percentage	Double	0.1	Input		
11 11 12 1it	11 library(raster) 12 library(maptools)					zipOutput	zip file con	File	absences.zip	Output		
13	time											
15 t0	<-Sys.time()											
10 ##	parameters											
18 li:	st= "species.txt"											
19 spe 20 #at	ecieslist<-read.table(1 ttach(specieslist)	ist,header=T,sep=",")	# my short o	dataset 2 sp	ecies							
21 re:	s=1;							Project Explo	rer			
22 ext	tent_x=180							L Cat Main		th Delevel		
23 ext	tent_y=90							+ set main	Open Delete	O Reload		
25 m=e	extent x*2/res;											
26 000	_percentage=0.1 #betwe	en 0 and 1										
27	comment for time filts								Maria		<u> </u>	- 4
29 #01	requirerer for time filte	T ANY							Name		Create	ea
30 #No	o time filter							Target		0	9 Mar 10:10 AM 201	6
31 Tin	neStart<-"";									-		-
33	nechas- ;							AbsencesS	peciesList-prod.r	2	4 Feb 11:59 AM 201	6
24 71	5	(the section fame)										

How SAI works

scripts to a Workspace area (lower- packages etc.). right panel) and indicates the starting (main) script to be executed by the The functional buttons operate 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.

A SAI user **uploads** (with drag-n-drop) the R interpreter version, the required

transformation into a service:

Create:generatesanas-a-Serviceversion of the script,

Publish: prepares the Cloud computing system to execute the script, **Repackage**: notifies the e-Infrastructure The other tabs allow setting **global** that the R code has been updated. variables and adding metadata to the process (e.g. a name and description,

35 36	TimeEnd<-gsub("(^ +) (+\$)", "", TimeEnd)			stat_algo.project
37	## opening the connection with postgres		-	
39	▲	► F		

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,

24 Feb 12:18 PM 2016

- 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).



ScalableDataMining Administration 💿	Members Data Miner Stat. Alg. Importer Data Catalogue							
DataMiner 🤇	o back Access to the Data Space Execute an Experiment Check the Computations							
Operators	Computations Execution							
🗄 🛃 🗶 User Perspective 🗸	Tools: 😢 Remove All Operators							
BAYESIAN METHODS (1)	•							
CLASSIFICATION (1)	Absence Generation From Obis							
CLIMATE (3)								
DATA CLUSTERING (4)	An algorithm to estimate absence records from survey data in OBIS. Based on the work in Coro, G., Magliozzi, C., Berghe, E. V., Bailly, N., Ellenbroek, A., & Pagano, P. (2018). Estimating absence locations of marine species from data of scientific surveys in OBIS. Ecological Modelling, 323, 81-78							
FILTERING (1)	Parameters							
Geo Processing (8)	list: list of species beginning with the speciesname header							
Maps (4)	res: 1 resolution of the analysis							
OBIS OBSERVATIONS SPECIES DATA (3)	Double Value							
OBIS OBSERVATIONS TRENDS (4)	Occ_percentage: 0.1 percentage of observations occurrence of a viable survey Double Value Double Value Double Value							
PERFORMANCES EVALUATION (2)	Start Computation							
PUBLISHING (2)								
SPECIES SIMULATION (4)								
STOCK ASSESSMENT (1)								
TAXA (3)								
TRAINING (3)								

Using SAI

1. RegistertotheD4ScienceWebportal:https://services.d4science.organdsubscribe 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



BlueBRIDGE receives funding from the European Union's Horizon 2020 research and innovation

programme under grant agreement No. 675680

