

## BlueBRIDGE - Task #9254

### Recurring issues when deploying / updating algorithms on Dataminer

Jul 12, 2017 12:36 PM - Julien Barde

<b>Status:</b>	Closed	<b>Start date:</b>	Jul 12, 2017
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Gianpaolo Coro	<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Sprint:</b>	WP08		
<b>Infrastructure:</b>	Production		

#### Description

As explained in #9190, we still face known issues when trying to update existing algorithms (or deploying new ones) with Dataminer. The specific issue (#9190) will be fixed at some point but it already happened and I would like to discuss here some more general concerns regarding interactions between users and administrators of Dataminer when trying to deploy or simply update algorithms: interaction between users and administrators and sometimes we just don't know why algorithms can be deployed or updated.

Since few weeks, we are trying to compile with Ifremer (working with both a supercomputer and a VRE) the new data and the related set of runs parameterizations for ICCAT eastern bluefin tuna. @taha.imzilen@ird.fr is stucked since three weeks trying to update the very first step of the workflow which was working since last year. His (minor) update failed and since we can't use the first step anymore which is obviously a blocking issue for all the workflow. There is a meeting in 10 days at ICCAT and we are not sure anymore we will be able to use the VRE for this meeting.

Moreover, few days ago, I understood that a critical issue happened again after republishing the algorithm: one of the algorithms deployed last year has been replaced by another one and it seems impossible to go back to a previous working versions. We all experienced the same problem in the past (our ongoing codes being replaced by other ones).

This is a major concern for the users at this stage of the project as the method to deploy algorithms (or for a simple update) is still not automated (we still need to share emails with administrators and it takes days between emails) nor reliable (as described above). Users don't know what is happening (no logs) in Dataminer while RStudio is working properly with the same codes which means that the problem has to be related to the deployment in Dataminer. We all spent days in such situations.

This is all the more surprising that if you use 52 North WPS server, there is a very simple GUI provide by 52 North server where users can upload their codes: either it works directly or they get logs if it fails. Dataminer GUIs are more sophisticated but we still find them very hard to use compared to 52 North ones.

It is really blocking and perhaps we could get both types of GUS and users could then choose / try this other way to deploy algorithms?

#### History

##### #1 - Jul 12, 2017 01:11 PM - Paolo Scarponi

- Priority changed from Urgent to Normal
- Status changed from New to In Progress
- Tracker changed from Incident to Task

The ticket is not assigned to me but I have to point out a couple of things:

1. Opening an urgent incident for a general complaint is not a correct practice
2. "at this stage of the project as the method to deploy algorithms (or for a simple update) is still not automated" --> We are currently working to automatize the algorithm publication procedure, moreover, for a simple update the repackaging works perfectly fine as long as the interface doesn't change
3. "we still need to share emails with administrators and it takes days between emails" --> I always try to promptly answer any user request in order to solve all the issues as fast as possible, ask @taha.imzilen@ird.fr if he ever had to wait more than few hours to get an answer
4. "As explained in #9190, we still face known issues when trying to update existing algorithms (or deploying new ones) with Dataminer" --> Issue #9190 probably arose because of an inconsistencies on the algorithms names generated on the user side. One algorithm (Step 1 Vpa Iccat Bft E Retros) has been republished with two different names ("STEP1\_Retros\_VPA\_ICCAT" and "STEP1\_VPA\_ICCAT\_BFT\_E\_Retros") and the R project of a second algorithm (Parallelized Step 1 Vpa Iccat Bft E Retros) has been wrongly created before publication so that it was nothing else than one of the previously mentioned algorithms with a different names. This combination lead to the current situation in which 2 algorithms are broken at the same time. This has nothing to do with the tools for integrating the algorithms or the interactions between users and administrators but with the usage of the tools themselves.

**#2 - Jul 12, 2017 01:17 PM - Gianpaolo Coro**

- Status changed from In Progress to Rejected

Dear Julien,

DataMiner has several differences with respect to the systems you cite, but also brings advantages. We have summarised these differences in a recent paper <http://onlinelibrary.wiley.com/doi/10.1002/cpe.4219/full>

For example, 52North brings easiness for some specific R integrations, but does not offer Cloud computing and cannot manage large communities that often use other programming languages than R. The supercomputer used by Ifremer might be a good solution for your cases, but usually the maintenance cost of these systems is high (around 30k euros per month), whereas ours is 0 euros.

We are trying to put good features together, offering something that goes into the direction of Open Science. This can be tricky for a developer at first stage and requires further enhancements on our side, which sometimes is difficult because we do not just solve your issues, but we also take into account Open Science requirements at the same time. We do believe that the other solutions you mention will gradually fade out in a long-term future, because reusable, reproducible and economically sustainable Cloud computing approaches will populate the next computational platforms (and funded projects). Indeed, I believe that you know this better than me.

We are currently working to simplify the SAI interface offering also very basic integration ways for many programming languages (including R). DataMiner is under enhancement into several directions and @[lucio.lelii@isti.cnr.it](mailto:lucio.lelii@isti.cnr.it) is now dedicated to this particular task.

As Paolo pointed out, the issue of Taha was due to some confusion in the names of the algorithms and in their implementation. These pointed, in the code, explicitly to some processes that Taha had substituted. We limited these errors as much as possible by manually checking the installations, but we are going to release a fully automatic deployment mechanism, which will highlight naming and process issues much more. We will not be responsible for these in the future.

**#3 - Jul 12, 2017 01:18 PM - Julien Barde**

Nothing personal there. It's not about this specific incident but as I said about recurring issues.

"Incident" and "urgent": I don't know the correct terminology in redmine.

We plan to attend a meeting in 10 days and we can't use the VRE. This is why I said it's urgent.

I used "incident" as it was the most appropriate term in the list from my point of view. Taha lost his code..I call it an incident.

You can reclassify it in "task" and put a "normal" priority but I don't think this is right either.

**#4 - Jul 12, 2017 01:21 PM - Julien Barde**

I disagree when you say this is the fault of Taha. We experienced similar issues in the past: our codes replaced by the wrong one...

Now I see this is "rejected". I was giving some feedback. I am thus done with this ticket.

**#5 - Jul 12, 2017 01:43 PM - Julien Barde**

- File *Re\_ [SAI] New software publication requested.eml* added

Taha "randomly" solved the problem by himself (see attached email):  
"I solved the issue, The problem was the initiation of a parameter of the enumerated type.  
It does not work when I initiate it like that "2014" but it works when I put '2014'. "

The code works perfectly in RStudio if you use "2014" or '2014' but not in Dataminer...and there is no logs.

Do you really think this is a normal situation for users and that all of them will be able to fix such issues ?

I can testify that most of them will give up in few hours.

**#6 - Jul 12, 2017 01:44 PM - Julien Barde**

- File Re\_ [SAI] New software publication requested.eml added

**#7 - Jul 12, 2017 03:28 PM - Julien Barde**

- Status changed from Rejected to In Progress

Can somebody explain the users where this kind of error comes from, why it works in RStudio and not in Dataminer and how they are supposed to understand what to do without logs in Dataminer?

**#8 - Jul 12, 2017 03:48 PM - Pasquale Pagano**

First let me say that I appreciate @julien.barde@ird.fr attempt to open a discussion on a specific topic. The issue is that the ticketing system is more for tracking a specific action to perform more than for a general discussion on things to improve. Maybe the social could work better for this kind of discussion. Said this, let me analyse a couple of things:

a) it is urgent that we find a solution since the community has a commitment in less than 10 days. By reading this discussion i did not get what we have to do, what has to be solved and which is the current status of the VRE. It is not needed that you make a summary for myself but it is important that the VRE is exploited as planned.

b) "2014" vs '2014'. I am not an expert and i don't know the implication of this different behaviour between RStudio and DM. However we should deliver a solution for the users. We always promote this development workflow: try your code in RStudio with the limited capacities of it; then use SAI to prepare it for the infra; finally run it through DM. Now, if it works on RStudio and it doesn't on DM we need to figure out how to identify the issue and then guide the users to solve them

c) about logs: with the new solution for automatic deployment of the algorithm, logs will become available at deployment time only. Can we provide a solution to collect and store the logs of the executions together with the output?

**#9 - Jul 12, 2017 04:26 PM - Gianpaolo Coro**

Hi, the gaps between RStudio and DataMiner are mainly these: 1) inputs/output of the process must be declared correctly, 2) care should be taken about the input files and the local paths because the names of input files can change, 3) the output files should possibly have a dynamic names to enhance the certainty of non-overlapping outputs between users. We have summarised common issues and solutions in the SAI FAQs (which is also linked through the help button in SAI): [https://wiki.gcube-system.org/gcube/Statistical\\_Algorithms\\_Importer\\_FAQ](https://wiki.gcube-system.org/gcube/Statistical_Algorithms_Importer_FAQ)

What we have observed is that point 1 can be tricky, because it is based on default values in the R code, which are then searched by DataMiner. For example, if the default value of a variable is declared as '2104' than DataMiner expects that the '2014' string (and not the 2014 number) is the default assignment of a certain variable, because the two values are different conceptual things. Point 2 and 3 do not occur often and are usually correctly

managed. As for the Wiki, we found it very hard to make people read it.

Thus, when something is working in RStudio but not in DataMiner, it is usually because of the three points above. Overall, I would not talk about "fault" when people have issues with this technicalities, since they are normal in software development.

Other issues include the fact that 1) the Workspace folder containing the SAI project gets manually deleted or copied to another project, 2) the name of the process has been altered and 3) the publication button is pressed without regenerating the code for DataMiner (i.e. the CREATE button). We are thinking about helping people to avoid these issues.

Apart from rare cases of issues on the Workspace, my impression is that the algorithms remain stable if not explicitly altered. We periodically check 100 algorithms automatically and there have not been algorithms corrupted because of WS issues this year. For example, the algorithm "Absence Points Generation from Obis" was created with SAI in 2015 and still works (depending on the availability of the OBIS service).

Finally, the logs of the computation are always returned to the user, both in the case of exception and in the case of successful computation. Those are the ones I use to debug the code.

#### **#10 - Jul 12, 2017 04:31 PM - Julien Barde**

About what to do: users can't help as they don't know what's happening behind the scene and they don't have access to logs (and not sure logs would help anyway).

As you said users try the code in RStudio but working in RStudio doesn't mean this is enough for Dataminer. There is an overlay of additional constraints to deploy the very same code in Dataminer.

By luck Taha has been able to find the reason this time but he had no help from logs and spent lot of time working on it.

This additional constraints are not explicit. This time it was because of ' instead of " but previous times it was for different reasons (and we helped to identify them). Either the users get help from logs or an exhaustive list of additional constraints but there is no way to solve these kind of issues if we don't share them with you. Without help we just can try to guess in order to fix issues by multiplying attempts. But If we don't share tickets, it will happen again and the solution would be lost for the next ones.

We are part of the projet and we don't expect tools to be perfect. By using the tools and reporting issues we are clearly helping to test and improve them before "real users". However this is not fair to consider that we are misusing the tools or misreading the wiki.

For my part, I was explaining here a reccurent problem in a ticket by clicking on "new issue" to fix it and improve the tools. I think this is what redmine is made for.

#### **#11 - Jul 26, 2017 11:44 AM - Gianpaolo Coro**

*- Status changed from In Progress to Closed*

We have acknowledged the request and have planned solutions. In other tickets, the detailed reason of the misunderstanding has been clarified and SAI will be modified to avoid them in the future (see #9328 for example). Other requests are under discussion.

**Files**

---

Re_[SAI] New software publication requested.eml	14.9 KB	Jul 12, 2017	Julien Barde
Re_[SAI] New software publication requested.eml	14.9 KB	Jul 12, 2017	Julien Barde