

D4Science Infrastructure - Task #9119

User Context Data Uploader

Jun 30, 2017 03:54 PM - Debhsish Bhakta

Status:	Closed	Start date:	Jul 05, 2017
Priority:	Normal	Due date:	
Assignee:	Emmanuel Blondel	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	WP07		
Infrastructure:	Development		
Description With reference to the original ticket (Project Activity #7554) I have some suggestion 1.User input as MPs (i.e shapefile). From Openlayer GUI(Clint side) we can directly read GeoJSON format from the shapefile instant of uploading or storing shapefile in the WP. Once we got the GeoJSON that we will pass to R environment/dataminer as a parameter(eg. userMPA). In R environment we can convert GeoJSON to spatilaDataframe and now it is available for any GIS operation. We have tested the same idea with sample cage.shp shapefile file (provided by Levi Westerveld) as input MPA's and it's working fine. In that challenge is shapefile must be topologically corrected so what we did initially we have refined the topology of cage.shp using ArcGIS topology tool separately before using as an input. 2.Reliability of shapefile (user input) The user input is a challenge for us because the shape file may have a lot of topological error and extra attributes same thing also address by the Emmanuel blondel in his previous update. The idea is we can define shapefile guideline for the user (such as no overlapping polygon, must have id field, file size, number of features etc.). Further in the client side (open layer GUI) we can take care of id/name by allowing the user to choose require Id/name from shapefile/geojsion which is a prerequisite for invoking the script in VRE. This is the first time I am replying in the VRE forum So, please provide you valuable feedback. Thanks !			
Subtasks: Task # 9189: Temporary area via Workspace			
			New

History

#1 - Jun 30, 2017 04:08 PM - Pasquale Pagano

- Assignee set to Emmanuel Blondel

By reading this ticket I cannot figure our which support do you need. I am tentatively assigning the ticket to Emmanuel. Please change it if inappropriate.

#2 - Jun 30, 2017 05:04 PM - Emmanuel Blondel

I thought i had replied, but my answer got stuck because Lino was faster to reply :-)

My first comment is about the exploitation of this. @levi.westerveld@gmail.com It seems that you are now considering it for the MPA web-app, while it was initially (AFAIK) only targeted to be a DataMiner process executed from the DataMiner UI, and not the web-app. IMHO, it is this choice that conditions or not what kind of output you want to defined for the uploader. From the last exchanges, i remember the envisaged road was to stick with a shapefile available through the workspace, and do not "upload" at all (by upload it could be either on the workspace, either as layer, in this case it's already uploaded to WS, and layer is not envisaged AFAIK). The "upload" is not really an upload, in the sense that you used an already uploaded file and push it to the analysis script.

For the latter analysis, you may indeed decide to package an output including the report + the equivalent of shapefile in GeoJSON, for visualization purpose (but again this is only in the case you would want to apply it within the MPA web-app).

I'm available to skype on it. Next week i will have some time to dedicate to PAIM VRE, and support GRID-ARENDAL in customizing the present algorithm (or create a separate one)

#3 - Jul 01, 2017 06:20 PM - Debhsish Bhakta

Thank you for your reply.

I apologize for creating a separate ticket apart from "Project Activity #7554" ticket because for me edit option is not visible(I don't know why).

Regarding VRE, I think to make entire process simple and more approachable, we could keep the two components of VRE i.e. "MPA reporting" and Data Miner separate for this instant (please comment).

1.It is quite possible and easy to upload data in Data Miner and the file available for process execution within the Data Miner.

2.But in my opinion "MPA reporting" is more toward WebGIS approaches rather file interaction system (please comment). It will more sense if user can upload data (may be in the form of geojason/layer/file url path) or push require parameter to the script and execute the WPS service directly.

3.Uploading a file into "Data Miner" interface and then returning to MPA reporting interface to select the require parameter may a quite confusing for the user to link the parameters. Also it is challenging for us to programmatically synchronize the parameter of Data Miner and MPA reporting such as one input come from "input folder" of data miner.

@Emmanuel Blondel, I think it is worth to discuss more during Skype. From your latest reply I know you are available for us this coming week, thanks! for that, could you please provide you prefer date and time for Skype. levi Westerveld is out of station right now (Kenya) but I will sure enquire his availability for skype in a separate email.

Thank You!

#4 - Jul 04, 2017 02:10 PM - Emmanuel Blondel

@massimiliano.assante@isti.cnr.it can you recall what @debhasish.bhakta@grida.no should do to have access to BlueBridge tickets? apparently he can't reply to it and can access only D4S tickets. Thanks in advance

@debhasish.bhakta@grida.no I'm available this week. Check with @levi.westerveld@gmail.com for his availability, i've no preferences.

We need to proceed in iterative steps:

- the plan was to stick with Dataminer (at least for the time being). We need to see with Levi if this changed or not
- on this basis, make the algorithm more generic (or just create a custom one) to accept a zipped shapefile already uploaded to workspace (or dropped to Dataminer dataspace), keeping the same output (report only).

I would not add any geojson within the output, except if the plan evolves to a direct exploitation through the GIS MPA app. Some comments:

- this is not directly related to the extension of algorithm with custom MPA input shapefile, but rather if you want to enrich the output of the process and its visualization, ie. display dynamically the output intersect to the web-map in addition of the report display. If you expect it, we should start doing it with the current algorithm (no custom MPA input)
- If you expect to let people choose their own data through the web-app, this is where the prior workspace upload is limitant, and where people should not have to upload the file to the workspace. This should rather be done by the algorithm himself, to a temporary area where the data is not stored in persistent way to the workspace. Again people wants to analyze their data only, not to upload it per se. For this, we need to have the capacity to use a local file through the Dataminer WS.

#5 - Jul 04, 2017 02:20 PM - Massimiliano Assante

Debhsish Bhakta is not member of the BlueBRIDGEProject VRE, this is why cannot access blueBRIDGE tickets.

#6 - Jul 04, 2017 04:19 PM - Debhsish Bhakta

Thanks for your reply and also thanks for linking me with the massimiliano.assante@isti.cnr.it for support.

I have inquired with Levi availability and he is not available for this complete week, but he can join us on next week from GRID-Adrenal. I am wondering is it possible for you to be available for the next week say Monday afternoon(10-07-2017).

Regards
Debhasish

#7 - Jul 04, 2017 07:36 PM - Pasquale Pagano

- Tracker changed from Support to Task

I am changing the tracker since this issue does not seem to me related to a support request.

As far as @emmanuel.blondel@fao.org comments, let me remember that the workspace is an application that uses the cloud storage. The cloud storage itself can be exploited without the workspace (even if we discourage it) and it offers a temporary area that is usually used by applications to store temporary files. This temporary area could be exploited by the web-app if needed but I need to know more about this use case to provide some hints on it.

#8 - Jul 04, 2017 08:17 PM - Emmanuel Blondel

Thanks Lino, the starting point of this use case is the capacity to input a local file (from user's machine filesystem), that is not intended to live in the workspace, but just for the sake of an analysis, with behind clear impacts in term of usability. You may distinguish here 2 sub use cases depending on the entry point:

- web application as entry point (here the MPA web app), where user select a file from his/her computer to analyze it. Here the web-application may live within a VRE, or not (with web-services calls made with application tokens instead of user tokens), so linking to workspace is too "strong" and too user's workspace oriented.

- the DataMiner as entry point. From last exchange with Gianpaolo (see #7554), i understood that the File selection pointing to user's filesystem was deactivated, in favor of WS only. So at now in SAI, we tag an input as FILE (but it means Workspace FILE). What we would need is to extend Dataminer to support both WS FILE (local for the infra) and EXTERNAL FILE (local file for the user machine). This would make it compliant with file types that we can already manage in APIs such as the gCube Data Transfer service or the GeoServer REST API.

Hence, if the workspace is used (but let's talk more generally of cloud storage), it's used the time of the analysis only: I upload it temporarily to storage, run my analysis on it, and then it's source data is deleted (immediatly or after some time). I don't know what would be the effort to make DataMiner compliant with such approach. Dataminer is fine for experiments strongly associated with workspace, but when coming to an application context (possibly living outside a VRE), its strong association to workspace becomes a constraint.

Some more general thoughts/ideas:

It is then up to the algorithm to provide option (eventually) to store/publish it: here 2 options maybe;

- either the data is deleted once the analysis finished, the user decides to publish it, it reruns the analysis saying publish=true, or
- output is uniquely identified in the cloud storage, such uuid is returned through DM output when the user finished the analysis. Based on the UUID, he can publish the output (with the publisher he wish, this can be a GIS layer, but it could be workspace resource), and the primary analysis output, still stored in the cloud storage is in any case delete after some time period (e.g. 1 day).

I don't know if the workspace could be used in such use cases, or if it is rather the cloud storage behind to consider. One important point is where this upload should(could) be done: Dataminer or early in the R script. If the direct use of cloud storage is discouraged (I agree, this must be managed through infra apps or services), it would be even more discouraged if storing is completely delegated to R user (besides the fact that that he/she may not know how to store/delete data there, without guarantee he/she will do it properly).

Let me know if it clarifies.. I hope these ideas will feed further the discussion.

Cheers

#9 - Jul 05, 2017 06:31 PM - Pasquale Pagano

Hi @emmanuel.blondel@fao.org, thanks for the explanation. I believe we need to find a solution first from the user perspective. It is not a matter of technology I believe since it is not complex to support. Let me add a couple of details.

- the cloud storage is accessible through an API. This API is used by the workspace clearly. However it is yet another API and it is a bit more complex than the workspace API. Moreover it requires more parameters to be used, such as the context, the exploitation of the temporary area or not, the proper token, etc. We discourage its exploitation because it allows to save data on the storage and this volume is accounted to the user but it is not accessible, visible, and manageable through the workspace making things more complex.
- the workspace manages more metadata about the stored data than the cloud storage and it

As a first step we could manage the temporary area via the workspace. This should represent already an improvement since the user could still use the same widgets (either the data miner or the workspace) or connector (like the one to access and use the workspace from RStudio) or workspace API but then the data stored there is automatically deleted after a time interval (I think that it is 7 days currently but we can easily change it).

Could this improvement be useful?

#10 - Jul 05, 2017 06:52 PM - Emmanuel Blondel

Yes, definitely useful, but please let me rehighlight that this goes in pair with the crucial and key requirement which is the need to have Dataminer supporting bot WS FILE and EXTERNAL FILE (for UI and backend engine), so temporary workspace area would be used behind the scene when the user would select a file directly browsed from his machine filesystem (and he would not be constrained to go through the workspace before for uploading files).

Having this would be a huge improvement, not only from DataMiner UI perspective, but also for target applications (like the MPA one running in the PAIM VRE). People could load their shapefiles and run a processing through a web-app, directly getting their results and consumed by the application (e.g. think in a map product that could be directly loaded into a web-map)

For application viewpoint: would a temporary storage area be accessible if I specify an application token to Dataminer Web-Service?

#12 - Oct 02, 2017 05:20 PM - Debhsish Bhakta

- Status changed from New to Closed

I am closing this ticket.

The purpose of the ticket is solved in the other ticket #9376. Moreover, the entire concept for data upload is revised.

Thank You.