

D4Science Infrastructure - Support #6148

GeoServer layer security

Dec 06, 2016 11:47 AM - Levi Westerveld

Status:	Closed	Start date:	Dec 06, 2016
Priority:	Normal	Due date:	Dec 12, 2016
Assignee:	Levi Westerveld	% Done:	100%
Category:	Data Management	Estimated time:	0.00 hour
Infrastructure:	Development		
Description			
<p>As part of the ProtectedAreaImpactMaps VRE we will be hosting a few layers on the GeoServer that will be queried by algorithms in DataMiner. Some of these layers need to be private, so that they cannot be (accessed) downloaded other than by the algorithm itself. Could you advise on the preferred approach using the current infrastructure?</p> <p>Please let me know if more information is needed.</p> <p>Many thanks,</p> <p>Levi</p>			

History

#1 - Dec 06, 2016 12:33 PM - Fabio Sinibaldi

Hi Levi,
our current policies doesn't allow to specify constraints at algorithm level, but at scope level i.e. we can make these layers accessible by only a specific VRE. One solution might be to have a specific VRE with only your algorithm, to be sure that it's the only one who can use these layers. However, if you're confident that the other VRE users won't use these layers as input for the other algorithms available in your VRE, you shouldn't need a new specific VRE for this purpose.

#2 - Dec 06, 2016 12:56 PM - Pasquale Pagano

- Status changed from New to Feedback
- Assignee changed from Gianpaolo Coro to Levi Westerveld
- % Done changed from 0 to 50

As @fabio.sinibaldi@isti.cnr.it reported, the security is supported at VRE level. To solve your request, we will first create a VRE. This VRE will generate a security context (with encryption based on a VRE sys key, token to access the data, etc.). Then, you will publish your private datasets in this VRE. Finally, the algorithms will be added to this VRE and they will get access to those private layers.
Is this solution clear and suitable for the case you need to support?

#3 - Dec 06, 2016 01:56 PM - Levi Westerveld

Thanks for the prompt reply.

The data will be used by the existing ProtectedAreaImpactsMaps VRE (we do not need a new one). I will upload the geospatial data to the dev. GeoServer which I believe serves data through portlets such as the GeoExplorer in the BiodiversityLab VRE. I just want to make sure that the original data cannot be downloaded (we do not mind it being viewed) anywhere. We need to host the data as it is not made available as a WFS by the authors of the data so that we can use it in our algorithm, but we need to make sure that we are not redistributing it through the infrastructure. Maybe this clarifies?

Perhaps the current settings in place do not enable users to download data from the GeoServer through the any parts of the infrastructure (I am not aware of all tools)? In which case this might be good enough.

Thank you,

Levi

#4 - Dec 06, 2016 03:00 PM - Pasquale Pagano

- Parent task set to #3835

#5 - Dec 06, 2016 08:22 PM - Emmanuel Blondel

@levi.westerveld@gmail.com IMHO, if some new layers involved in the algorithm are not to be exposed publicly (WMS/WFS), the straightforward way would be to handle this as not published, that is you make it available in some workspace, we download in R from there (and not from WFS).

Otherwise, we complexify if we want to use it as WFS in the algorithm and prevent it externally (this would need to pass VRE security token to WFS, but i'm not sure if CNR extends WFS with such information)

#6 - Dec 07, 2016 11:50 AM - Fabio Sinibaldi

@emmanuel.blondel@fao.org has a point here. In the meantime I've been checking on GeoServer security features and it seems to let user specify access rules for layers and geo interfaces (WFS/WMS/WCS). However, since we'd need to evaluate possibilities and test solutions, for the time being I think it would be preferable to follow Emmanuel suggestion.

@levi.westerveld@gmail.com please, let us know if Emmanuel's suggested approach is feasible.

#7 - Dec 07, 2016 01:07 PM - Pasquale Pagano

Emmanuel Blondel wrote:

@levi.westerveld@gmail.com IMHO, if some new layers involved in the algorithm are not to be exposed publicly (WMS/WFS), the straightforward way would be to handle this as not published, that is you make it available in some workspace, we download in R from there (and not from WFS).

Differently from @emmanuel.blondel@fao.org and @fabio.sinibaldi@isti.cnr.it, I think that we should stick to standard protocols for accessing data and solve the issue with the deployment in the VRE. We can create a dedicated instance of GeoServer that is accessible only by the authorised VRE and not accessible outside the CNR network. This will allow to use the standard WFS protocol to access the data while ensuring secure access to the data.

What do you think?

#8 - Dec 07, 2016 01:38 PM - Emmanuel Blondel

Agreed in principle (always better to rely on standards..), but as usual it should not block the use case. Then it will be very easy to switch from one data access (read data from workspace, eg. ESRI shapefile) to another (WFS / GML). If you can work on it, it's good.

But if the need is to read a datasource in an R algorithm, and do not publish this source on the web, then Geoserver WMS/WFS is not strictly needed, we can take advantages of i-marine workspace (IMHO).

#9 - Dec 07, 2016 02:40 PM - Pasquale Pagano

I was highlighting another option based on the flexible deployment that we can offer in the VRE.

I am ok also with your proposal of exploiting the workspace, even considering that starting from next Tuesday (rollout of release 4.2) you will be able to read and write to the workspace directly from R.

#10 - Dec 07, 2016 02:45 PM - Levi Westerveld

Thank you for your input @emmanuel.blondel@fao.org , @fabio.sinibaldi@isti.cnr.it , and @pasquale.pagano@isti.cnr.it . From your suggestions, it seems that for now we should explore using the workspace of the VRE as a host for the geospatial data that should not be shared externally. Would this have any implication for the speed in terms of fetching the data (WFS from geoserver versus ESRI shp from workspace?). Could you please advise how to load geospatial data to the workspace and call it to an R algorithm? Maybe a tutorial is already available somewhere? Many thanks.
Levi

#11 - Dec 07, 2016 03:16 PM - Emmanuel Blondel

- For the download time from workspace, i never tried it. IRD has highlighted in the past some problems, that i could not verify. We should try it, and compare with a WFS call in R (in the meanwhile CNR can allow security at VRE level with the objective WFS)
- You need to zip the shapefiles and upload the ZIP to the workspace
- For reading in R:

** use the public link of this workspace ZIP file, by precaution use the long URL (data.d4science.org/...), not the short 'goog' URL (don't have clear details but we have been facing issues with the short URL). Also use http link instead of https (following CNR advice collected in parallel activity)

** unzip from R the downloaded file

** read the Shapefile, by preference use mapproj package / readShapePoly (or other method depending on the geometry type). This package is more robust for ESRI Shapefiles than the rgdal package / readOGR.

Hope this helps

#12 - Dec 08, 2016 02:28 PM - Levi Westerveld

- Status changed from Feedback to Closed

- % Done changed from 50 to 100

Dear @emmanuel.blondel@fao.org , thank you for your response. We will try to use the workspace for some of the data used in the MPA Intersect algorithm. I have currently uploaded the data to the geoserver (for the duration of the development of the web interface portlet to go on the VRE) but eventually we will need to move it to the workspace. Thank you all. Levi