D4Science Infrastructure - Task #5543

Questions about CKAN and CSW support (with possible application to FAO)

Oct 20, 2016 07:45 PM - Emmanuel Blondel

Status:	Closed	**Start date:**	Oct 20, 2016
Priority:	Normal	**Due date:**	
Assignee:	Francesco Mangiacrapa	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	Data Publishing		
Infrastructure:	Development		

Description

I have some questions about CKAN and its support of OGC Catalogue Service for the Web (CSW):

- would it be possible to **test** a direct harvesting of FAO Geonetwork CSW (with same filtering based on 'FIGIS' tag) by D4science CKAN? (this because i'm considering this really useful, but especially as possible alternative in case we would face issues of CSW harvesting between Geonetwork of different versions - considering that FAO Geonetwork will move to GN3, and that i already experimented issues of CSW harvesting across GN of different versions)
- Is the D4science CKAN already providing the CSW endpoint extension? (http://docs.ckan.org/projects/ckanext-spatial/en/latest/csw.html?highlight=CSW) If yes can we use such endpoint publicly? (from CSW clients)
- If the above is possible: what would be the impact on data discovery applications like the GeoExplorer? And, last but not least: If Geoexplorer relies on CSW: could we envisage to make this application relying directly on CKAN's CSW (for OGC geospatial products) instead of Geonetwork?

Many thanks in advance!

History

#1 - Oct 20, 2016 08:01 PM - Pasquale Pagano

- Assignee set to Francesco Mangiacrapa

- Target version changed from UnSprintable to Data Publishing

Hi Emmanuel,
the Product Catalogue was presented as a tool to collect data from other domain specific catalogues. GN is the one authoritative for GeoSpatial data and so we would keep the workflow as it was presented. However, the issues about different versions of GN is real and so we could investigate a different workflow that I would consider a temporary solution only.
Said that, I do not see CKAN as good as GN for geospatial data. It harvests ISO19139 standard metadata and it makes questionable mappings. Some of the information is lost in the harvesting procedure and therefore I think it cannot replace GN even considering the CSW extension.
Assigning the ticket to Francesco since he knows better than me the details.

#2 - Oct 20, 2016 08:10 PM - Emmanuel Blondel

Thanks Lino for these clarifications. To be honest, I don't know anything about CKAN, and then i was not aware of these limitations you highlight. On FAO side, i will need to test harvesting between GN3 and GN2 (in the past i already had CSW harvesting issues between different minor versions of GN2, so despite the fact we are talk of a standard CSW protocol and that we should not have issues *in theory*, this deserves to pay attention to it). In case you are interested, i will share my findings with your team.

#3 - Oct 21, 2016 09:02 AM - Pasquale Pagano

- Tracker changed from Support to Task

Yes, please keep us informed on the issues you will find.
I am changing the issue type since this is a task for Francesco to check and enable CSW on Ckan.

#4 - Nov 18, 2016 01:01 PM - Francesco Mangiacrapa

- Status changed from New to Closed

Hi Emmauel,

I apologize for the delay in replying.
As Lino said.. only GN is the one authoritative for GeoSpatial data, CKAN is a data catalogue that has a CSW support but:

1 - During harvesting from CKAN to GN some ISO 19139 metadata fields are lost (e.g. the "ISO Category");
2 - In order to provide CSW support over CKAN, We must install a new service "ckan-pycsw" (one for each CKAN hosted) and then create a dedicated database for pycsw. Well, this last is a waste of resources.

For the reasons explained at the moment only GN will be our authoritative CSW web-service.
I'm going to close this ticket. Thanks for you question