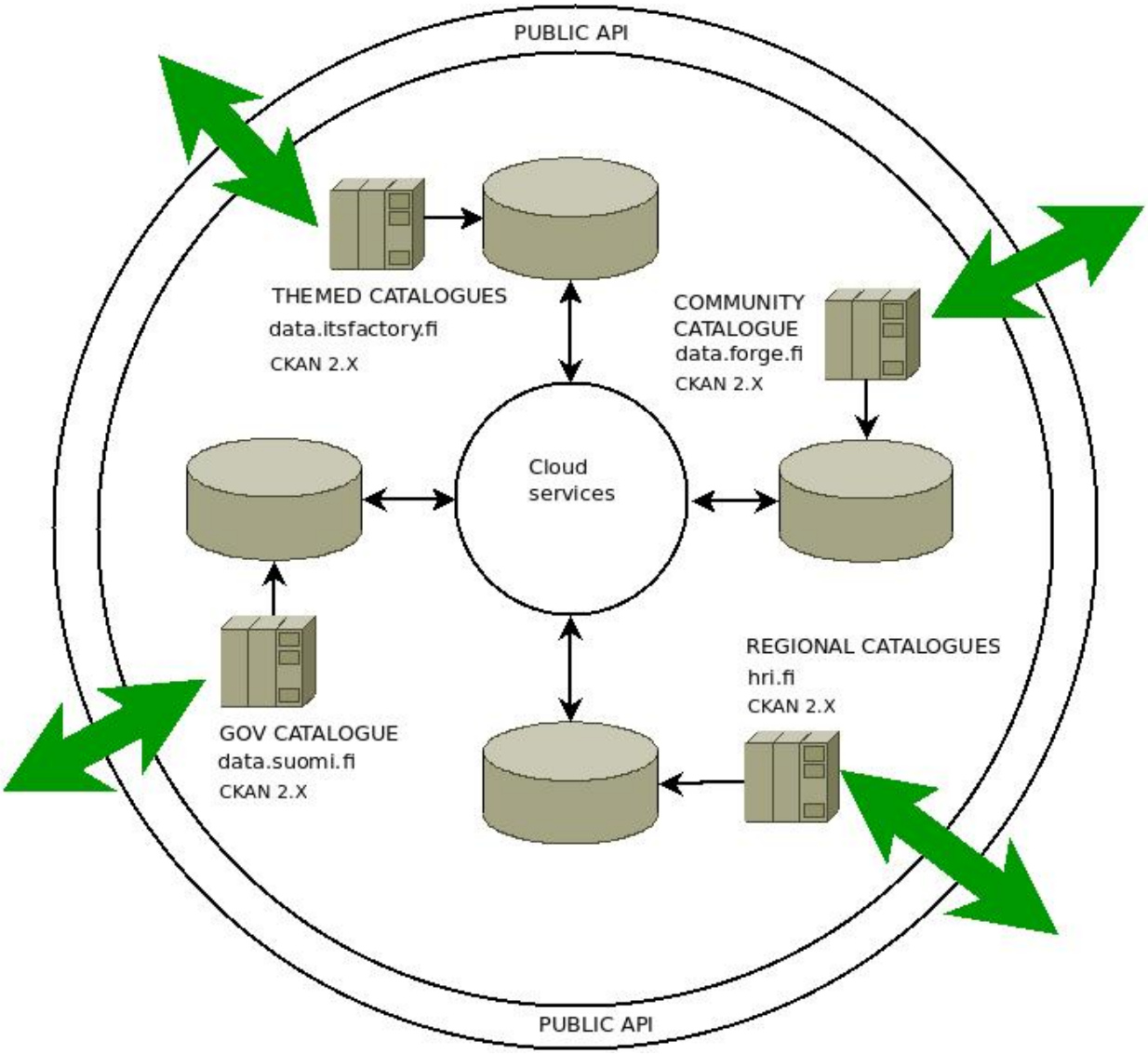
Mapping open data catalogue archtypes

Submitted on 03 Jun 2013 by Jarkko Moilanen

For the past few months I've had discussions with around two dozen people who are engaged in building Open Data movement in Finland. These people represent Open Knowledge Foundation Finland, Helsinki Region Inforshare, FORGE Service Lab, Ministry of Finance, ForumVirium, Finnish Tax Administration, State Treasury and Ministry of Transport and Communication, to mention a few.

Missing the big picture

During those discussions I found it a bit disturbing that the big picture was often neglected. The reason might be that others are more familiar with various projects and efforts that are underway. The other option is that there is no big picture (worst case scenario). Whatever the reason is, I started to give a little time to the subject and drafted a picture that might represent the direction in which we are heading. I've named it 'Future Open Data catalogues map' (see below). The map is not all inclusive and still lacks some clarity and details.



Map of major open data catalogue types

The idea is to build around cloud services which can be used to extend data storage almost infinitely. Of course initial data storage can be local hard drives, but if data sets will contain map data, data size will grow to a totally new level. Besides, cloud storage can be more easily scaled (with money). Cloud services do not need to be understood as Amazon only, Finnish operators are building more and more cloud based services.

Around the cloud services, you'll find 4 different kinds of archetypes of open data catalogues:

- Government catalogue,
- Themed catalogues,
- Regional catalogues and
- Community catalogue.

Government catalogue

Government open data catalogue is government-funded and a controlled 'central' open data

catalogue. It will be 'the' place to find open data in Finland. However, it will not (in an initial phase) contain data itself, but rather serve as metadata base. It can harvest and list open data from all over Finland. It is planned that the Gov Catalogue will provide a 'data hotel' in the future, but the timetable for that is not known. That in turn causes pressure to establish other places where different kind of parties can store open data.

Themed catalogues

Themed catalogues are one example of such a response to unavailable data hotel services. Different research and development-focused efforts such as ITS Factory have an instant need for data storage solutions. **ITS Factory** is a new innovation, experimentation and development environment, where companies and individual developers can develop, test and produce traffic solutions. As far as I know, ITS Factory does not have an open data catalogue yet, but that has been on the table.

Regional catalogues

Helsinki Region Infoshare is a practical example of **Regional catalogues** and aims to make regional information quickly and easily accessible to all. The site provides data that may be used by citizens, businesses, universities, academies, research facilities or municipal administration. Currently HRI contains over 1,000 data sets at least from Helsinki, Vantaa and Espoo. Other regions might be following the example since the model offers a chance to survive until the Gov catalogue will offer data storage as well.

If HRI is followed and CKAN-based systems are established for example in the Tampere region (including for example Tampere, Sastamala, Nokia, Pirkkala, Kangasala), transition to the Gov catalogue can be done with little efforts. In fact, establishing regional catalogues will leave regions more options in the future. 1) They can make the transition to Gov catalogue and 'abandon' the regional catalogue or 2) then just allow data harvesting and keep own system and still get all data listed in Gov catalogue and retain all control to data. Using cloud-based storage would make the options even more broad and flexible.

Community catalogue

The fourth type is a **Community catalogue**. This could be established for example in FORGE Service Lab. A community catalogue would not have the organizational and bureaucratic burden of the Gov catalogue. It would be a rather open catalogue for startups, developer communities and even individuals (my data approach). Putting a CKAN-based open data catalogue side by side with FORGE tools, it would make an ideal test bed for service and application development. FORGE is also (or remains to be seen) a reliable partner and most likely has a long-term future.

A community catalogue could also act as a home for municipal open data. The regional catalogues discussed above, would be formed, funded and controlled by groups of municipalities. That is the natural approach if such groups can be formed more or less naturally — i.e. they already have joint IT or other activities and businesses. However, not all municipalities are in such a good situation. There are around 300 municipalities in Finland and some of those will most likely have to survive 'alone' with open data. For them, a Community catalogue could offer a 'home', affordable joint

platform.

Discussion

The above archetypes are based on a many discussions with Finnish Open Data efforts-related people and it is likely that there are flaws and gaps. I will continue the model development as time goes by and projects like Gov Catalogue and FORGE reach operational level. Until then it's just speculation and represents a possible model for the future Open Data catalogue map for Finland. I'd love to hear your thoughts on that.