



NATURAL RESOURCES CANADA - INVENTIVE BY NATURE

Arctic SDI catalogue harvesting

*Polar Data and Systems Architecture
Workshop, Geneva Switzerland*

Presented on Nov 29th 2018 by Sampo Savolainen / Spatineo



Natural Resources
Canada

Ressources naturelles
Canada

Canada

Background

Natural Resources Canada has contracted Spatineo to produce an automatically harvested catalogue of Canadian services

- Replaced a manually maintained catalogue
- Cheaper and more complete

This method is now being applied for the Arctic SDI as well

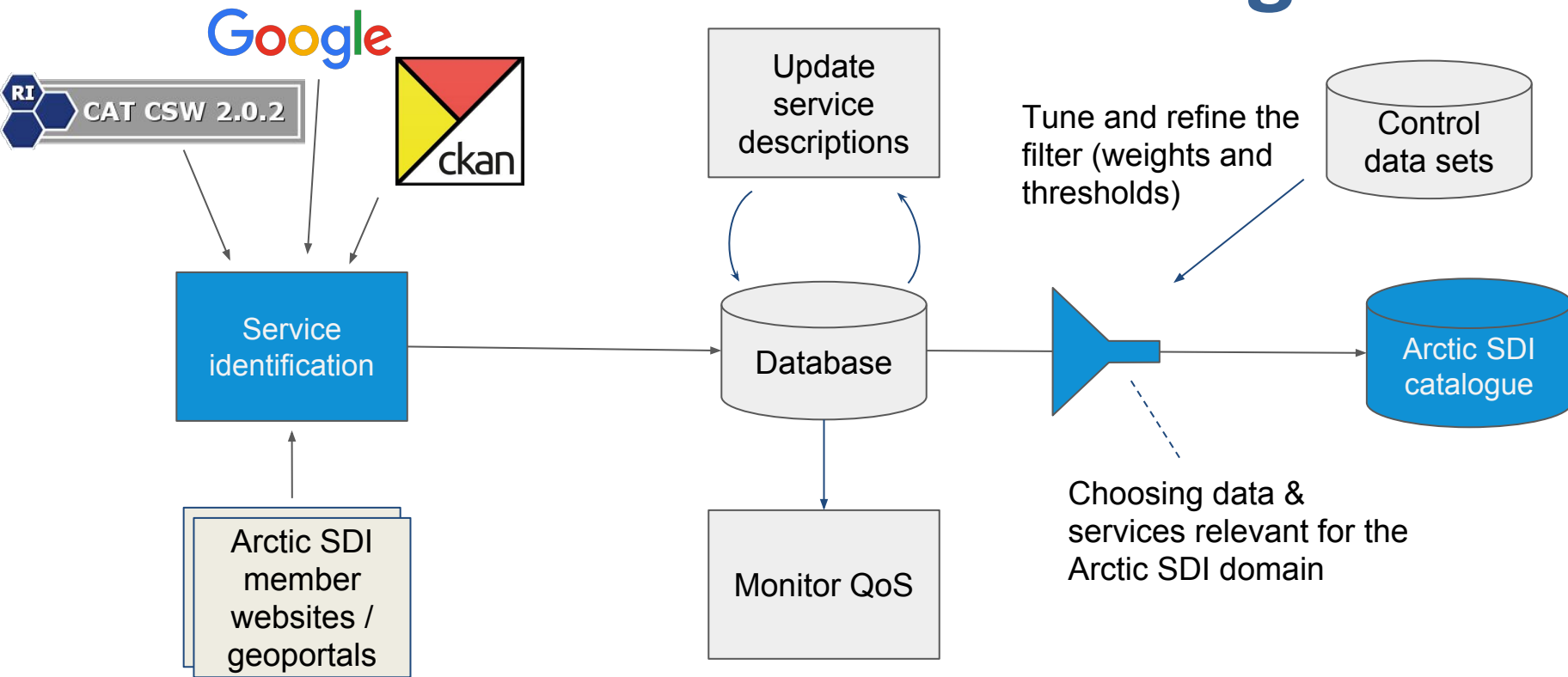


GIS software company based in Finland

- SaaS products for quality assurance (usage analytics, reporting, availability monitoring, performance testing)
- Expert services

Spatineo maintains a world-wide catalogue of spatial web services (currently 99k services, 1,8M layers)

Automated harvesting



© Her Majesty the Queen in Right of Canada, as represented by the Minister of Natural Resources, 2017



Natural Resources
Canada

Ressources naturelles
Canada

Canada

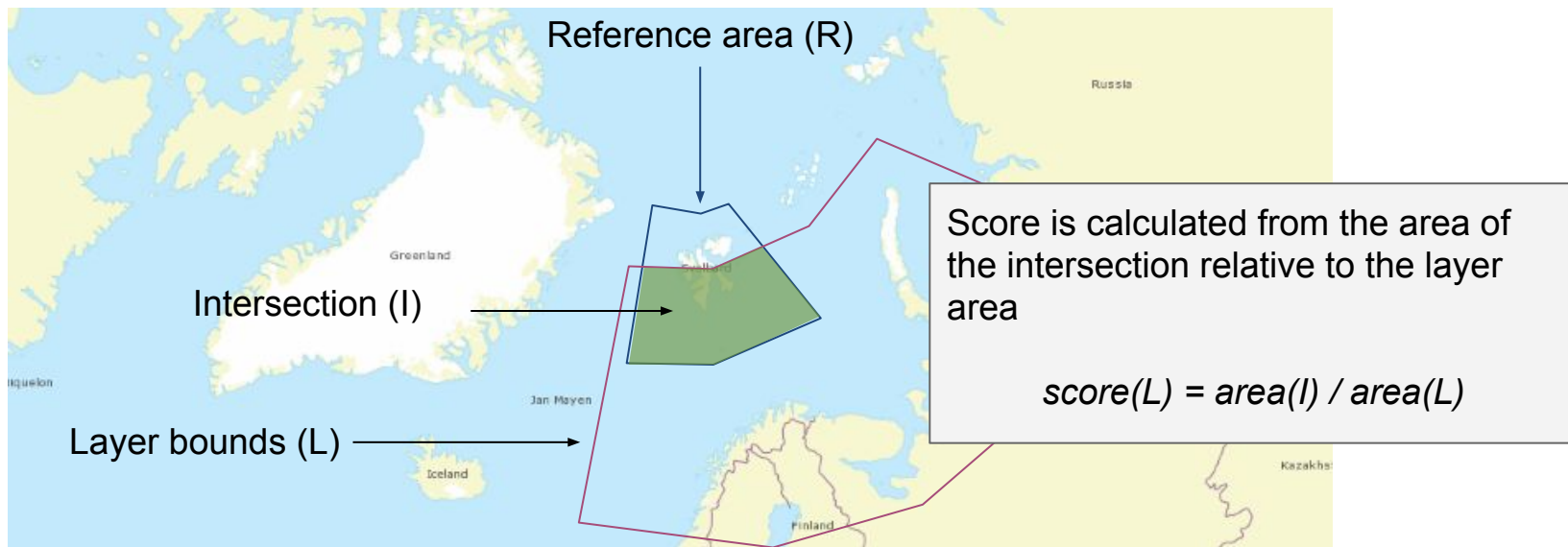
How to determine relevance?

1. Text analysis
 - Keywords: ArcticSDI, arctic, albedo, sea ice extent
 - Geographical names unique to the Arctic
2. Technical attributes
 - Polar projection support
 - Internet domain name (identify certain agencies)
3. Geographical scoring

Each scorer component produces a score (0.0 - 1.0)

Geographical scoring

All layers are scored by calculating the relative area of the intersection between the reference area and the layer. This score gives an accurate number on how much of that dataset is relevant to the reference area.



© Her Majesty the Queen in Right of Canada, as represented by the Minister of Natural Resources, 2017

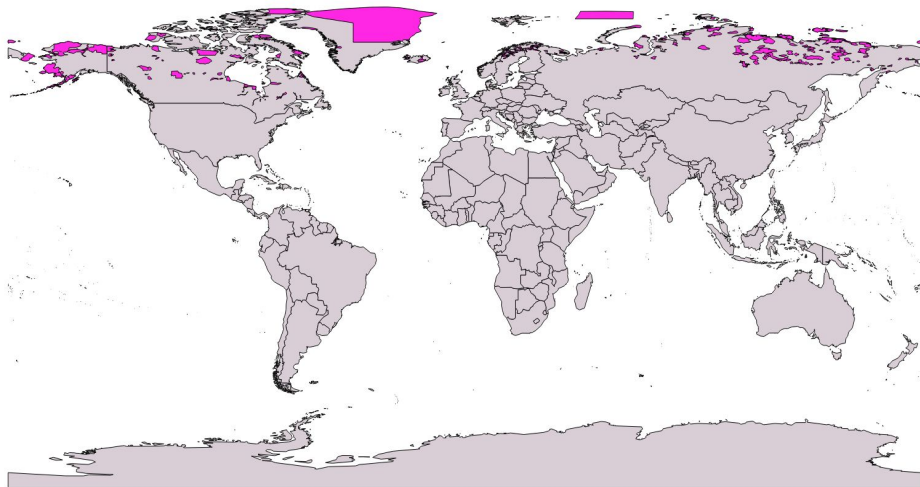


Natural Resources
Canada

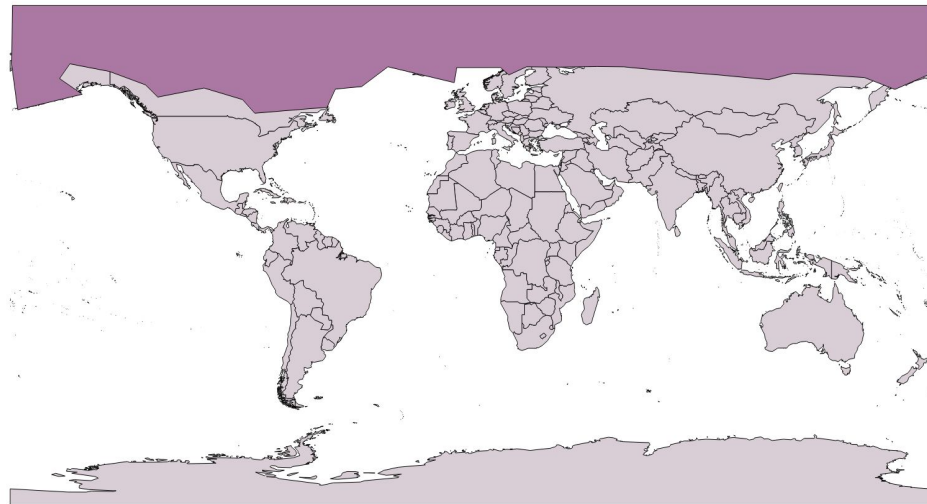
Ressources naturelles
Canada

Canada

Reference areas used



Protected areas in the Arctic



Arctic treeline (simplified)

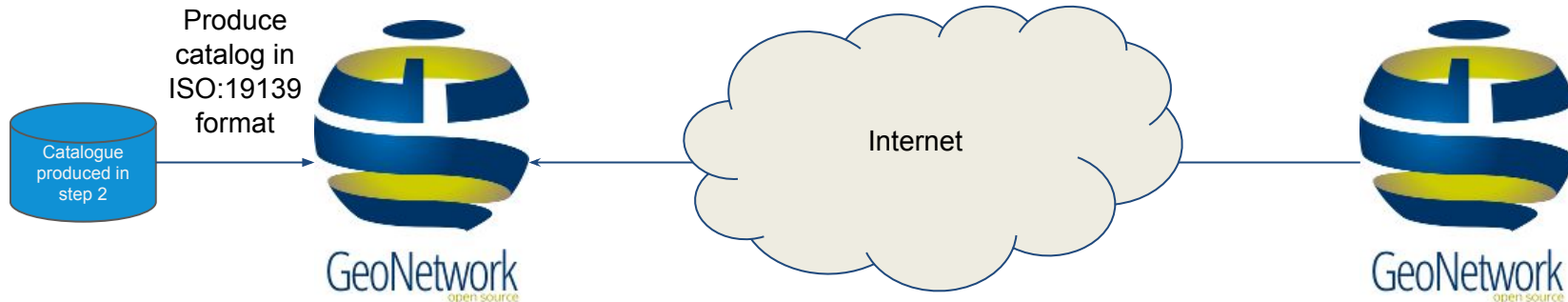
Evidence to evaluation

A training set of services has been collected and evaluated using the different scorers

Linear logistic regression was used to produce a model to produce an evaluation from the individual scores.

Resulting catalogue: **6389 services, 333390 datasets/series**

Transferring the catalogue to Arctic SDI



The GeoNetwork hosted by Spatineo will be refreshed from the catalogue information periodically

The Arctic SDI GeoNetwork will be set up to periodically harvest the GeoNetwork service hosted by Spatineo

Metadata in the catalogue

Catalogue has metadata for services and data sets:

- Use metadata produced by the service provider when the service or dataset has working link to it
- Otherwise Spatineo produces metadata based on the service/offering description (Capabilities document in most cases)

Catalogue access

<https://arctic-sdi-catalogue.spatineo-devops.com>

Username: arcticsdiuser

Password: sdi4arctic

Note! This is the source catalogue provided by Spatineo, not an official Arctic SDI catalogue.



Search over 339798 data sets, services and maps, ...

Browse by topics



Browse resources



Example links

Service Provider service metadata:

<https://arctic-sdi-catalogue.spatineo-devops.com/geonetwork/srv/eng/catalog.search#/metadata/da77377c-f271-44fc-9951-c795f9f96af1>

Service Provider data set metadata:

<https://arctic-sdi-catalogue.spatineo-devops.com/geonetwork/srv/eng/catalog.search#/metadata/fr-120066022-jdd-9a591d92-0f9d-40a9-88e7-bd68810cfc5f>

Generated service metadata:

<https://arctic-sdi-catalogue.spatineo-devops.com/geonetwork/srv/eng/catalog.search#/metadata/e803c2e8-fd61-4c2f-9df9-8df2dd554301>
<https://arctic-sdi-catalogue.spatineo-devops.com/geonetwork/srv/eng/catalog.search#/metadata/d2d2642c-c4db-4918-ae43-9d2d22071af9>

Generated data set metadata:

<https://arctic-sdi-catalogue.spatineo-devops.com/geonetwork/srv/eng/catalog.search#/metadata/d33e2d40-cda3-4777-b405-fdc8993d7580>

What's next?

- Improving the training set and splitting it into a training and validation set (currently no metrics on quality)
- Currently a service and its datasets are evaluated as a whole, this should be separated
- Harvesting to Arctic SDI (status unclear)
- Resolve performance issues with Geonetwork
- Metadata by the service provider is used as-is, unclear how to amend it to include QoS and scoring information without making a conflicting version of the canonical record

Thanks!

Cameron Wilson, Simon Riopel, Eric Wright / Natural Resources Canada
Sampo Savolainen / Spatineo

