

NATURAL RESOURCES CANADA - INVENTIVE BY NATURE

# A method for producing a catalogue for Arctic SDI

*Polar Data Forum III, 2019-11-21*

*Spatineo: Sampo Savolainen*

*NRCCan: E. Wright, S. Riopel, C. Wilson*



Natural Resources  
Canada

Ressources naturelles  
Canada

Canada

# Catalogues are essential for SDIs

Aggregating information about services and data is essential:

- Reduce duplicated effort, in both data collection and distribution
- Understand what information is available and what needs further collecting / research - “a map of data”
- Discovery of relevant data provided outside of the SDI

Without a catalogue, an SDI cannot reach its full potential



# Catalogue maintenance

Canadian CGDI catalogue was maintained manually, but this approach abandoned due to quality, sustainability and cost:

- Records need to be added, updated, removed per request
- However not all changes were reported: active maintenance was required
- No process to add supplemental information important for users (for example QoS)

Solution was to automate the catalogue production.

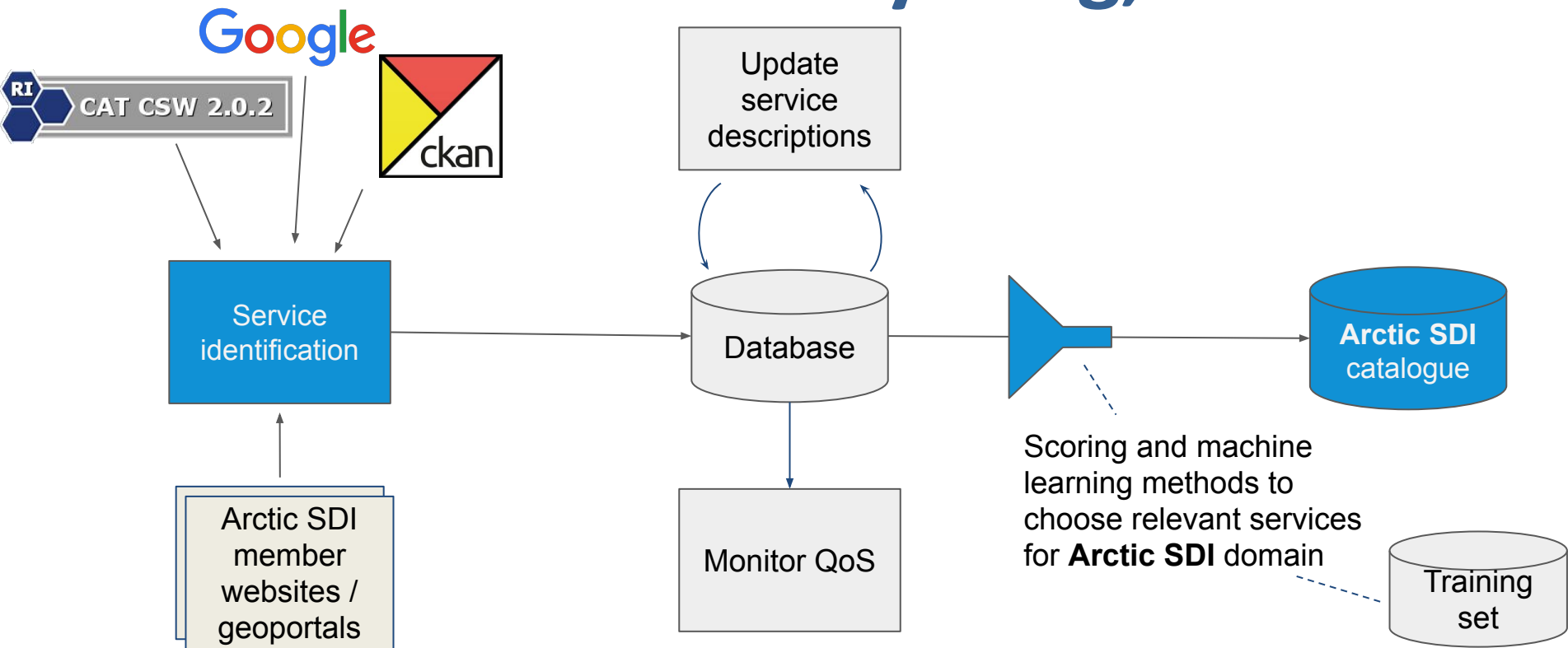
The same approach has been investigated for the Arctic SDI.

# Efficient cataloguing of “everything” Arctic?

- Arctic data and services are spattered across a wide array of research fields, organisations, domains
- An ever-increasing amount of data products and services providing that data
- No standardisation body, no harmonized metadata conventions

=> Advanced methods of determining the relevancy to the Arctic are required

# Solution: Harvest everything, then filter



© Her Majesty the Queen in Right of Canada, as represented by the Minister of Natural Resources, 2019



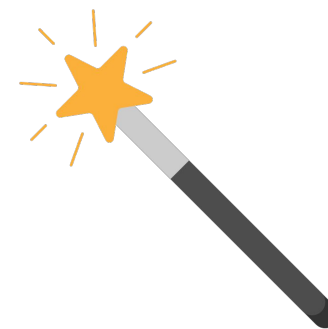
Natural Resources  
Canada

Ressources naturelles  
Canada

Canada

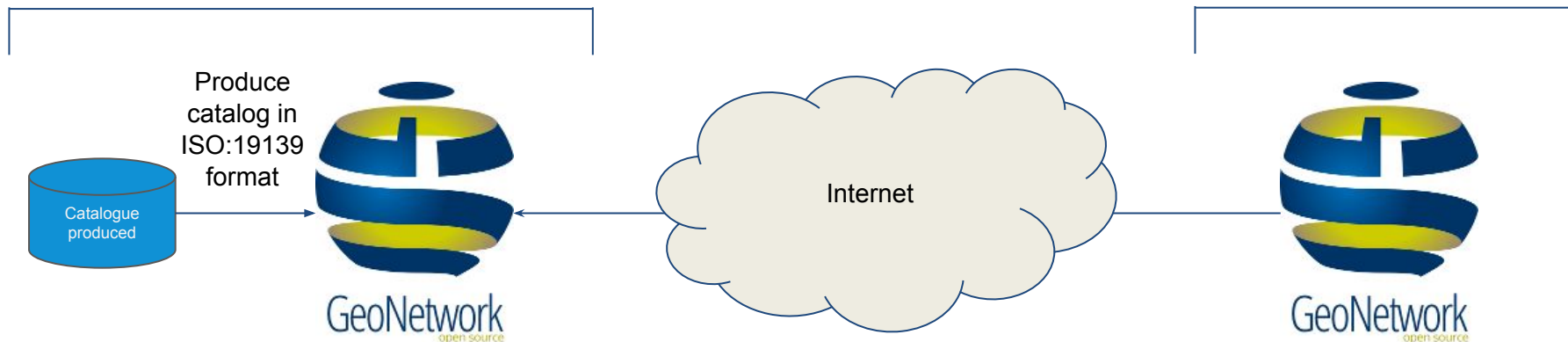
# Magic Sauce for filtering

1. Multiple heuristics are used to extract signal from service and dataset descriptions
2. Training set of positives and negatives are required
3. Machine learning (linear regression) is used to train a model to evaluate relevance



(Similar approaches tested in OGC Testbed 15)

# Transferring the catalogue to Arctic SDI



The GeoNetwork hosted by Spatineo to be used as a source for the official catalogue

The Official Arctic SDI GeoNetwork will harvest the GeoNetwork service hosted by Spatineo

© Her Majesty the Queen in Right of Canada, as represented by the Minister of Natural Resources, 2019



Natural Resources  
Canada

Ressources naturelles  
Canada

Canada

# What's next?

Work was paused waiting for contract renewal,  
now we are ready to continue

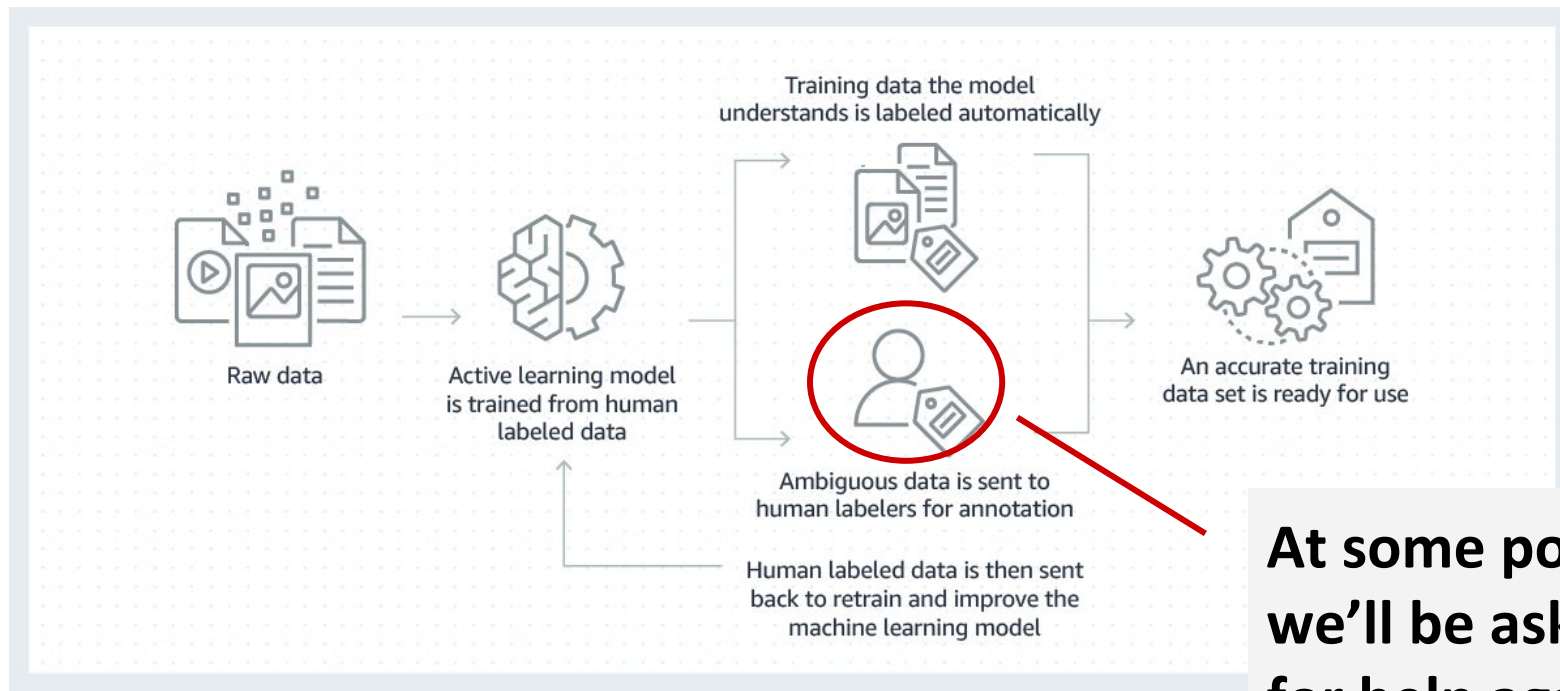
Improve operational capabilities

Improve training set





# Building the training set



**At some point,  
we'll be asking  
for help again!**

Image © 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

© Her Majesty the Queen in Right of Canada, as represented by the Minister of Natural Resources, 2019



Natural Resources  
Canada

Ressources naturelles  
Canada

Canada