

The Automated Data (layer) Harvester

Adding Scalability
to the
NBN System



THE UNIVERSITY
of EDINBURGH



What was this project about?

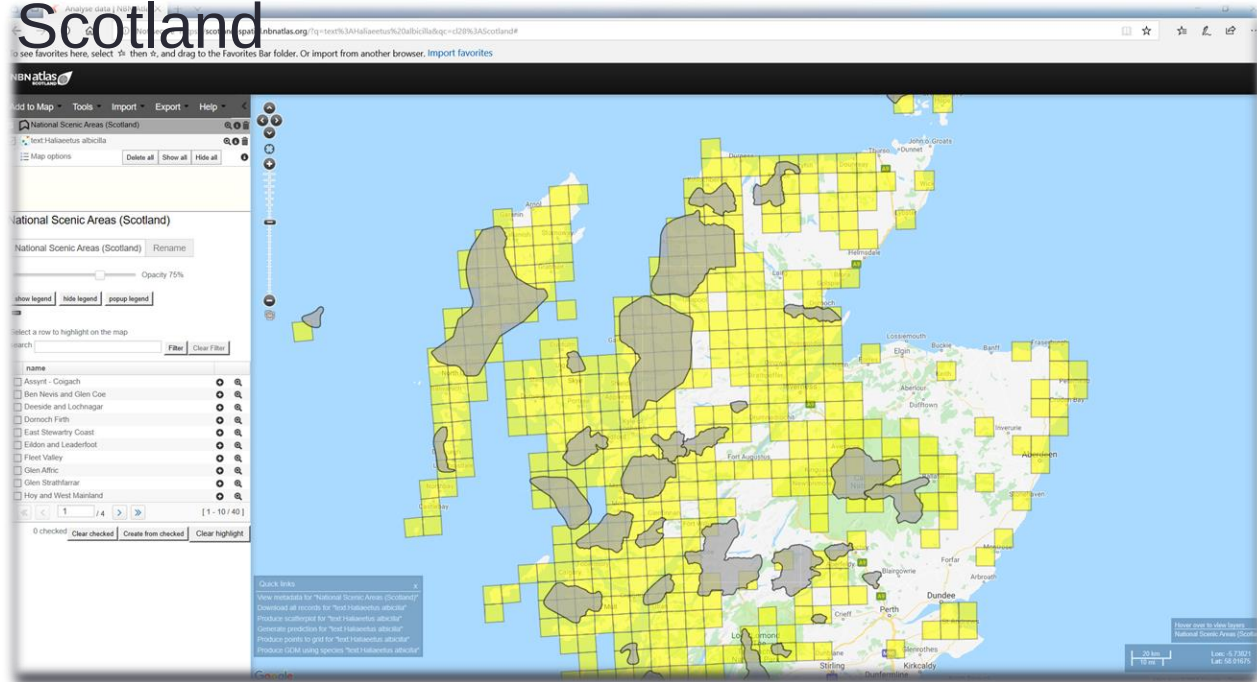
- Data engineering project

- NBN Atlas Scotland

- SEPA

- SNH

- EPCC



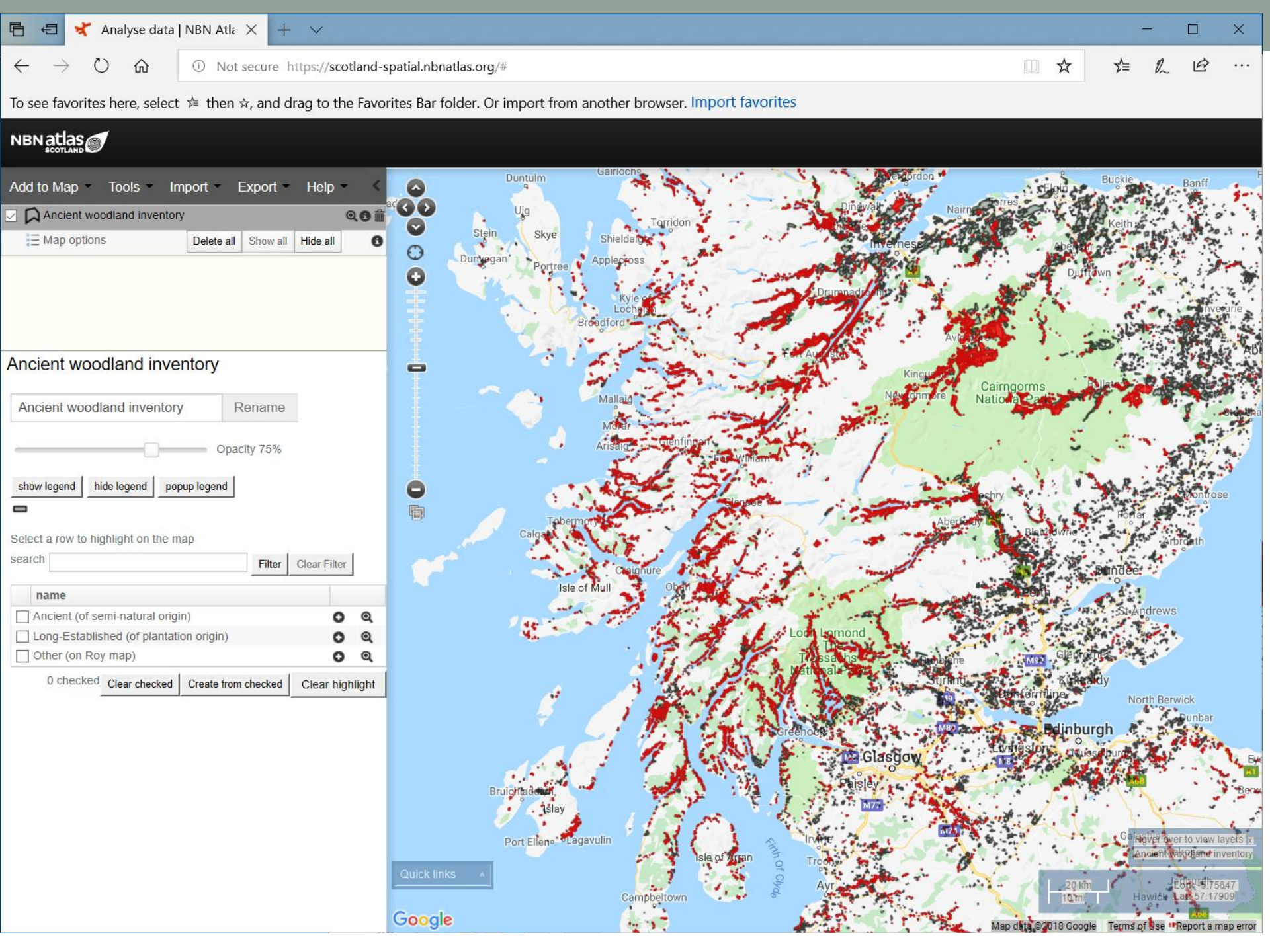
- Examine system architecture

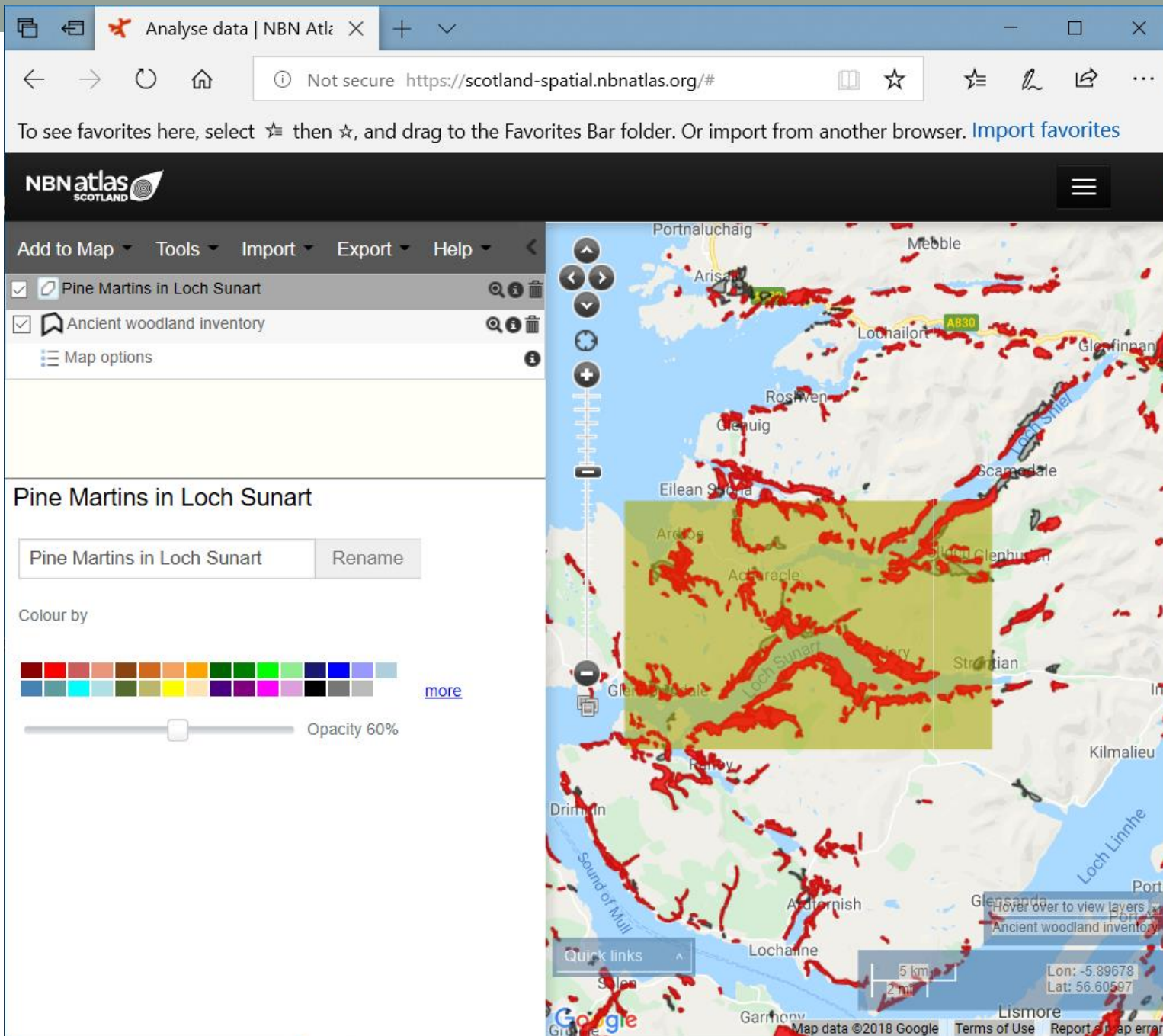
- Come up with ideas to improve the NBN Atlas Scotland system

Who are EPCC?









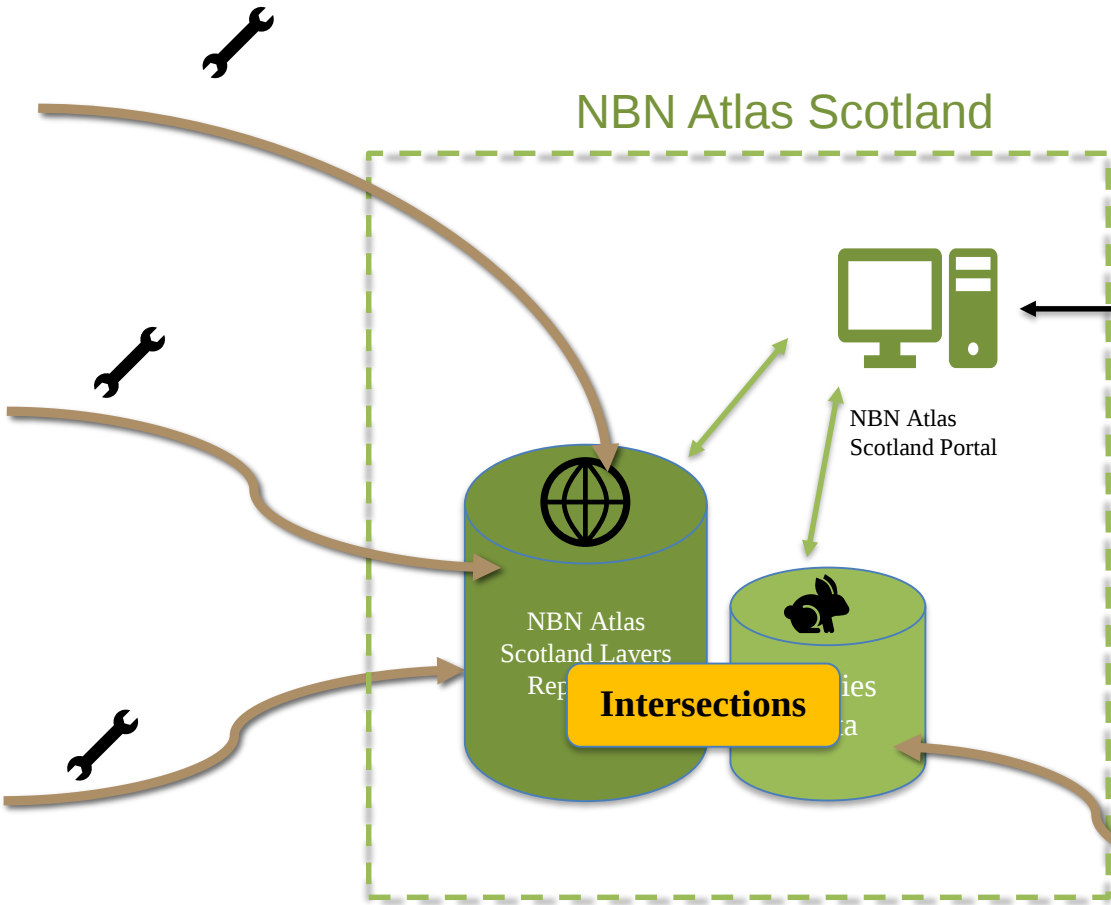
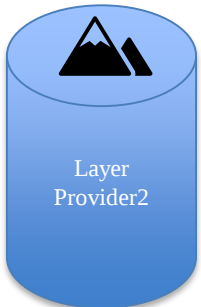
- [illegible]

- Create a system that we can build upon in the future.

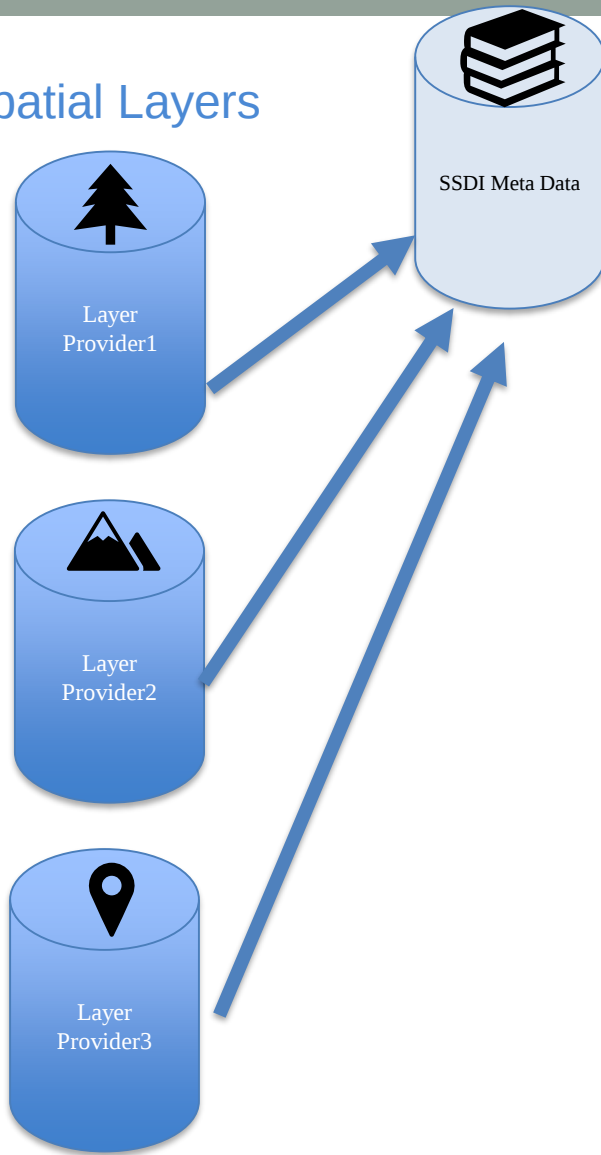
NBN Atlas Scotland High Level View



Spatial Layers Providers

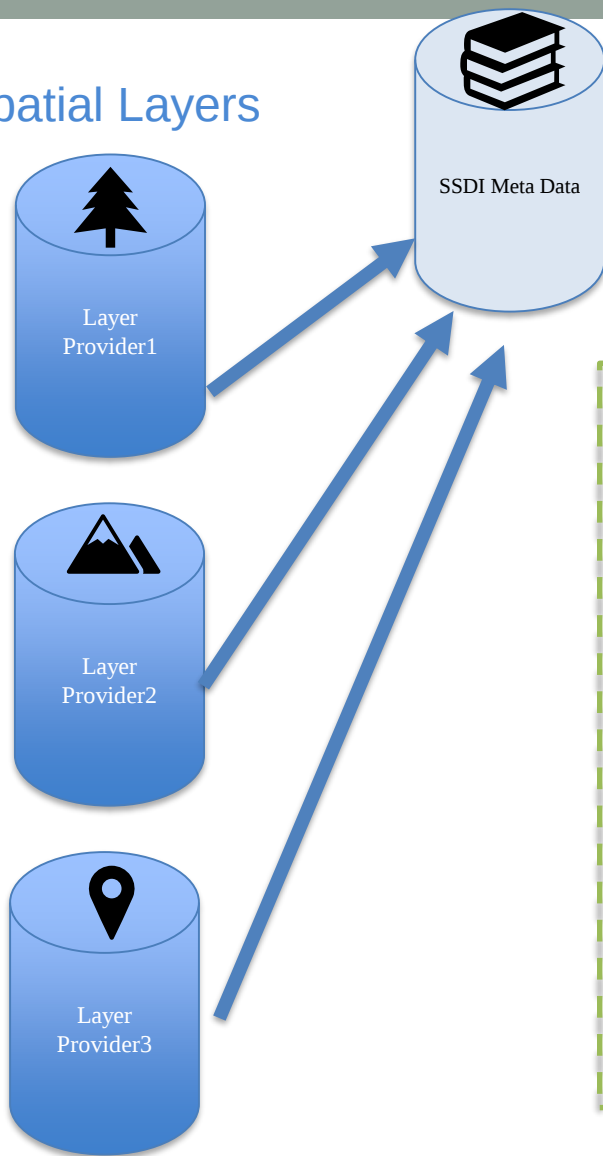


Spatial Layers



SSDI • Scottish Spatial
• Data Infrastructure
• Metadata Portal

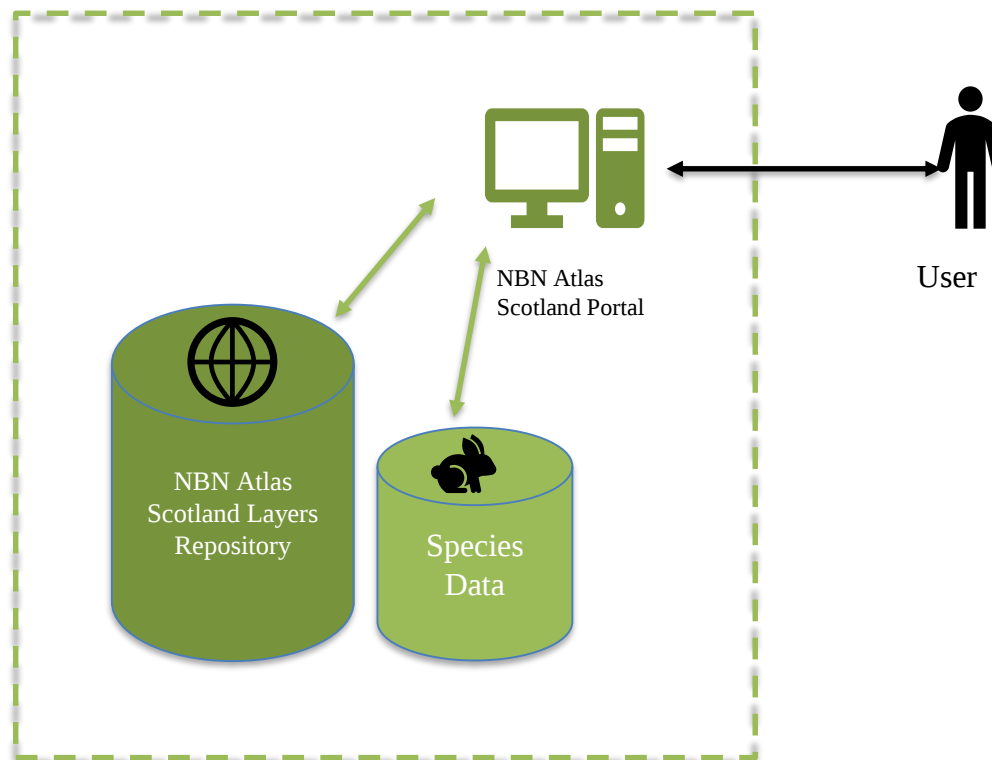
Spatial Layers



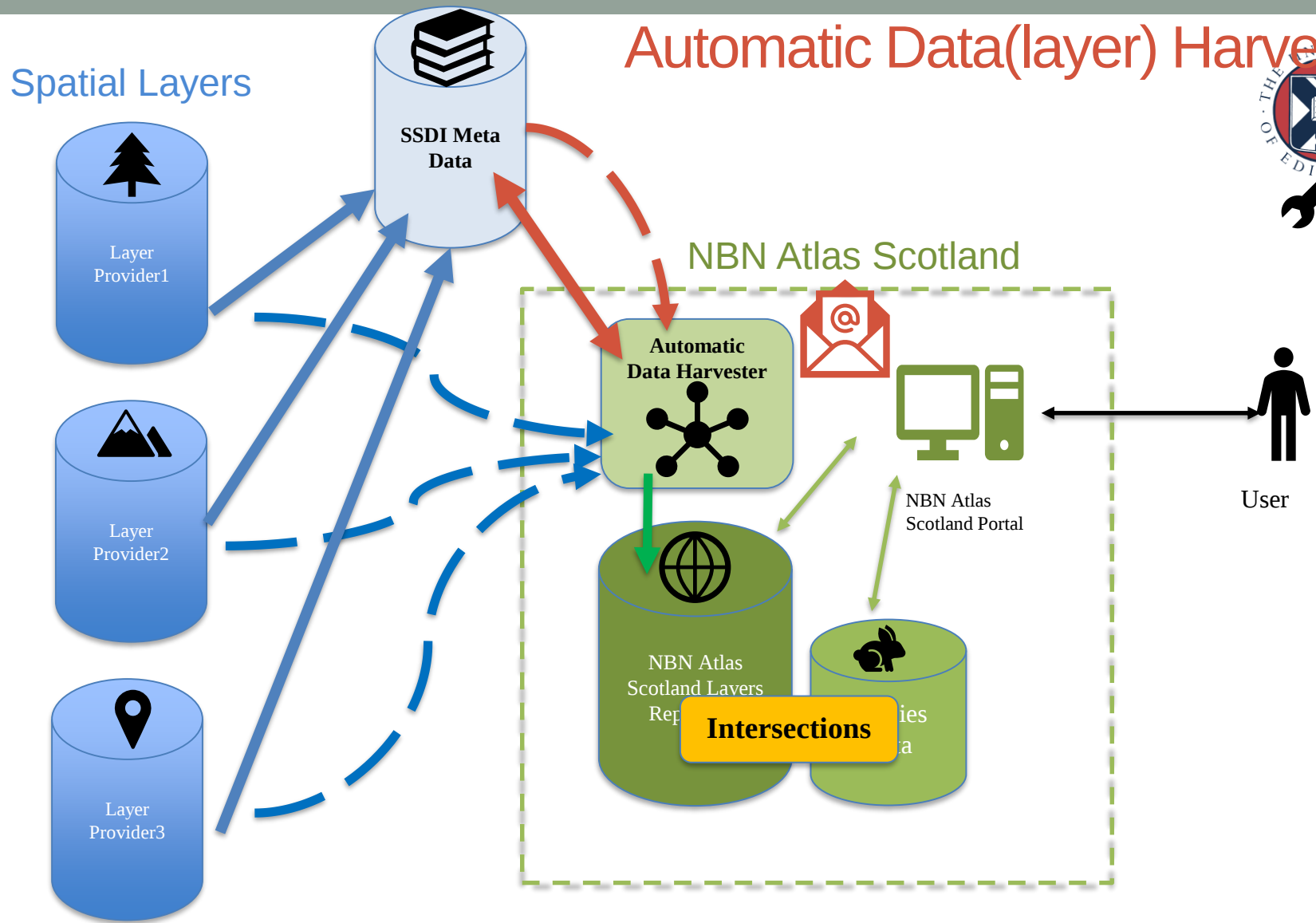
SSDI

Scottish Spatial
Data Infrastructure
Metadata Portal

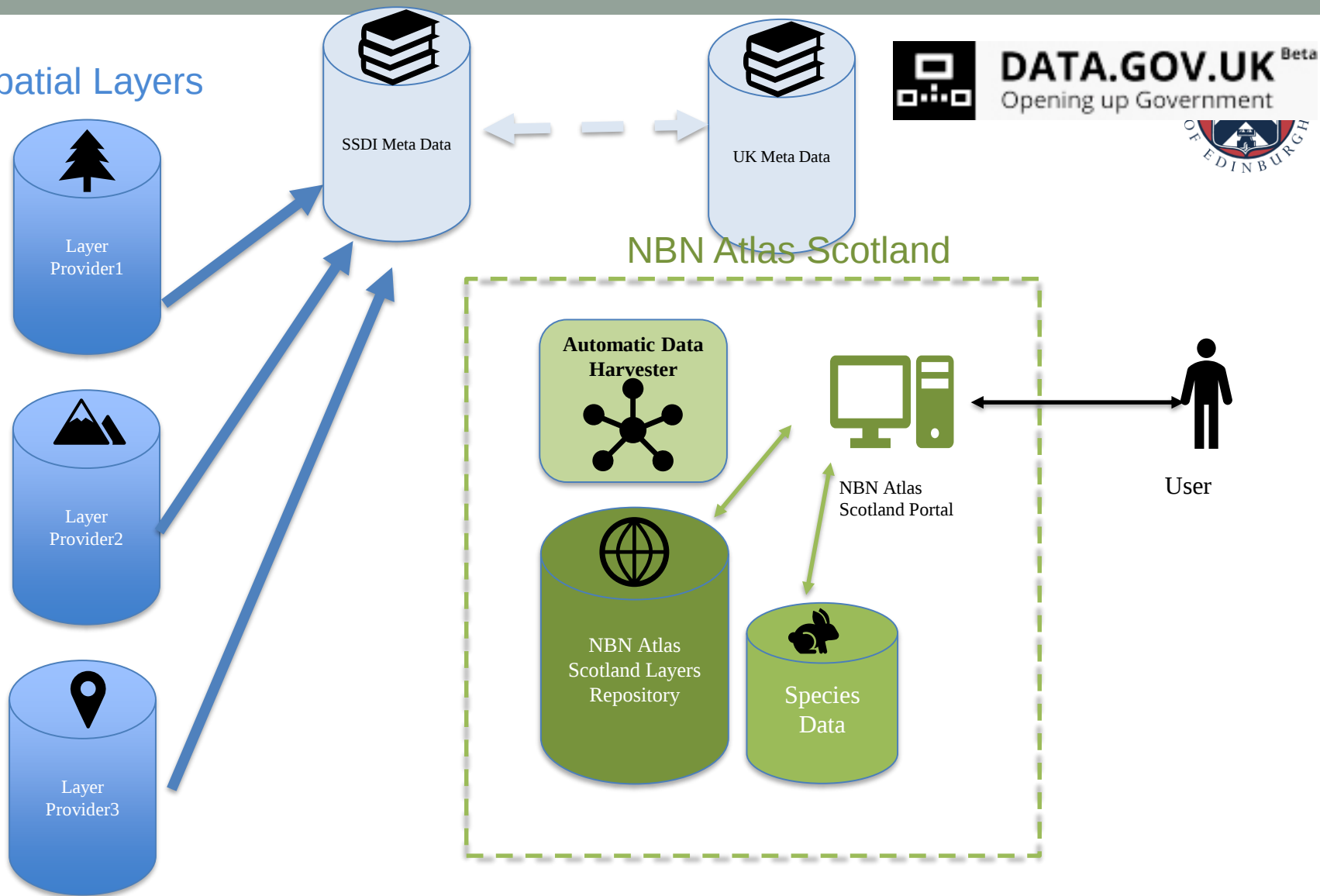
NBN Atlas Scotland

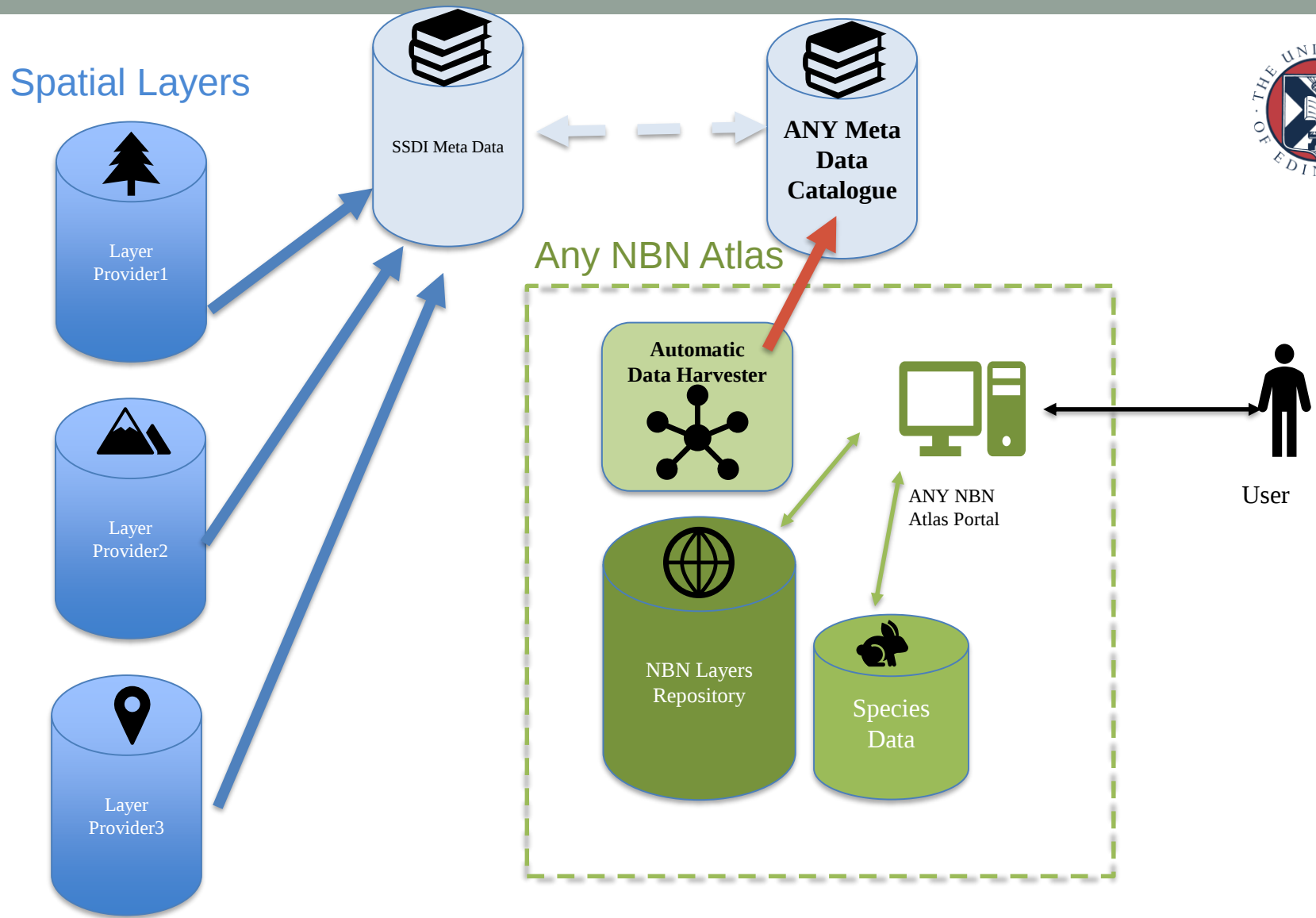


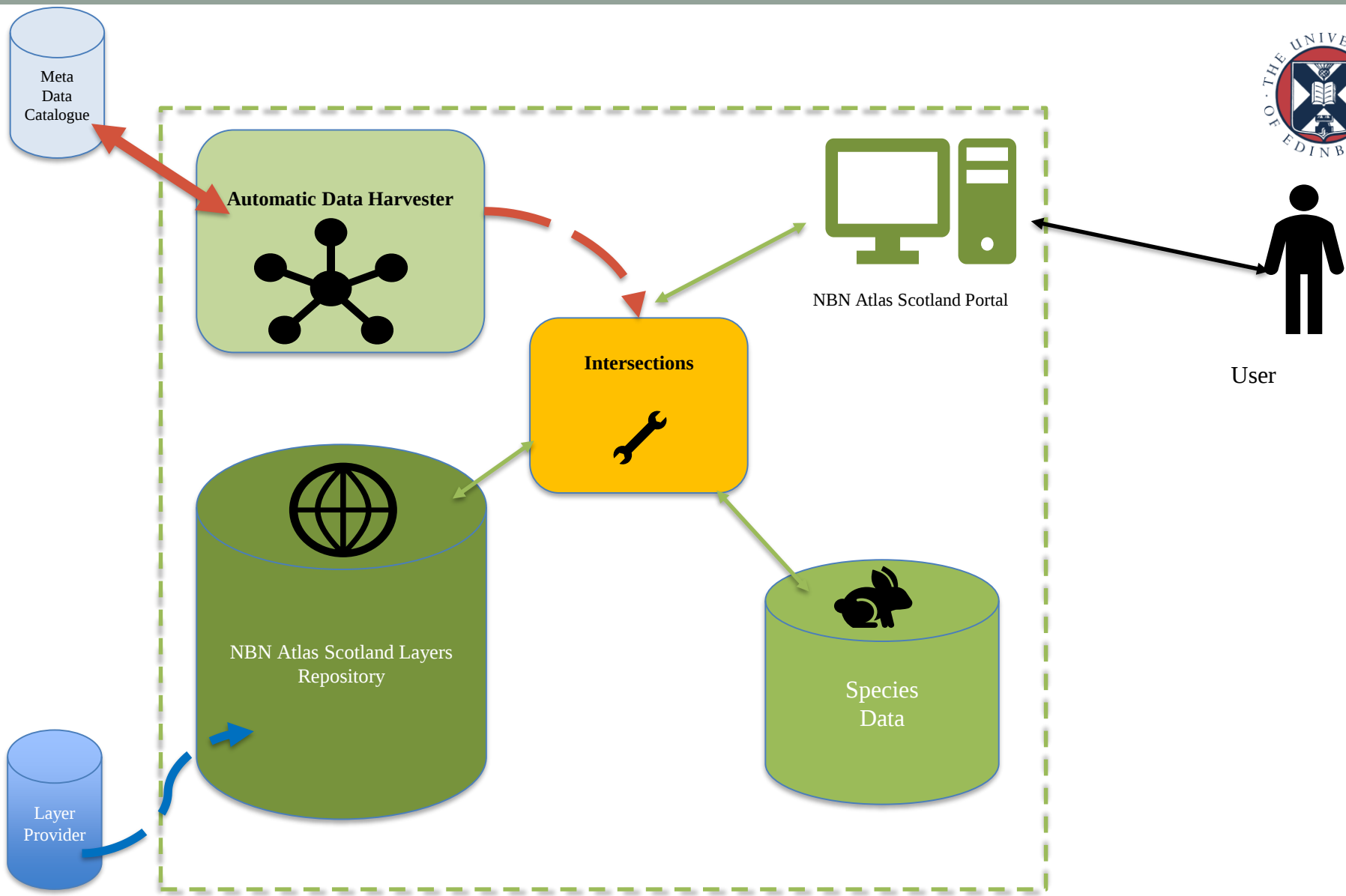
Automatic Data(layer) Harvester



Spatial Layers







Summary of ADH functionality:

- Queries meta data service
- Keeps and tracks status and age of current layers in repo.
- Pulls layer data to local place (directory)
- Uploads to repo for Admin to “switch on”
 - can be automated – currently manual
- Emails admin to tell that layer is ready to be processed and ready to be “switched on”

Making NBN Scotland Atlas (more) dynamic



Short-medium

- Automated intersections
 - Functionality exist, just needs to be
 - 1) optimised
 - 2) automated
 - Currently some layers take 8+ hours
- Automated quality checks
 - Currently checks layer is “complete” and valid: syntactic, not semantic

Making NBN Scotland Atlas (more) dynamic



Long term

- Turn into a proper, authorised ALA component – fully configurable, “plug ‘n’ play”.
- Making a generic Automatic Data Harvester – just spatial layers at the moment – hook up to species meta data catalogues
- Generic credential management to allow an ADH to query any meta data service, and pull data from any source

ADH as a basis for predictive analytics



- Looking for data outliers in species data, quality control
- Habitat changes - knock on predictive changes to species?
- And other way around – invasive species impact on habitat using predictive analytics?
- Current infrastructure seems sufficiently “elastic” – however components do need to be optimised for efficiency as data grows.



Impatiens glandulifera –
Himalayan Balsam on a Highland shore

Contributors

Sophie Ratcliffe, NBN

Rueben Roberts, NBN

Christine Johnston, NBN

Olivia Gill, SEPA

Paula Brown, SEPA

Ed Mackey, SNH

Rob Baxter, EPCC

Alistair Grant, EPCC

Ally Hume, EPCC

Amy Krause, EPCC

Davy Virdee, EPCC

Questions?

