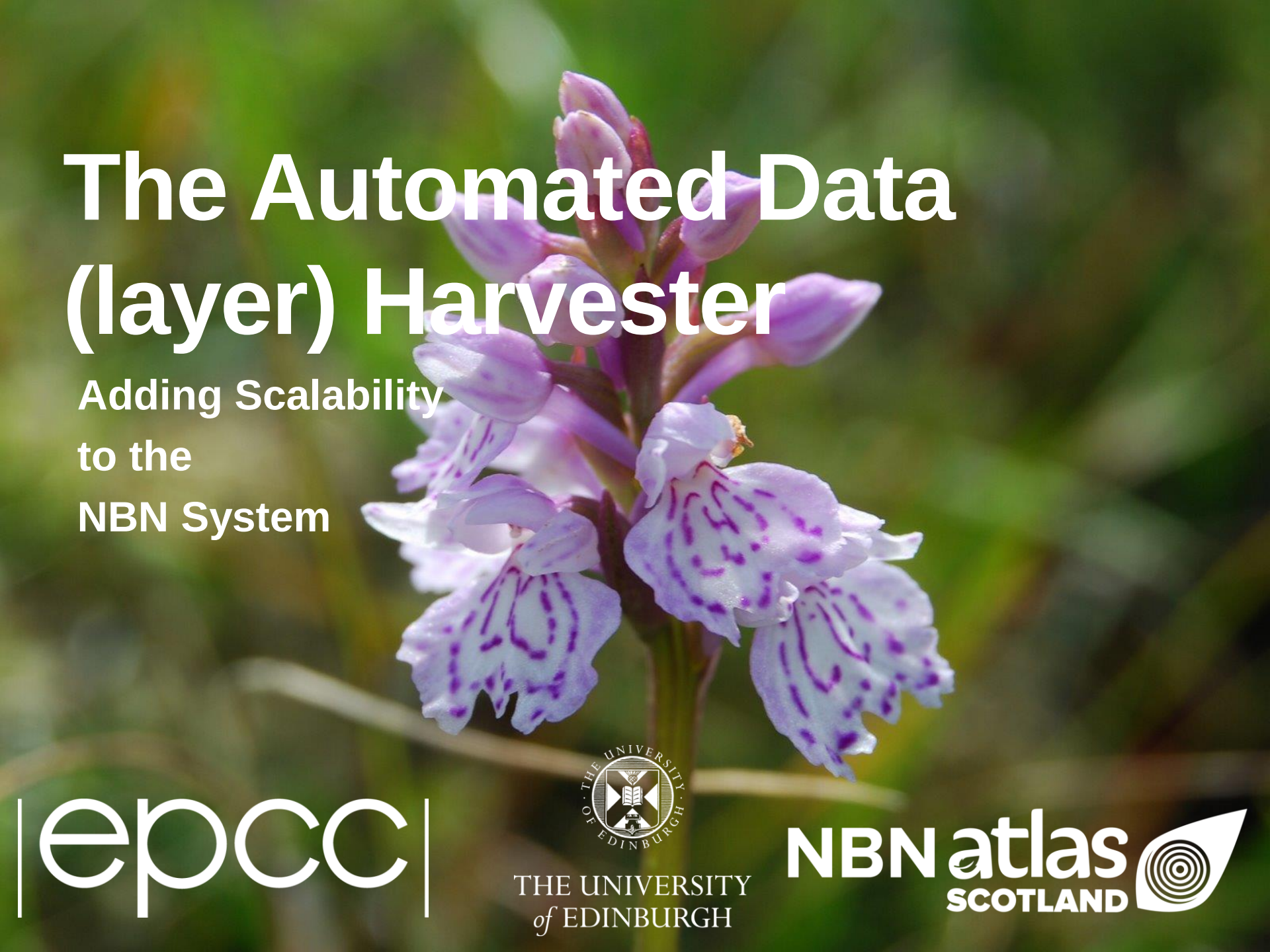


The Automated Data (layer) Harvester

Adding Scalability to the NBN System

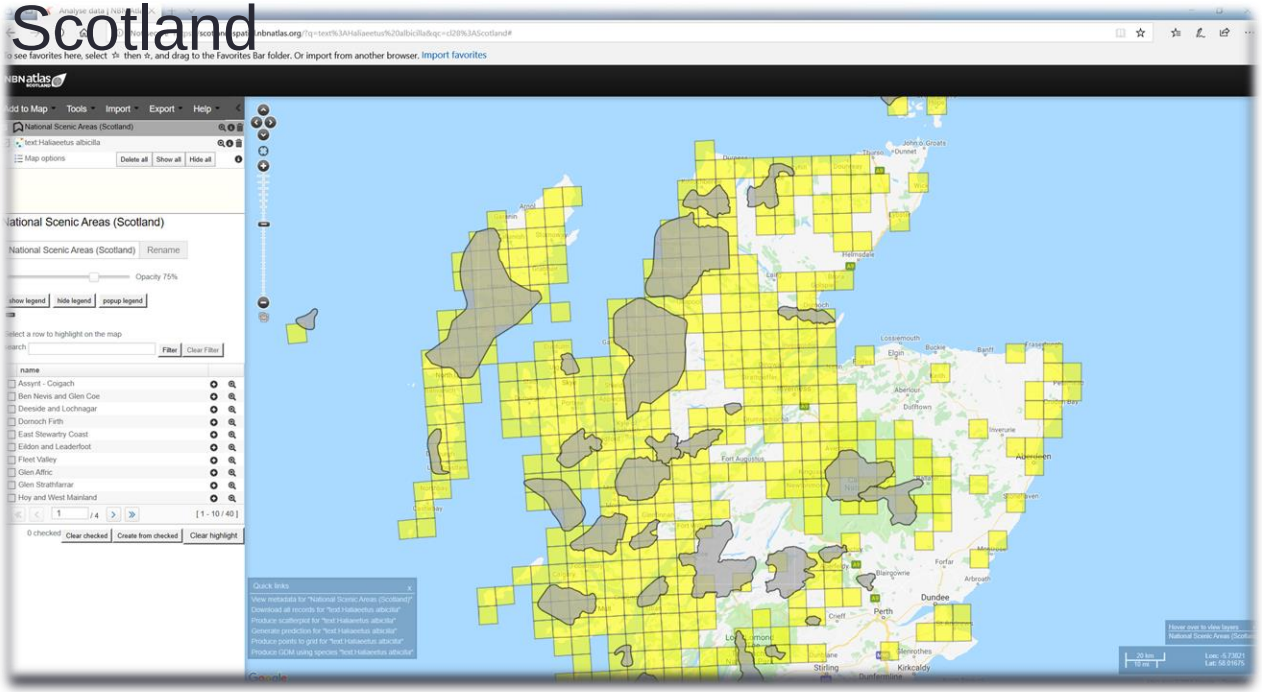
epcc

THE UNIVERSITY of EDINBURGH

NBN atlas SCOTLAND

What was this project about?

- Data engineering project
  - NBN Atlas Scotland
  - SEPA
  - SNH
  - EPCC

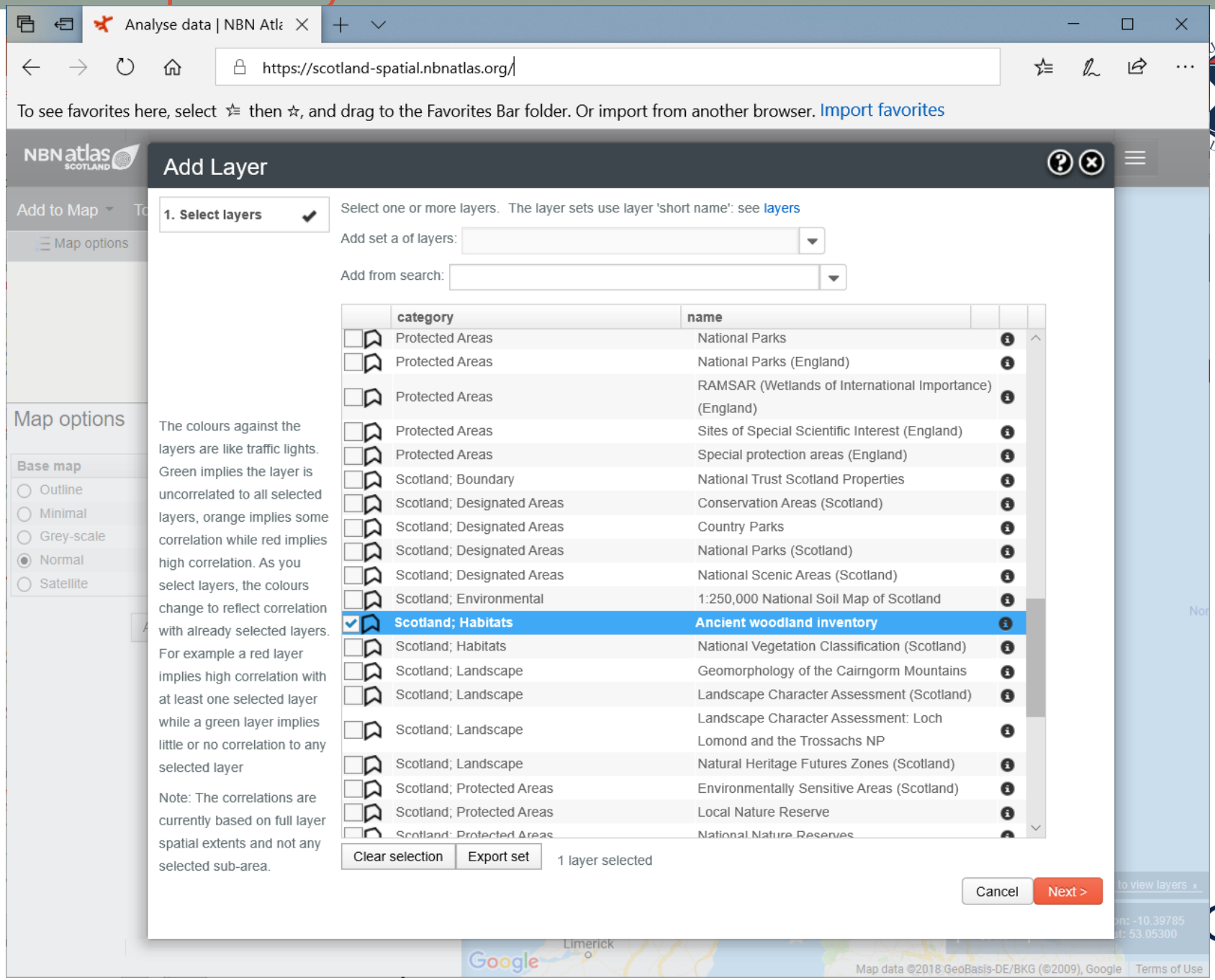


- Examine system architecture
- Come up with ideas to improve the NBN Atlas Scotland system

Who are EPCC?

EPSRC
Engineering and Physical Sciences
Research Council
archer
NATURAL
ENVIRONMENT
RESEARCH

How are Spatial Layers used?



Analyse data | NBN Atla

Not secure https://scotland-spatial.nbnatlas.org/#

To see favorites here, select ☆≡ then ☆, and drag to the Favorites Bar folder. Or import from another browser. Import favorites

NBN atlas SCOTLAND

Add to Map | Tools | Import | Export | Help

Ancient woodland inventory

Map options | Delete all | Show all | Hide all

Ancient woodland inventory

Ancient woodland inventory | Rename

Opacity 75%

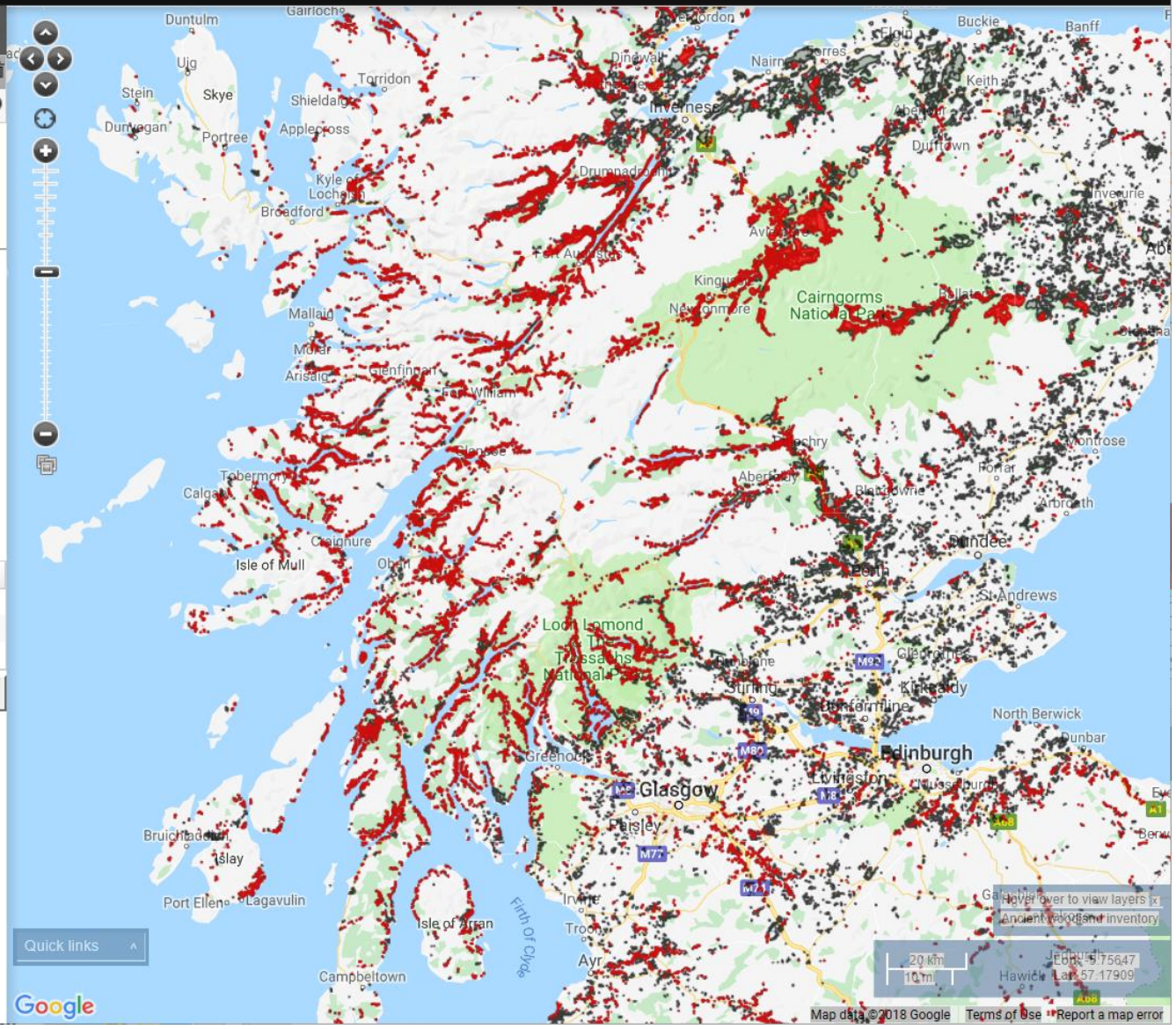
show legend | hide legend | popup legend

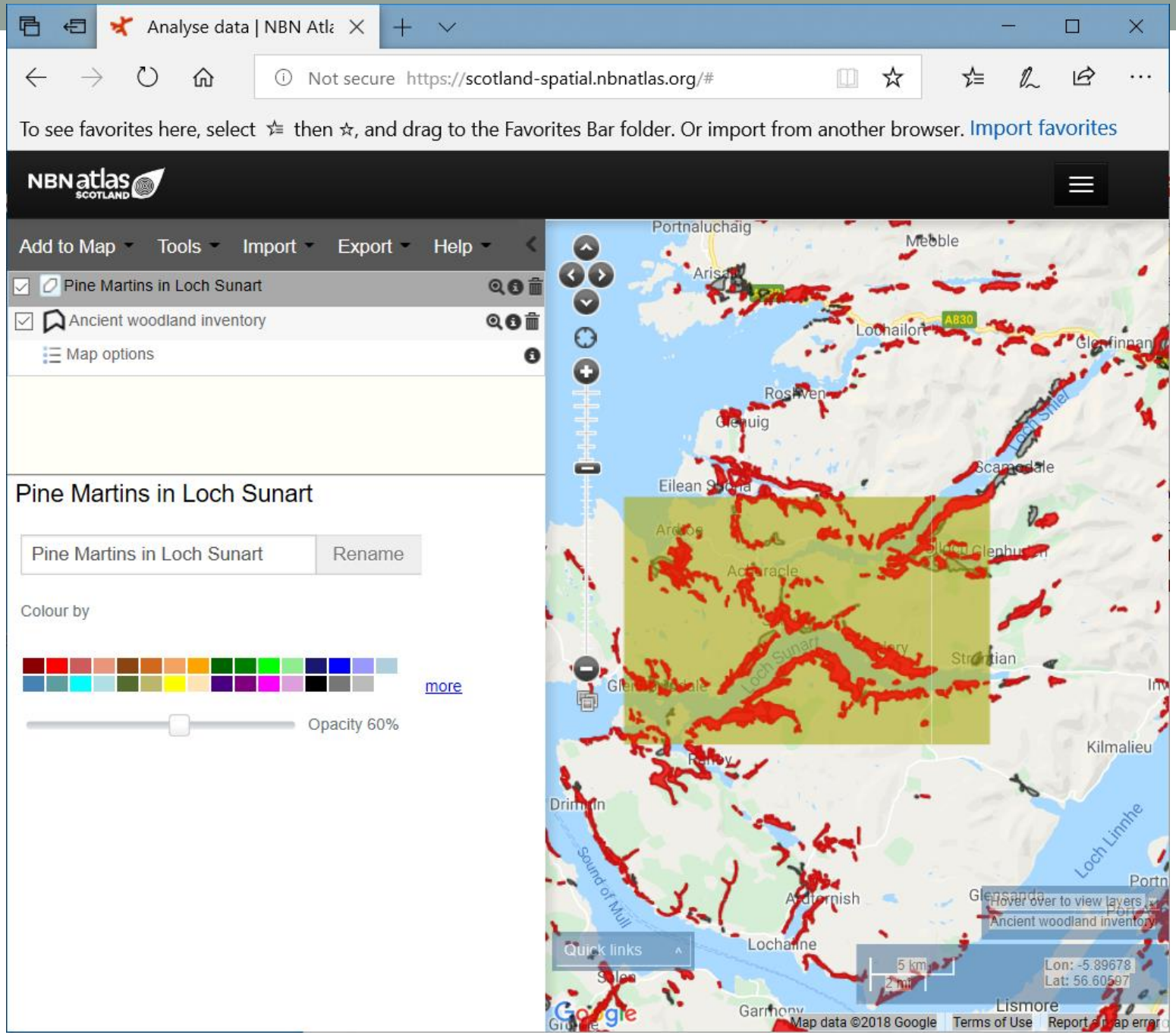
Select a row to highlight on the map

search | Filter | Clear Filter

name	
Ancient (of semi-natural origin)	
Long-Established (of plantation origin)	
Other (on Roy map)	

0 checked | Clear checked | Create from checked | Clear highlight



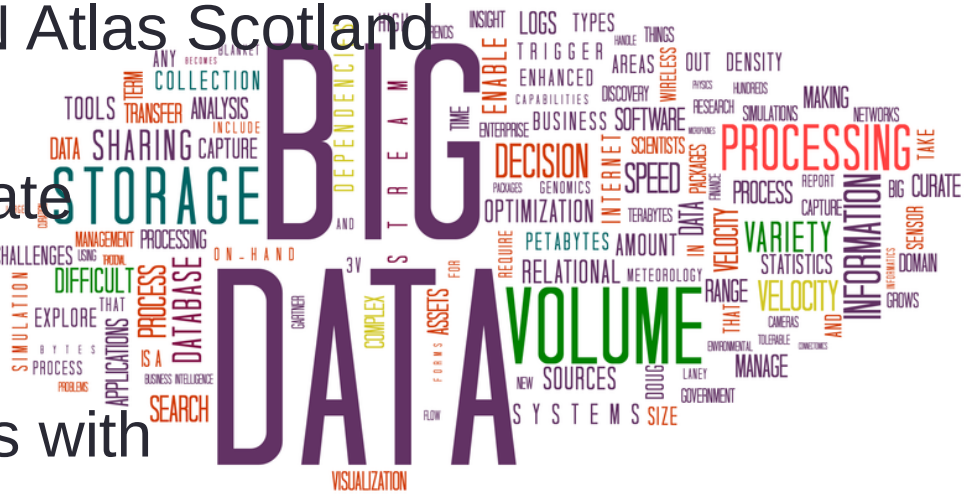

Analyse data | NBN Atla
Not secure https://scotland-spatial.nbnatlas.org/#
To see favorites here, select then , and drag to the Favorites Bar folder. Or import from another browser. Import favorites
NBN atlas SCOTLAND
Add to Map
Tools
Import
Export
Help
Pine Martins in Loch Sunart
Ancient woodland inventory
Map options
Pine Martins in Loch Sunart
Pine Martins in Loch Sunart
Rename
Colour by
more
Opacity 60%
Portnaluchaig
Meoble
Lochailort
A830
Roshven
Glenuig
Scamodale
Eilean Shona
Ardrog
Acharacle
Loch Sunart
Strontian
Kilmalieu
Drimnin
Sound of Mull
Loch Linnhe
Ardtornish
Lochaline
Quick links
Hover over to view layers
Ancient woodland inventory
5 km
2 mi
Lon: -5.89678
Lat: 56.60597
Lismore
Garmony
Map data ©2018 Google
Terms of Use
Report a map error

Challenge -

- NBN Atlas Scotland **spatial layers** data
  - Currently 30+ layers in NBN Atlas Scotland
  - 10s of mega bytes in size
  - How to maintain an up-to-date repository of layers?
  - How to download them?
  - How to provide intersections with species data?
  - How to minimise system admin time?
  - To ultimately create an dynamic record of spatial layers for users

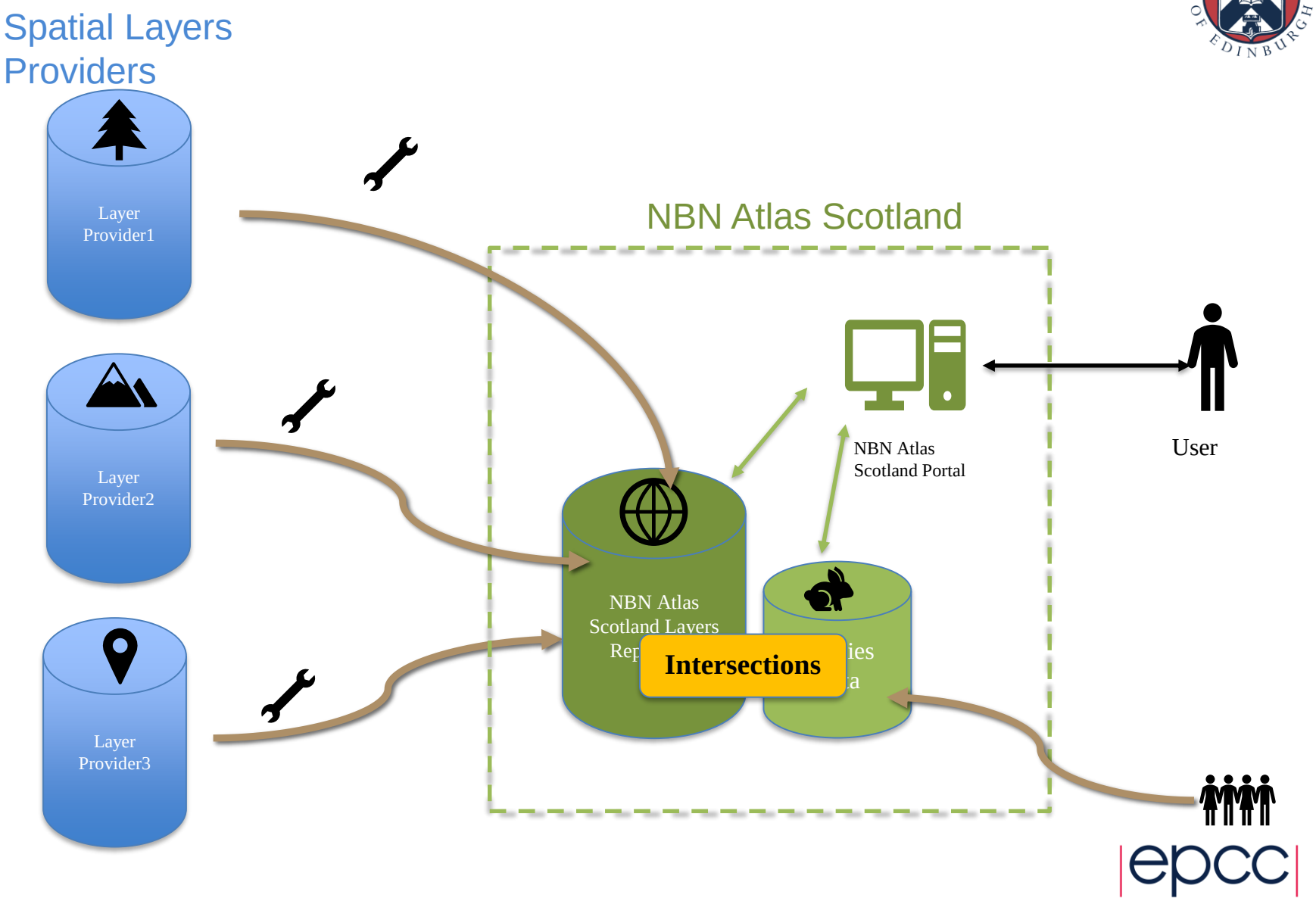
AND

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

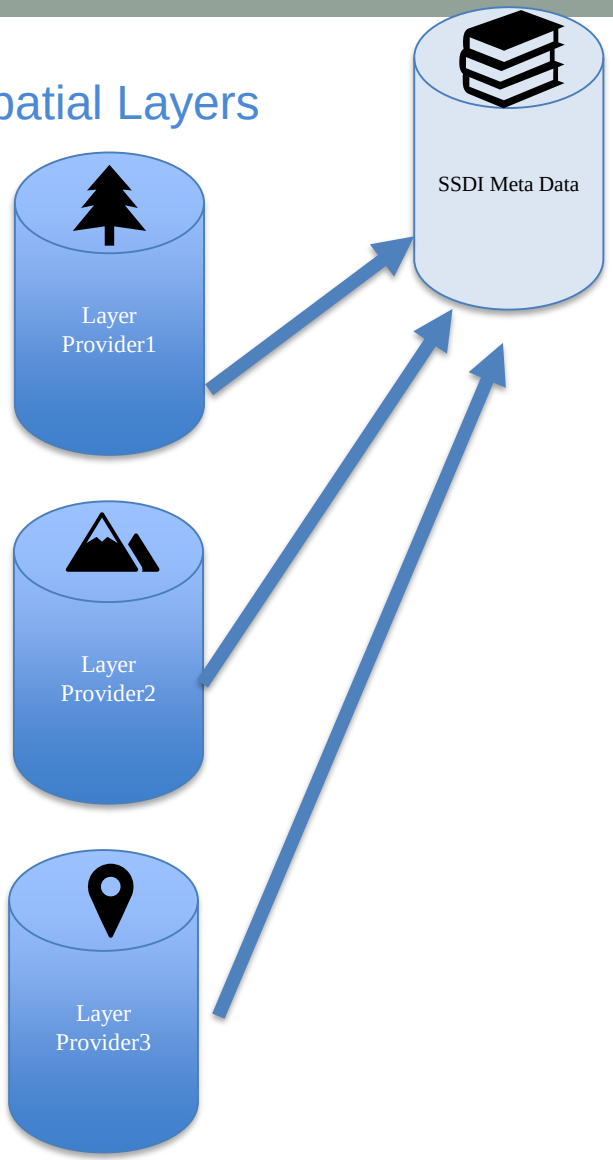


epcc

NBN Atlas Scotland High Level View

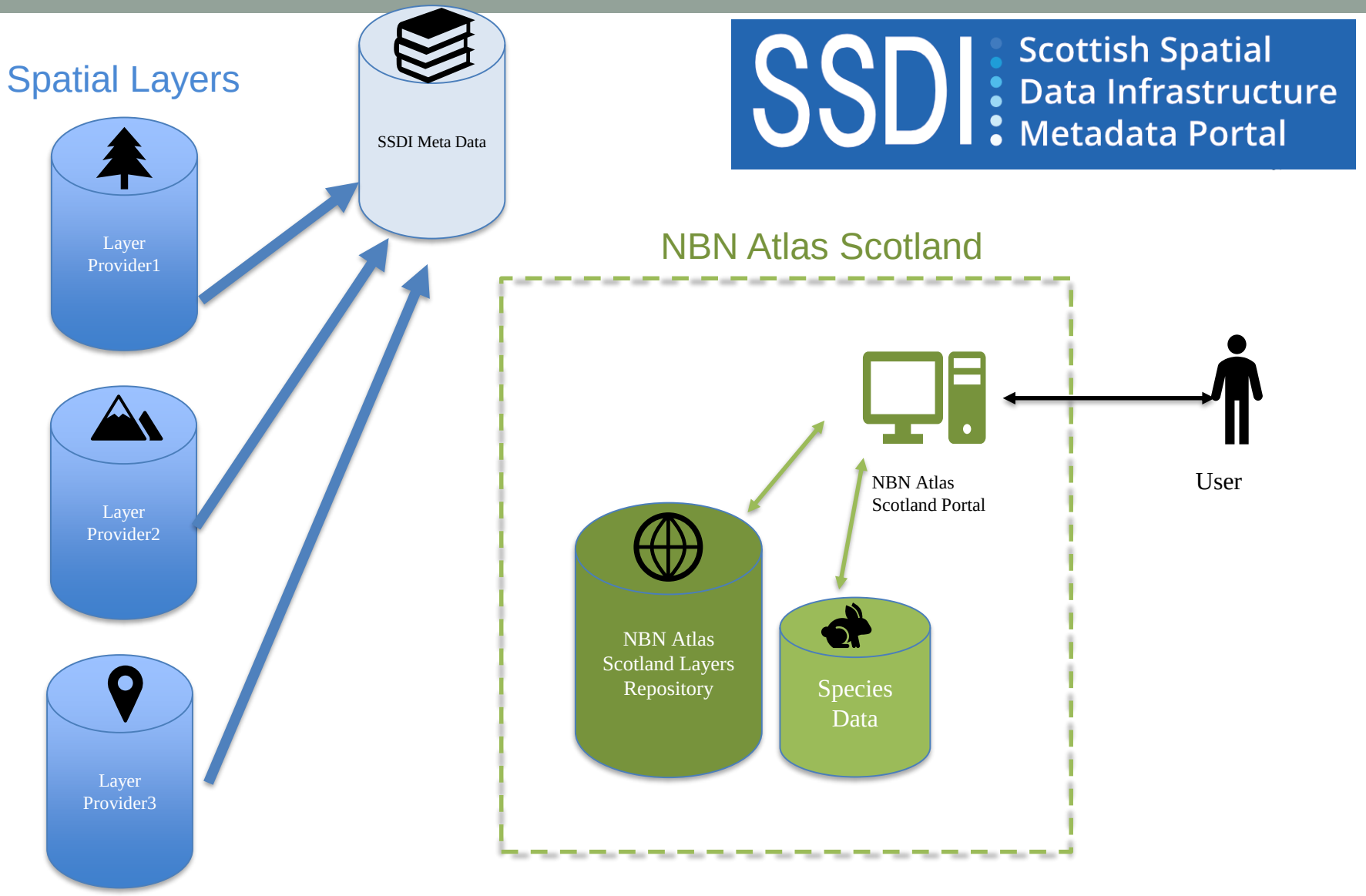


Spatial Layers

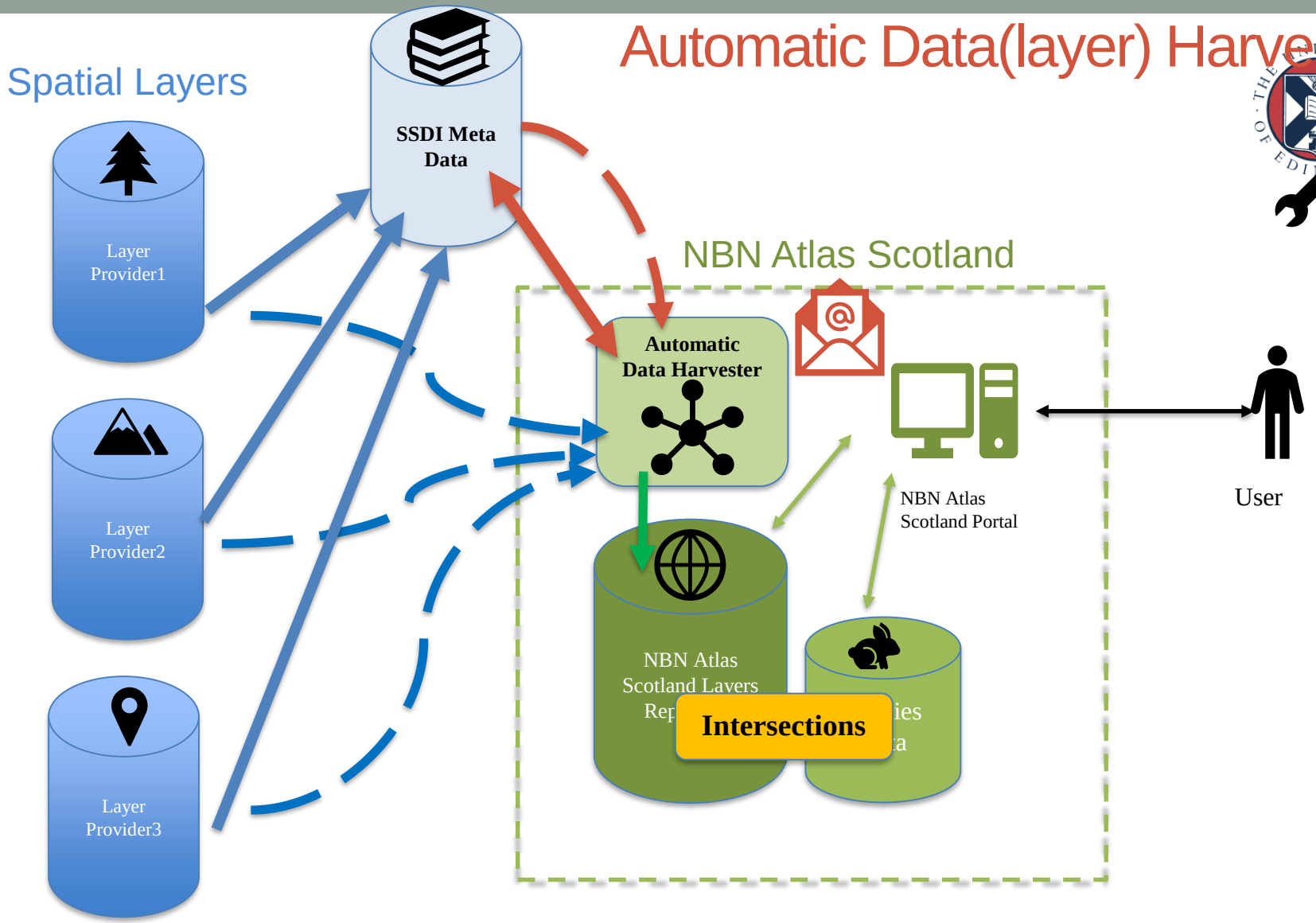


SSDI Scottish Spatial Data Infrastructure Metadata Portal

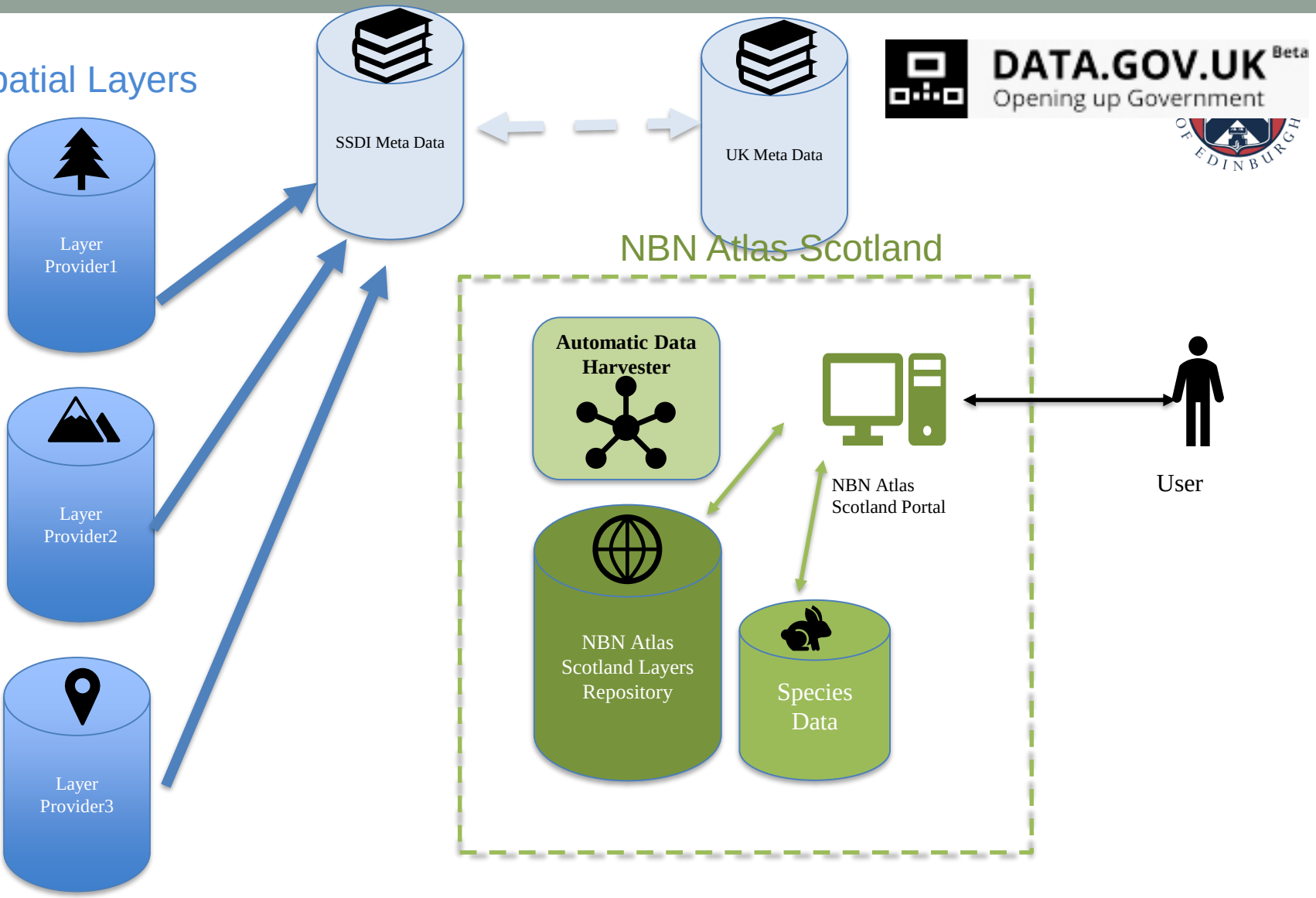
epcc

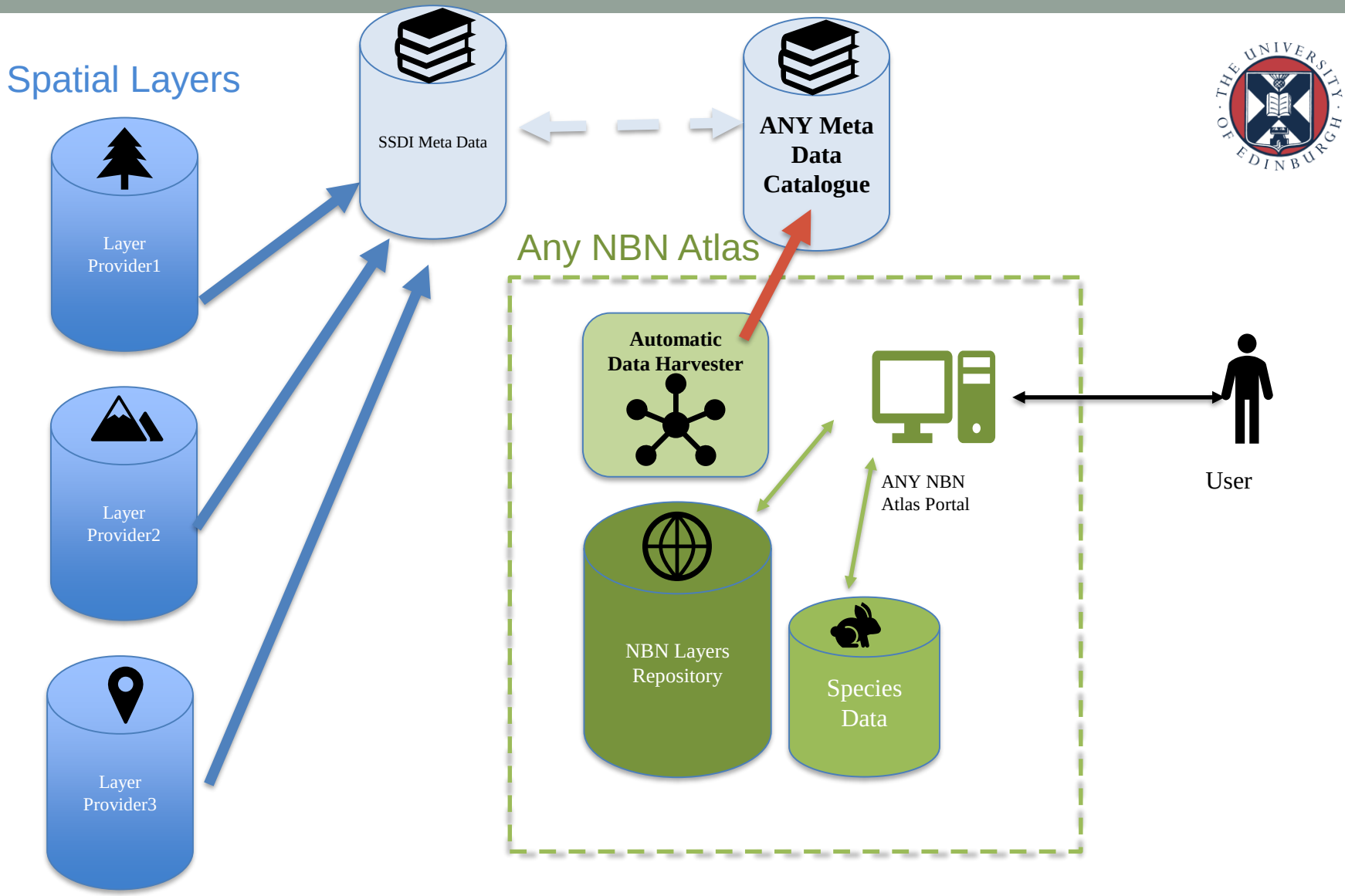

Spatial Layers
SSDI Meta Data
Layer Provider1
Layer Provider2
Layer Provider3
SSDI
Scottish Spatial Data Infrastructure Metadata Portal
NBN Atlas Scotland
NBN Atlas Scotland Portal
User
NBN Atlas Scotland Layers Repository
Species Data

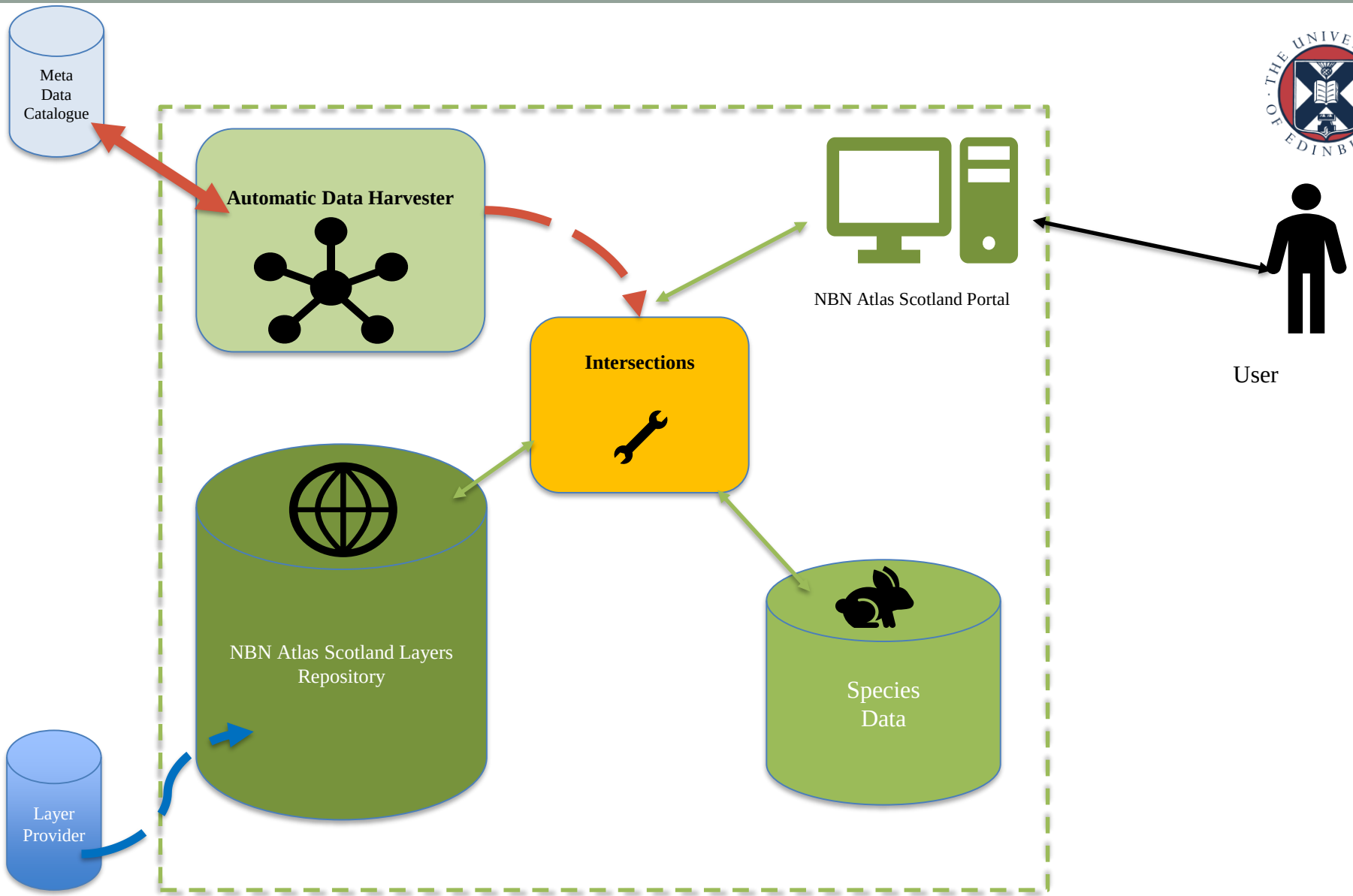
Automatic Data(layer) Harvester



Spatial Layers




Spatial Layers
SSDI Meta Data
ANY Meta Data Catalogue
Layer Provider1
Layer Provider2
Layer Provider3
Any NBN Atlas
Automatic Data Harvester
ANY NBN Atlas Portal
User
NBN Layers Repository
Species Data


Meta
Data
Catalogue
Automatic Data Harvester
NBN Atlas Scotland Portal
User
Intersections
NBN Atlas Scotland Layers
Repository
Species
Data
Layer
Provider

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?

epcc