

# Seeing through clouds

David Blamire-Brown relates how satellite Earth Observation data, Open technologies and cloud-based solutions are being exploited to change the way that Defra monitors the environment

In the UK public sector, several factors have combined to drive an urgent need for new ways of serving and exploiting Earth Observation (EO) data. These include reduced budgets, a focus on digital services and improved customer service, and how Brexit potentially changes the responsibilities of many central government departments.

The Department for Environment, Food & Rural Affairs (Defra) family<sup>1</sup> is particularly affected by these forces. The Defra family monitors England's environment in many different ways, for example:

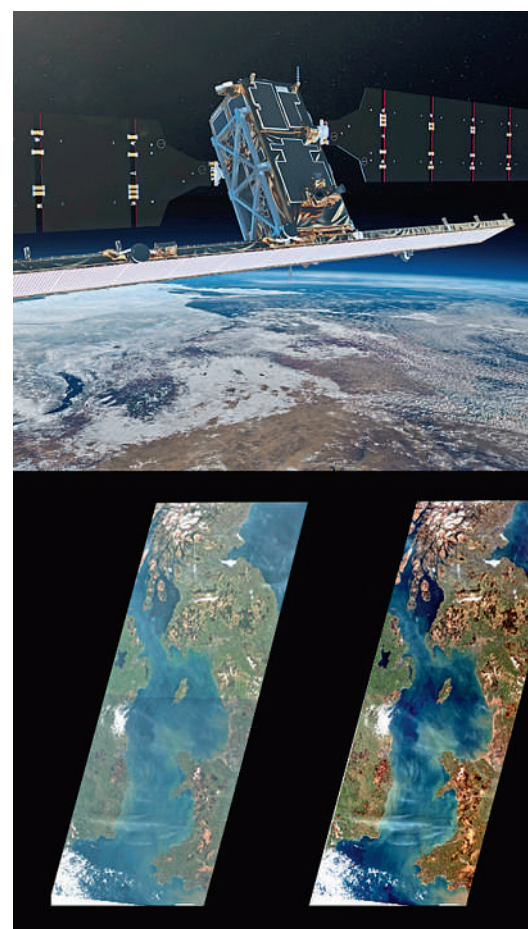
- Rural Payments Agency: land use and validating Common Agricultural Policy (CAP) Basic Payments data.
- Forestry Commission: forestry cover and health.
- Environment Agency: land use, habitat stress, flooding, pollution, etc.

EO data is already vital for Defra family members to effectively and efficiently perform their duties.

Will better satellite data change the game?

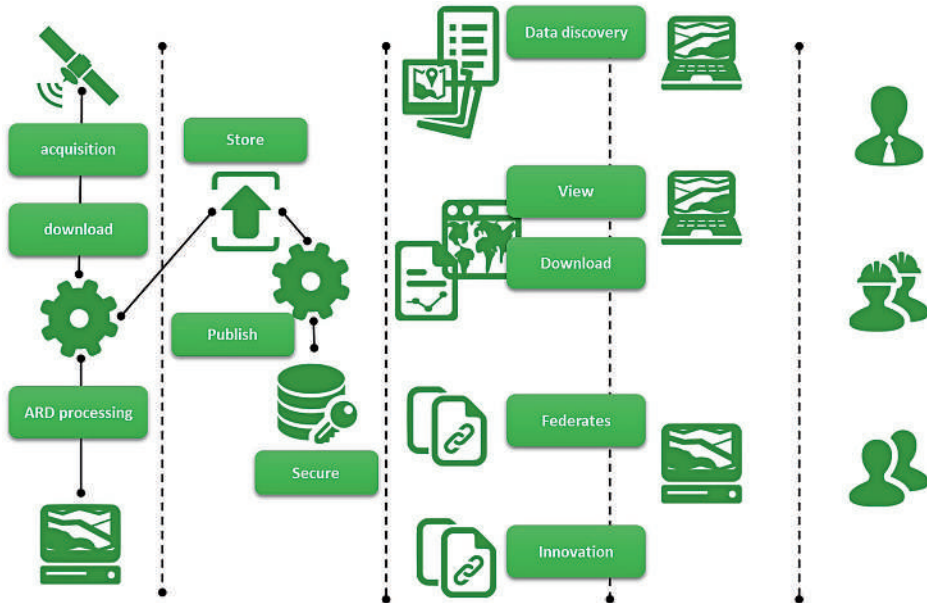
## Enter Copernicus

The last decade has witnessed an explosion in the quality and availability of EO data. For example, the European Space Agency (ESA), through its Copernicus Space Agency (formerly GMES) Programme<sup>2</sup>, has introduced the Sentinel family of satellites. These capture multiple types of fine resolution, freely

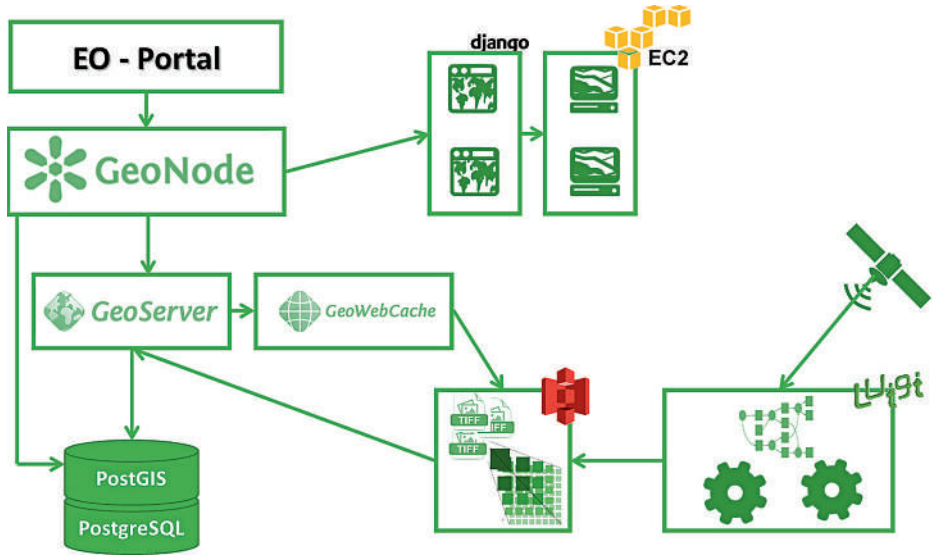


Top: Sentinel-1, carrying an advanced radar, scans Earth's surface no matter what the weather and regardless of whether it is day or night. Image: ESA/ ATG medialab. Lower: Processing single "grains" of Sentinel data can produce boundary artefacts, like the line across the middle of the left-hand image. Images: Defra

The Sentinel-2A satellite takes us over part of northwest England. Contains modified Copernicus Sentinel data (2017), processed by ESA. CC BY-SA 3.0 IGO



Schematic overview of the processing chain and data flow for Defra's EO ARD production facility



How key components of the solution fit together

available imagery for land, ocean and atmospheric monitoring.

Three constellations of Sentinel satellites are in active operation, with Sentinel 1 and 2 being most important for land use monitoring. Sentinel 1A and 1B deliver all-weather, day and night radar images, while Sentinel 2A and 2B deliver high-resolution, multi-spectral optical images for land services.

Sentinel data has a high refresh frequency, meaning that quite subtle changes can be rapidly detected. This can be used, for example, to respond to the unexpected removal of trees or instances of upland burning before much damage is done.

The new Sentinel data provide massive opportunities to improve policy-making and operational services and will be key to delivering Defra's 25-year plan for managing land across the England.

To maximise this potential, how can

Defra reduce the cost of using EO data for existing users, while making consistently high-quality data available to new users?

**Delivering analysis-ready data**

Although the Sentinels' valuable data are open and freely available, there are significant barriers to widespread use. Data volumes are large (tens of TB per year), and data sets need to be processed to a common standard and consistent quality (known as Analysis Ready Data or ARD).

Defra estimates that up to 70% of the cost of using EO data for analytical purposes is expended in processing, cleansing and transforming raw data. The Department's requirements for processed data ranges from simple visualisation right up to the application of more complex machine learning algorithms. With nothing in the marketplace that currently meets this need, processing costs impose

a limit on how groups in Defra can exploit available data.

With a remit to improve the use of data, Defra's central Data Programme and its Earth Observation Centre of Excellence<sup>3</sup> identified the need for a central facility that would automate the generation of ARD and adhere to the principle of "process once, use everywhere".

To ensure the idea delivered real benefits, Defra and JNCC conducted an initial discovery project that validated the feasibility of automating the production of ARD from Sentinel data.

**Meeting the challenge**

Since December 2017, Defra, JNCC<sup>4</sup> and SCISYS<sup>5</sup> worked together to build the initial version of a central ARD production facility. This project has addressed three main challenges:

1. Automating the processing of Sentinel data to deliver ARD,
2. Cost-effectively storing the data, and
3. Making a single source of ARD easily accessible to a wide audience, with different needs.

By designing and implementing a solution based on Open standards, Open architecture, Open Source software and public cloud hosting, the project was able to meet these challenges and fit with both Defra and UK Government's technical and data strategies.

The first challenge was to automate the processing of raw Sentinel data, including correcting for issues such as observing angle, atmospheric effects, cloud and cloud shadow, and surface reflectance. This requires a combination of mathematical modelling, atmospheric science and software engineering. Rather than develop a solution from scratch, Defra chose to use the Python-based ARCSI (Atmospheric and Radiometric Correction of Satellite Imagery) Toolbox developed at Aberystwyth University to automate the processing of Sentinel-2 data, and the ESA SNAP Toolbox to process Sentinel-1 data.

Once the automated processing chain had been implemented, the next challenge was to find a cost-effective way of storing the ARD.

**Cloud-first policy**

Defra, as part of the UK Public Sector, has a cloud-first policy for software architectures. The applications likely to make use of ARD would naturally create a spiky demand, e.g. retrieving complex result sets. For this purpose, a public hyperscale cloud provider, such as Amazon Web Services (AWS), Azure or Google would be a natural fit. After evaluating with Defra's Enterprise Architects, AWS

was chosen.

There are plenty of storage options in AWS, but with large volumes of Sentinel data (up to 20TB per year), the big challenge is that of cost-effective storage. While Elastic File Storage offered the best performance for this purpose, it was deemed prohibitively expensive. Defra and SCISYS then designed a solution that used S3 instead. This slashed the cost of data storage by more than 90% ... an innovation that was vital to the viability of the overall project.

Finally, there was no point in having done all this hard work if the results weren't accessible to the target audience.

To support the plethora of potential uses and users, the answer focussed on implementing software that supported and exploited OGC Open Standards: WMS, WCS, etc. This meant that the solution could integrate seamlessly with a wide variety of systems within Defra's mixed estate of geospatial and non-geospatial information systems, including both commercial and non-proprietary software.

It was also clear that supporting data discovery was critical. To provide a user-friendly front-end and discovery portal, SCISYS and JNCC integrated Geoserver, GeoNode and Django. With only minimal configuration, these tools make it easy for users to search a catalogue of data products, as well as create and publish

their own new products. An added benefit of using the open source OSGeo suite is that there is visibility of the ongoing development roadmap, providing confidence in the long-term suitability of the software.

### Quick results

Building a technically demanding solution to support a wide range of uses is no small challenge. But by using modern tools and cloud-based infrastructure, the project delivered the first "Minimum Viable Product" in just three months. This will facilitate timely and reliable access to analysis-ready EO data across the Defra family.

The next phase will bring this initial solution into full production operation, so that Defra can start to capture feedback and lessons from real use. As Defra begins to realise the potential of EO data, new uses will be discovered.

There is already a score of projects that could use the ARD to either dramatically reduce the cost of fulfilling statutory duties or even achieve outcomes that have not been previously possible. Examples include:

- Monitoring forest stewardship and new tree growth,
- Assessing diffuse pollution (e.g. water run-off from roads, houses or farmland) and impacts on water quality, and

- Mapping drought stress and impact on water resources.

The EO ARD tool has an Open architecture to allow rapid and flexible future development to meet new requirements. The intention is that the whole of the development will be made available as Open Source. This will allow others to benefit from what is learned and, potentially, adopt an identical approach, including building commercial services.

This project demonstrates how public investment can drive broader benefits at the same time as meeting a specific priority - very much in line with the underlying philosophy of Government's approach to Digital.

<sup>1</sup>. <https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs>

<sup>2</sup>. [www.copernicus.eu](http://www.copernicus.eu)

<sup>3</sup>. <https://defradigital.blog.gov.uk/2017/09/07/process-once-use-everywhere/>

<sup>4</sup>. <https://www.gov.uk/government/organisations/joint-nature-conservation-committee>

<sup>5</sup>. [www.scisys.co.uk](http://www.scisys.co.uk)

**David Blamire-Brown is the Business Development Manager at SCISYS PLC, based in Chippenham, Wiltshire ([www.scisys.co.uk](http://www.scisys.co.uk))**



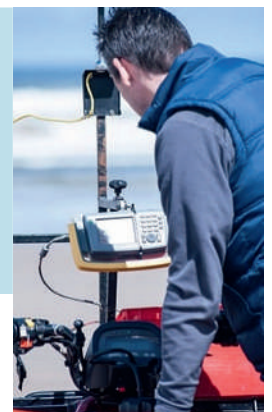
## Established provider of online distance learning postgraduate courses in GIS

### Geographic Information Systems by Online Distance Learning Study for a PG Certificate, PG Diploma or Masters, full-time or part-time

With over 20 years' experience in teaching GIS, and 15 years teaching via online distance learning, Ulster University's School of Geography and Environmental Sciences has a significant track record in GIS education. We offer a broad-based postgraduate course in the principles and practice of GIS covering core principles and concepts and providing hands-on experience using a variety of software, including ArcGIS, ERDAS Imagine, Excel, SPSS and a number of GIS extensions, plug-ins and open source applications such as OpenLayers, GeoServer, PostgreSQL and PostGIS. Our extensive experience in online course delivery means high student satisfaction levels and very positive feedback from graduates.

#### Why study with us?

- Flexibility of location – you can study from almost anywhere and there is no need to travel to classes
- Study part-time or full-time and choose the times you study each week to suit yourself
- Support and advice from experienced lecturers, tutors, librarians, e-learning and IT staff
- Access a wide range of online resources such as e-books, digital lectures and podcasts, discussion boards and video-conference tools all within a dedicated e-learning platform
- Free student copies of GIS, remote sensing and statistical software (ArcGIS, Erdas Imagine and SPSS)
- Two intakes per year for the part-time course (September and January)
- Fees may be paid by instalments
- Relevant work experience may be accepted in place of standard academic entry requirements
- Study for a Master's degree, a PG Diploma, PG Certificate or enrol for individual modules that interest you



*"The course met my expectations and more. I was put out of my comfort zone a lot with learning new areas... but thoroughly enjoyed every minute. Doing this course has already had a positive impact on my career and a big thank you to all the staff and fellow students who helped with great support. I have thoroughly enjoyed the experience."*

Email: [s.cook@ulster.ac.uk](mailto:s.cook@ulster.ac.uk)

[www.ulster.ac.uk/ges](http://www.ulster.ac.uk/ges)