Emmett, B., Gurney, R., McDonald, A., Ball, L., M, B., Blair, G., ... P, Z. (2014). Heads in the cloud: innovation in data and model dissemination. International Innovation 141 (pp. 82-85).

Heads in the cloud

Professor Bridget Emmett summarises a UK-based collaborative project providing a model framework to make better use of complex data and models for more sustainable soil and water management

Can you explain why 'sustainable use of soils and water' was identified as the focus of the Environmental Virtual Observatory (EVO) pilot?

The sustainable soils and water community was chosen for its experience working in cross-disciplinary teams due to the UK's long tradition in catchment science, where different communities come together to deliver integrated land-water management solutions. Looking to the future, this type of cross-disciplinary work will be even more critical if environmental challenges are to be successfully tackled – stepping away from silo management towards more holistic solutions. The community is also experienced in focusing on end-user problems and, more importantly, helping users to develop their own solutions.

Have recent advances in cloud technology facilitated the development of the EVO platform?

Absolutely. This step change has already been well exploited by the astronomy and climate science communities, but is underutilised in land and water sciences. This is perhaps because the land-water community does not rely on large infrastructures such as satellites, ships and particle accelerators, or large powerful models that require major computing power. To a large degree, the benefits of using the new cloud technologies have therefore been missed.

Can you discuss some of the problems related to how climate and environmental data are currently collected and consolidated? Have you faced challenges in the quest to bring together data from a number of different sources?

A major issue is the fragmentation of our current data holdings. The Natural Environment Research Council (NERC)'s National Data Centres are helping to solve this problem but the robust integration of these data, often collected at different

Innovation in data and model dissemination

Utilising environmental data to better support decision making is a complex task. The **Environmental Virtual Observatory** pilot project has demonstrated that a multidisciplinary approach to making integrated data and modelling services available via new cloud technologies offers a way forward

DESPITE THE VOLUME of environmental data available at local, national and global scales, the likely impacts of land management and climate change, as well as their interactions and the management options available to protect against natural disasters like flooding can remain obscure. Interpreting the complexity of findings presents a broad range of challenges. For example, from the perspective of industries such as water utilities, which use large amounts of natural resources, determining how to ensure sufficient supplies for short- or long-term needs while minimising the carbon impact of their activities and maintaining profitability is a difficult juggling act. For governments and regulators, creating policies and management approaches that will prove to be practical and effective to implement under different environmental conditions is a fundamental requirement but relies on greater predictive certainties than can currently be achieved. Improved transparency to reassure the public of the evidence base for coping with environmental change and need for new policies is also becoming a priority. New cloud technologies for the data-intensive world can provide the ability to analyse and integrate the vast and complex historic and current environmental information to manageable levels, as well as visualisation and presentation services to make it easier to gain creative insights and build collaborations.

A number of initiatives are underway worldwide to increase access, integration and exploitation of environmental data and knowledge by scientific and end-user communities through these improved information sharing technologies. This requires both community-led cyber-infrastructure development and new approaches to scientific workflows that describe, compose, model and execute ensembles of data, models, tools and visualisations on distributed resources with global access. A cloud-based approach where resources are all made available as services offers economies of scale for users who require access to large datasets and computing power to manipulate or run complex procedures against them, alongside the benefits of greater accessibility, efficiency and transparency.

CLOUD-ENABLED ENVIRONMENTAL SCIENCE

For environmental specialisms such as climate change science and land management, a key challenge is the need to integrate the currently fragmented data, models and tools from different disciplines. This is needed to reflect the complex interaction of internal biogeochemical, ecological and hydrological processes which underpin the emergent properties we observe at a catchment or landscape scale. "We need to develop new integrated data-model systems which reflect these interactions if we are to develop more sustainable integrated land and water management systems" states co-Principal Investigator (PI) Professor Bridget Emmett of Bangor University. These issues are at the core of the Environmental Virtual Observatory (EVO) pilot project, which Emmett recently led alongside co-PIs Professor Robert Gurney of the University of Reading and Professor Adrian McDonald of the University of Leeds. Funded by the UK Natural Environment Research Council (NERC), EVO represented a two-year proof-of-spatial and temporal resolutions, is a constant challenge. Looking to the future more integrated and efficient monitoring programmes are in place for some parts of the UK, eg. in Wales, with other activities such as the Environment Research Funders' Forum (www.erff.org.uk) ongoing to try and increase integration of environmental monitoring.

How does EVO make vital research resources more accessible?

The principle at the heart of the EVO approach is to encourage and facilitate researchers to share and make more accessible their data, models and tools through their delivery as a 'service' in the cloud which is visible and accessible to all. Following the successful proof of concept in the EVO pilot project, this idea is being taken forward and made operational as part of the NERC Environment Research Workbench, that will allow researchers to share and develop data and models while allowing private workspace, thus enabling both transparency of approaches and reproduction by others. Over time, this will provide a robust library and evidence base of data and model applications for identifying commonly used models and evaluating their performance. This idea of transparency and efficiency in having 'thought experiments' online for others to learn from is at the heart of the benefits of EVO for the research community.

Looking to the future, in what direction do you see the EVO concept developing?

By early 2012, it became evident that delivering a full EVO to underpin such a complex range of activities would require a mix of capital, research and private sector investment along with international and stakeholder collaboration. NERC therefore opted to continue under the Environment Information Initiative with Environmental Big Data Investments such as NERC's Environment Research Workbench and the Environmental Big Data Capital Call: Innovation Activities. These were a precursor to the joint NERC/UK Technology Strategy Board call, 'Solving Business Problems with Environmental Data'; and the Belmont Forum (an eInfrastructure Knowledge Hub programme).

Other organisations are also taking up the EVO approach with, for example, the UK Department for Food, Environment and Rural Affairs and Environment Agency (Defra) releasing funding calls for delivering cloud and web-based data and modelling platforms.

To what extent will the data interrogation and analysis strategies implemented by EVO be applicable to the environmental contexts of countries beyond the UK?

Many models are used internationally and thus the availability of a platform that hosts models and conditioned data will be helpful beyond our national borders. One example would be for collaboration with EU projects on climate change impacts.

Beyond the technical and scientific challenges, EVO encountered a host of other barriers to fulfilling the ambitions at this time; white papers addressing legal and security considerations were key outputs of the project. We also recognised the value of other initiatives and their role in the development of standards. Progress across all these areas will be essential to underpin new technology solutions, which will undoubtedly facilitate and enhance international collaboration.

concept undertaking within which 33 scientists from 12 environmental research centres and institutes across the UK joined forces to answer two key questions concerning data use in the context of sustainable soil and water usage: what facilities would enable scientists, policy makers, industrial bodies, regulators and members of the public to access the data/information they need? And how might a culture be developed of greater openness and rigour in testing and evaluating current models and interpretations? In doing so, the researchers improved their understanding of the issues and barriers a community resource of this type might encounter.

KEEPING IT 'REAL'

The EVO team brought together a range of expertise (including environmental and computer scientists) and resources (data, models, uncertainty and visualisation tools) in order to address the task in hand. With guidance from a Project Advisory Board representing the interests of all potential end-users and stakeholders – from industry, academia and government agencies to regulators and the general public – the team set about creating a cloud-based test bed for exploring scenarios concerning the sustainable use of soils and water. The Agile method of end user-orientated system design and development was used to ensure end-user needs were built into the cloud infrastructure and data services from the very beginning. Storyboards were used to elaborate user requirements to test the efficacy of the cloud-computing enabled infrastructure to evaluate environmental behaviour and its likely response to changing climate and land use at a range of spatial scales.

Emmett explains that the team combined a 'broad and shallow' approach for certain aspects such as intellectual property rights, licencing and regulatory issues with a 'narrow and deep' method for developing scenarios covering river discharge predictions, uncertainty about climate change projections, land management interventions and diffuse water pollution at three spatial scales (local, national and global). These approaches ensured that EVO's exemplars were tightly focused on real end-user questions; an approach that was welcomed by organisations dealing with these complex problems on a daily basis.

STORYBOARDING APPROACH

To assess the efficacy of the cloud computing approach in addressing environmental questions at the local, national and global scales, storyboards were created for each. At the national scale, the storyboard focused on testing uncertainty in river flow predictions and the likely rates of nutrient (phosphorus and nitrogen) flux to inland and coastal waters from land-based sources under a range of land management scenarios. Currently, more than 60 modelling tools exist for simulating the complex chemical, physical and biological interactions involved for nutrient pollution of waters alone, and more again for discharge predictions. The team selected FUSE, a state-of-the-art modelling framework that supports the combination of a number of well-established conceptual models to produce time series of river discharges. FUSE

By consolidating multiple instances of fragmented and widely distributed public and private sector data, expert knowledge, modelling tools and visualisation services, EVO has demonstrated that cloud technologies can improve the effectiveness, efficiency and transparency of such resources

allows daily discharges from many catchments to be modelled simultaneously and additionally, a comprehensive assessment of prediction uncertainties for a range of flow behaviour.

The researchers also developed the first physically-based national biogeochemical modelling framework for the UK to allow the application of nutrient flux models at 4 km^2 grid scale across the whole of the UK landscape, regardless of whether the area was regularly monitored or not. The team implemented the National Export Coefficient Model within the cloud, and used it to test the National Biogeochemical Modelling Framework as one example of the range of models which could be implemented to provide ensemble modelling predictive capability in future.

The model produces spatial predictions of nitrogen and phosphorus flux from land to water at 4km^2 grid scale, relative to landscape character, land use and management, which can be lumped to translate nutrient export predictions to an appropriate scale to inform policy, from small catchments to River Basin Districts to reporting units on UN nutrient flux to coastal waters under the International OSPAR Convention. The model is capable of producing information based on historical data and current conditions, and of assessing future mitigation options based upon different theoretical scenarios. Further work to implement a range of models within the framework would allow assessment of the performance of different modelling approaches for the range of catchments across the UK and thus uncertainty in current projections.

At the local scale, the researchers tested visualisation approaches for both data and model output focused on tackling local flooding problems in three locations in the UK; the Dyfi (Wales) the Eden (England) and the Tarland (Scotland). Flooding can be a societal problem, with a complex and tangled array of 'causes' and 'effects'. The downstream effects can result in costly damage to property, human health impacts and, in extremes, even loss of life. Attributing, understanding and mitigating cause can be incredibly difficult but some of the solutions can come from custodians of the land – commonly farmers, foresters and catchment managers. The challenge is coordinating and developing a sense of common ownership of the solutions at the catchment community scale. To address this challenge, in each catchment a community was developed and nurtured in order to help define the problems and clarify the potential solutions. Having listened to the feedback, the researchers developed a series of web-based portals that allowed the catchment managers to learn about the nature of the flooding in their catchment, make available real-time flow information, both stage and visual, and determine ways to solve the problem, perhaps by modifying choice of land use.

At the global scale, the team focused on uncertainty concerning likely impacts of climate change projects on carbon and water fluxes. Modelling climate change impacts relies on projections derived from Global Circulation Models (GCMs). For the UN Intergovernmental Panel on Climate Change (IPCC) Assessment Report in 2007, 22 GCMs were used, each focusing on different parts of the environment. As expected, results from respective models differed and so were represented as a range of values. The solution was to display the results of all 22 GCMs on a common grid, which then led to refinement of IMOGEN modelling systems, which link circulation patterns to land and atmosphere interactions in accordance with the Joint UK Land Environment Simulator (JULES) model. The EVO portal made the enhanced IMOGEN framework available for modelling scenarios of greenhouse gas emissions for a particular set of years and fundamental parameters – soil, vegetation carbon and water – either as a new cloud service for the user, or by extraction of the results archived from an equivalent scenario previously run. Output maps indicate the locations around the globe where uncertainties have the greatest impact on, for example, soil carbon change.

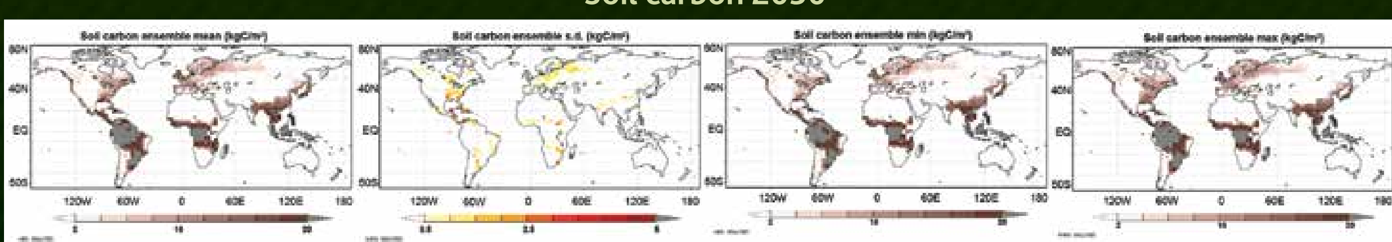
CHALLENGES AND SUCCESSES

The EVO project succeeded in showing the added value of using new cloud technologies for environment monitoring and decision making; by consolidating multiple instances of fragmented and widely distributed public and private sector data, expert knowledge, modelling tools and visualisation services, EVO has demonstrated that cloud technologies could improve the efficiency, effectiveness and transparency of such resources. In addition, the EVO concept enables extraction of more holistic environmental information, and clearly demonstrates how making projections and other results more openly available supports transparency when translating scientific evidence into policy making decisions. A range of resources, including videos of interviews with the project team, their final technical report and two-page flyers, are all available to download from the EVO website (www.evo-uk.org).

EVO LEGACY

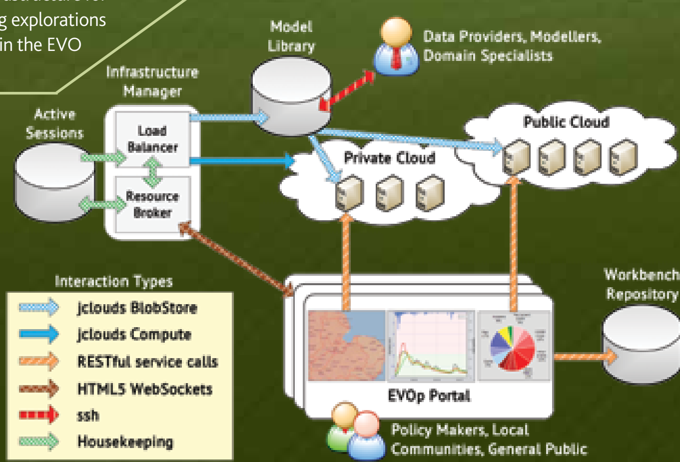
The project was always intended to provide proof of concept only; the EVO portal currently offers restricted access to data, analysis, modelling and visualisation tools as a demonstrator to end-user communities, as well as near real-time weather, rainfall, river level and water quality data for the local communities in the three case study regions. "The vision was to build a test platform

Soil carbon 2050



Outputs from the global exemplar illustrating the locations where uncertainty in climate projections between 22 Global Circulation Models have greatest impact globally on soil carbon by 2050 according to the Joint UK Land Environment Simulator (JULES) model using the CBL-IMOGEN system. Many other outputs are available including runoff, evaporation and vegetation production. Figures from left to right indicate soil carbon ensemble mean; s.d; min; max all as kgC/m^2.

Proposed cyber infrastructure for a full EVO following explorations of different option in the EVO pilot project.



and highlight the opportunities and barriers as preparation for taking a full EVO forward," explains Emmett.

Following on from the EVO pilot project, NERC is now developing an Environment Research Workbench which is an operational version of the cloud platform developed in the EVO pilot project. An Environmental Big Data Capital Call has also enabled large models such as the JULES land-atmosphere model which the EVO tested in its global exemplar to be cloud-enabled to increase access to a broader range of scientists and end-users. Beyond NERC, the lessons learned from EVO are also proving constructive for other scientific community projects, such as the US National Science Foundation (NSF)'s EarthCube geoscience project, and have also been taken up by a number of end-user organisations. For example, the UK Department for Environment, Food and Rural Affairs (Defra) has recently adopted the EVO concept of an integrated data and modelling platform for their diffuse pollution work with the aim of linking/expanding to include greenhouse gases and ecosystem work. NERC is also involved in developing international standards through its involvement in the Belmont Forum in this area.

Emmett expects that greater exploitation of a cloud-based approach with everything available as a service will become more widely used by the scientific community and that this will inevitably result in greater attention to making research data, results, models and other tools more accessible, interoperable and robust. Platforms such as that established in the EVO project also have the potential to foster the kind of international cooperation required to ensure the future sustainability of food, ecosystems and water on a global scale: "Many models are used internationally, so the availability of a platform which brings together models and conditioned data that are fit-for-purpose will be helpful beyond our national borders," Emmett states.

Visualising live data

The use of new technologies in environmental sensors allows researchers to measure environmental variables at increasingly high frequencies using automatic monitoring systems. These systems provide a stream of data which can be sent via a telemetered link and viewed in near to real-time. The potential for alerts for industry and regulators by combining such data with models projections in the cloud could lead to early warnings of potentially damaging conditions enabling mitigating actions to be implemented, avoiding cost and environmental damage. The following data types are currently available to view in this way in three test catchments:

River level • Rainfall • Weather • Webcam images of river conditions • Water quality information

INTELLIGENCE

EVO

ENVIRONMENTAL VIRTUAL OBSERVATORY

OBJECTIVES

A proof-of-concept project to develop new cloud-based applications for accessing, interrogating, modelling and visualising environmental data. By developing local and national scale exemplars, EVO demonstrates how cloud technologies can make environmental monitoring and decision making more efficient, effective and transparent to the whole community.

PROJECT TEAM

UK: **Professor Bridget Emmett** (co-PI) • **Professor Robert Gurney** (co-PI) • **Professor Adrian McDonald** (co-PI) • **Dr Lucy Ball** (Project Manager) • **Professor Keith Beven** • **Dr Mesude Bicak** (Lead for Global exemplar) • **Professor Gordon Blair** (Lead for Cloud platform) • **Dr John Bloomfield** • **Dr Wouter Buytaert** (Lead for Modelling in the Cloud) • **Julie Delve** (Project Officer) • **Dr Yehia Elkhatib** • **Professor Dawn Field** • **Prof Jim Freer** (Lead for National hydrology exemplar) • **Dr Alistair Gemmell** • **Dr Sheila Greene** • **Professor Phil Haygarth** (Lead for Local exemplar) • **Dr Chris Huntingford** • **Professor Penny Johnes** (Lead for National Biogeochemistry exemplar) • **Dr Eleanor Mackay** • **Professor Mark Macklin** • **Dr Kit Macleod** • **Dr Nick Odoni** • **Barbara Percy** • **Dr Paul Quinn** • **Dr Sim Reaney** • **Gwyn Rees** (Lead for Data) • **Dr Marc Stutter** • **Dr Bholanath Surajbali** • **Professor Doerthe Tetzlaff** (Lead for International activities) • **Dr Nicola Thomas** • **Claudia Vitolo** • **Dr Mark Wilkinson** • **Bronwen Williams** • **Przemyslaw Zelazowski**

FUNDING

EVO is one of a portfolio of Natural Environment Research Council (NERC) Research Programmes – grant number: NE/I002200/1. As a pilot project, an important remit of EVO was to investigate and develop funding models and approaches for future initiatives.

CONTACT

Professor Bridget Emmett
Co-Principal Investigator

Centre for Ecology and Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd LL57 2UW
UK

T +44 124 837 4500
E bae@ceh.ac.uk

www.evo-uk.org

BRIDGET EMMETT is a biogeochemistry and ecosystem scientist working for the UK public environment research body the Centre for Ecology and Hydrology (CEH) owned by the NERC. She is Head of CEH's Bangor Research Station in Wales with 60 staff and students from over 10 countries.

NERC SCIENCE OF THE ENVIRONMENT
EVO Environmental Virtual Observatory