Virtual labs and digital environments: can virtual lab technology support a paradigm shift towards a more open, collaborative and integrative environmental science?

Gordon Blair and Michael Hollaway
What are virtual labs?

“Transdisciplinary collaboration spaces hosted in the cloud that allows stakeholders to access a range of environmental data, analytical methods and assessment tools, and to execute these analyses using the elastic capacity of a cloud.”
Why virtual labs?

• Openness and transparency
• Support for the 4th paradigm of data-intensive science
• Collaboration, collaboration, collaboration
• Integration, integration, integration
Cf. Data centres

- Builds on the strong legacy of environmental data centres
- Aim of value add in a changing world
- Cf. Ongoing work in NERC’s Strategic Need Advisory Working Group (SNAG) on Environmental Data Services
Exploiting digital innovation: cloud computing and big data

- Scalable and elastic infrastructure capable of handling big data
- Service-oriented philosophy supports re-use of a wide range of existing and new services
- Increasing support for portability and interoperability
- Delegation of responsibility for system maintenance and management
Experience of virtual labs: the Data Labs Project

- Implemented on JASMIN
- Tailorable
- Support for data science
- Support for end-to-end analyses
- Increasing interest in publishable notebooks
Changepoints for a changing planet: a CDE feasibility study

- Use of the UK Environmental Change Network (ECN)
- Focus on methodologically enhanced virtual labs
  - Multivariate changepoint detection
  - Clustering
- Transdisciplinary collaboration in action
Data Labs in action: a live demo
Virtual labs: research challenges

- Improved data architectures
  - Use of semantic web, linked data concepts
  - FAIR assets
  - Towards an environmental data commons

- Data science for the natural environment
  - Tailored data science and AI techniques that meet the demands of this domain
  - Urgent need for more case studies
  - Co-existence of data understanding and process understanding
Virtual labs: research challenges (continued)

- Raising the level of abstraction
  - Exploit the *right* abstraction techniques from computer science and software engineering
    - E.g. containers, serverless execution, software frameworks, domain specific languages
  - Allow scientists to do science

- Applying a socio-technical lens
  - Building communities of practice around virtual labs
  - Enhance trust in environmental decision-making
  - Supporting decision making in the face of uncertainty
Virtual labs: research challenges (continued)

• **Modelling in the cloud**
  • Moving environmental models to the cloud, incl. migration of legacy models
  • Support for integrated models
  • Support for model coupling

• **Towards a digital twin of the natural environment**
  • Builds on all of the above
  • A truly grand challenge for the digital environment community
  • Cf. Beven’s models of everywhere
Concluding remarks

- Virtual labs *do* have the potential to support that paradigm shift towards a more open, collaborative and integrative environmental science
- We have *significant experience* of virtual labs in action across a range of environmental science domains
- A lot remains to be done and to be successful this has to become a *community effort*
Thank you for listening

Any questions?
Resources and links

- CEEDS website - https://ceeds.ac.uk/
- Blair’s senior fellowship - https://www.ensembleprojects.org/
- Data science of the natural environment - https://www.lancaster.ac.uk/data-science-of-the-natural-environment/
- Virtual labs/DataLabs - https://doi.org/10.1016/j.patter.2020.100103