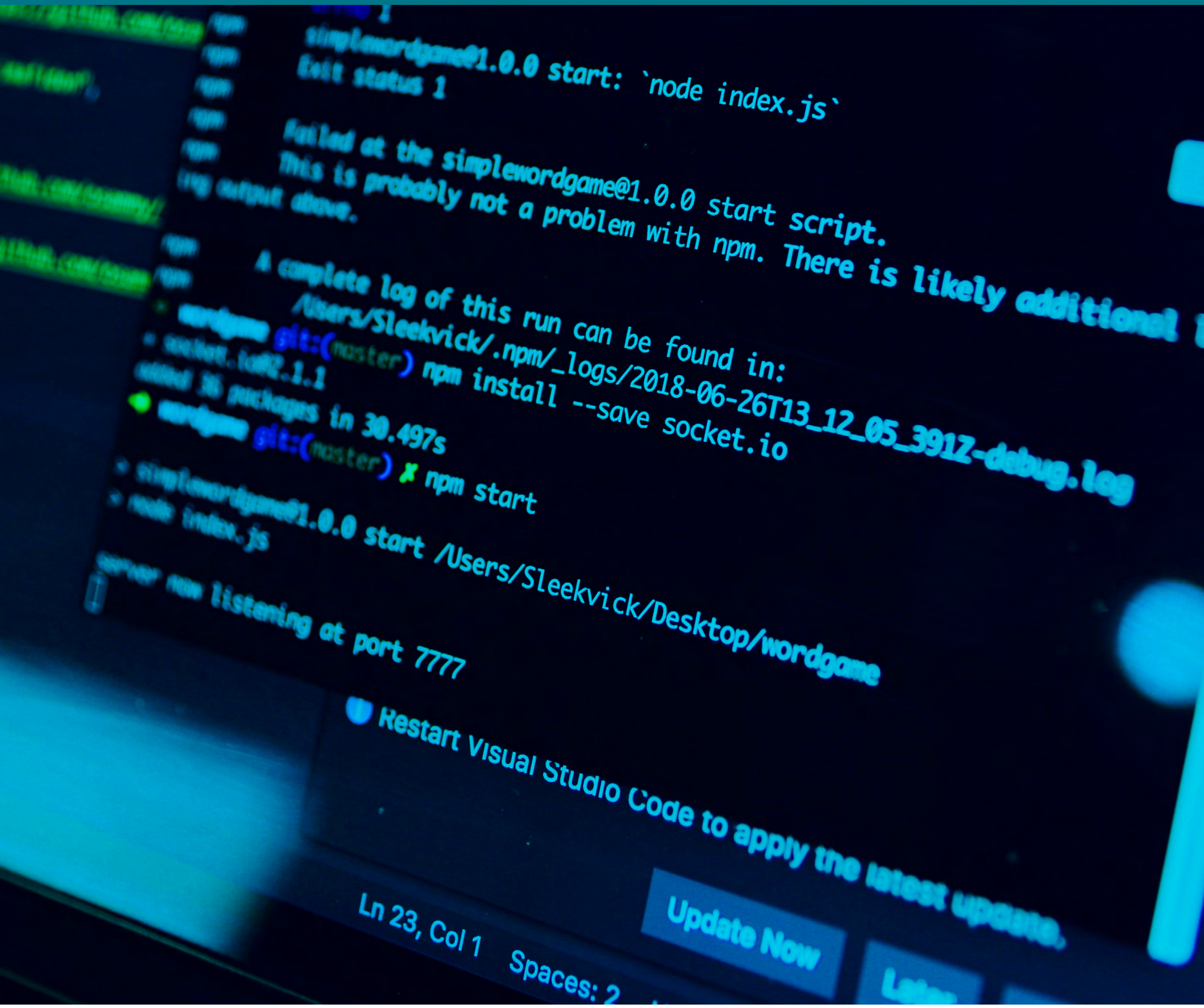
Learning from the Big Picture: Applying Responsible Innovation to the Net Zero Research Infrastructure Transformation (ARINZRIT)

Summary briefing

January 2023



University of Bristol | Lancaster University | UKRI Science and Technology Facilities Council | epcc

Overview of the research

UK research increasingly relies on ‘digital research infrastructure’ (DRI): digital technologies and computational facilities from laptops to high-performance computing and large-scale data archives. DRI has an energy and carbon impact as a result of performing computational work (emissions scope 1), the energy needed to drive them and how this is generated (scope 2), and their manufacture and disposal (scope 3). Their combined environmental impacts depend on decisions which are not only technical but also organisational and social, linked to the science and scientists they support. These range from how research and DRI are reviewed and funded, to adherence to policies such as the FAIR data principles (findability, accessibility, interoperability and reusability) or open science, as well as the practices and expectations of research. This all leads to a lasting infrastructural legacy with implications for the energy, climate and resource footprint of projects and infrastructures.

We consider DRI facilities to be comprised of:

1. Individual specialist equipment needed to meet specific academic research needs
2. Institutional DRI including High Performance Computing (HPC) and data storage
3. DRI provided through Higher Education Institutions (HEIs) but owned / managed elsewhere e.g., data storage, cloud hosted services
4. Tier 2 shared (regional) HPC e.g., Isambard, Sirius
5. Research council / national DRI e.g., ARCHER2, JASMIN, Monsoon
6. Other remote computation including direct access cloud, distributed remote (global) computers

DRI is used across a wide range of disciplines and subject areas, and its use and footprint is growing. As an example, the ARCHER national HPC system, which was active between 2014 and 2020, had around 250,000 jobs run on it in the last full quarter of its operation (quarter 4 2019). In contrast, the successor, ARCHER2, had approximately 450,000 jobs run on it in the last quarter of 2022. This expansion in usage and users has been evident for a long time during and across the lifespans of DRI, with DRI becoming more embedded in research and data analysis workflows, and becoming key tools for a wider range of users over time.

To explore the wider socio-cultural influences affecting DRI provision and use, 25 interviews were conducted across stakeholder groups of users, providers, commissioners, and senior management relating to DRI. The participants interviewed spanned multiple disciplines, institutions, and research councils and, whilst not fully representative of the sector, the interviews capture a variety of experiences that point towards some of the complexities and relationships to be considered in moving towards net zero ambitions for DRI. These experiences also demonstrate that interventions should not just be technical. In this report, we outline our key findings and offer recommendations for more sustainable DRI policy and practice.

Key findings

The growth of DRI, and its resulting environmental costs, means that it is important to understand the wider social and organisational factors that shape, and are shaped by, its use and provision. Users of DRI individually have relatively little agency in addressing net zero and are – generally – embedded in a 'publish or perish' culture, making it critical that research funders and institutions work collectively to establish practice for DRI that is net zero compliant. There is currently little to limit this growth in DRI, or its use. Targeted funding and continued embedding of DRI and computationally intensive methods in all disciplines may even be accelerating this.

Our interviews showed that:

DRI use is increasing due to advances in science and increased adoption of machine learning techniques in many research domains, including those not previously considered to be computationally intensive (e.g., computational chemistry, synthetic biology). In parallel, the social sciences are developing new approaches using digitised artefacts, and examining social life as it occurs in online spaces. Concurrently (undergraduate and postgraduate) students are increasingly expected to use DRI as part of their studies, further embedding DRI use in academic tools and practices. (Recommendation 6).

"... Students read papers, right, that's part of the work. So, they investigate the field, and [...] the majority of those papers are results from like, big models, [...] they want to compete with that, and publish in a certain conference or journal..."

Increasing demand drives DRI growth as computation use and the provision of more services creates new opportunities for innovation and knowledge creation. Nationally, demand for services outstrips supply, resulting in DRI expansion and upgrade in line with UKRI funding cycles. At a local (HEI) level, provision is shaped by the need to spend budgets in funding cycles as well as researchers' desire for ownership and control of 'own' facilities (e.g., local workstations, compute clusters). (Recommendation 3, 5 and 6)

"... It's more a case of the more we add, the more it would be used: there's basically more demand than supply. ..."

DRI use can be inefficient due to poor software development practice (e.g., lack of code optimisation, codebases poorly attuned to leading edge hardware), decisions made because of policy, practice and culture (e.g., researchers using more resources or higher priorities/speed than required, replicating computation to meet reviewer expectations), and because of unnecessary compute (e.g., results already exist, research questions are poorly formulated or redundant). This, ultimately, decreases DRI's efficiency and increases hidden environmental costs, without resulting in an increase in the knowledge created. (Recommendation 3, 7, and 10)

DRI provision can be unsustainable. This can be the result of new DRI being provided too soon (wasting hardware) or too late (creating inefficiency), and by constraining procurement protocols (e.g., ringfenced clusters for health applications). DRI may also be used at lower capacity, resulting in overprovision or duplication, or impacted by provisioning for peak demand periods (e.g., student term times, academic paper deadlines). DRI is underutilised in other ways too (e.g., lack of waste heat reuse) due to wider cost, logistical, organisational or estate related reasons. DRI procurement can also be driven by funding availability (funders finding they have money to use before the end of a financial period and keen to spend it on something that can be procured quickly), as well as planned need. (Recommendation 4, 6 and 9)

"... If there's duplication of platforms or resources, then that's wasteful. If there are inefficiencies, because there aren't ways of joining up different platforms, that's wasteful. ..."

The growing trend to FAIR data has implications for net zero DRI. Easier findability and greater potential for reuse of data could avoid unnecessary duplication of DRI usage. However, the process of FAIRification of data requires an ecosystem of supporting tools and services that in turn require DRI resources, such as long-term data preservation, machine processing of data, training and upskilling. The balance between benefits and costs is currently poorly understood: a variety of influencing factors interact together to drive the generation and use of FAIR data, and the proportionality (in terms of net zero) between effort put in and benefits derived is still unclear – for example, whether there are any factors that dominate over the long term. (Recommendation 2)

Academic practice shapes the unnecessary and inefficient use and provision of DRI as noted above, e.g., through expectations by reviewers and academic communities on the quantity of DRI use implicated in producing work acceptable for publication; local academic norms and 'lock in' to old, possibly inefficient, software; and a desire for unrestricted use and control leading to the provision and inefficient use of extra local DRI. However, care needs to be taken to balance against innovation in academic practice using DRI that can be said to increase the quantity and quality of research. (Recommendation 3 and 10)

"...the fundamental thing for me, is that academia is still culturally, very much rooted in this sort of 19th century independent scholar, slightly artisanal approach to research. And that's increasingly at odds with the reality of how work is done on the ground, but it's still the culture. ..."

There is a lack of awareness and training toward achieving net zero DRI, with its environmental impact being hard to find or contextualise, and therefore, largely unknown among researchers (including Research Software Engineers). Where awareness does exist, net zero DRI is a low priority and actions are usually grounded in informal knowledge that does not always lead to sustainable, evidence-based change. Time and cost are frequently used as proxies for carbon, but these are inadequate. Users say that better information would help to inform their decision-making on how to use DRI facilities. (Recommendation 1, 5, 6, 8, and 10)

"... So for me, the speed of the run is a proxy for how energy efficient it is. And that might not always have been true, and it probably isn't going to be true going forward. But I don't know, at the moment, I have no way of measuring that. As far as I know, there are no tools that I can access to do that. ..."

There is no clear ownership, oversight, or resource for progressing towards net zero DRI, resulting in little overt or cultural pressure to consider or embed net zero in DRI activities. Where cost is incurred, this is often decoupled from the researchers at the point of use, and funders, practitioners and academic communities are not typically holding each other to account for DRI's environmental impacts. Resultantly, there is no obvious ownership, human or financial resource set aside to address this issue. If net zero DRI proves more expensive or complex than the default DRI, or there are no resources provided to address the cost and complexity, it is unlikely to be addressed. (Recommendation 1, 8, and 9)

"...there really is zero external pressure. You know, the university when they made the Net Zero announcement. Pretty much nothing changed..."

Recommendations

These recommendations, synthesised from the views supplied by our interviewees, are shown at the level of 1) overall governance and management, 2) how DRI is provisioned and supported, and 3) how the use of DRI can be enhanced to address net zero ambitions.

Governance and management

1. **Make informed net zero DRI policies from transparent evidence of its environmental impacts**, involving sector-wide policy to ensure all research institutions share DRI environmental data with UKRI which follows a consistent carbon calculation method and considers DRI's full lifecycle impact.
2. **Establish and promote sector-wide FAIR data and code protocols** to maximise visibility and re-use of existing data and code, and minimise duplication or unnecessary processing and storage.
3. **Formalise net zero research incentives to reshape academic practice**, promoting research which truly embeds a sustainable approach to DRI (e.g., by assessing DRI's full lifecycle in peer-review processes and funding applications and calls, during project execution and review, offering best paper awards for delivering results with minimal environmental impact).
4. **Support flexible sharing of sector-wide DRI** for researchers to utilise available computational resources when required, avoiding underutilised DRI and the expansion of new and unnecessary DRI elsewhere.
5. **Publicise and resource mandatory net zero and climate emergency policies** so that low-carbon options are the default choice by ensuring appropriate funding and regulations are compatible with addressing the ambitions and cost of net zero.

Provision and support

6. **Establish clear decision processes in funding applications** for whether DRI use is required, based on consistent processes for defining type and use of DRI and methods for determining its environmental impact.
7. **Offer training to researchers on sustainable DRI use** and better software engineering practices to ensure best choice and use of appropriate DRI hardware e.g., via specialist research software engineers (RSEs), costed into or shared across projects, and supported beyond the project lifetimes to avoid inefficient use of DRI and DRI duplication.
8. **Recognise DRI's role as part of the wider infrastructure** and embed in institutional policy and practices, ensuring valuable outputs (e.g., heat) are integrated into institutions' estates and beyond (i.e., local, regional, national) to maximise value and avoid waste.
9. **Address barriers to maximise the operational lifetime and reuse of equipment**. For example, enabling hardware manufacturers, suppliers or their agents to offer longer warranties; ensuring sufficient resources are available to maintain equipment.

Use of DRI

10. **Ensure researchers follow best practice in the sustainable use of DRI** whilst recognising the need to advance knowledge, e.g., by reusing DRI, data and code where possible, ensuring new code is optimised, embedding FAIR data practices, and considering whether the proposed research or new DRI is really required.

Further information

This project is a sub-project of the UKRI Net Zero Digital Research Infrastructure Scoping Project, a major project managed by the Centre for Environmental Data Analysis (CEDA) working towards a roadmap for carbon neutrality in the UK's digital research infrastructures (https://net-zero-dri.ceda.ac.uk/).

Contact the researchers

Primary contact: **Adrian Friday** (a.friday@lancaster.ac.uk)

Caroline Bird (Caroline.Bird@bristol.ac.uk),
Chris Preist (Chris.Preist@bristol.ac.uk)
University of Bristol;

Adrian Friday (a.friday@lancaster.ac.uk),
Carolynne Lord (c.lord@lancaster.ac.uk),
Kelly Widdicks (k.v.widdicks@lancaster.ac.uk),
Lancaster University;

Adrian Jackson (a.jackson@epcc.ed.ac.uk),
EPCC, University of Edinburgh;

Gabin Kayumbi (gkayumbi@turing.ac.uk, formerly STFC),
Simon Lambert (simon.lambert@stfc.ac.uk),
Science and Technology Facilities Council.