

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

January 2023

```
code ILIFECYCLE (module.js:458:3)
error 1
simplewordgame@1.0.0 start: `node index.js`
Exit status 1

Failed at the simplewordgame@1.0.0 start script.
This is probably not a problem with npm. There is likely additional logging output above.

A complete log of this run can be found in:
  /Users/Sleekvick/.npm/_logs/2018-06-26T13_12_05_391Z-debug.log
simplewordgame git:(master) npm install --save socket.io
simplewordgame git:(master) x npm start
simplewordgame@1.0.0 start /Users/Sleekvick/Desktop/wordgame
node index.js
server now listening at port 7777

Restart Visual Studio Code to apply the latest changes: 2
```

Authors

Caroline Bird (Caroline.Bird@bristol.ac.uk)

Chris Preist (Chris.Preist@bristol.ac.uk)

University of Bristol



Carolynne Lord (c.lord@lancaster.ac.uk)

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

Kelly Widdicks (k.v.widdicks@lancaster.ac.uk)

Lancaster University



Gabin Kayumbi (gakayumbi@turing.ac.uk, formerly STFC)

Simon Lambert (simon.lambert@stfc.ac.uk)

Science and Technology Facilities Council



Adrian Jackson (a.jackson@epcc.ed.ac.uk)

EPCC, University of Edinburgh



* Primary Contact

Cover photo by [Desola Lanre-Ologun](#) on Unsplash.

The ARINZRIT project focused on mapping the social and cultural factors driving the need for Digital Research Infrastructure (DRI) in the UK. ARINZRIT is one of seven ‘sandpit’ projects contributing to the UK Research and Innovation (UKRI) Net Zero DRI Scoping Project, run by the Centre for Environmental Data Analysis (CEDA), working towards a roadmap for carbon neutrality in the UK’s DRI (<https://net-zero-dri.ceda.ac.uk/>).

The aim of ARINZRIT is to develop insights into what drives the carbon emissions and climate impacts of data generation, analysis, storage, and dissemination arising from DRI; contributing to understanding how UKRI DRI estate can deliver carbon neutral computing by 2040 or earlier.

From interviews with carefully selected stakeholders, and analysis of the usage data from specific high performance computing resources over the past 10 years, this report highlights how current cultural and academic practices, including a lack of policy and tools for decision making, are impeding the path to net zero. We offer diagrammatic representations of interacting forces (‘causal loop diagrams’) that show how existing practice leads to the growth of DRI and its resulting environmental footprint through inefficiency, redundancy and a confluence of existing processes. We conclude with our recommendations for stakeholders to address net zero more effectively by stimulating a necessary cultural shift in all areas in response to the climate emergency.

Contents

1. Introduction and background	3
2. Methodology	5
2.1 Semi-structured interviewing	5
2.2 Thematic analysis	7
2.3 Presentation of findings: Construction of stakeholder personas and causal loop diagrams	8
2.4 Quantitative analysis of DRI usage over time	9
2.5 Analysis of FAIR approaches and methods	9
2.6 Benefits and limitations	9
3. Analysis of factors influencing progress to net zero DRI	10
3.1 Stakeholder personas	10
3.1.1 Persona 1: The Academic User	11
3.1.2 Persona 2: The Research Software Engineer	12
3.1.3 Persona 3: The University HPC Manager	13
3.1.4 Persona 4: The Research Council	16
3.2 Causal loop diagrams	17
3.2.1 Demand for DRI	18
3.2.2 Efficiency and waste in DRI	19
3.2.3 Using ‘own facilities’ versus shared HPC	22
3.2.4 Provision of DRI	23
3.2.5 Dynamics of FAIR in relation to net zero	25
3.3 UK HPC usage data	27
4. Intervention points and barriers to net zero	32
5. Recommendations	36
References	38
Appendix 1: Glossary	39
Appendix 2: Interview questions	40

1. Introduction and background

Academic research increasingly relies on ‘digital research infrastructure’ (DRI). DRI refers to any specialised, IT devices and facilities that enable research, spanning from laptops and workstations to high-performance computing and large-scale data archives. How DRI comes to be used and provisioned, however, depends on decisions which are not only technical (e.g., driven by large or complex data, simulation or models), but which are organisational and social too (e.g., the result of policy, practice, culture). Open science policies, reviewer expectations, and availability of funding are just some examples of factors which have influence over the use and provisioning of DRI. Decisions made intentionally or even unintentionally as a direct or indirect result of policy, practice and culture can produce waste, for instance, with researchers using more resources, or infrastructure, than strictly required. Waste (in the use of resources, or the provisioning of infrastructure), ultimately, increases DRI’s resulting environmental and economic cost, and it does so *without* necessarily producing an improvement in the attendant research outcomes. However, the environmental costs of DRI use and provisioning result from issues that go beyond ‘waste’ and ‘inefficiency’. In this work, we explore wider issues of practice, policy and culture, to understand how these produce environmental impacts at different levels in the UK DRI ecosystem.

Large-scale computing infrastructure accounted for an estimated 2% of the US electricity supply in 2016 (Arman, S. et al 2016), with a further 2% of emissions potentially attributable to embodied carbon costs for such hardware (IRISCAST 2023). Very large high-performance computers, such as the Fugaku system in Japan, consume upwards of 30MW for powering the computers and provide other services such as cooling and data storage/movement (Fujitsu, 2020).

There has been growth in computing usage in all areas of academic research and industrial processes, with 7 out of 10 researchers reporting that computing was essential to their research as far back as 2014.

As an example, the ARCHER national high performance computing (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. Whilst computing systems have undoubtedly become more efficient in their use of resources, this rapid increase in the need for DRI inevitably has an impact on the environment through both the materials needed to construct DRI and the energy used in its operation.

"There has been growth in computing usage in all areas of academic research ... with 7 out of 10 researchers reporting that computing was essential to their research as far back as 2014."

The growing environmental costs of DRI suggest that efforts must be made to better understand the wider social and organisational factors driving its use and provision. This is especially important as we work towards net zero DRI (reducing the climate impact of DRI effectively to zero). We take a primarily interview-led approach to mapping this landscape to identify points of leverage and intervention. We also highlight opportunities for DRI demand reduction whilst providing insight into how DRI's use, and provision connects with the wider research culture. This is key, given the potential for unintended negative consequences of net zero initiatives on practice and policy that could occur should the interplay between social and organisational factors not be considered. For example, improvements in efficiency (and thus availability) of computation have already led to growing demand across more disciplines and levels, an example of the rebound effect in practice (Widdicks, 2023).

Our objectives were to identify:

1. The ways in which research practice, policy and social expectation shape the environmental impacts of DRI. This includes, given UKRI's statement on Open Research, understanding the possible implications of the Findable, Accessible, Interoperable and Reusable (FAIR) data principles for the long-term achievement of net zero.
2. How these (non-technical) factors might be addressed to meet net zero targets in future DRI use and provision.
3. Where intervention points in the DRI system provide opportunities for impact reduction e.g., by reducing infrastructure demand; whilst also mitigating the risks of unintended consequences arising from net zero initiatives.

We first outline our methodology for achieving this in section 2. Section 3 develops 'personas' synthesised from the stakeholder interviews that have helped us map and navigate this complex landscape from various perspectives. We then offer diagrammatic maps of influencing and interacting factors that lead to DRI growth. In section 4, we look at intervention points and barriers to change. Section 5 draws this all together to offer 10 key recommendations for addressing and embedding net zero decision-making in the DRI landscape.

2. Methodology

We conducted 25 interviews across stakeholder groups of users, providers, commissioners, and management to establish the high-level oversight of research practices and culture guiding UK DRI infrastructure establishment and use.

The participants spanned multiple disciplines, institutions and research councils. Two researchers conducted the interviews, before thematically analysing them – first individually, and then collaboratively with the wider project team during an in-person workshop. With the major themes then selected, and triangulated, the analysis could be used to construct personas and causal loop diagrams to bring the issues to life. In parallel, figures on the usage of DRI were analysed and the FAIR (Findable, Accessible, Interoperable, and Reusable) approach to storage of data and code discussed within the context of the interview findings.

Further details of this approach and its stages are given below.

2.1 Semi-structured interviewing

a) Participant selection and sampling

The research team identified the key stakeholder groups in UK academic DRI, deciding that: users, providers (e.g., HPC managers, IT network managers, research development, research software engineers), management / culture representatives (e.g., ‘meta-research’, senior academic management), and commissioning (e.g., research councils, university procurement) were important actors who would be able to speak to the research questions posed. A spreadsheet was then created of known contacts from these groups, drawing from the initiating institutions.

For expedience, given the short timeframe and limited resources, a convenience sampling approach was taken – with most of the participants being selected due to their proximity to the professional networks of the research group, and the remainder then snowballing based on recommendations from participants.

The sample that resulted was: ten interviewees from the University of Bristol, five from Lancaster University, one from the Science and Technology Facilities Council (STFC). Others were from three other universities and three research councils (see [Table 1](#) for details). Some interviewees’ roles crossed over between stakeholder groups but broadly, there were twelve users drawn from disciplines such as climate science, plasma physics, radio astronomy, computer science and computational chemistry; eight provider / support roles including HPC manager, procurement, research software engineering (RSE); four commissioners and four in management / culture positions.

The resulting sample is thus potentially not fully representative. It does, however, capture a wide-ranging and detailed view of how DRI procurement and use work across two specific research-intensive institutions (University of Bristol and Lancaster University) whilst also including participants from other institutions to capture other issues and influences that may not generalise from these two alone. The themes and stories that are presented in this report should thus be read as illustrative vignettes that point towards many of the complexities and relationships that should be considered as we move towards net-zero ambitions for DRI; also, critically, that points of interventions should not just be purely technical nor limited to tuning software and increasing efficiency.

ID	Role and Domain	Stakeholder Group
P1	Computer scientist	<i>user</i>
P2	Climate modeller	<i>user</i>
P3	Director of Advanced Computing	<i>management / provider</i>
P4	Lecturer, Computer Vision and AI	<i>user</i>
P5	Plasma Physicist	<i>user</i>
P6	Research Software Engineer	<i>provider / support</i>
P7	Scientific Computing / Particle Physicist	<i>user</i>
P8	Radio Astronomy/Computational Imaging	<i>user</i>
P9	HPC chief technical officer	<i>management / commissioning</i>
P10	IT Procurement Manager	<i>provider</i>
P11	Computational Chemistry (RSE)	<i>user / support</i>
P12	Research Development Officer	<i>provider / support</i>
P13	Storage and Virtualisation Manager	<i>provider / support</i>
P14	Professor of particle physics	<i>management / user</i>
P15	Lecturer, Natural Language Processing	<i>user</i>
P16	HPC Manager	<i>provider / support</i>
P17	Senior Data Scientist	<i>user</i>
P18	Research council HPC senior programme manager	<i>commissioning</i>
P19	Head of Research Computing	<i>provider / support</i>
P20	Senior Research Associate, Computational Chemistry	<i>user</i>
P21	IT Network Manager	<i>provider / support</i>
P22	Research Council Data and Infrastructures Team	<i>commissioning</i>
P23	Professor, metaresearch	<i>management / culture</i>
P24	Senior Portfolio Manager, Digital Research Infrastructure	<i>commissioning</i>
P25	Associate Professor, Systems and Synthetic Biology	<i>user</i>

Table 1: Participants (ID, Role and Stakeholder Group)

b) Interviews

Guided by the research questions, background reading and broader expertise in the research team, an interview schedule was created for each of the key stakeholder groups. The areas of questioning are broadly captured by the following headings:

- DRI use and provision
- Sources of waste and inefficiency
- Acquisition and upgrade of DRI
- Sustainability and net zero

The majority of questions were chosen to be open-ended, and probes were provided for each, to elicit responses that were as detailed as possible (see [Appendix 2](#) for interview schedules).

The two researchers who conducted the interviews are predominantly social scientists, with expertise in qualitative methods and research surrounding technologies, society, and environmental sustainability. Until the inception of the project, both were relatively inexperienced with DRI use and provision, allowing them to approach the interviewing process with an ‘anthropological mindset’ of “*making the strange familiar*” (Nelson, 2019). This meant that little could be taken for granted during the interviews, with the researchers having to ask further questions for their own understanding. This approach produced – arguably – more detailed interviews than if a specialist in the field had undertaken them instead.

The interviews were conducted either in person or online and lasted from 30 minutes to nearly 120 minutes. The recordings were transcribed through Otter.ai (<https://otter.ai>), and the resulting transcripts were then reviewed to ensure their accuracy.

2.2 Thematic analysis

A thematic analysis approach was used (Braun and Clark, 2006; Nowell et al, 2017) to identify and analyse themes within the interviews. As noted by Braun and Clark (2006), a theme ‘*captures something important about the data in relation to the research question*’ although it does not necessarily have to be more prevalent in the data. The two interviewing researchers first took the broader themes as outlined in the proposal and questions (e.g., use, provision and upgrade, waste, culture, net zero) and broke these down into sub-themes as they emerged during the process, in order to fully analyse the interviews and provide the basis for the findings below. The ‘outcome’ themes of potential solutions, barriers and points of intervention were then added and again broken into sub-themes as necessary. The two researchers each took a subset of the interviews for analysis, presenting the raw findings in different but complementary ways based on their established practice.

To triangulate the insights that emerged through this process, a project workshop was held with the other six researchers from the group reading a selection of the interviews and reviewing them

in advance of the event. After presentations in which the two interviewing researchers outlined their initial findings, the research group discussed points of confluence and convergence across their readings, noting and confirming the major themes as they emerged.

2.3 Presentation of findings: Construction of stakeholder personas and causal loop diagrams

a. Personas

With the major themes having been selected, the interviewing researchers selected the participants' transcripts in which these themes were best depicted. From these, stakeholder personas were constructed, building on the interview of a specific participant from each stakeholder group (i.e., as the protagonist) and combining this with the details of other interviewees from that same group. Personas are, more generally, understood as "*archetypal users who embody the goals and aspirations of real [interviewees]*" (Haines and Mitchell, 2014). Though the method was first developed in software development, it has since been used in other contexts including policy design (e.g., Haines and Mitchell, 2014). We adapted this method somewhat, using this as a means of creating a personable representation of each stakeholder group, pulling illustrative details from participants across that group. This contrasts with its usual presentation, in which the personas show a range of different types of 'users' in the same stakeholder group – for instance, Khaled Al-Shboul & Abrizah (2014) use it to demonstrate the different informational needs of humanities scholars. For our purposes, the method allows us to create an illustrative narrative of the barriers and points of intervention for each stakeholder for net zero ambitions in DRI use and provision. Because of the use of this method, the stories that follow are semi-fictional as they do not relate solely to the interview of any one participant. However, we have created one story for each type of stakeholder interviewed that draws together the main themes from the interviews for each stakeholder group as these speak to the barriers to net zero ambitions, and their attendant intervention points.

b. Causal loop diagrams

Causal loop diagrams (CLDs) were constructed by identifying significant issues (demand, efficiency, type of DRI and provision) and using the interview analysis to identify factors influencing them in both negative and positive ways. These diagrams were constructed using the free software 'loopy' available at <https://ncase.me/loopy/> and are qualitative and not quantitative in nature, based on interpretive coding of the data. This means that they indicate *areas of influence* for the different issues identified but not the relative size or impact of these different factors. Causal loops are used in systems thinking to explore dynamic relationships between different parts of the system (e.g., Haraldsson, 2004) with the 'loop' part showing reinforcing elements. In this case, we use this type of approach to show both loops and linear influences.

2.4 Quantitative analysis of DRI usage over time

We also undertook an analysis of the usage data from a restricted set of DRI resources over the past 10 years to identify patterns in the exploitation of large-scale HPC resources as part of research use of DRIs. We analysed the usage of two Tier 1 HPC systems, ARCHER and ARCHER2, and one Tier 2 HPC system, Cirrus, processing the batch system data from those systems to gather insights into how the different types of resource are used by research and how that use has changed over time.

2.5 Analysis of FAIR approaches and methods

The dynamics between FAIR and net zero were examined from a causal perspective, highlighting the structural links that produce reinforcement or counter effects. The same approach of building causal loop diagrams was adopted. This preliminary work began by gathering a wide-ranging list of factors relevant to FAIR data. The interviews conducted in ARINZRIT were used as one source. Causal influence modelling was used to link these factors together, with particular attention paid to the possibility of ‘hidden’ factors mediating between the more obvious ones.

The diagram helps us obtain a better understanding of the ‘energy proportionality’ between costs and benefits of FAIR considered in terms of the drive to net zero, particularly which factors dominate over longer periods of time, where there are feedback loops, and where effective points of intervention might be found.

2.6 Benefits and limitations

In summary, by adopting these complementary approaches, we take different perspectives on the same qualitative dataset. The persona-based approach gives us insight into the *practices* of stakeholders, and how these affect, and are affected by, the technical, social and cultural environment in which they are situated. The causal-loop based approach provides insight into the wider *systems* in which these practices are embedded, and how different factors interact to reinforce or dampen their (environmental) impacts, including of FAIR data practices. Meanwhile, the quantitative analysis of DRI usage allows us to connect these different perspectives to broader trends and insights around the environmental impacts of DRI use and provisioning. When taken together, these lay the foundations for the recommendations that are presented at the end of the report in [section 5](#). Further work would be needed to complete and validate these initial observations.

3. Analysis of factors influencing progress to net zero DRI

As outlined above, the findings of this research are presented in a number of complementary ways: (1) a series of personas capture the experiences of the stakeholders with whom we spoke, highlighting some of the issues and influences in engaging with net zero DRI. (2) causal loop diagrams show the influences and impacts of different themes identified. (3) A further layer of analysis zooms into one policy area and reflects on FAIR approaches to data. (4) Finally, these insights are contextualised with a quantitative analysis of DRI usage to better understand its environmental impacts.

We note our interviewees described different types of DRI. Broadly these are considered as a group depending on scale. Larger HPC systems are categorised by Tiers. The Tiers are defined as: Tier 0 national / transnational systems such as those provided through PRACE (Partnership for Advanced Computing in Europe (<https://prace-ri.eu>)) – petaflop systems (i.e., a system capable of performing one quadrillion floating-point operations per second (UITS, n/d). Only one of our interviewees talked about a Tier 0 facility (i.e., CERN in Switzerland). Tier 1 systems are nationally-leading centres e.g., ARCHER2, Jasmin, Monsoon, of which there tends to be a few per country. Tier 2 are traditionally regional facilities, though increasingly less so (e.g., Isambard, Cirrus). They tend to be highly specialised and there to fill the gap between Tier 3 and 1. Tier 3 are University supercomputers, including HPC and data storage, aimed at helping researchers within given institutions only (EPCC, n/d). Outside of the larger HPC systems, participants also spoke about individual specialist equipment they needed for specific academic research (laptops, workstations, etc.), as well as remote computation such as direct access cloud, and distributed remote (global) computers.

Though findings cut across the different DRI facilities described above, the wide range of ideas of what actually constitutes DRI suggests that further definition and specification of the term is required in net zero planning and implementation to avoid confusion.

3.1 Stakeholder personas

The following section presents four different stakeholder personas (i.e., a Research Software Engineer (RSE), an academic researcher, a DRI manager and a research funder). Each persona is intentionally fictitious and synthesises representative findings from our interviews with stakeholders. We refer back to the source material using participant identifiers beginning with (P), see [Table 1](#).

3.1.1 Persona 1: The Academic User

Lisa is an Academic. Her research focuses on ‘*a branch of artificial intelligence*’ (P4) which requires the use of institutional DRI for the high-performance computing needs of her projects. Alongside her own work, she supports the research of ‘*nine [PhD students], who all want GPUs*’ (P4). The type of DRI used by each student depends not only on the project itself (i.e., funding, intensity of computational task), but on the skills of the individual researcher. ‘*Some of them don’t [use the institutional DRI] because they think it’s a bit complicated [...], they have to do a lot of setup. [...] Only the really good programmers [use the institutional DRI], they do complain that it’s a bit fiddly, but they’ll use it. Because it’s powerful in the end*’ (P15).

Commenting on how needs for computational power had changed in recent years, ‘*usage [...] is ramping up*’ (P7). ‘*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. And when I say big models, that’s models that require a lot of computational power*’ (P15). The increasing demand and expectation of computationally intensive tasks in academic research, then, was not just changing what it meant to do academic research, but also the expectations of students about what they would be taught and what their own research projects should include. This was not just the case with PhD students, but even for undergraduate students who were completing their final-year projects. The integration of computationally intensive tasks into different degree schemes and university courses, at times, caused issues because of their temporal patterning. In remarking on how demand peaks could be handled in the future, Lisa suggested commercial cloud could offer mitigation of the congestion that deadlines could cause on institutional systems: ‘*[...] you can simply increase or decrease the number of machines you have available depending on how many jobs you want to run*’ (P20).

In terms of the environmental impact of Lisa’s research, she felt that she ‘*was personally usually [...] mindful about how [she] used [resources]*’ (P8). But when asked about whether her usage could ever reach ‘net zero’, she said: ‘*[it’s] something that I will never achieve. There’s really no question about that*’ (P4). There were ways, however, to make use of resources and computational power more efficiently. Computational resources, for instance, could be shared to ensure that power, heating, and ventilation were also shared – though she noted that security and privacy were issues with some research contexts (e.g., medicine, security, personal data) (P2, P4, P17). Other forms of waste included: ‘*duplication of platforms or resources, [...] inefficiencies, because there aren’t ways of joining up different platforms. And if there are computing resources, for example, being devoted to projects that ultimately aren’t worth doing, then that’s wasteful*’ (P23).

Efficient coding, too, was important in minimising the impact of DRI use. However, the nature of research (i.e., as a mode of systematic inquiry) meant that ‘*[sometimes] you need to redo [a simulation or run]. And that’s science, isn’t it?*’ (P20). In other moments, it was clear that a code could be made more efficient, but the solution was not always so clear: ‘*It’s always going to be the*

case with our research and some things might be quite obvious [...] but most of the time, it's stuff you could spend three or four weeks looking for, and still not find' (P2).

The academic 'publish or perish' paradigm, at times, limited time that could be spent in producing efficient code, or using DRI efficiently: 'It's kind of a time to solution that we're interested in. You know, how quickly can we get that back?' (P5). Academic structures, like journal, conference, and grant applications too, influenced how much experimentation may be done: 'You do a heck of a lot more experimentation [for publishing] than for a real-world application [...] all of the experimentation that you're doing, [...] some of it is essential. A lot of it isn't. But then again, if I want to keep my job, I better get published' (P4). Despite the power of these structures in shaping how much experimentation was done and when, they were not leveraged to influence researchers to work in a more environmentally friendly way. 'There's lots of sort of schemes and policies within the university around being conscious of our use of resources, and trying to be sustainable in our, our activities. Generally, there's nothing that's sort of specifically directed at these sorts of, you know, resource usage' (P5). This was despite there being external pressure (from funders and conferences) around other elements of academic work: 'The advice that I see is far more related to travel. I can't think of anything related to computational equipment, to be honest' (P4). Lisa believed that these structures could act as lever points to influence researchers, 'like a best paper award, with that in mind would motivate a lot of people' (P15). As could the competition between commercial companies and academic work: 'if one of the companies create something with net zero in mind, [that could] motivate researchers to do something similar. It will become the norm' (P15). Despite this, at times the structures of universities appeared to motivate a lack of sustainability rather than encourage it: 'we get very frustrated here, with the university about how much they demand we replace things [...] as soon as the server is out of warranty, they're like you have to have a new one. And it's like this computer works absolutely fine [...] often the university says, 'Well, we're striving to be a Top 50 World University. So, we always had that the latest tech'. It doesn't work very well with sustainability' (P2).

3.1.2 Persona 2: The Research Software Engineer

Though the category of Research Software Engineer (RSE) did not initially form a part of our stakeholder groups, it did emerge as a crucial group when thinking of intervention points in net zero DRI ambitions. As such, we have drawn the experiences of the RSEs that we spoke with into a separate persona than that of 'academic researcher', despite RSEs often being involved in academic research of their own.

The RSE role has been increasing in significance in recent years, in line with the growth in computationally intensive research. Eric had received an RSE fellowship (P6, P11) and was working to train and guide researchers in how to improve their code, alongside his own research activities. 'It's not just about making the code faster [...] if it's a really low efficiency [even if it's quick], then it's not making a good use of [the machine]' (P6). He recognised the growing significance of his area, suggesting that universities should '[...] employ more people that do my role [...] because scientists are not trained to develop software [...] they're trained to do science' (P6).

Despite this, there were times when he was unsure whether he was providing the best advice to the researchers under his training. This is because using DRI facilities efficiently required knowing how it was already being used, as well as how specific hardware features were being exploited. This could differ across different facilities too. *'If only half the cores are doing something, we can power off half the cores, then suddenly, me using those cores to do 10% of the work probably is now a bad idea. And I don't necessarily know that. [...] And even within particular computer chips, there are areas of the chip used for very specific jobs. And in some chips, if you don't use them, it powers them off. But some of those are high performance pathways that we feel like we ought to be using. But if we use them, then they have to be powered on. [...] And I say [this] as someone who this is kind of my area of expertise, but actually, it's quite complicated, and it can be very difficult'* (P11).

Eric suggested two ways to counteract this problem. The first was for different vendors to create libraries that could automate some of this process (P11). The second was for users to run test calculations on different numbers of cores to work out whether the run was *'acceptably faster'* (P11). He did note, however, that this was difficult to define, and that the answer would depend *'[...] on whether you were looking at a PhD student, and [...] whether they were going to get enough chapters for their thesis in time, or a postdoc getting papers, or whether you were an HPC facility going, well, how many scientific simulations can we get through in the lifetime of the machine?'* (P11).

This challenge around interpreting the efficiency of a run was evident in other ways too. Not all facilities provided, by default, a breakdown of how carbon-intensive a run had been. But even in cases where *'[this information is provided and made visible to] to researchers [...] okay, this simulation costs 20 grams of carbon. What does that mean? How much is 20 grams of carbon?'* (P6).

Despite the challenges of knowing whether uses of DRI and the coding that this depended on was as efficient as it could be, Eric suggested that there were some intervention points that could be mobilised to work towards net zero ambitions, *'I'm really keen on pushing people to share [...] data. And, you know, setting up the facility so that people can share that data with a wider audience so that other people can use it, and the kind of shared cost in terms of CO₂ for that data goes down'* (P6). His own research was climate-based modelling, so he often questioned *'do I really need to do 20 simulations of a model? Or can I limit it to 10?'* (P17). *'You can [also] make the option of, well, I want my result fast. So, I'll use a high carbon intensity processor. Or you could have the option of going, well actually, I'll use the slower processers, but know full well it'll take me longer to get my data or my result, but my carbon footprint for that result will be lower'*, though he noted that *'sometimes you've got no choice but to get the result, quickly'* (P17).

3.1.3 Persona 3: The University HPC Manager

Nick runs the HPC facilities for a University in London. He manages two separate facilities, one for *'high throughput computing. So that is where you have to run 1000s of similar workloads, but on different data'*, and another *'machine [that] is a large parallel machine, [...] which is mostly used*

by chemistry and physics [...] they have very different user bases and somewhat different design requirements' (P19). Though he intended in the future to bring these systems together, they had initially developed separately due to high land costs and space limitations in London (*'when space becomes available [it] kind of decides where they get put'*), as well as the different ways that they handled network traffic (between nodes: east west; from and to disk: north south) *'if you don't have a lot of money, and you never have enough money, you can optimise for one or the other of those usually, but not both'* (P19).

There are various inefficiencies he can't remove from the systems he manages; for instance, consolidating resources is important as it allows for the heating, cooling, and ventilation of resources to be shared across projects and facilities. *'Consolidating facilities makes them more efficient to run. We can control where the power comes from, all that stuff'* (P19). This is sometimes limited, however, due to the security and privacy barriers of specific research contexts. Where equipment is purchased for medical research, for example, a VAT exemption is given but this means that these parts of the DRI must be siloed and only used for medical research to comply with their attendant policies. *'So that's a separate system that is kind of completely ring fenced, and it's got very harsh security. And it's not just security, it's like process because it needs to be auditable, right?'* (P19). Where clusters are separated, *'[they're] not effectively used. Like it [can be] really, really low usage. But you know that the machine's got to be on the whole time. And even when it's idle it still [...] draws a reasonable amount of power'* (P6).

Poor coding is another such inefficiency. Different institutions have different access policies around who can use the HPC – at his institution, anybody (i.e., staff or student) conducting research was allowed to use it *'free at the point of use, through a fair share model', 'everybody at the university [...] if they've got a computational need [...] Obviously, I'd caveat with that they need to be suitably trained on the infrastructure as well, so that they're not wasting it'* (P3). Nevertheless, users, at times, write code that demands more storage or memory than is required. *'If we can pinpoint who were the biggest wasters, then we can sit down with them and educate them more, or if it is their code, then have discussions about how we can rewrite their code and make it more efficient. But that's where the research software engineers come in'* (P3). Nick was aware that his university had invested in RSEs more than other institutions, *'universities tend to have several of them. But [redacted university name is] playing catch up with this. So [they've] started with [their] first one at the start of this year'* (P16).

One of the facilities has been renewed in recent years, upgrading it from a more antiquated system. The upgrading of the system, however, was not just in relation to its hardware though, *'we're currently going through a restructure within IT services and in doing so, we're bringing in some [...] sort of like engagement in order to improve our relationship between IT departments and the end users so we can capture their needs and understand them better'* (P21). This was a particularly important part of managing the system, as capturing user needs would allow them to invest and improve the DRI facilities accordingly. Despite the recent upgrade, in general *'we have what we call a rolling refresh. So rather than buy a single HPC system, run it for 5, 6, 7 years,*

and then replace it with the new one, [...] our strategy is to buy a few compute resources every year. Install that as new and decommission the oldest' (P16).

Though he accepted that there were sometimes frustrations around the lifetime of computing components, with some – like Lisa – being frustrated about the retirement of equipment that is out of warranty, *'I do not like running things beyond the warranty life [...] it's okay to run compute nodes a bit longer if you have to. But [...] storage is where our users' data lives, and that's the most important thing and we cannot lose that' (P19).* *'The answer there is not black and white, it falls in a different place for different people at different times. [...] If you can no longer get updates for a system, and that is critical security level, then that system, you know, is better off being recycled, because if we were compromised, then that compromises all the researchers' data. [During COVID] many people were at home, you know, and that data was really interesting to a lot of people. [...] So, we spent a lot of time protecting that data' (P13).* *'Aspirationally, we always have aimed to cycle network equipment on something between a seven-to-eight-year life cycle. [...] quite often, the vendors will stop releasing security patches after [that]' (P21).*

When it comes to the purchasing of new components, this is limited by the funding available, rather than the current needs of the users, *'it's more a case of the more we add, the more it would be used, there's basically more demand than supply from that point of view' (P16).* Funding for upgrades comes from a portion of the central pot, which is then topped up by money from research grants. *'We don't have [...] a flat rate [...] instead of doing it that way, what we do is we ask the person to cost in to buy a server, for the high-end computing [...] they buy a node [...] the rate starts off about 10,000 pounds [...] as equipment in a research proposal, and then that buys the node' (P12).* Not all research projects which use the institutional DRI cost this into their grants though *'people on smaller grants might be using the [system] for free' (P12).* Where projects had costed nodes, Nick would then have choice over what type of equipment to purchase to add to the institutional DRI. *'Energy consumption, for obvious reasons is very important to us because we're in London, and [...] those costs] are included, [they] are part of the total cost of the solution. It's not just the actual hardware' (P19).* He would also work with his institutional procurement team and go through *'an options appraisal, we'll go through various standard things [...] the history of the requirement and then there's a set of topics for us to discuss [for example] sustainability and disposal' (P10).*

Nick was aware of a range of design innovations that made net zero DRI provision more possible, but funding, local context, and planning can limit these options. Some systems involve *'liquid cooling, so water will go around the machines, extract all of that heat and then [...] that hot water [will] essentially feed into a department who needs hot water' (P2).* But *'if it's new technology [...] the greener more cutting-edge things might be more expensive [...] there isn't that additional budget to pay for something that might be a more sustainable option' (P10).* Other institutions, like those in the North, were able to make use of their local environments to provide more sustainable cooling: *'at this time of year, [they] don't have to bother cooling it, [they] literally just pump it out to the other side of the road, [they] take out sort of mid-20s, 30-degree water and it comes back*

again, nice and cold' (P13). He had planned for the new 'onsite data centre [to be] built with liquid cooled doors. It's in the basement of a building where there are students and offices. They could have taken that hot water and used it [...] as the first stage of heating [...] before it goes into the boilers. And Estates never hooked up those pipes, like it was in the design spec. It's really frustrating' (P19). He believed this lack of joined-up thinking existed at a broader level too, 'we could as a country, say we're going to build a data centre for university HPC, [...] we'll build a solar farm and a wind farm next to it, we will power it through that, [and] we will recover the heat' (P19). Universities did hold some power in pushing for net zero ambitions, 'what a university could and should be doing is lobbying the government' (P2) and with technology companies too, 'the spend cumulatively, for the nation, for universities, is big enough that we can have a voice at the table' (P13).

3.1.4 Persona 4: The Research Council

Marie works for a UKRI research council that funds research in topics in which computationally intensive tasks are common. The growing demand of compute resources represents a concern, but she recognises the importance of compute power for researching important topics, and the complex problem that this produces. 'That's one of the arguments that I think we've been trying to have is that using, like, yeah, energy hungry, hungry, compute to do climate modelling [...] the thing is, without something like [Tier 2 system] that research either just surely wouldn't happen, or it would take a lot longer on something smaller. [...] It's that kind of potential, how much will you save going forward? Versus how much have you used in [...] the moment?' (P22).

This problem leads her to consider effectivity of funds and resources as a potential solution. 'The financial incentive often kind of aligns quite well with the environmental incentive here [...] to drive forward efficiency of use of our funds. And from the environmental perspective, we just need to make sure these resources are being used as effectively as possible [...] individual researchers aren't incentivised to do things efficiently, I think. They're incentivised to produce as much high-quality research as [possible]' (P22).

One way to ensure effectivity of funds is 'a clear drive towards FAIR data' (P18). 'Making those data sets more available and able to be reused' (P18). Catalogues are one way in which this can be done to: 'store simulation code and various models or software to [...] reuse, I think, I guess that is a key thing in the future, you know, to actually be able to reuse rather than spending extra effort to replicate some code or something, to use that as a key resource for future reuse for our community' (P18). The FAIR principles go beyond more general data retention policies, which represent a challenge in thinking about net zero ambitions. 'You can have a principle there of 'oh well, we'll just keep every single thing. And if it's useful someday, then that's great' [...] When you're getting to this point of, you know, basically massive information generating machines [...] then you have to start choosing, don't you? Not just [...] because the financial cost of storing these things, but also the environmental ones' (P22). There has been 'a massive increase in requests for data storage' (P18), with strategies (e.g., decision-making processes, data value checklists) having to

emerge to *'assess the value of that data, whether and how long it should be held for. If stuff, you know, doesn't have the right [...] useful metadata to be reused, they will make that assessment and not store the data'* (P18). A tension exists, therefore, in ensuring that the FAIR principles are followed to reduce the duplication of computationally intensive work, and that this is not misunderstood as simply *'setting the policy around [...] storing pretty much everything'* (P22). Where data should be stored, some technological solutions – like tape storage (P18, P22) – were emerging as lower carbon possibilities. But there was space for other solutions too: *'communities in the arts and humanities, in particular, in practice research [are] crying out for good platforms to allow them to archive those intermediate research objects that capture the process of practice research, but currently lack anything that would be adequately searchable'* (P23).

Looking forward to thinking about future research and the systems that will support this, efforts are being made to imagine the DRI required. *'We would go out and try and engage with the community about what are the actual requirements for the next stage supercomputer? What kind of technologies might people want to use? What kind of scale might people want to use?'* (P24). This involves a certain level of *'uncertainty'* and *'a risk attached to it'* (P24). It involves imagining not just the kinds of future technologies researchers will need, but the problems that they might want to research. As daily life evolves and changes, and more of social life moves online, other domains – that sit beyond STEM subjects where such work is more typical – may begin to make use of computationally-intensive methods. *'There's definitely potential for like a bit of a big bang in the use of [...] compute [in the humanities]'* (P22). *'So suddenly in social science [...] if there's data a lot more available, [...] a lot more people could be showing interest in doing these things. [...] Everyone will do their little bit, as much effort as they feel they want to, and then they will start coding. And [...] there's the likelihood that the code that is written will be less efficient'* (P22). Marie noted how this would contribute to the already growing significance of RSEs. *'Over the last couple of years, research infrastructure has funded a lot more software engineers type people's like, fellows. But I guess the trick is, how are they recognised in their careers within universities?'* (P24).

3.2 Causal loop diagrams

Beyond the personas, we have used causal loop diagrams (CLDs) to explore the breadth and variety of our interviewees' experiences within wider organisations and communities of practice. CLDs illustrate diagrammatically key issues that emerged from our interview analysis by capturing the causes, influences and impacts between sets of elements. We consider the factors illustrated by the diagrams as influences rather than necessarily 'causes', as the sample used was not, in our opinion, representative enough to ascribe causality. It is possible the diagrams may not capture all possible factors, nor do they indicate any dimension of scale. CLD models are best thought of as a method for exploring factors and how they might interrelate or be co-dependent. We have created a series to explore particular key areas relating to DRI.

Nodes within the diagrams are colour coded as follows:

Orange – central issue under exploration in the CLD and / or connecting to another CLD

Red – primary socio-technical factors which impact on DRI

Green – HPC technical / provision factors

Yellow – wider research culture

Purple – socio-cultural factors

3.2.1 Demand for DRI

Increasing demand for DRI sits at the core of our findings. DRI use is increasing due to advances in science, increased capability and ease of use of HPC. Increased adoption of compute intensive research techniques such as machine learning is also seen in domains not previously considered to be computationally intensive and where ‘...many people don’t view them as computers but scientific instruments’ (P9), (e.g., computational chemistry, synthetic biology). Additionally, the social sciences are developing new approaches using digitised artefacts, and examining social life as it occurs in online spaces with one interviewee noting that ‘... this is an evolution and a cultural shift ... when you’re using computational techniques to study social science that’s ... more alien to the social science community than it is to the ... environmental or the engineering or STFC communities (P22). These advances across disciplines and HPC capability have meant that demand exceeds supply and each replacement results in a growth in capacity.

‘...the systems are getting more and more complex. The size of the system is increasing. The timeline on the simulation is increasing. So obviously we need more resources. And we need faster computers and we’ll need more optimized software’ (P20).

The increase in demand across different disciplines is also reflected in students’ expectations and they are increasingly being expected to use HPC as part of their studies, further embedding DRI use in academic tools and practices.

Increases in demand can also come from inefficiencies as reflected in the next section, for example through academic, disciplinary and funder expectations or through inefficient code or redundant research.

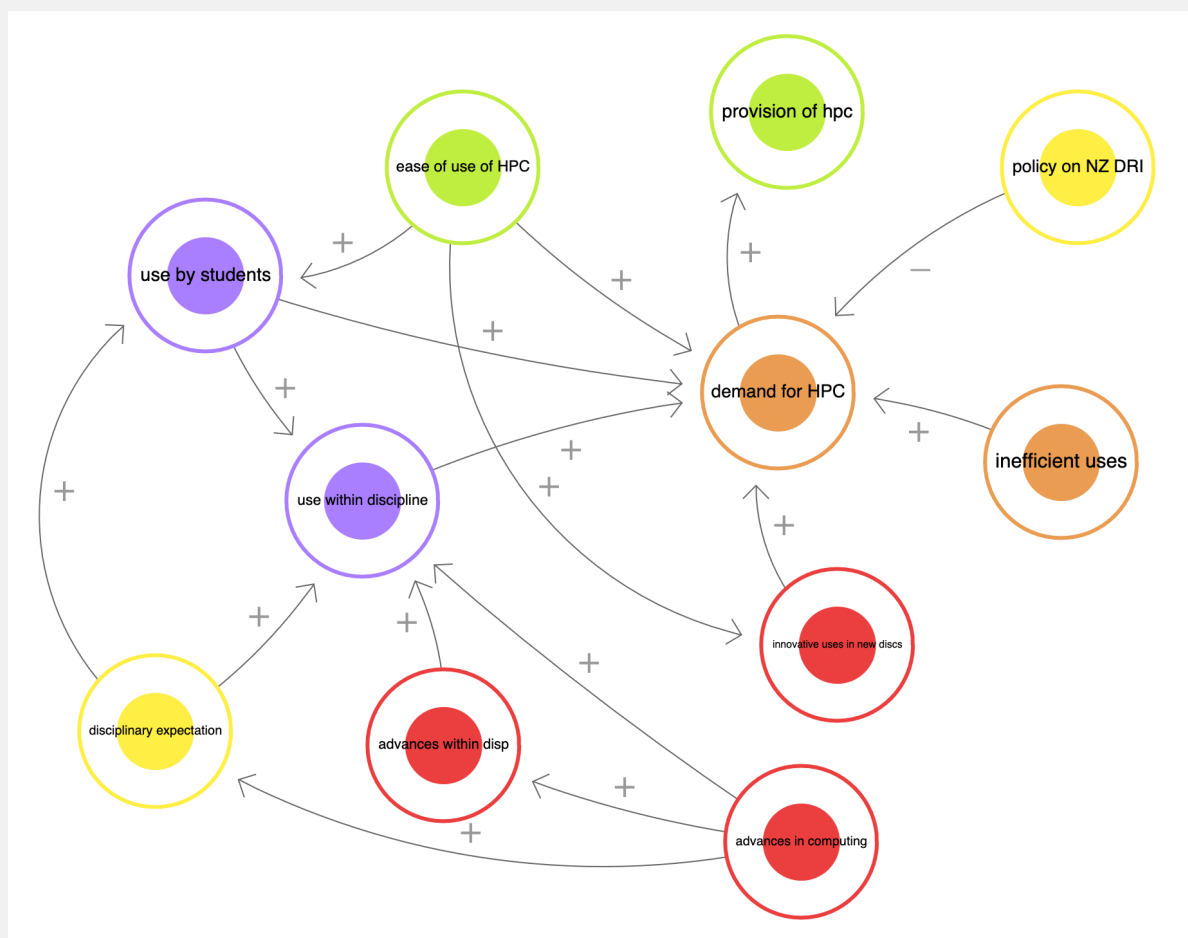


Figure 1: Causal loop diagram showing the influences on demand for HPC.

See <http://bit.ly/3ZSPXz3> to further explore the factors.

Data storage is a further demand issue for DRI, increasing the need for storage facilities at institutional and national level as models become more complex and data volumes increase rapidly.

'...they have to store that data coming from that research in ... data centres. ... that's probably the biggest use of our DRI portfolio ... when the researchers put in a proposal [to research council], they have to include a data management plan, where they will specify kind of what type of data will come out of that research, how much data and which data centre it might go to... there's definitely been a massive increase in requests for data storage' (P18).

3.2.2 Efficiency and waste in DRI

Our exploration of efficiency and waste as it related to use of DRI might be loosely grouped into 1) socio-technical and 2) socio-cultural with attendant cross-cutting factors that further impact efficiency of use. Inefficiencies result in over-provision and / or under-use of existing DRI with associated impacts on the environment. It's important to recognise that we are not commenting on the efficacy of a focus on efficiency for achieving net zero. DRI has a significant impact at

all stages of its lifecycle and requires significant energy just to operate at all. Climate impacts are therefore in part a relation to the size, extent, and replacement of the DRI itself. Efficiency gains are also linked to rebound effects (Hilty et al., 2006; Widdicks et al., 2023). Efficiency and wasted computation did feature significantly in our interviews, as it does directly relate to many stakeholders' daily experiences.

Firstly, there are software development practices such as lack of code optimisation (either through inefficiencies or bugs which slow runs: *'...in corner cases, of course, things might fail... everything works, a change is made. And then it mostly works or you know, things break.'* (P7), or codebases poorly attuned to the hardware being used (*'...as new architectures and technologies come on, we do need to keep the codes in line with the new practices coming through'* (P3)) which impact directly on efficiency in use. These inefficiencies can be further impacted negatively by the expectations and culture of the research group – for example in using legacy code or where time pressures mean there is little time to improve or check:

'“Oh, this is what I got from, you know, the PhD student before me and this is what I'm going to use”, kind of way of thinking, it's a lot of repetition of mistakes' (P7).

'...it always comes back around to what are you trying to maximise as your efficiency? Is it cost? Is it time? Or is it total compute time? And essentially it always comes down to human time. That's what we want to maximise ... and that doesn't tally very well often with efficiency of compute time' (P2).

As we have noted in our recommendations, these issues can be addressed with the support of RSEs whose specific job is to help optimise code and to provide support and training for the wider user community, for example *'helping the physicists to write efficient algorithms. ... these people, they have a repertoire of initiatives that they could bring into some piece of code to make it more efficient, very quickly'* (P1). RSEs *'enable more research and better research, and we'd get more bang for our buck, really, on our return on investment'* (P3).

The second area of efficiency and waste is in doing more than might be necessary because of policy, practice and culture. For example, researchers using more resources or higher priorities / speed than required or replicating computation to meet reviewer expectations / timescales; or running more simulations or experiments than might otherwise be required. Superfluous simulations could be the result of an unsuccessful grant application, attempts to meet funder or publisher demands for more justification, poorly formulated or redundant research questions, or where they ultimately duplicate work already done elsewhere:

'There's growing evidence that poor quality research is just inherently wasteful. You know, if you do research that doesn't need doing because we already know the answer, and you're just interested in publishing a paper, or you do research badly so that it doesn't give you a clear answer. So, you may as well not have done it in the first place. All of the resources that went into that were effectively wasted' (P23).

Local academic norms such as a desire for unrestricted use and control of local DRI (see section 3.2.4) and a culture ‘rooted in this sort of 19th century independent scholar, slightly artisanal approach to research’ (P23) further impact on the efficiency of DRI use. Having said this, it is not easy to define in advance what might be wasted and, of course, it is inevitable that some DRI use will turn out to be a dead-end or not result in further funding, so care needs to be taken to balance efficiency demands against innovation in academic practice using DRI that could increase the quantity and quality of research overall.

‘...there’s a lot we can do in terms of educating and disseminating best practices. But also just recognising that researchers work in the wrong way. They don’t stick to nine to five, they will run some wacky things and we want to enable cutting edge research as well. We don’t want to say right, everybody needs to fit the certain size T-shirts. And you can only do your science if you can wear the right T-shirt.’ (P3)

Better use of common protocols such as FAIR and open source / shared data and code should improve efficiency and avoid unnecessary duplication of DRI usage but this will take time and resourcing to be properly implemented.

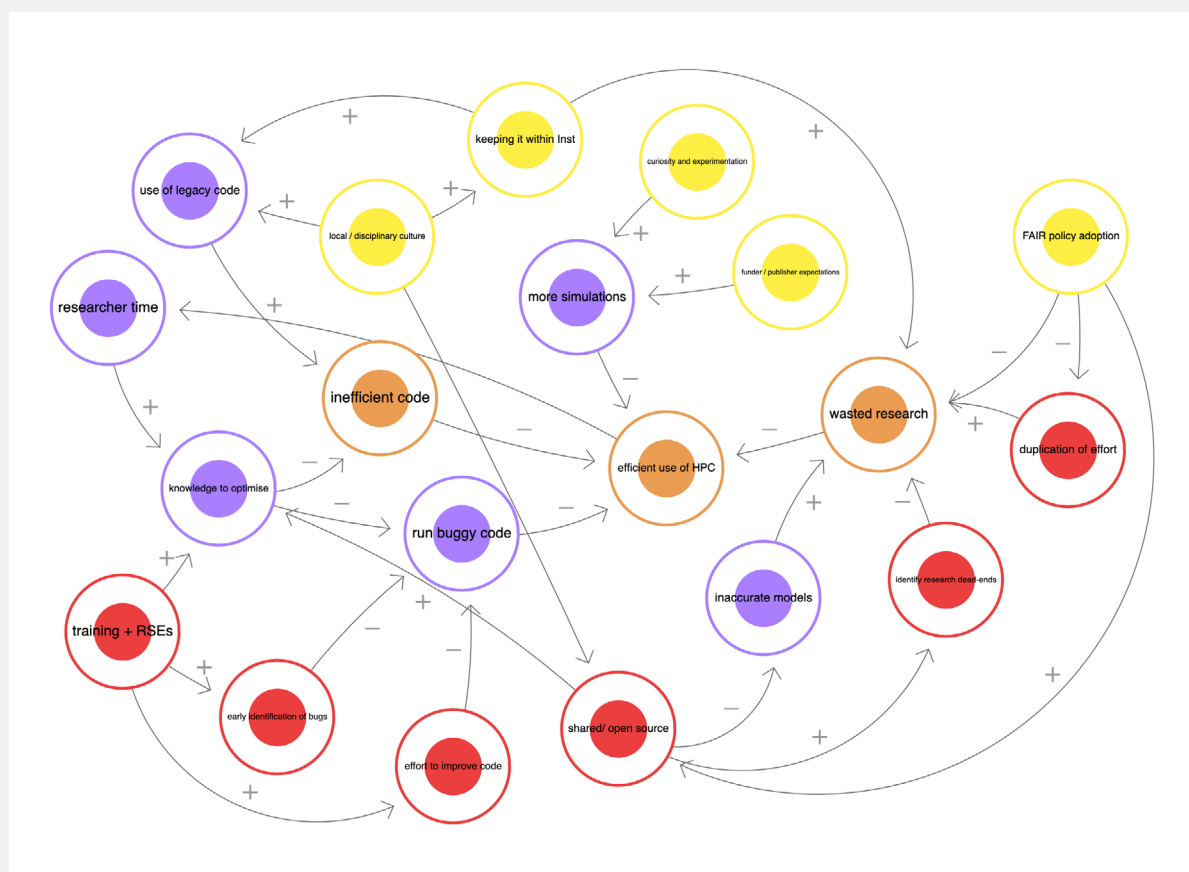


Figure 2: Causal loop diagram showing factors of influence on the efficiency of DRI usage. Explore here: <http://bit.ly/3ZCoVfG>

3.2.3 Using ‘own facilities’ versus shared HPC

Drivers of own vs. shared compute facilities are shown in [Figure 3](#). In parallel with the increase in HPC capacity and demand, ‘own facilities’ are also in demand for a variety of uses. For example, as P20 notes, having instantly available compute facility facilitates testing out code and software before trialling it on bigger datasets and machines:

‘...if you do software development, or if you do methods development. You do need to try things very often to see if it runs: does it work? Does it break? Where it breaks? Fix it, well, you need to debug it, basically. And doing that in HPC machine, imagine, if you need to submit a job just to see if something runs and then you wait one day, two days, three days, a week and then you realize it, I have the code is broken and I need to debug it and then you debug it and then you find a new bug. Okay, imagine, it makes no sense. I think it from the point of view of optimizing time, this would make no sense. So that’s why we tend to have machines that can do some of the work [locally].’ (P20).

There is also a cultural influence on whether ‘own facility’ is chosen and at a local (HEI) level, provision of DRI is shaped by the need to spend budgets in funding cycles / within financial years as well as researchers’ desire for ownership and control of ‘own’ facilities (e.g., local workstations, compute clusters). Academic practice, local academic norms and a desire for unrestricted use and control, and concerns over security of shared HPC further leads to the provision and inefficient use of extra local DRI. However, this is changing as HEIs discourage it and provide training to move to shared machines which can be used more efficiently:

‘...at [HEI], they’ve really clamped down on that [own machines] but in an effective way. ...not that you absolutely cannot do that, but it would be much better if you use our machine instead, let us help you get on that machine. ... it was much more the norm for not just the groups, but also individuals to have that sort of hardware... great big desktop server, things that, you know, you can hear from, from down a corridor ... that’s not an effective use because these machines are basically on all the time. And they’re not being used to 100% so they’re wasting carbon.’ (P6).

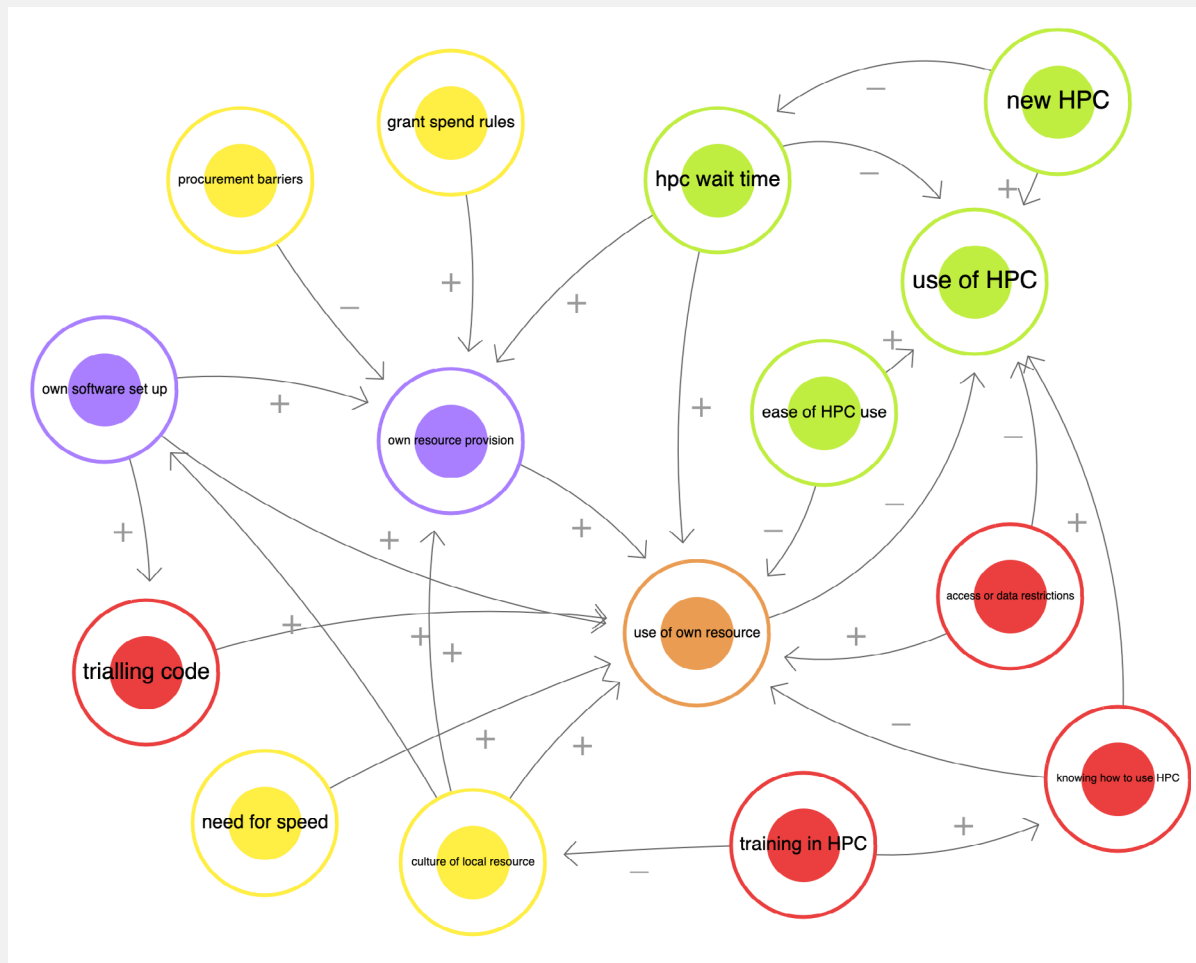


Figure 3: Causal loop diagram showing the influences on own DRI v shared HPC use.
Explore here: <http://bit.ly/3l3tH72>

3.2.4 Provision of DRI

The other side of the DRI is in the process of provision and, in pursuit of net zero DRI (Figure 4), there is currently no clear ownership, oversight, or, in many cases, resource, resulting in little overt or cultural pressure to consider or embed net zero in DRI activities or provision. Where additional cost might be incurred, this is often remote 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, with one academic saying, '...there really is zero external pressure. You know, the university when they made the net zero announcement. Pretty much nothing changed' (P2). If net zero DRI proves more expensive or complex than the default DRI, there are often no resources provided to address the cost and complexity and it is unlikely to be addressed:

'... there isn't that additional budget being made available to pay for something that might be a more sustainable option. To factor in that maybe it's a lower energy costs or whatever, that through life cost of it. That's not really being made a priority at the moment ... [so] having a little bit more support from the senior levels in terms of, there is money to fund that, that drives things' (P10).

More generally in the context of providing DRI resource at HEI level, budget and procurement constraints determine what can be provided. If additional facilities are needed there can be insufficient time to properly explore the options:

‘...we often don’t get a lot of time to respond to the need. And that, I would say, is quite a big driver of our decision making. For example, somebody comes to us and they need something very soon, that will prevent us going out and doing something like a tender or mini tender where we could really evaluate sustainability options. Instead, in order to procure equipment quickly, we rely on the reuse of something we’ve already got from elsewhere, or we’ll go out through one of our framework contracts and buy from an existing supplier’ (P21)

Finally, as with the other areas addressed here, there is a need for training and consistent methodologies for assessing carbon and 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. From a funder perspective *‘...in order to mandate carbon budgets, you need to have some understanding of what is realistic, what is appropriate. Right now we have none, we’re just having cost. And the reviewer says, Oh, you’re asking for this sort of money. I’ve got some sort of feeling that this is reasonable or not. With carbon, we don’t have this’ (P1).*

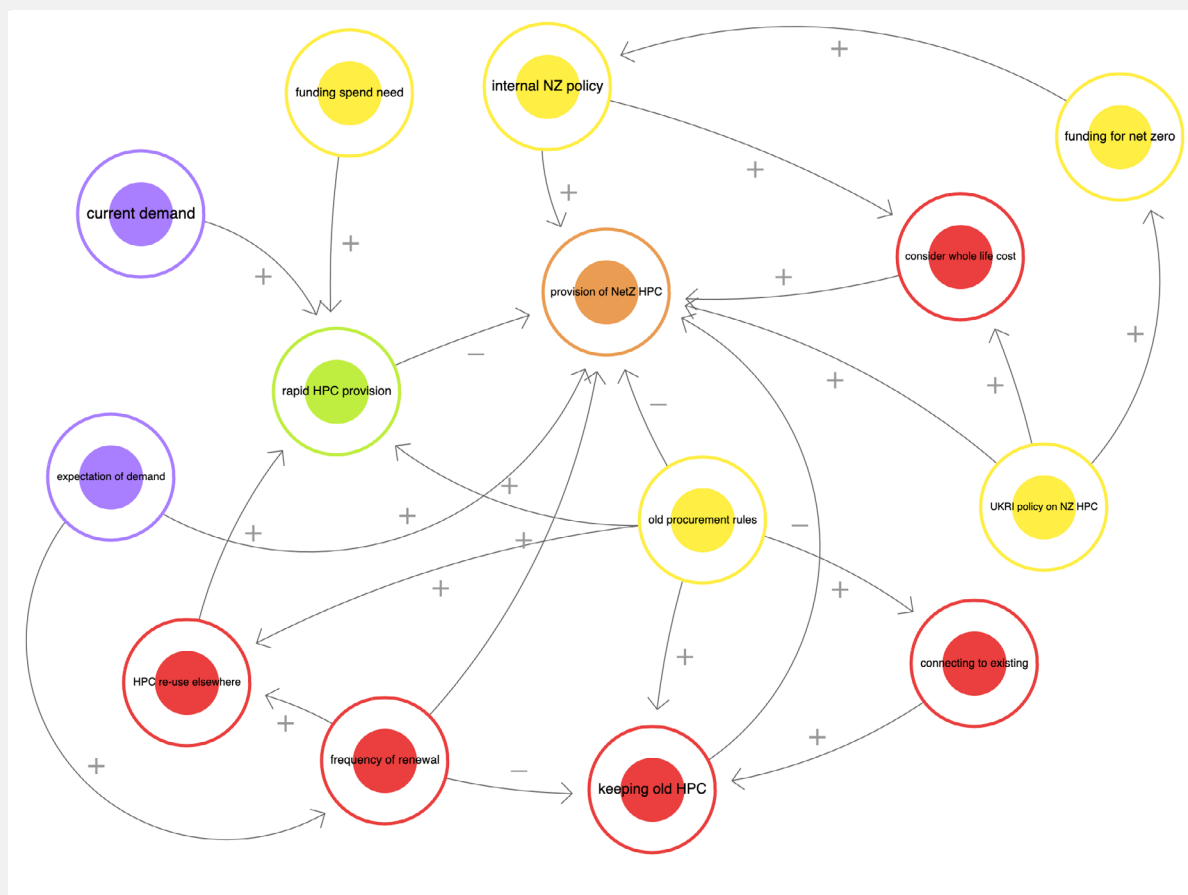


Figure 4: Causal loop diagram showing the influences on provision of net zero HPC
 Explore here: <http://bit.ly/3JqIjGV>

3.2.5 Dynamics of FAIR in relation to net zero

FAIR practices in relation to net zero are broadly influenced by four groups of factors (Figure 5): 1) data generation (red) representing the broad generation of research data; 2) FAIR processes (blue) including identifiers, meta-data, preservation and reuse; 3) data infrastructures (yellow); and 4) ancillary data activities such as transfer, sharing, integration and quality (brown). Carbon emissions linked to net zero are represented with a green node.

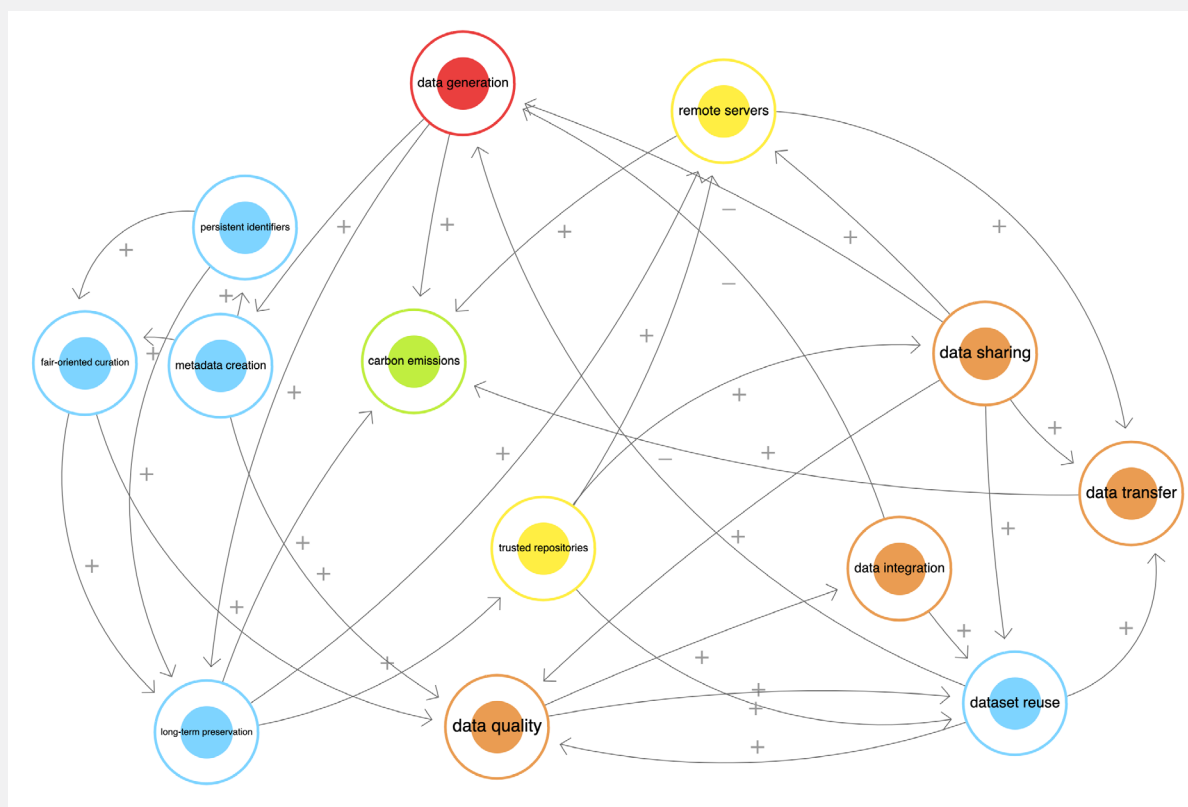


Figure 5: Causal loop diagram illustrating a model of FAIR to net zero. Blue nodes represent FAIR-related processes, green carbon emissions, yellow infrastructures and red data processes. Explore here: <http://bit.ly/3F4ZaMH>

While we cannot be conclusive, our initial work suggests FAIR processes do not appear to show direct, substantial links to net zero in terms of either carbon emissions increase or decrease. However, when implicit factors are brought into the equation then the forces in play between FAIR and net zero become more visible, including the role of data quality, data sharing and the adjacent data transfer, and reuse. The influencing factors mediating between the two categories reveal the connections between FAIR and net zero (Figure 6). FAIR-oriented curation, metadata curation, and dataset reuse are influential on carbon emissions when data quality, data sharing and data transfer factors implicitly involved in FAIR processes are included. For example:

- FAIR processes influence data quality, which in turn stimulates data reuse, triggering more data transfer, which ultimately influences (increases) carbon emissions.

- Remote servers where data is stored for long-term preservation of FAIR data are linked to data transfer, then in turn to carbon emissions.
- Growth in data sharing allows more data reuse which improves data quality. Better data quality increases dataset reuse possibly reducing redundancy of data or enabling computation and data transfer, which are ultimately linked to carbon emissions (likely reducing if the redundant data was expensive to compute, or increasing if this enables computationally expensive new science).
- Improved sharing of codebases which are better optimised and ideally exploiting low energy and specific hardware features (not shown) may lead to reduced runtime energy consumption.

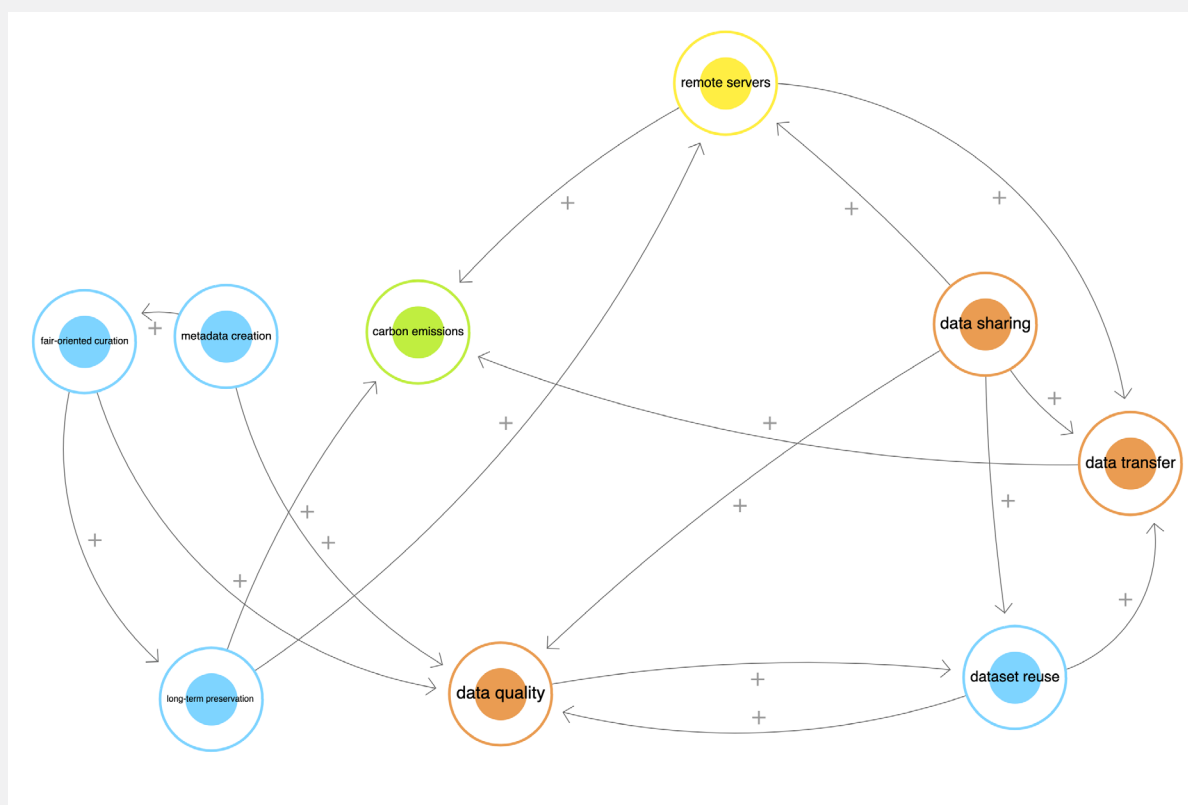


Figure 6: **Subset diagram highlighting influences through mediating factors in the system FAIR – net zero: data quality, data transfer and data sharing reveal the underlying connections between FAIR processes and carbon emissions.** Explore here: <https://bit.ly/3L0gcQ5>

We note contrasting effects: data transfer grows with data reuse, and potentially pushes up carbon emissions. However, data reuse reduces data generation, the latter being a factor in causing carbon emissions to increase. Follow-on work would enable capture of the magnitude of forces at play. Further steps should include: extracting more mediating factors between nodes; explicating variations over time of quantities represented by the nodes; and understanding of the differences between long- and short-term consequences of these influences.

3.3 UK HPC usage data

The following graphs outline usage on the ARCHER2 service, <https://www.archer2.ac.uk>, one of the UKRI's Tier 1 services. ARCHER2 was commissioned in two phases, initially around a sixth of the overall compute and storage capacity was installed for early access users (1,024 compute nodes, 131,072 CPU cores). This was followed by an upgrade to around 5,860 compute nodes (750,080 CPU cores).

From the time series for ARCHER2 use (Figure 7), we see an increase in usage (the number of users) from late 2021 (green dashed line) onwards when a larger set of users were given early access and from the full system opening (dashed red line). The jobs run on the system scale commensurately with the number of active users (Figure 8), showing the demand for this new, larger, HPC system.

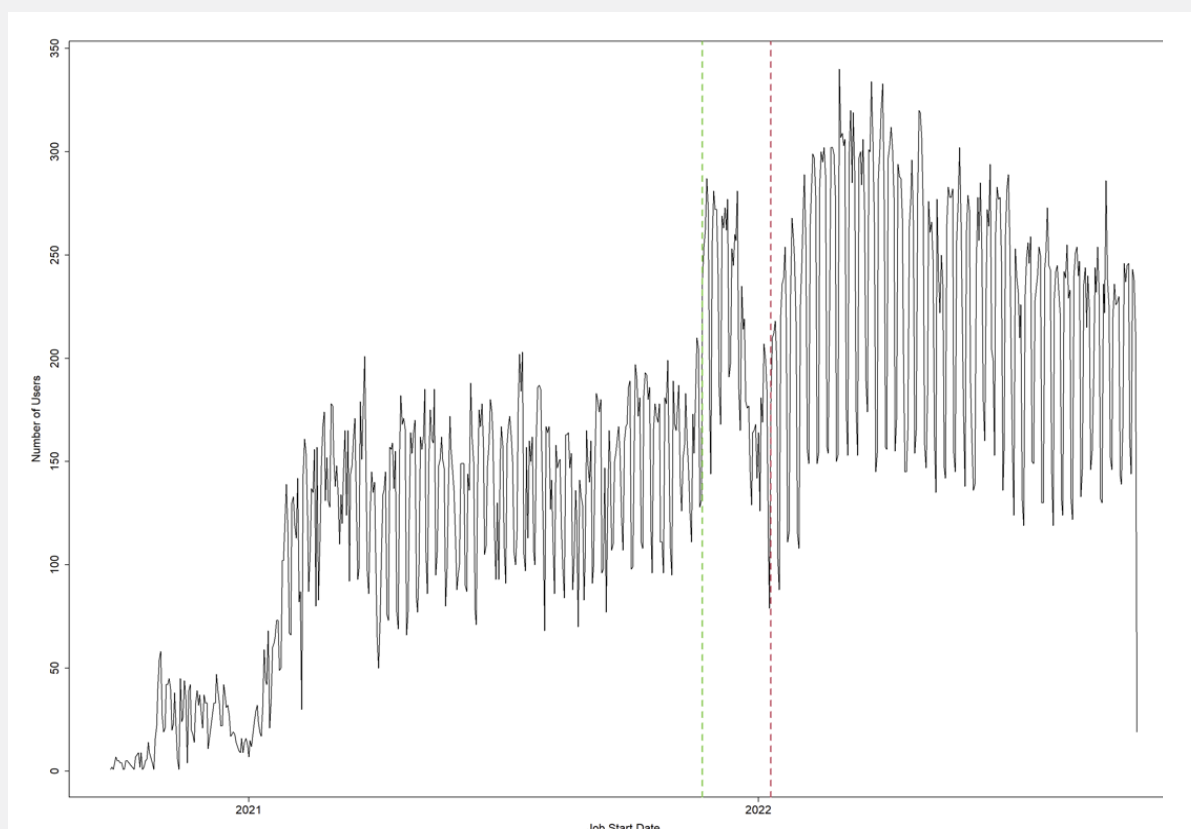


Figure 7: Daily active users on the ARCHER2 system during the early access, upgrade, and production phases of the system.

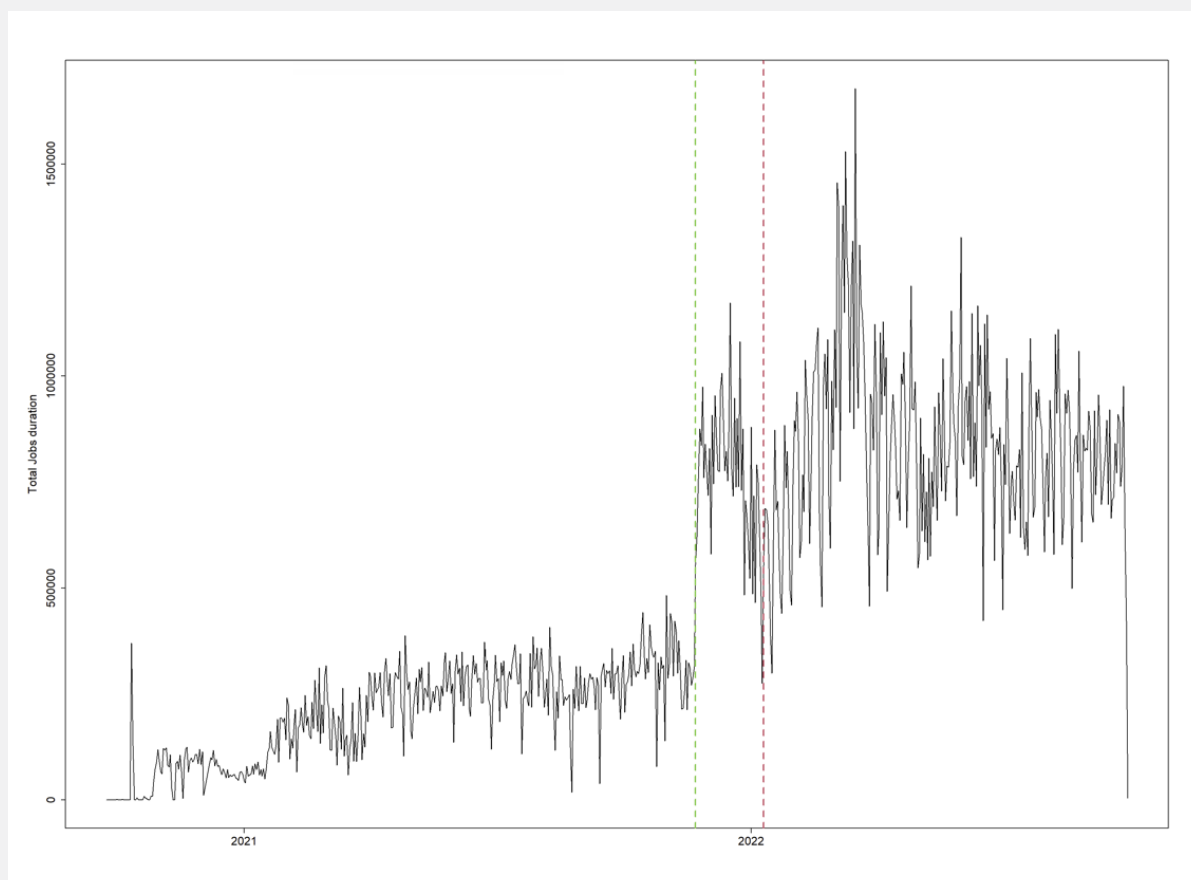


Figure 8: **Total compute job duration on the ARCHER2 system during the early access, upgrade, and production phases of the system.**

Figures 9 and 10 outline the breakdown of the number of compute nodes used versus the time spend on jobs on the system for both the new ARCHER2 and the previous ARCHER service. ARCHER had a similar number of nodes as ARCHER2, but with 24 cores per node, compared to the 128 cores in each ARCHER2 node.

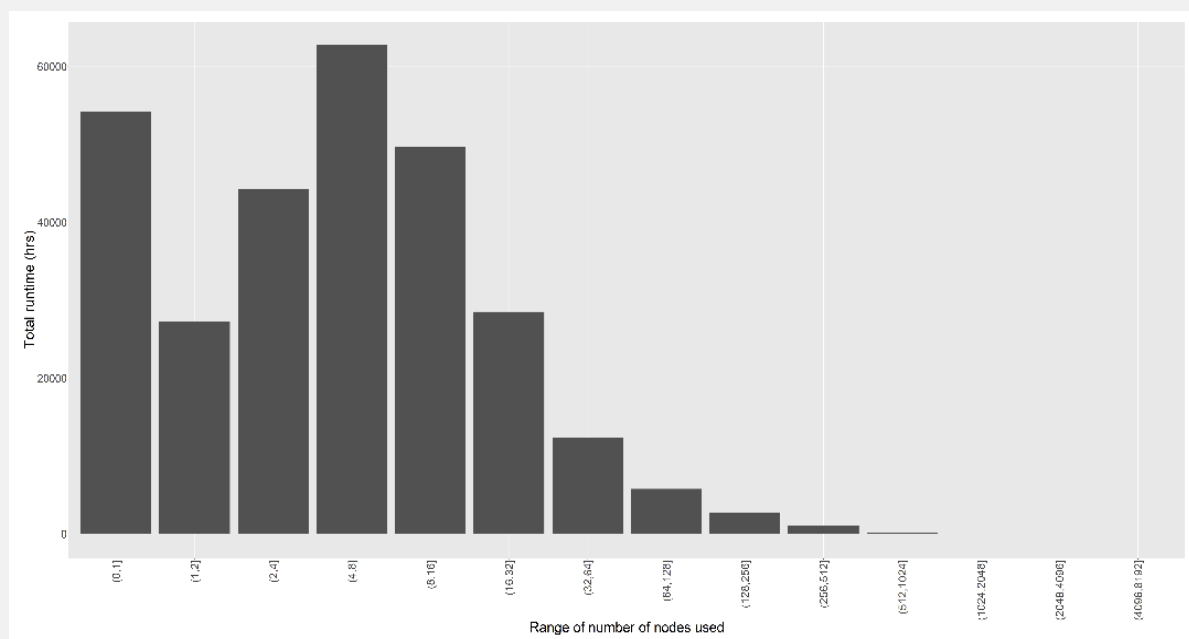


Figure 9: The amount of time used on ARCHER grouped by different node sizes (number of CPU cores).

Users obtained over 5 times more compute cores per node used when moving from ARCHER to ARCHER2. Even with this significant increase in available computational power we see user behaviour patterns preserved, with users consistently moving to use a range of resources above what is available in a single node. For instance, one of the largest node count ranges observed on ARCHER was the 4 to 8 nodes, or 96 to 192 compute cores. The straightforward mapping going from the old system (ARCHER) to the new system (ARCHER2) would be for these jobs to use 1 to 2 nodes. However, we see significant use of the system all the way up to 256 nodes (32,768 cores) and beyond, illustrating the growth in demand.

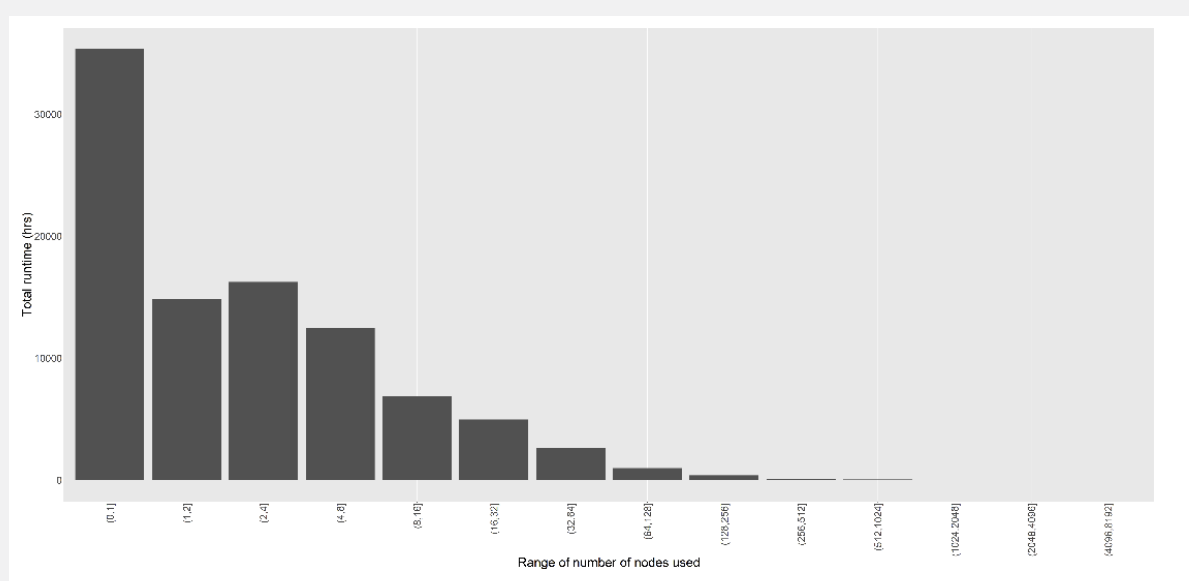


Figure 10: The amount of time used on ARCHER2 grouped by different node sizes (in hours).

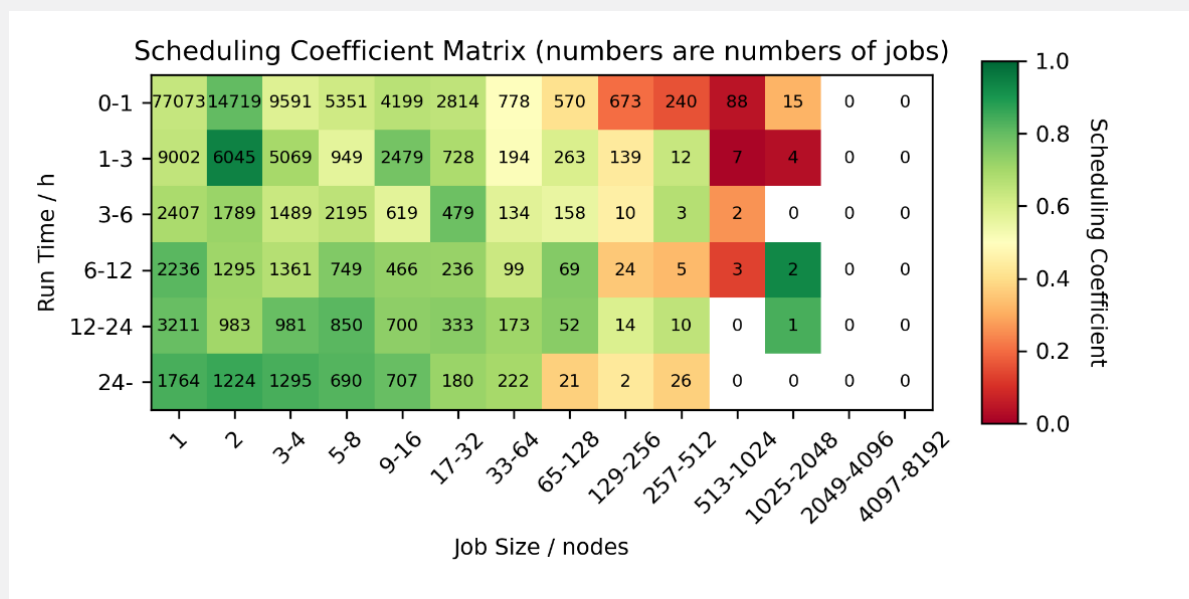


Figure 11: **ARCHER2 scheduling coefficient data for February-March 2023.** Taken from <https://www.archer2.ac.uk/support-access/status.html>. The colour indicates scheduling coefficient which is computed as [run time] divided by [run time + queue time]. A scheduling coefficient of 1 indicates that there was zero time queuing, a scheduling coefficient of 0.5 means that the job spent as long queuing as it did running.

These figures demonstrate the demand for computing that DRI is satisfying, and the progress over time that users exhibit moving from smaller to larger resources. The average compute power per job has consistently increased as users have exploited newer systems. It is also worth considering that the practical constraints and experiences of users will affect their behaviours on the systems. Jobs requiring larger numbers of nodes will queue for longer before running, a fact that can lead users to optimise their application runs to get the largest amount of work through the system that is possible for them. This is evidenced by [Figure 11](#), that shows average queue wait time (scheduling coefficient) by requested number of nodes from ARCHER2:

Finally, **Figure 12** demonstrates the ongoing demand for such systems, showing the active usage of ARCHER2 nodes over a snapshot two-day period in May (although the general trend is the same across the lifetime of the service).

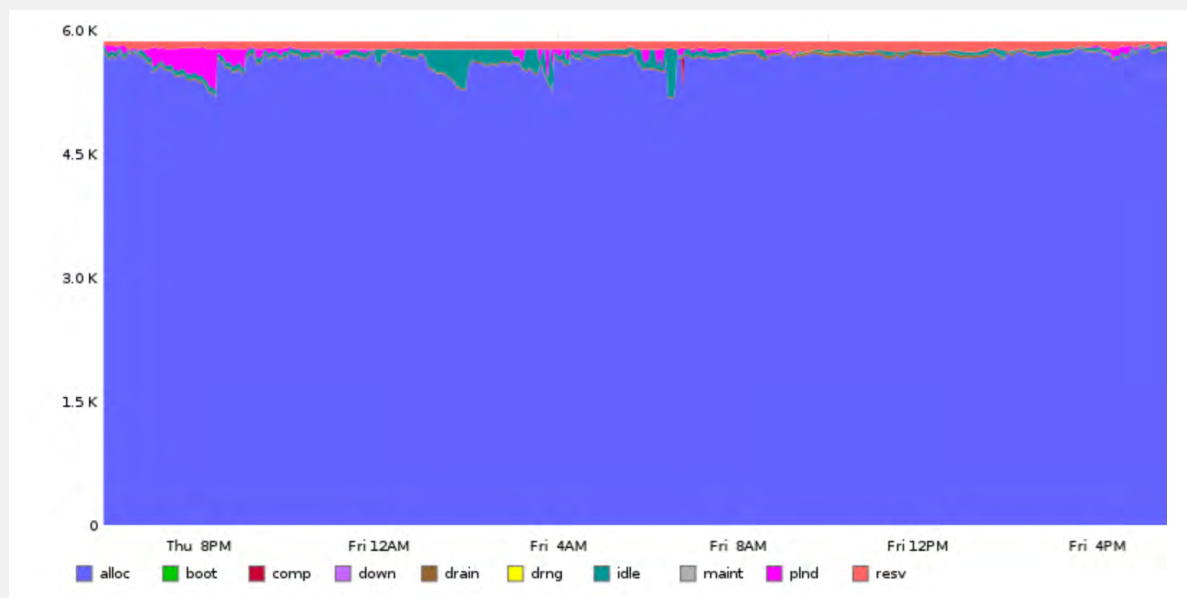


Figure 12: **ARCHER2 system load taken from May** <https://www.archer2.ac.uk/support-access/status.html>. The y-axis is the number of nodes, and the x-axis the day/time

This outlines that the system, in line with other similar systems in the UK and around the world, is effectively running at capacity. The small amount of idle time for compute nodes is largely due to scheduling constraints for larger jobs (jobs that use a large number of nodes simultaneously) on the system.

4. Intervention points and barriers to net zero

Having identified many of the factors influencing DRI growth and related emissions, we now move on to discuss points where action could be taken to improve the situation. Here we consider reducing emissions impact from DRI and the potential for reductions in the growth and use of DRI. It is important to note that not all of these interventions are equal in terms of difficulty or potential significance in emissions or energy reduction terms.

Training and information

Throughout the stakeholder interviews, it was repeatedly suggested that there was a lack of useable information on the environmental impacts of DRI and an inability to make meaningful comparisons across different DRI options to choose which to use. Without this information, it is impossible for users to assess their impacts of DRI usage and make decisions about future acquisition and use. This information could be made available to both users and providers (also see below regarding procurement) contributing to an easily accessible transparent evidence base and dashboards to evaluate runtime energy and carbon performance.

Eco-feedback (i.e., information on the environmental implication of a user's action) however, has been shown to have little lasting impact on how resources of different types are used in other domains (e.g., household energy use (Pereira et al., 2013)). Initially, users may make changes in relation to the information they are provided with, but over time these same users tend to interact and engage less with it. It is crucial that any information provided is complemented by training and advice that not only allows users to fully understand the data being presented to them, but to better understand the environmental effects of different forms of coding and compute in the longer term. RSEs, with their integrated knowledge across systems and disciplines would be well placed to support users and build expertise.

This type of intervention could be integrated into undergraduate training, specification of DRI or of DRI use in grants, and the building of new research teams.

Sharing resources: time and security

It is generally accepted that shared resources, properly optimised, are the most efficient way to provide DRI for most uses – but it has to be available within a reasonable timescale. There are a number of approaches to this, for example using smart job scheduling to place jobs more efficiently to better exploit available hardware with lower runtime energy cost, incentivising slower runs if this leads to a better performance/ energy trade-off, or timed to exploit cleaner available energy sources. Research into this smart scheduling approach could extend to

optimising usage across different facilities and scales e.g., automatically balancing usage of HPC and distributed resources, and/or between DRI at different ‘tiers’, while taking account of differences in hardware architectures and their appropriateness for given jobs.

Concerns over shared DRI must be addressed e.g., in relation to security of data where local facilities had previously been preferred but shared DRI can now accommodate this need.

Procurement for net zero

Frustrations over procurement processes took several different angles. Firstly, there are issues relating to knowledge of the best options and a lack of evidence base / transparency of data for giving good advice about procurement decisions, together with the knowledge of how to calculate impacts and make meaningful comparisons between facilities. This extends to being able to consider wider impacts such as buying into a shared data storage facility compared with developing new facilities on-site.

In some cases, it has not been possible to carry out full procurement tendering processes to secure the most environmentally friendly option. Coupled with a lack of leadership or drive from senior management and no additional funding to secure facilities which would have longer term benefits, this can mean that net zero falls far down the priorities when securing new facilities.

The embodied emissions from manufacturing (as well as wider impacts of resource use such as the social and environmental impacts of rare earth metal mining) are a significant contributor to DRI impacts. Tendering processes should place pressure on suppliers to assess and publicly report on these (scope 3) emissions.

Reducing scope 3 emissions requires using equipment as efficiently as possible, for as long as makes sense environmentally. One barrier to this is the practice of replacing useable equipment as soon as it is out of warranty for support and security patches. Suppliers should be encouraged to provide longer warranties, and this made part of the tender assessment process. DRI procurement, and public procurement more generally, could exert pressure for change which would have wider positive benefits than reducing impacts of DRI alone.

Where useable equipment needs to be retired at end-of-warranty, consideration should be given to repurposing it for less mission-critical roles.

Where possible, purchasing of local DRI systems for project use should be discouraged. Barriers to shared use (such as real or perceived concerns about security) should be further explored and where possible removed. Finally, DRI should be properly considered as part of HEI infrastructure and developed as such, fully integrated into wider systems where waste heat and other secondary impacts can be best utilised. It thus becomes part of university estates planning processes.

Efficiency

Efficiency takes different forms, including making code more efficient, making data more available (e.g., FAIR principles), and avoiding duplication and unneeded re-computation – making each use of DRI count in the pursuit of knowledge. Each instance of ‘inefficiency’ leads to waste in DRI use and provision – in contradiction with net zero ambitions. There could be an increasingly significant role of RSEs in supporting academics to optimise code and store it efficiently and in a manner that promotes findability and reuse.

However, it is important to provide a cautionary note on the topic of ‘efficiency’. It is natural given our stakeholders and their research context, that efficiency and making the optimal use of facilities would emerge from our research. Intuitively, improving efficiency of computation, and of different forms of HPC can contribute to the decrease of waste and, thus, to the pursuit of net zero. However, increasing efficiency alone is not a solution to the increasing demand for HPC in academic research. As has been repeatedly demonstrated in other contexts, increasing efficiency can lead to rebound effects where overall demand increases (i.e., Jevons Paradox). For example, the average internal temperature of UK homes has increased by 4 degrees since the 1970s as home heating has become more efficient but also more prevalent and convenient. The pursuit of efficiency for net zero ambitions is thus likely to have unintended consequences: increased efficiency will reduce the completion time of compute intensive jobs, reducing overhead and job wait times, potentially leading to more compute and use of DRI overall. Focusing on efficiency alone detracts from the longer-term growing expectation for research to depend on more and more computationally intensive methods, thus, detracting from questions of ‘sufficiency’ (i.e., using just enough) (Shove, 2017; Santarius, 2022) in HPC use.

There is, therefore, a further ‘upstream’ intervention needed in recognition of sufficiency – how much computation is really needed for different academic domains and questions? How can funders and managers act on this? And how might journals and conferences tackle any increase in wasteful, unnecessary or duplicated research?

Beyond business as usual: Cultural change

Tackling the growing demand for DRI through measures that move beyond efficiency alone, requires a cultural change in terms of how DRI is used and provisioned in academic research. It was repeatedly suggested that there was little managerial or commissioning oversight or buy-in in terms of net zero ambitions, particularly as these relate to the DRI use.

The current and highly competitive system of academic research means that researchers are valued for publications, research funding and conference attendance, with little oversight as to the value and (environmental) harms of the work being conducted. Because of the nature of the system, and the systemic changes required, interventions need to go beyond individual researchers caring or being conscious of the environmental impacts of their work. Instead, there

is a need for a cultural shift, accountability, and pressure for the academic community to better consider the harms and values of their work.

To achieve this, there must be recognition at a senior level, and governance solutions in line with net zero ambitions. This will entail a shift in funding, procurement and policies to support this. Such a shift is not possible without senior policymakers at universities (and beyond) better understanding net zero and the paths towards it. Recognising the interplay between policy and practice is essential. For instance, we found University financial systems can force spending at specific times of the year (e.g., end of financial year) and can result in wasted spend and resources. Further work is needed to develop this understanding, and particularly how shifts in practice and policy might occur without detracting from important research, nor the value of experimentation and exploration that depends on DRI usage. However, it is essential for net zero that research that is compute intensive is not duplicated or wasteful. Better understanding this, and the changes required, will take time.

In the meantime, individual domains, funders, journals and conferences could begin to incentivise researchers to make net zero considerations in their work. Writing efficient code, or leveraging efficient submissions of jobs means that research can take more time. This should be incentivised, given the temporal pressures of academic work such as publication deadlines and short projects. ‘Best paper’ awards, environmental rationalisation of DRI usage in funding applications and careful considerations of whether research results are simply duplications of existing work represent some examples of how individual entities could begin to add pressure for the aforementioned cultural change. At the very least, these entities must consider if, or how, they are currently incentivising wasteful DRI usage.

While this clearly goes beyond a specific technology and process, a lack of sector-wide sharing protocols (e.g., FAIR) may be contributing to unnecessary duplication of code and data, the use of inefficient code or code badly matched to hardware features. Addressing sharing protocol issues whilst recognising the issues related to FAIR noted earlier in this report is another important area for progress.

We should also recognise that tackling issues such as inefficient code or inefficient usage of resources will require time investment from researchers, which will reduce time available for other work. This should be considered when defining policies and incentive practices for researchers.

5. Recommendations

Our recommendations are synthesised from the views and analysis above and address various levels. Level 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. Where possible, we have made reference to specific personas and causal loop diagrams. Our main recommendations are:

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 follow a consistent carbon calculation method and consider DRI's full lifecycle impact. *Personas 1, 3 and 4; Figure 6.*
- 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. *Persona 4; Figure 2.*
- 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). *Personas 2 and 4; Figures 1 and 2.*
- 4. Support flexible sharing of sector-wide DRI** for researchers to utilise available computational resources when required, avoiding underutilised DRI (e.g., due to funding constraints, or false assumptions about inability to use shared facilities) and avoid the expansion of new and unnecessary DRI elsewhere. *Persona 3; Figures 1 and 5.*
- 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. *Personas 2, 3 and 4; Figures 1 and 5.*

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. *Personas 2 and 4; Figures 1, 5 and 6.*
- 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. *Personas 1 and 2; Figure 2.*
- 8. Recognise DRI's role as part of the wider infrastructure** and embed in institutional policy and practice, ensuring valuable outputs (e.g., heat) are integrated into institutions' estates and beyond (i.e., local, regional, national) to maximise value and avoid waste. *Persona 3; Figure 6.*
- 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. *Persona 1; Figure 6.*

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. *Personas 1 and 4; Figures 1 to 4 and 6.*

References

- Arman, S. et al (2016) 'United States Data Center Energy Usage Report' <https://eta.lbl.gov/publications/united-states-data-center-energy>
- Braun, V. & Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- EPCC. (n/d) 'Introduction to High Performance Computing for Life Scientists: The HPC landscape in the EU and UK', EPCC.
- Fujitsu (Oct 2020) 'Fugaku and associated technologies help to solve major challenges in the world today', Interview between Satoshi Matsuoka and Naoki Shinjo <https://www.fujitsu.com/global/about/innovation/fugaku/pickup/interview01/>
- Haines, V. & Mitchell, V. (2014) 'A persona-based approach to domestic energy retrofit', *Building Research & Information*. 42(2).
- Haraldsson, H. (2004) 'Introduction to System Thinking and Causal Loop Diagrams.' *Reports in ecology and environmental engineering*, (1). KFS AB, Lund, Sweden.
- Hilty, L. M., Köhler, A., Von Schéele, F., Zah, R. & Ruddy, T. (2006) 'Rebound effects of progress in information technology'. *Poiesis & Praxis*, 4(1):19–38.
- IRISCAST project report (2023) <https://zenodo.org/record/6787626#.YtVQCzfMI2w%2F>
- Al-Shboul, M. K. & Abrizah, A. (2014) 'Information Needs: Developing Personas of Humanities Scholars', *The Journal of Academic Librarianship*, 40(5): 500-509.
- Kim, D. H. (1999) 'Introduction to systems thinking' (Vol. 16). Pegasus Communications Waltham, MA.
- Miaskiewicz, T. & Kozar, K. (2011) 'Personas and user-centered design: How can personas benefit product design processes', *Design Studies*, 32(5): 417-430.
- Nelson, K. (2019) 'Doing Fieldwork: Methods in Cultural Anthropology', *Perspectives: An Open Invitation to Cultural Anthropology*. American Anthropological Association.
- Nowell, L. S., Norris, J. M., White, D. E., Moules, N. J. (2017) Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods*, 16, 1.
- Palmer, J. & Cooper, I. (2013) *UK Housing Fact File*. DECC: London.
- Pereira, L., Quintal, F., Barreto, M. & Nunes, N. (2013) 'Understanding the Limitations of Eco-feedback: A One-Year Long-Term Study', *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured Big Data*. 237-255.
- Santarius, T., Bieser, J. C., Frick, V., Höjer, M., Gossen, M., Hilty, L. M., Kern, E., Pohl, J., Rohde, F. & Lange, S. (2022) 'Digital sufficiency: conceptual considerations for ICTs on a finite planet', *Annals of Telecommunications*, pp.1-19.
- Shove, E. (2017) 'What is wrong with energy efficiency?', *Buildings Research & Information*, 46(7).
- UITS. (n/d) 'Understand measures of supercomputer performance and storage system capacity', *UITS*. Available: <https://kb.iu.edu/d/apeq>
- Widdicks, K., Lucivero, F., Samuel, G., Croxatto, L. S., Smith, M. T., Holter, C. T., Berners-Lee, M., Blair, G. S., Jirotko, M., Knowles, B., Sorrell, S., Rivera, M. B., Cook, C., Coroamă, V. C., Foxon, T. J., Hardy, J., Hilty, L. M., Hinterholzer, S., and Penzenstadler, B. (2023) 'Systems thinking and efficiency under emissions constraints: Addressing rebound effects in digital innovation and policy'. *Patterns* 4, 2.

Appendix 1: Glossary

CLD Causal loop diagram

DRI digital research infrastructure, as a category, DRI refers to any specialised, research-oriented IT devices and facilities, spanning from laptops and workstations to high-performance computing and large-scale data archives

FAIR data and code that is ‘findable, accessible, interoperable, and reusable’

HPC High Performance Computing for processing large amounts of data, generally also stores data for users.

RSE Research Software Engineer

Tier 0 / 1 / 2 / 3 HPC facilities

Tier 0 international facility (e.g., CERN Large Hadron Collider for particle physics)

Tier 1 national supercomputer infrastructure (e.g., ARCHER2)

Tier 2 regional e.g., EPSRC-funded Tier-2 HPC services include i) Isambard – a supercomputing collaboration between the GW4 universities (University of Bath, Bristol, Cardiff and Exeter) and the Met Office. ii) JADE iii) Cirrus and others

Tier 3 institutional HPC – i.e., within universities

Appendix 2: Interview questions

INTRODUCTORY TEXT (COMMON TO ALL INTERVIEWEES):

Aim: This project is exploring how creating a net zero digital infrastructure needs an understanding not just of the technical factors, but wider academic, social and organisational factors drive demand for it and to map the broader landscape of, for example, the

- Interplay between policy and practice (e.g., funder or peer/reviewer expectations)
- Factors affecting growth for DRI (and barriers for shared infrastructures)
- Risks and opportunities for answering questions about net zero
- How DRI is shaped by socio-technical forces

Process of interview: check reading of info sheet, any questions before we start, check it's ok to record and transcribe.

All respondents: Please tell me about your role in relation to the provision, management and / or use of DRI facilities.

Questions for ACADEMIC DRI USERS:

In relation to your use of DRI facilities:

1. Who's involved, processes involved:
2. What do you use compute facilities for / in your research? Roughly how 'much computation' and 'how often' / what 'scale'?
3. Do you make use of shared computation or data storage in conducting your research (HPC, cloud)?
4. How much choice/ freedom over which DRI do you have?
5. Can you describe the role of data in your research, especially as this links to generation, processing and use (or possible reuse) of large datasets on an ongoing basis?
6. To what extent do you consider the environmental cost of your work on DRI?
 - a. And how do you justify / explain this?

Tell me about waste in using DRI, are there times when you have felt that (some or all of) your use of the DRI was a 'waste'?

- a. Why? Could you have known in advance?
 - i. e.g., A partial failure that could have been anticipated or halted earlier?
 - ii. Or a run that was unnecessary from your perspective, but you were required to do for some reason?
- b. Are there any software or support capabilities that could have helped avoid this?

Tell me about choosing / buying DRI

1. Have you commissioned dedicated compute resources for a project (e.g., your own servers)?
 - a. What and why?
 - b. Where does it live (e.g., remote, hosted, local)?
 - c. What factors influence your decisions?
 - i. Technical capabilities, availability
 - ii. Probe – environmental, cultural, personal
 - d. What policies have an impact?
 - i. Internal, organisational e.g., access to central HPC
 - ii. External: UKRI, govt e.g., restrictions on access to DRI
 - e. Are there restrictions on using central HPC? How does access work in your organisation?
 - f. What about decommissioning? To what extent do you have a say?
 - g. Where does **funding** for this come from?
2. Do you feel your need for computation or data storage is stable, increasing or decreasing over the years?
 - a. If so, what do you think causes this?
3. Tell me about any waste or inefficiency in choosing DRI, please describe where you see this arising. Probe: lifecycle phases (embodied, use, disposal).
4. What happens/happened to it at the end of the project?
 - a. how is your DRI decommissioned?

Tell me about the environmental impacts of your DRI

- a. Including e.g., lifecycle – embodied, use, disposal?
- b. To what extent do you consider this when using DRI?
- c. Do you have to consider the needs of any external bodies / stakeholders (e.g., waste disposal companies, funders?)
- d. Are there any other external pressures (e.g., conference sustainability statements) around environmental impacts of DRI?
- e. How could / do you consider 'net zero' in your use or procurement / purchase of DRI?
- f. How visible is energy / carbon in enabling you to make decisions about kit and usage?

Questions for

P – PROVIDERS / SUPPORT

C – COMMISSIONERS AND FUNDERS

M – INSTITUTIONAL MANAGEMENT / CULTURE

P: Tell me about the use of your facilities

1. Who uses it, who's involved – e.g., in approving users – how are decisions made? What restrictions are there on access?
2. Are your facilities used 'to the max', or are there times when parts of them are idle?
 - a. Do you have jobs queuing? Or complaints about waiting times?
 - b. How do you manage the queue? What are the criteria for prioritising?
 - c. Do you use any **energy or carbon** criteria in your job scheduling algorithms?
3. What sort of role if any do you have in managing shared facilities or cloud access?

OR

C: Tell me about the sorts of DRI that are requested by researchers, what gets asked for in different types of research projects?

Tell me about how the need for computation gets costed into research projects?

- a. Who are the stakeholders involved?
- b. What factors are considered in costings?
 - i. E.g., energy, space?
 - ii. Reuse of current equipment?
 - iii. Future re-use / upgrades
 - iv. carbon, environment, net zero?

OR

M: Could you tell me about how research culture might drive demand for compute facilities?

- a. Probe on both formal policies
- b. ...and informal culture.
- c. ...and any other factors? E.g., FAIR

P/C: Tell me about the drivers for decision making in new facilities:

1. What drives expansion of your compute facilities?
 - a. e.g., Capacity increases? Changing technology? Research priorities?
 - b. How long does your infrastructure meet demand before it needs an increase in capacity – how fast is demand increasing (if it is)?
 - c. Who are the funders of any expansion?

2. How do you commission new facilities?
 - a. How do you determine what capacity they should have?
 - b. Tell me about any pressures to consider environmental impacts in your commissioning? And net zero? (If they are pressures, how do you consider this)
 - c. Are there any restrictions in commissioning – e.g., organisation policy?
3. To what extent are you able to reuse or upgrade existing facilities to meet demand?
4. What drives retirement of existing facilities?
 - a. When do you consider a resource out of date?
 - b. How do you consider environmental impacts? E.g., waste disposal / recycling, whole life cost / embodied energy

P / C / M

5. How does **overprovision and/or waste of energy/resources occur?**
 - a. Probe on formal policies for data storage and retention
 - b. Probe on wider research cultural issues, social drivers, expectations
 - c. Implications for net zero?
6. **What policies or guidance shape the commissioning and use of facilities e.g., shared vs. individual equipment?**
 - a. Where are the policies from? e.g., funders, internal organisation,
 - b. (M) and how does this vary across disciplines?
 - c. To what extent is there encouragement to maximise the use or investment in shared facilities?
 - d. Are there pressures to consider environmental costs that you take into account?
 - i. How do you consider environmental sustainability?
 - ii. Any net zero policies?
 - iii. Energy, space
 - iv. Re-use of equipment
 - v. Future use / upgrades?
7. (P/C) **Could you tell me about how research culture might drive demand for compute facilities?**
 - d. Probe on both formal policies
 - e. ...and informal culture.
 - f. ...and any other factors? E.g., FAIR

All, final question:

Is there anything else relating to the conversation and topic of today's conversation that I haven't already asked you, but that you think we'd like to know?

