

# The environmental sustainability of data-driven health research: A scoping review

Digital Health  
Volume 8: 1–11  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076221111297  
journals.sagepub.com/home/dhj



Gabrielle Samuel<sup>1,2</sup> and A.M. Lucassen<sup>2,3</sup>

## Abstract

Data-Driven and Artificial Intelligence technologies are rapidly changing the way that health research is conducted, including offering new opportunities. This will inevitably have adverse environmental impacts. These include carbon dioxide emissions linked to the energy required to generate and process large amounts of data; the impact on the material environment (in the form of data centres); the unsustainable extraction of minerals for technological components; and e-waste (discarded electronic appliances) disposal. The growth of Data-Driven and Artificial Intelligence technologies means there is now a compelling need to consider these environmental impacts and develop means to mitigate them. Here, we offer a scoping review of how the environmental impacts of data storage and processing during Data-Driven and Artificial Intelligence health-related research are being discussed in the academic literature. Using the UK as a case study, we also offer a review of policies and initiatives that consider the environmental impacts of data storage and processing during Data-Driven and Artificial Intelligence health-related research in the UK. Our findings suggest little engagement with these issues to date. We discuss the implications of this and suggest ways that the Data-Driven and Artificial Intelligence health research sector needs to move to become more environmentally sustainable.

## Keywords

Environmental sustainability, environmental impacts, sustainability, digital technologies

Submission date: 21 February 2022; Revision date: 14 June 2022; Acceptance date: 15 June 2022

## Introduction

Data-Driven and Artificial Intelligence (DDAI) technologies are rapidly changing the way that health research is conducted.<sup>1</sup> New technological capabilities to store and process vast quantities of clinical data, and new collections of health data from non-traditional sources,<sup>2</sup> have led to new opportunities for large-scale data analytics. These in turn have led to the explosion of health data repositories, containing troves of clinical and genomic data,<sup>1</sup> as well as swaths of self-tracking data from wearables, biosensors and/or environmental data.<sup>3,4</sup> Meanwhile, social media and other data are being mined for health research,<sup>5</sup> and machine learning techniques are being used to help predict health conditions.<sup>6</sup> The storage and processing of health-related data are set to become the fastest growing sector in the datasphere.<sup>7</sup>

Expansion of DDAI technologies inevitably results in adverse environmental impacts. These include heavy carbon dioxide emissions linked to the energy required to

generate and process large amounts of data. Approximately 100 megatonnes of carbon dioxide emissions are produced from the digital sector per year,<sup>8</sup> and the yearly electricity usage of data centres is over 205 TetraWatts per hour,<sup>9</sup> which already exceeds the consumption of countries such as Ireland and Denmark (see, e.g.<sup>10</sup>) Furthermore, this consumption fails to account for indirect electricity usage (and likely carbon emissions) associated with data centre supply chains.<sup>11</sup> DDAI technologies also have adverse impacts on the material environment (e.g. *where* data centres are

<sup>1</sup>Department of Global Health and Social Medicine, King's College London, London, UK

<sup>2</sup>Wellcome Centre for Human Genetics, Oxford University, Oxford, UK

<sup>3</sup>Clinical ethics, law and society (CELS) Faculty of Medicine, University of Southampton

### Corresponding author:

Gabrielle Samuel, Department of Global Health and Social Medicine, King's College London, Bush House, London WC2B 4BG, UK.

Email: gabbysamuel@gmail.com



constructed); the unsustainable extraction of minerals for technological components;<sup>8,9</sup> and e-waste (discarded electronic appliances) disposal.<sup>10</sup> Rautela and colleagues (2021) note that while rates of e-waste production vary per continent and per capita, as do rates of recycling, in 2019, 53.6 million metric tonnes of e-waste were generated globally (for more information and breakdown, see<sup>12</sup>). Only about one-fifth of e-waste is formally collected and recycled; the fate of the remainder is uncertain, though it is likely disposed in dumps and landfills with other waste, or traded through illegal markets.<sup>13</sup> These are important concerns: carbon emissions can lead to drastic health impacts<sup>14</sup> and unsustainable mineral extraction and e-waste practices can have detrimental health consequences – especially affecting individuals in low-and-middle-income countries where mining and waste storage predominantly occurs.

While likely improvements in energy efficiency and the move to renewable energy will no doubt relieve at least some of these concerns,<sup>15</sup> the pace of data-driven innovation raises concerns that information and communication technologies could outpace the world's renewable energy sources, leading to increases in carbon emissions when other sectors are decreasing their energy use.<sup>11,16</sup> Furthermore, data-driven solutions have rebound effects, meaning that while digital solutions in the near term may appear to offer environmental advantages, in the long run, this may not be the case. For example, the move to centralise health research data, biobanking data, and/or the move to open science will not *necessarily* lead to less overall reduction in health research data being collected and/or a reduced impact on the environment.

There is a compelling need to consider the environmental impacts of DDAI technologies.<sup>17</sup> This includes, for example, who is, and who should be responsible for these impacts and how these responsibilities should be enacted and distributed. Given that concern about environmental sustainability in the health research sector is relatively recent,<sup>2</sup> this paper provides a scoping review to explore how the academic literature is engaging with these issues. Our research question was: what discussions pertaining to the environmental impacts of data storage and processing during DDAI health-related research have been discussed in the academic literature? Using the UK as a case study, we also asked: what policies or initiatives consider the environmental impacts of data storage and processing during DDAI health-related research? Our findings suggest very little engagement with these issues to date.

## Methods

### Literature searches

In October 2021, Web of Science, PubMed and Google Scholar were searched for relevant articles using the keywords in Table 1. Inclusion criteria included documents

(book chapters, preprints, conference proceedings, articles, etc.) that focussed on the negative environmental impacts of data storage and processing during DDAI health-related research and care. Health research and care are becoming increasingly interlinked through learning health care systems (see, e.g. the UK's '100k Genomics Project', and 'Our Future Health'), and widening the search to health care ensured all pertinent articles were included. Relevant exclusion criteria were applied (Table 3). Reference lists of included articles were checked for further relevant documents/authors via snowballing. Following removal of duplicates, 25 documents remained. Six documents – all identified during snowballing – could not be accessed (either because they could not be found in web searches or because they required a fee to access), and so were excluded. The 19 remaining documents were deductively coded into the following categories: title, date published, type of publication (book chapter, journal, etc.), place of publication (journal name, etc.), first author name and country of institution of the first author. The documents were read thoroughly, and additional inductive codes were added, including the type of environmental impact addressed/mentioned; whether the document was framed in terms of green information technology (IT); use of the term 'sustainability' as an overarching concept for discussion; and mention of unaddressed challenges and perceived solutions. Documents were also qualitatively reviewed (read in depth) for relevant concepts.

### Web searches

In October 2021, Google was searched using the keywords in Table 2 to identify relevant initiatives and/or policies pertaining to the research question: what policies or initiatives consider the environmental impacts of data storage and processing during DDAI health-related research specifically in the UK? Inclusion criteria included UK web pages associated either with sustainability and the health sector (as above, health care was also included to ensure all relevant initiatives were identified); with research laboratory sustainability more generally; or with sustainability, health sector and DDAI technologies specifically. For each keyword search, all returned pages were checked for at least the first five pages. If at least one relevant link was identified on page 5, the search was continued until two consecutive pages returned no relevant links (the maximum number of pages reached was 10 pages). Exclusion criteria included published journal articles or links to scholars whose work had been identified in the literature review, and multiple links from the same organisations. For each retrieved link ( $n = 104$ ), information on the weblink's institutional origins, and a short description of the weblink (purpose of the web page; blog, pdf document, policy statement and other information) were collected. All retrieved links were checked for content pertaining to

**Table 1.** Search strategy for literature review, including database searched, keystings used and number of relevant articles retrieved.

Key-string used	Type of article searched	Site searched	Articles searched	Articles deemed relevant
(bioinformatics OR e-health OR m-health OR "health research" OR "health care" OR healthcare OR genom* OR neuroimaging OR radiology OR "medical imaging" OR "electronic health records" OR "health data*" OR "clinical data") and (digital OR AI OR "big data" OR "big-data" OR "app" OR "tech*" OR "artificial intelligence" OR "machine learning" OR "ICT") and (sustainab* OR "environment* impact*" OR "environmental* sustainab*" OR "climate change" OR "carbon emissions" OR "e-waste" OR "green").	Abstract	Web of Science	5172	Following the checking of title, and abstract if needed, 149 which had some relevance. Checking the full article in more detail = 7 of relevance.
-laborator* OR "lab " OR "labs " (title) AND "environment* impact" (abstract) -("lab " OR "laborator*")(title) AND green AND "climate change" (abstract) "sustainab* lab*" (title), -lab " OR "laborator*" (title) OR environment* OR "sustainab*"	Titles or Abstracts	Web of Science	436	0 articles
(e-health OR m-health OR "health research" OR "health care" OR healthcare OR genom* OR neuroimaging OR radiology OR "medical imaging" OR "electronic health record*" OR "health data*" OR "clinical data") AND (sustainab* OR "environmental impact*" OR "climate change" OR "pollution" OR "carbon emissions" OR "greenhouse" OR "waste") AND (digital OR AI OR "big data" OR "big-data" OR "artificial intelligence")	Titles and Abstracts	PubMed	523	Following the checking of the title, and the abstract/full article if needed, 1 additional article was deemed relevant.
((("lab "[Title/Abstract] OR "laboratory"[Title/Abstract]) AND ("digital"[Title/Abstract] OR "data"[Title/Abstract])) AND ("sustainab*"[Title])	Titles and Abstracts	PubMed	48	0 additional articles
A whole range of keywords with various permutations, for example ("digital health" sustainability "environmental impact") and ("health database" "environmental impact") and (e-waste health)	Titles, then abstracts if relevant	Google Scholar	For each keystring, 5 pages of google scholar were checked (in all cases by the fifth page, articles were no longer deemed relevant)	5 additional articles
-sustainab* lab	Titles, then	PubMed	For each keystring, 5 pages of google	0 additional articles

(continued)

Table 1. Continued.

Key-string used	Type of article searched	Site searched	Articles searched	Articles deemed relevant
-“science lab” “environmental impact” -Research laboratory* environmental impact -Research lab environmental impact -Science lab and environmental impact	abstracts if relevant	Google Scholar	scholar were checked (in all cases by the fifth page, articles were no longer deemed relevant)	
Snowballing				12

Table 2. List of key strings used to search Google search engine

Key strings for Google search
“sustainable lab”
“environmental impact” sustainability health
“environmental impact” sustainability “health research”
“environmental impact” “digital health”
“environmental impact” “digital tech” health
“environmental impact” health “big data”
“health data” sustainability “environmental impact”
“environmental impact” sustainability “health app”
“environmental impact” sustainability health software
sustainability health digital
sustainability “environmental impact” “big data” research
sustainability “environmental impact” AI research
sustainability “environmental impact” digital research
sustainability “environmental impact” “data-driven” research
genomics sustainability “environmental impact”
imaging sustainability “environmental impact”
biobank sustainability “environmental impact”
“electronic medical records” sustainability “environmental impact”
radiology sustainability “environmental impact”
“biometric data” sustainability “environmental impact”
“clinical data” sustainability “environmental impact”
“health data” sustainability “environmental impact”
“genome sequencing” sustainability “environmental impact”
“environmental impact” sustainability health “working group”
“environmental impact” sustainability health “interest group”
“environmentally sustainable” health
environmentally sustainable health society
green “digital health” climate
green digital health climate
green data health climate

**Table 3.** Exclusion criteria for literature searches.

Documents focusing on health and sustainability, or lab sustainability, but not DDAI
Documents that reported appropriate energy sources for wearable biosensors
Documents that related to the positive impacts of data-driven/digital health research or technologies
Documents that explored the environmental impacts of neuroimaging but focused on non-data aspects
Documents that explored power saving in laboratories, where computers were just one component of measures.

DDAI, Data-Driven Artificial Intelligence.

the unintended environmental impacts of data storage and processing for DDAI-associated health research or care.

### Limitations

While literature and Google searches were kept broad to ensure all relevant articles and UK initiatives/policies were identified, key articles and policies may have been missed. This is because practices pertaining to reducing the environmental impacts of data storage and processing may not be written into policies or initiatives – for example, if they are being considered tacitly and from the ground up. Furthermore, DDAI technology-associated environmental impacts may be being considered as part of broader initiatives that were not detected in our searches. For example, a recent UK Medicines and Healthcare products Regulatory Agency (MHRA) consultation document comprised a section on the environmental sustainability of medical devices that was not detected in our searches. Arguably this was not specifically related to DDAI technologies, but the section highlighted how the health technology sector is beginning to engage with these issues. At the same time as noting these limitations, we emphasise that the aim of a scoping review is to offer insight and summary of an emerging body of scholarship and identify gaps, and to this end, the methodology used is appropriate.

## Findings

### Literature analysis

All analysed documents (n = 19) were published between 2010 and 2021 and included peer-reviewed articles and pre-prints (n = 7), commentary pieces (n = 3), conference presentations/proceedings (n = 6) and book chapters (n = 3). All conference presentations/proceedings and four articles/commentaries were written for an IT sector audience

(n = 10 out of 13, as defined by the journal article or discipline of conference). First, the authors were affiliated with institutions in India (n = 5), the UK (n = 2), Spain (n = 2), South Africa (n = 2) and the United States (n = 2), as well as with institutions in Canada, France, Pakistan, Botswana, Greece and West Africa (n = 1 each). Half of the documents (n = 9) were written by three research groups. Nearly all 18 documents (n = 15) placed their focus on addressing or raising awareness about the environmental impacts of e-health, healthcare and/or hospitals rather than health-related research (n = 3) or health apps (n = 2).<sup>3</sup>

**Types of environmental impact mentioned.** All documents highlighted the importance of reducing energy use and/or carbon emissions to address the environmental impact of digital technologies in the health sector.<sup>4</sup> Nearly all documents (n = 14) also stressed the need to attend to the environmental impacts associated with the technology's component materials in terms of used mineral resources and e-waste. Scott and colleagues (2012) categorised these environmental impacts collectively into 'upstream impacts' (extraction, processing, or synthesis of raw materials, the manufacture of components and the packaging and distribution of these components), 'mid-stream impacts' (design, implementation and use) and 'downstream impacts' ('end-of-life' aspects of disposal or recycling).<sup>18</sup>

**Promoting technological solutions.** Most documents were technical in nature, meaning that documents focussed on developing *software and hardware solutions* to decrease the environmental impacts of DDAI technologies. Many of the authors developing such solutions defined their approach under the umbrella of green IT (n = 11), that is, an approach to IT that produces minimal waste during its development and operation and promotes recyclability. In fact, green IT approaches were considered by many of the authors as a key approach to addressing the environmental impacts of digital technologies in the health sector.<sup>19</sup> For example, the authors described their use of green IT to develop efficiency-increasing software, as well as to improve the energy efficiency of data centres serving the health sector. They reported their research on designing the digital infrastructure of smart hospitals<sup>20</sup> and medical Internet of Things (IoT) systems.<sup>21</sup> Novel cloud approaches that were more energy efficient were also recommended as solutions.<sup>22</sup> Saiyeda and Hamdard (2020) provided a roadmap for green IT healthcare that included not only aspects of design and manufacturing, but also purchase, use and disposal of IT technologies.<sup>23</sup>

**Beyond technical solutions.** Few articles considered the *amount and/or type of data collection and processing* as factors associated with the unintended environmental



impacts of DDAI technologies in the health sector. One study examined the American College of Radiology (ACR) Appropriateness Criteria for the type of imaging recommended for a specific clinical condition. The authors determined whether imaging modalities used in the US healthcare radiology departments could be switched to those more energy efficient without affecting patient care.<sup>24</sup> Another separate article raised questions about whether carbon emissions should be a factor in determining the most appropriate video image resolution used during virtual clinical appointments in the UK National Health Service (NHS) because more data is required for higher resolution images leading to a higher environmental.<sup>25</sup> This same author also described the environmental impacts associated with the exponential increase in data collection and processing in the UK NHS and called for more differentiation between useful and redundant data when considering which data to store: *'all this data uses energy and infrastructure to store and access. There is a need to start asking the question of whether we need to store all data and how long the data needs to be kept in healthcare settings'*.<sup>25</sup> Writing in the Lancet, Chevance and colleagues (2020) called for 'digital temperance' rather than 'overconsumption and overpromotion' of data. These authors described three guiding principles they believed researchers and clinicians should incorporate into their data-relevant practices: restraint in production, use and promotion of digital technologies; lifecycles instead of waste (c.f. the circular economy); and complex systems approaches through interdisciplinary collaboration.<sup>26</sup>

In fact, *collective discussion*, action and responsibility were emphasised by several authors as important for meeting the challenges of environmental impacts. These authors called for health sector workers to collaborate with those in IT, as well as other sectors and industries<sup>23,25,26</sup> so that together they could properly assess both the state of data use in the health sector, as well as its short-term and long-term direct and indirect impacts. This recommendation has already been followed up, with a group of UK scholars assessing the carbon footprint of bioinformatics in one recent 2021 article.<sup>10</sup> They found that biobank-scale analyses emitted substantial carbon emissions. These emissions could be reduced by, for example, software upgrades, faster/more efficient computer processing and appropriate data centre choices. Chevance and colleagues (2020) were particularly concerned with rebound effects, that is, where improvements in the technological efficiency of energy use led to greater direct or indirect energy consumption. A need for greater awareness of the environmental impacts of digital technologies in the health sector was stressed.<sup>23,26,27</sup> Consultation and training on green IT practices (discussed above) were suggested as ways to increase awareness,<sup>27</sup> as was the creation of standards for assessing and auditing environmental impacts.<sup>19,23,25,26</sup>

*Sustainability as a normative framework for action.* Half of the authors (n = 10) explicitly used the concept of *sustainability* to frame their research. Sustainability was articulated as a valued normative principle and defined as the need to embed economic, social and environmental considerations into the design and use of digital technologies (also referred to in its similar guises of 'sustainable development' and 'the triple bottom line'<sup>5</sup>). Scott and colleagues (2012), for example, developed a responsible response hierarchy for addressing e-waste in e-health<sup>6,18</sup>. They argued that respect for the environment is necessary and that *'every person or business whose action might impact the environment has an obligation to...act in a manner that maintains a balance between the economy and the ecosystem, and that benefits society at large'*. In some documents, sustainability was defined as being synonymous with green IT (*'green software often refers to the environmental dimension of software sustainability...In this paper, green software means sustainable software'*).<sup>28</sup> Several challenges were identified with the implementation of sustainable systems, including cost,<sup>22,23,29,30</sup> and the need to balance sustainability against other normative values such as privacy, for example, when the local data storage is considered as more privacy enhancing compared to more energy-efficient cloud-based systems.<sup>29</sup>

### *Policies and initiatives on the environmental sustainability of DDAI health research or care*

Alongside the academic literature scoping review, web searches were conducted to identify relevant policies or initiatives associated with the unintended environmental impacts of DDAI technologies in the UK health research or care sector. Retrieved web pages focusing on laboratory research sustainability (n = 34) or health sector sustainability initiatives more generally (n = 70) were checked for such policies or initiatives. In the former, many institutions and groups have developed resources to assist researchers' efforts to maintain environmentally sustainable laboratory practices. For example, decreasing the environmental impacts of fume hoods and freezers, reducing water consumption, as well as reducing emissions and waste practices (refuse, recycle, repurpose, reuse and reduce). In these resources, the energy consumption of computers is mentioned, though this is limited to statements concerning the need to turn off digital devices regularly. We were unable to find statements addressing the environmental concerns specifically associated with the unintended environmental consequences associated with DDAI technologies specifically.

In health sector sustainability initiatives, high-level sustainability principles were often presented, with little information about specific practices, policies or initiatives. Because of this, it was sometimes difficult to ascertain how much DDAI-associated environmental impacts were

included in these strategies, or whether specific initiatives or policies existed. Several documents found on the searched web pages considered issues associated with the adverse environmental impacts of digital technologies more generally and focused on better assessments of energy use, as well as various practices that could decrease energy consumption. Similar to the sustainable research setting described above, this usually amounted to calls to switch off computers when not in use, or a better estimation of the energy expenditure from digital (and other) technologies. Very few considered the *amount* of data being collected, though, for example, the UK Great Ormond Street Hospital Sustainable Development Management Plan did call for the need to ‘*make visible the emissions for key high-carbon GOSH activities...e.g. data heavy use...*’.<sup>7</sup> The UK National Institute of Health Research (NIHR) also made reference to avoiding ‘*unnecessary data collection*’ in its carbon reduction guidelines.<sup>8</sup> NHS Digital, which has a specific digital sustainability programme,<sup>9</sup> discussed the need for *technological* solutions such as low-carbon platforms, consolidating and decommissioning ‘legacy’ data centres and commissioning more efficient data storage. Finally, a Sustainability and Environment Action Special Interest Group formed by the Organization for Human Brain Mapping (OHBM) in 2020 described their engagement with assessing the environmental impacts of neuroimaging DDAI research, and communicating this more broadly to the sector to drive changes in practices. This latter group was the only one we could identify in the UK considering DDAI-specific issues in health-related research.<sup>10</sup>

## Discussion

While the adverse environmental impacts of DDAI-associated technologies are well established, our analysis shows that to date there has been relatively little engagement with them in the academic literature specifically focusing on health, and even less in health research. This might be because most efforts to address the environmental sustainability of DDAI technologies are seen to be an issue for the computing sector rather than as a concern for those who use computational and IT technologies.<sup>31</sup> It might also be because of current difficulties with calculating the exact environmental impacts of DDAI technologies. For example, there is disagreement and uncertainty in the digital sector about specific carbon emissions associated with different digital technologies, and many of the assessments include varying criteria (e.g. some include embodied carbon or indirect effects, most do not<sup>11</sup>).<sup>32,33</sup> Furthermore, private corporations are often unwilling to release data necessary for these assessments, requiring the need to develop – at best – rough estimations of such environmental effects. Finally, specific upstream impacts associated with embodied carbon, biodiversity, mining and manufacturing

are particularly problematic to calculate because the health research sector only accounts for a small fraction of global use of digital products. Therefore, while steps are being made to assess various environmental impacts of health-related computational research,<sup>1012</sup> much uncertainty remains about how to do so effectively.

At the same time, although environmental impacts are difficult to measure or compare, inaction is not a solution. We need to ‘*shift from treating uncertainty as a temporary issue, as something that more research and more funds can fix, into something that is unavoidably part of science and of policy making*’.<sup>34</sup> We are left with the question of *how* to consider such impacts. Our findings show that most of the work associated with thinking about the environmental impacts of DDAI technologies in the health sector revolves around providing technological solutions and e-waste management systems through green IT. These are important steps towards addressing many of the issues associated with the unintended environmental impacts of these technologies. At the same time, technical fixes cast situations ‘*as neatly defined problems with definite, computable solutions*’,<sup>35</sup> masking the complexity of social problems. Such complexity necessitates multidisciplinary approaches that draw on technical fixes, but also on a range of associated social, economic, political and cultural factors. As such – and as some of the authors in our analysis emphasised<sup>23,25,26</sup> – if we want to consider the environmental impacts of DDAI technologies in health research, we need to look beyond technical fixes to problematise our social practices.<sup>36,37</sup> The drive to collect and analyse ever-increasing amounts of data is one example.<sup>1113</sup>

Problematizing ever-increasing data collection and analysis practices requires being attentive to not only the associated potential benefits of these practices, but also reflecting on the economic and political drivers of this ‘datafication’<sup>38</sup> culture (what are the assumptions that drive us to collect and analyse this data, and where do these assumptions come from), as well as any potential harms that may come from such a culture (including addressing the commonly held misperception that digital technologies do not have an unintended environmental impact). Underlying the ‘more data is better’ culture is the underlying assumption that individuals can be completely knowable if enough data is collected and analysed.<sup>39</sup> In the health research field, this is seen to lie with more accurate and detailed information that can help improve the health of patients and/or individuals (c.f..<sup>4,35,40</sup>). At the same time, many of these assumptions are tied to economic and political interests: data collection is driven by a cycle of capital accumulation that comes from the expected economic value attributed to the data.<sup>39</sup> One notable example is genomics, in which the value of health-related data goes beyond health benefits in that genomic technologies attract investments as a way to stimulate the economy towards economic growth.<sup>4114</sup> Coupling genomics and

economic growth, while providing much-needed investment in the field, is problematic because it portrays health conditions in technological solutionist frames, amplifies genetic determinism discourses in inappropriate ways, promotes a datafication discourse (if only we had more data our health issues could be addressed; as such store now and if we can think of a use, analyse later), and fuels sociotechnical imaginaries that minimise the complexities of genomic data interpretation.<sup>38,42–44</sup> While genomics research is vital for a range of specific genetic conditions, coupling genomics with datafication cultures and economic growth also promotes the continued and ever-increasing use of this technology in potentially environmentally unsustainable ways. Datafication in genomics (and health more broadly) also creates a path dependency for digital infrastructures, as well as the epistemic authority that comes from the elevation of certain logics, techniques and imaginaries associated with digital technologies. A sunk cost mentality, that is, the tendency to continue a process after investment and effort, then sees this infrastructure repeatedly built upon.

This does not mean that we should cease collecting and analysing data for health (including genomic) research – especially when analysis could lead to obvious health benefits. Rather, a nuanced approach is required. There is a need to acknowledge that while many health researchers are trying to help the diagnosis, prognosis and/or care of patients and/or individuals with health-related illnesses, their research is embedded within a sociopolitical climate that unquestioningly views datafication as intrinsically good, with little reflection on what this assumption means in practice. Reflecting on this requires those conducting health research to carefully consider the rationale for data collected (who are we helping, but also who (individuals) and what (the environment) are we *not* helping) and refrain from collecting data in such a way that only promotes health and/or societal value for a minority of potential research recipients, or which has no foreseeable value. There is a need to also ensure research *plans* contain pathways and *infrastructures* that allow research impacts to reach wider communities, societies and/or more diverse groups rather than viewing such aspects as a long-term imaginary. This will ensure that while some environmental impacts are unavoidable, research benefits will be as wide ranging and inclusive as possible. To do so, health researchers must consider that many will not be able to afford their potential research outputs, such as drugs or health technologies.<sup>45</sup> Furthermore, many will not be able to access health benefits because they do not live within the society within which the research is being conducted. These are likely to be those most likely to be harmed by the environmental impacts of DDAI technologies.<sup>15</sup> Researchers must not be given a ‘free pass’<sup>31</sup> to defer responsibility away from these issues.

While as stated earlier, it is difficult to understand which choices will have the largest impact on the environment,

there are steps that can be taken. At an institutional level, green IT can drive improvements in energy consumption at data centres, e-waste recycling and computer product design and labelling.<sup>46</sup> At an individual level, several hyperscalers, including Google and AWS (Amazon), now provide carbon emission data to track carbon footprints.<sup>17,18</sup> Other behaviours, such as turning off computers when not in use, can also have an impact on energy use, and various guidelines have been developed to help individuals and institutions align themselves with eco-friendly computing. For example, a University of Cambridge factsheet<sup>19</sup> notes that leaving a computer on overnight for a year creates enough carbon dioxide to fill a double-decker bus, and while such facts are likely based on contested assumptions and predictions (ref me), they can still be useful for building awareness and driving change. Other helpful guidelines include <https://www.greenimpact.org.uk/GIforHealth> and LEAF, which is a standard for sustainable laboratory operations.

Lannelongue and colleagues (2021) have proposed a series of 10 rules for health researchers to make computing more environmentally sustainable.<sup>47</sup> These include calculating the carbon footprint of their research and including this in a cost–benefit analysis; choosing a computing facility and any associated hardware carefully; increasing the efficiency of the code used to analyse the data; and being a frugal analyst during this process (pilot energy-hungry algorithms on smaller data sets first). They also remind researchers to be aware of rebound effects, that is, such as that increases in IT efficiency will lead to increases in demand for data storage and analyses, not a reduction.<sup>48,49</sup> Taking these rules on board takes seriously the concept of ‘digital temperance’<sup>26</sup> – it forces us to slow down and think carefully about the specific data we need to collect for our research (rather than collecting as much as possible), and give due consideration to how this data is stored and analysed (c.f. an ‘ethic of care’.<sup>50,51</sup>)

Finally, due consideration should be given to more environmentally sustainable ways of achieving health benefits. This is, of course, not always possible, as we have witnessed during the COVID-19 pandemic, during which decision-making about the most appropriate health interventions has benefited from the collection and analysis of large swaths of data. At the same time, we already know that the biggest improvements to health can be made by addressing its social, economic and political determinants – improving people’s lives and well-being (education, economic livelihoods, work-associated stress, etc.), re-balancing economic and social inequality, and improving the quality of our living environments (air and water pollution).

Concluding, health-related DDAI research has tremendous potential to bring health benefits to many, but DDAI technologies also have adverse environmental impacts that must now be addressed.<sup>52</sup> The health sector has



engaged little with the issues that emerge from these impacts. As more and more data are collected for health research – and as more and more societal issues fall under the umbrella of being a health concern and therefore as having an intrinsic value that requires attention ('healthitisation'<sup>16</sup> c.f. medicalisation), it is crucial that we explore how to negotiate these challenges. Making explicit decisions about how we do so is the first step to developing shared understandings across the entire research ecosystem about the need to consider these issues within our own research practices. Such work can become a part of the wider sustainability agendas of the UK health sector and/or research institutions.

**Acknowledgements:** There are no acknowledgements for this article.

**Contributorship:** GS conducted the review of the literature, web searches and wrote the first draft of the manuscript. AL reviewed and edited several follow-up drafts and approved the final version of the manuscript.

**Declaration of conflicting interests:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval:** Not applicable, because this article does not contain any studies with human or animal subjects.

**Funding:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Wellcome (grant number 222180/Z/20/Z).

**Guarantor:** Dr Gabrielle Samuel.

**Informed consent:** Not applicable, because this article does not contain any studies with human or animal subjects.

**ORCID iD:** Gabrielle Samuel  <https://orcid.org/0000-0001-8111-2730>

**Trial registration:** Not applicable, because this article does not contain any clinical trials.

## Notes

1. For example, UK Biobank, UK national disease registration service, UK health research data alliance; with similar initiatives all over the world.

2. The environmental impacts of DDAI technologies have been known for decades, but have received little attention in the health/research sector until recently (exceptions include, e.g.). Childs S. Editorial: Green Computing. *He@lth Information on the Internet* 2008; 62: 1–2. This is slowly changing, with the need

to consider digital impacts comprising aspects of sustainability reports in health. For example, see Health Care without Harm. *Global Road Map for Health Care Decarbonization, Annex C: interventions*. 2021. Health Care without Harm. Furthermore, environmental sustainability has become an increasing concern more generally in health-related academic literature, with special issues being dedicated to this topic. See, for example, [https://journals.sagepub.com/topic/collections-hsr/hsr-1-environ\\_sust/hsrb](https://journals.sagepub.com/topic/collections-hsr/hsr-1-environ_sust/hsrb).

3. One document covered both research and care, which is why the total number of documents is greater than 18.

4. Including issues associated with the excess heat produced during data storage and processing.

5. The 'triple bottom line' is a sustainability framework that requires businesses to measure their social and environmental impact alongside their financial profit.

6. Modified from Ahluwalia, Poonam Khanijo & Nema, Arvind K., 2007. "A life cycle based multi-objective optimization model for the management of computer waste" Resources, Conservation & Recycling, Elsevier, vol. 51(4), pp. 792–826.

7. [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fmedia.gosh.nhs.uk%2Fdocuments%2FFOIRQ5888\\_SDMP\\_2020\\_short.pdf&clen=1379688&chunk=true](chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fmedia.gosh.nhs.uk%2Fdocuments%2FFOIRQ5888_SDMP_2020_short.pdf&clen=1379688&chunk=true).

8. <https://www.nihr.ac.uk/documents/the-nihr-carbon-reduction-guidelines/21685>.

9. <https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/sustainability-reports/sustainability-annual-report-2020–2021/content>.

10. <https://ohbm-environment.org/>.

11. Embodied carbon is the carbon associated with the production of a technology rather than the carbon associated with the technology use. Indirect impacts are non-direct sources of impacts associated with the technology, for example, the environmental impacts associated with e-waste.

12. Also see <https://ohbm-environment.org/>.

13. This is because Moore's law – which has described an ever-increasing IT system efficiency that has allowed less resources (and emissions) over time for equal efficiency gains – is now slowing down as IT infrastructure reaches its minimal size limits. Rotman D. We're not prepared for the end of Moore's Law. <https://www.technologyreview.com/2020/02/24/905789/were-not-prepared-for-the-end-of-moores-law/> (2020). It is also because technological solutions fail to consider Jevrons' paradox – that when efficiency is improved it leads to *more* consumption rather than *less* consumption. Alcott B. Jevons' paradox. *Ecological Economics* 2005; 54: 9–21. <https://doi.org/10.1016/j.ecolecon.2005.03.020>.

14. Also see [https://www.gov.uk/government/publications/genome-uk-2021-to-2022-implementation-plan/genome-uk-2021-to-2022-implementation-plan#:~:text=NHS%20England%20and%20NHS%20Improvement,to%20specific%20groups%20of%20patients](https://www.gov.uk/government/publications/genome-uk-2021-to-2022-implementation-plan/genome-uk-2021-to-2022-implementation-plan#:~:text=NHS%20England%20and%20NHS%20Improvement,to%20specific%20groups%20of%20patients and in particular Lord Bethell of Romford's statement, 'I truly believe that the UK is in a fantastic position to take advantage of our heritage in genomic research and innovation to transform healthcare in this country, to benefit patients and drive our economic recovery') and in particular Lord Bethell of Romford's statement, 'I truly believe that the UK is in a fantastic position to take advantage of our heritage in genomic research and innovation to transform healthcare in this country, to benefit patients and drive our economic recovery'. This relationship between health and the economy is also evident in other areas, including, for example, stem cell research. See Gardner J, Webster A and Mittra J. The "Entrepreneurial State" and the Leveraging of Life in the Field of Regenerative Medicine. In: V. P and J. G (eds) *Bioeconomies*. Palgrave Macmillan, Cham., 2017.

15. . Finally, this includes minority groups whose data may be removed from analysis to ensure algorithms remain optimised for specific functionality.

16. Using the term ‘healthitisation’ we mean that increasingly, ever greater forms of data generated in our daily lives (e.g. how we type, how we look at digital screens, how we communicate through digital devices, how we move around social places and in and out of social relationships, etc.) are being framed using health discourses, and as having predictive health value.

17. See <https://cloud.google.com/blog/topics/sustainability/reduce-your-cloud-carbon-footprint-with-active-assist> and <https://aws.amazon.com/blogs/aws/new-customer-carbon-footprint-tool/>

18. While cloud computing is often the most environmentally sustainable option, this needs to be balanced against other issues, such as public-private relationships and social justice issues (in terms of hyperscalers approaches to addressing social and environmental concerns).

19. <https://www.environment.admin.cam.ac.uk/facts-figures>.

## References

- Subrahmanya SVG, Shetty DK, Patil V, et al. The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish Journal of Medical Science (1971)* 2021. DOI: 10.1007/s11845-021-02730-z
- Vayena E, Dzenowagis J, Brownstein JS, et al. Policy implications of big data in the health sector. *Bull World Health Organ* 2018; 96: 66–68.
- Birk R and Samuel GN. Can digital data diagnose mental health problems? A sociological exploration of ‘digital phenotyping’. *Sociology of Health and Illness* 2020; 42: 1873–1887. DOI: 10.1111/1467-9566.13175
- Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. *Crit Public Health* 2013; 23: 393–403.
- Sharpe JD, Hopkins RS, Cook RL, et al. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. *JMIR Public Health Surveill* 2016; 2: e161–e161.
- Insel TR. Digital phenotyping: technology for a new science of behavior. *Jama* 2017; 318: 1215–1216.
- Reinsel D, Gantz J and Rydning J. *The Digitisation of the World*. 2018. Report from IDC and Seagate. <https://resources.moredirect.com/white-papers/idc-report-the-digitization-of-the-world-from-edge-to-core> (accessed February 2022).
- Lannelongue L, Grealey J and Inouye M. Green algorithms: quantifying the carbon footprint of computation. *Advanced Science* 2021; 8: 2100707. DOI: 10.1002/advs.202100707
- Koot M and Wijnhoven F. Usage impact on data center electricity needs: a system dynamic forecasting model. *Appl Energy* 2021; 291: 116798.
- Grealey J, Lannelongue L, Saw W-Y, et al. The carbon footprint of bioinformatics. *bioRxiv* March 2022; 39(3), msac034. DOI: 10.1093/molbev/msac034.
- Freitag C, Berners-Lee M, Widdicks K, et al. The real climate and transformative impact of ICT: a critique of estimates, trends, and regulations. *Patterns* 2021; 2: 100340
- Rautela R, Arya S, Vishwakarma S, et al. E-waste management and its effects on the environment and human health. *Sci Total Environ* 2021; 773: 145623
- Forti V, Baldé CP, Kuehr R, et al. *The global E-waste monitor 2020: quantities, flows and the circular economy potential*. Bonn/Geneva/Rotterdam: United Nations University (UNU)/United Nations Institute for Training and Research (UNITAR) – co-hosted SCYCLE Programme, International Telecommunication Union (ITU) & International Solid Waste Association (ISWA), 2020.
- European Environmental Agency. Safeguarding people from environmental risks to health, <https://www.eea.europa.eu/soer/2015/synthesis/report/5-riskstohealth> (2020).
- Giles M. Is AI the next big climate-change threat? We haven’t a clue, <https://www.technologyreview.com/2019/07/29/663/ai-computing-cloud-computing-microchips/> (2019).
- Blair G. A tale of two cities: reflections on digital pollution. Oxford University, May 13, Digital Pollution Workshop, 2020.
- Samuel G, Lucivero F and Lucassen A. Sustainable biobanks: a case study for a green global bioethics. *Global Bioethics* 2022; 33(1): 50–64.
- Scott RE, Palacios M and Maturana T. Electronic waste – A growing concern for the health sector. In: Mathias IMA (ed) *GoldBook - technological innovation in education and health*. Brazil: EdUERJ, 2012, pp. 621–649.
- Thompson M. The environmental impacts of digital health. *Digital Health* 2021; 7: 20552076211033421.
- Spyropoulos B, Alexandropoulos A, Boci N, et al. Toward the data-driven “smart” and “green” hospital-care. In: 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K) 27–29 Nov. 2017 2017, pp. 1–9.
- Sodhro AH, Sangaiah AK, Pirphulal S, et al. Green media-aware medical IoT system. *Multimedia Tools Appl* 2019; 78: 3045–3064.
- Godbole NS and Lamb J. Research into making healthcare green with cloud, green IT, and data science to reduce healthcare costs and combat climate change. In: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) 8–10 November 2018, pp. 189–195.
- Saiyeda A. Challenges and opportunities with green and sustainable computing in healthcare. *Green automation for sustainable environment*. Boca Raton: CRC Press, 2020, pp. 19–35.
- Alshqaqeeq F, McGuire C, Overcash M, et al. Choosing radiology imaging modalities to meet patient needs with lower environmental impact. *Resour Conserv Recycl* 2020; 155: 104657
- Tongue B. Designing sustainability into the digital healthcare revolution. *British Journal of Healthcare Management* 2019; 25(4): 1–3. DOI:10.12968/bjhc.2019.0038
- Chevance G, Hekler EB, Efoui-Hess M, et al. Digital health at the age of the anthropocene. *Lancet* 2020; 2: E290–E291
- Gala I and Doss S. Scrutinizing the level of awareness on green computing practices in combating Covid-19 at institute of health science-gaborone. In Anandan R, Suseendran G, Balamurugan S, Mishra A and Balaganesh D (eds.) *Human communication technology* 2021. DOI:10.1002/9781119752165.ch14.
- Ouhbi S, Fernández-Alemán JL, Idri A, et al. Are mobile blood donation applications green? In: 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA) 20–21 October 2015, pp. 1–6.

29. Godbole N and Lamb JP. The triple challenge for the healthcare industry: sustainability, privacy, and cloud-centric regulatory compliance. 2013 10th International Conference and Expo on Emerging Technologies for a Smarter World (CEWIT) 2013: 1–6.
  30. Godbole NS and Lamb J. Calculating a hospital's IT energy efficiency and determining cost effective ways for improvement. In: 2014 11th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT) 29–30 October 2014, pp.1–6.
  31. Samuel G, Hardcastle F and Lucassen A. Environmental sustainability and biobanking: a pilot study of the field. *New Genetics and Society*, in press.
  32. Lucivero F, Samuel G, Blair G, et al. Data-driven sustainability - an interdisciplinary perspective on governing the environmental impacts of a data-driven society, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3631331](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3631331) 2020.
  33. Samuel G. Towards an environmentally sustainable responsible innovation of data-driven technologies and AI. *AsSIST virtual* (UK) 2021.
  34. Kovacic Z, Strand R and Voöker T. *The circular economy in Europe: critical perspectives on policies and imaginaries*. London: Routledge, 2019.
  35. Morozov E. *To save everything, click here: the folly of technological solutionism*. New York: PublicAffairs, 2013.
  36. Cribb A. Managing ethical uncertainty: implicit normativity and the sociology of ethics. *Sociol Health Illn* 2020; 42: 21–34.
  37. Mosco V. *To the cloud: big data in a turbulent world*. Boulder, CO: Paradigm, 2014.
  38. Ruckenstein M and Schüll ND. The datafication of health. *Annu Rev Anthropol* 2017; 46: 261–278.
  39. Crawford K. *Atlas of AI*. New Haven: Yale University Press, 2021.
  40. Kitchin R. *The data revolution. Big data, open data, data infrastructures and their consequences*. London: Sage Publications, 2014.
  41. Datta Burton S, Kieslich K, Paul K, et al. Rethinking value construction in biomedicine and healthcare. *Biosocieties* 2020. DOI: 10.1057/s41292-020-00220-6
  42. Tutton R. *Genomics and the reimagining of personalized medicine*. Farnham: Ashgate Publishing, 2014.
  43. Dheensa S, Carrieri D, Kelly S, et al. A 'joint venture' model of recontacting in clinical genomics: challenges for responsible implementation. *Eur J Med Genet* 2017; 60: 403–409.
  44. Horton R and Lucassen A. Consent and autonomy in the genomics era. *Curr Genet Med Rep* 2019; 7: 85–91.
  45. Sterckx S, Dheensa S and Cockbain J. Presuming the promotion of the common good by large-scale health research. In: Beers B, Sterckx S and Dickenson D (eds) *Personalised medicine, individual choice and the common good*. Cambridge: Cambridge University Press, 2018, pp. 155–182.
  46. Shuja J, Ahmad RW, Gani A, et al. Greening emerging IT technologies: techniques and practices. *Journal of Internet Services and Applications* 2017; 8: 9.
  47. Lannelongue L, Grealey J, Bateman A, et al. Ten simple rules to make your computing more environmentally sustainable. *PLoS Comput Biol* 2021; 17: e1009324–e1009324.
  48. Hilty LM, Köhler A, Von Schéele F, et al. Rebound effects of progress in information technology. *Poiesis Prax* 2006; 4: 19–38.
  49. Takahashi KI, Tatemichi H, Tanaka T, et al. Environmental impact of information and communication technologies including rebound effects. In: IEEE International Symposium on Electronics and the Environment, 2004 Conference Record 2004 10–13 May 2004, pp. 13–16.
  50. Tronto JC. An ethic of care. *Generations: Journal of the American Society on Aging* 1998; 22: 15–20.
  51. Held V. *The ethics of care: personal, political, and global*. Oxford: Oxford University Press, 2007.
  52. Vinuesa R, Azizpour H, Leite I, et al. The role of artificial intelligence in achieving the sustainable development goals. *Nat Commun* 2020; 11: 33.
-