ETHICAL ISSUES OF E-INFRASTRUCTURES: WHAT ARE THEY AND HOW CAN THEY BE ADDRESSED?

Damian Eke, Simisola Akintoye, William Knight, George Ogoh, Bernd Stahl

De Montfort University (Leicester, United Kingdom)

damian.eke@dmu.ac.uk; simi.akintoye@dmu.ac.uk; william.knight@dmu.ac.uk; george.ogoh@dmu.ac.uk; bstahl@dmu.ac.uk

ABSTRACT

E-infrastructures are emerging as novel and effective ways of increasing creativity and efficiency of research. As technological innovations, these virtual, ubiquitous, pervasive infrastructures offer possibilities of international collaborations through open, data-driven and high-quality computing environments. Particularly in Europe, the aim is to create an ecosystem of e-science where multiple disciplines converge to foster interoperable and open collaboration with the help of significant data processing and computing capacity. While most agree that these research infrastructures are crucial to scientific reproducibility and rigor, e-infrastructural literature lacks critical discussions on the ethical concerns they raise or potentially can raise. This paper argues that e-infrastructures can raise a number of ethical, legal and social concerns. Some of these relate to data privacy and data security but they also include issues around animal welfare, data bias, intellectual property rights, environmental sustainability, digital divide and other unintended uses/misuses. This paper also presents a practical way of thinking about ethics in e-infrastructures. The underlying argument here is that addressing e-infrastructure ethical issues should start from the design of the infrastructure and continue through to its lifecycle. It requires the integration of relevant ethical principles into its design to foster responsible use/application. We then propose that this can be done through the Responsible Research and Innovation approach as an ethics-by-design tool.

KEYWORDS: Research, E-infrastructures, Infrastructures, Ethics, RRI, Ethics-by-design.

1. INTRODUCTION

In the last few decades, there has been an increasing move towards collaborative research to provide common solutions for shared concerns which has led to the development of research infrastructures, particularly e-infrastructures, that are intended to transform scientific research and practice. Backed by huge funds, national (like the NSF Office of Advanced Cyberinfrastructure) and regional (such as the European Strategy Forum on Research Infrastructures) programmes are being used to facilitate the establishment and operation of such research infrastructures (Schroeder, 2007;Pollock and Williams, 2010; Andronico et al., 2011). There is a long tradition to develop such infrastructures in Europe with over 700 million euro of budget allocated to close to 100 e-infrastructure projects in the EU's research framework programme Horizon 2020 (H2020) (Versweyveld, 2019). These infrastructures cover most aspects of science and research such as biological sciences, health/medicine, good, energy, environmental sciences, Physical sciences/engineering, computing sciences, social sciences, arts

and humanities. The European Union also showcases some of larger e-infrastructure initiatives that include EPOS: Viable solutions to tackle solid Earth grand challenges (Jeffery and Bailo, 2014), ELIXIR: A distributed infrastructure for Life science information (Crosswell and Thornton, 2012) or SHARE: A survey on Health, ageing and retirement in Europe. (Börsch-Supan et al., 2013).

The international ecosystem of e-infrastructures is extensive and different in that they have evolved along different geographic, disciplinary and application dimensions. These often-disparate e-infrastructures aim to provide open, flexible and competitive national and international services to advance scientific collaborations. In addition to enabling collaborations across disciplines and national boundaries, these ICT-based, data-driven and computer-intensive infrastructures are also being established to promote scientific reproducibility, improve research efficiency, creativity and rigor. In many cases, they provide digital/virtual open access to resources and services that provide answers to complex scientific questions that require global reflection. This means that they attract users from different jurisdictions who access multimodal and multidimensional resources (including a significant amount of data) and services (such as super computing services). E-infrastructure services often depend on significantly increasing amounts of data raising novel ethical, privacy and security challenges. The e-IRG White paper of 2013 also highlighted the technical, legal and ethical challenges involved in creating a collaborative data infrastructure in light of the volume, variety and velocity of data involved (e-IRG, 2013.).

The above emphasis on data issues and the recent EU's legislative emphasis on Data protection by design and by default (DPbD) in article 25 of the EU General Data Protection Regulation (GDPR) can be read as suggesting that privacy and data protection are the only data related concerns to be considered in innovations associated with human data such as e-infrastructures. Such a reading of the current discourse would clearly be wrong considering the possibilities of commercial research via these e-infrastructures which raise some uncertainty about the rules and formalities at national, European and international levels. According to the EU Horizon 2020 Expert Advisory Group on *European Research Infrastructures including e-Infrastructures,* research infrastructures go beyond datasets and scientific instruments because they play significant roles in shaping the intellectual and cultural dimensions of the innovation ecosystem; enabling smart specialization strategies. E-infrastructures improve national and regional scientific capacity through multifaceted relationships between data, computing resources and researchers from different scientific and regulatory backgrounds. These pose a number of novel ethical, social and legal challenges, depending on a lot of things including the nature of the e-infrastructure involved, the type data that it engages with and the breath of its applications. But as the construction and operation of these e-infrastructures continue to advance, there is an evident dearth of literature on the ethical issues they raise or can raise. Therefore, this paper seeks to answer two research questions: what are the ethical issues research e-infrastructures raise? And how can these ethical issues be addressed? To address these questions, we critically reflect on the e-infrastructure ecosystem.

This paper makes two contributions. First, it provides a general overview of the ethical concerns associated with e-infrastructures. As e-infrastructures continue to transform research and innovation, it is pertinent to be aware of the ethical, legal and social concerns they raise or exacerbate. These issues depend on the type of e-infrastructure and the nature of the services/resources it offers. Second, it is easy for the conversations around ethics to become too abstract to offer practical insights into achieving its objectives. Therefore, we also present a

practical way of thinking about ethics in e-infrastructures through ethics-by-design. The underlying argument here is that addressing e-infrastructure's ethical issues should start from the design of the infrastructure and continue through its lifecycle. It requires the integration of relevant ethical principles into its design to foster responsible use/application. We propose that this can be done with the Responsible research and innovation approach as an ethics-by-design tool.

Therefore, this paper starts with a brief overview of what we mean by research e-infrastructures and how they raise ethical concerns. We then provide a brief description of Ethics by Design and how it can positively shape the design and application of e-infrastructures. Responsible Research and Innovation (RRI) is then identified as a possible ethics-by-design tool that can help to achieve Ethics by Design in e-infrastructures. An overview of how this can be applied in the development of an e-infrastructure is then outlined. We conclude by discussing the challenges of applying RRI as an Ethics by Design tool to e-infrastructures.

2. A NOTE ON CONCEPTS AND STRUCTURE

The general discourse on the ethical issues associated with e-infrastructures requires the explanation of our understanding of what constitutes an e-infrastructure. It is important also to describe how we arrived at the ethical issues. Therefore, this section starts with a brief clarification of what is referred to as e-infrastructures in this paper and how they are different from other research infrastructures. Then it provides clear insights on what ethical issues are associated with e-infrastructures which e-infrastructure developers can use to take proactive actions to build ethically responsible platforms for research and innovation.

The identification of ethical issues was undertaken through a critical exploration of emerging literature on e-infrastructures. A critical literature review was essential for identifying the landscape of e-infrastructure operations and applications that can raise some concerns. Issues were adjudged ethical issues given that they create conflicts with established moral principles; requiring a person(s) to choose between alternatives justified or evaluated as right or wrong. As Stahl (2012) observed, these are issues that are associated with moral intuition or explicit morality. This paper identifies these issues in the context of e-infrastructure operations and in the next section suggests a possible way of addressing them. The contributions of this paper are based on non-empirical critical reflection on literature and personal observations of the authors as members of the Ethics support team of the EU Human brain project building a neuroscience e-infrastructure - European Brain Research InfraStructure (EBRAINS).

3. E-INSTRASTRUCTURES

According to the Directorate-General for Research and innovation of the European Commission, research infrastructures are ''facilities that provide resources and services for research communities to conduct research and foster innovation''. In our increasingly digitized world, these facilities are becoming more electronic with all ICT based resources – interconnecting computer science and specific research disciplines. These are called e-infrastructures combining digital technology, computational resources to support research collaborations; creating new virtual research communities that share, federate and access scientific tools and resources (including but not limited to data and computing facilities). They are defined in this paper as *computing facilities and electronic resources that facilitate research and innovation* which are

differentiated from other research infrastructures due to their ability to provide digital services, resources and tools for scientific research.

E-infrastructures are innovations that harness knowledge from multiple disciplines for the advancement of research and the benefit of society. From the online etymology dictionary, 'innovation' is a 1540s word that comes from the latin word *innovationem,* meaning ''*a novel change, experimental variation, new thing introduced in an established arrangement*''. Therefore, the key meaning of innovation relates to a change of what is existing; it does not necessarily mean a new invention but involves a new approach. As an innovation, e-infrastructure is revolutionizing the way scientific research is done through a combination of some or all of these critical elements: supercomputing, cloud computing, big data, networking, collaboration and open access. They are not only changing the landscape of scientific relationships but also raising new ethical challenges or exacerbating old concerns.

According to the European e-infrastructure Reflection Group (e-IRG), the pan-European e-infrastructure landscape includes *networking infrastructure*, *supercomputing facilities, Cloud Infrastructure* and *data infrastructure*.(van Rijn and Vandenbroucke, 2017) An example of a networking infrastructure is GÉANT that provides interconnectivity between national research and Education networks (NRENs) across many European and non-EU countries with estimated 50 million users. Super computing infrastructures in Europe include European Grid infrastructure (EGI) and Partnership for Advanced Computing in Europe (PRACE). Both provide high-level computing services but EGI focuses on large-scale federated High-throughput computing (HTC) solutions and PRACE provides access to high-performance computing services and facilities. In recent years, ELIXIR has become a key data infrastructure that manages and safeguards the increasing volume of data generated by publicly funded research worldwide. Other data infrastructures include EUDAT, ZENODO and OpenAIRE. A further element of e-infrastructure is the increasing use of cloud technologies. The *Helix Nebula Initiative* is one example creating research partnerships between industry, space and science through open cloud services. These are indeed becoming core platforms for e-science, education and innovation (Andronico et al., 2012).

E-infrastructures have one or more of these elements: data, networking, HTC, HPC or cloud computing. For instance, as a network infrastructure, GÉANT also has a cloud service platform but cannot be classified as a High-performance computing (HPC) infrastructure. PRACE is a HPC infrastructure but can also offer data services. However, EGI is a High-throughput computing (HTC) infrastructure but also can offer both cloud and data services. The type of infrastructure determines what research is supported and the type of research community created. These e-infrastructures empower researchers to explore scientific questions in different ways. They involve a lot of complex interrelationships between researchers, the technology, organizations, networks and variety of human and animal datasets (in most cases). The technical and social dynamics of these interrelationships present a number of ethical, legal and social risks because persons and data are involved. Some of these concerns can generally be associated with all e-infrastructures or indeed all research infrastructures but some are peculiar to certain infrastructures.

3.1. E-infrastructure ethics

E-infrastructures raise different ethical concerns depending on the type of resources and services they offer and the relationships they foster. For Data e-infrastructures, concerns are mainly related to data sharing, security, access and interoperability. While these e-infrastructures attempt to foster scientific discoveries through sharing of data, there is also the need to ensure the subject's privacy and confidentiality (Gagliardi and Muscella, 2010). These are fundamental ethical and legal requirements for research collaboration in any research involving human data. A virtual platform hosting terabyte of human data surely amplifies concerns related to privacy and data protection. Identifiable research data-related issues include; informed consent issues, incidental findings and possibilities of positive re-identification. Informed consent (D'Abramo, 2015; Wolf et al., 2018) and incidental findings (Viberg et al., 2014) have been acknowledged in literature as a major challenge facing human subject research and infrastructure operations. An e-infrastructure that facilitates a borderless sharing and usage of research data only exacerbates these problems. Fundamentally these raise the questions of autonomy and the practical questions of accountability and whose responsibility it is to address these issues. They also raise the technical question of how to protect the rights (to privacy and confidentiality) of the research subjects in the face of the changing landscape of technology making it difficult to achieve effective de-identification/anonymization of human datasets (Rocher et al., 2019).

Additionally, there is the critical issue of data controllership- who owns and or controls the data. This is a concept defined by the EU General Data Protection Regulation as any organizational entity ''which alone or jointly with others, determines the purposes and means of the processing of personal data''. A clear identification of the data controller therefore has important influence on the sharing of data as well as the apportioning of liabilities in cases of data breach. E-infrastructures involve data processing at different levels and by multiple institutions which are sometimes located in different countries with different data protection provisions. Determining the data controllership and data processor roles at the various levels of the 'research grid' becomes complicated. Even the EU GDPR does not provide sufficient clarity to this issue given the different national interpretations of its provisions.

The collaborative nature of the e-infrastructure ecosystem also complicates the issues of copyright and intellectual property rights. For instance, the collaborative work in or for e-infrastructure can lead to the development of a bespoke software with the possibility that no licence or intellectual property agreement was made prior to development as was observed by the 2013 e-IRG Task Force Report on Legal issues. Determining the owners of the intellectual property and what the software can be used for especially when state funds are involved become a critical challenge. Also raised in the context of data e-infrastructures are deeper questions of data misuse which may include, but are not limited to using findings out of context or distorting research findings. There are also the issues of dual use which are gaining more prominence in EU and international policy literature. This refers to the use of research findings for unacceptable military or commercial purposes. This is a concern already identified in the EU HBP where an opinion on Responsible dual-use has been developed (Aicardi et al., 2018). Dual use and misuse also featured prominently in the recently published OECD recommendation on Responsible innovation in neurotechnology. For instance, a neuroscience e-infrastructure that increases the integration of brain data and technology could be leveraged for military purposes. A recent article published in the US's National Defense University's journal PRISM, observed that China's military strategy is informed by the belief that brain science is key to the effectiveness

of future battlefields (Kania, 2020). This raises the possibility of brain data misuse and highlights the level of national security challenges brain science research is likely to present.

Another issue that does not receive deserved attention in the data e-infrastructure discourse is animal welfare. Animal experiments that generate animal data are regulated differently in different parts of the world. What is ethically permissible in one jurisdiction may be illegal in another. A data e-infrastructure designed to curate and share all types of animal data from every part of the world without any form of governance implicitly endorses all forms of animal experiments including potentially unethical research. As a virtual platform with users in different parts of the world, data e-infrastructures for animal data could potentially be used to share data from unethical animal experiments which exacerbates animal welfare concerns. Without any form of governance, e-infrastructures can thus facilitate low quality animal data which can ultimately affect the quality of e-science they foster.

There are also complex ethical issues that are associated with supercomputing infrastructures. Addressing them as ethics in HPC, Lawson et al., (2019) identified three major ethical issues surrounding the use and application of HPCs. These include misuse, inequality and environmental concerns. The underlying argument here is that HPCs could potentially be misused for unacceptable military purposes or to develop a destructive pathogen. It can also be argued that, as with any other high-resource intensive technology, HPCs can indirectly perpetuate a digital divide between developed and less developed countries; men and women and other minority groups. Of the Top500 list of supercomputing sites published in 2019, only 1 is in South America and none is in Africa. The capacities offered by these systems can increase a country or region's competitive advantage while thickening the digital lines that divide nations which creates an ethical imperative. There are also the environmental impacts of HPCs. The large amount of power consumption and its associated carbon emission are well documented critical challenges of supercomputing services (Yang and Chien, 2016).

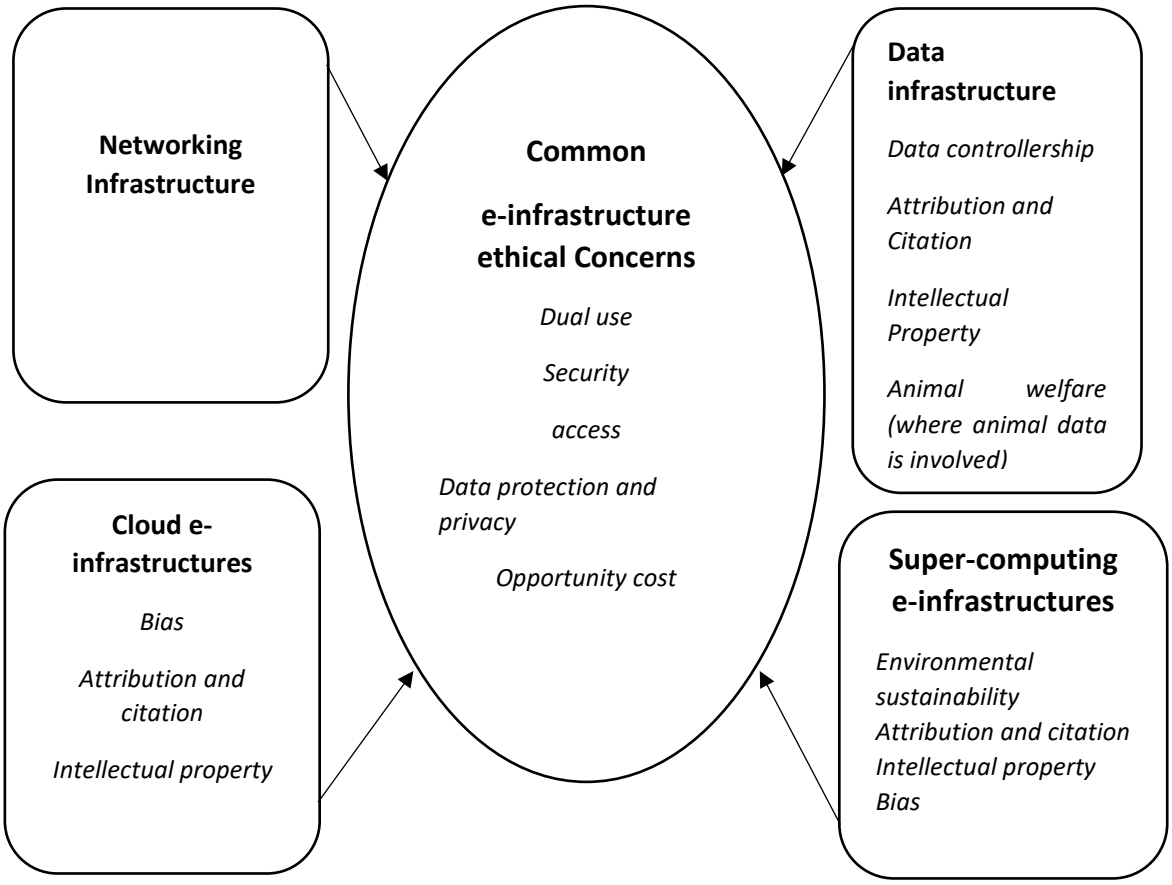
Furthermore, the cloud e-infrastructure ecosystem in Europe provides support for the creation of many artificial intelligence (AI) systems. Many of these systems (machine learning, deep learning) are supported over distributed e-infrastructures in the European Open Science Cloud (EOSC). An example is the DEEP-Hybrid-Datacloud that helps users to develop, train, share and deploy models. The EOSC is the European Commission's platform created to federate emerging e-infrastructures in a way that will unlock the value of big data and foster innovation (including advancements in artificial intelligence) towards achieving the digital single market strategy. The ever-growing relationship between these infrastructures, data and deep/machine learning raises the possibility of discrimination/bias which can creep in at different stages of the analysis, exacerbating intersectional biases that might be evident in the curated data. While network e-infrastructures are not primarily designed to store research data, a significant amount of personal data is processed through it. Consider the examples of eduroam and eduGain that provide a communications platform across research and education. This requires that organizations be able to communicate and collaborate with enhanced privacy. There is still the danger of malicious attacks that can lead to privacy breaches. Network security therefore is of great importance.

A further concern about e-infrastructures relates to the value they represent to the scientific community and to society at large and the often-significant costs and opportunity costs they incur. Most of the infrastructures introduced earlier have large financial value and cost significant amounts of money to maintain. These are resources that cannot be used for other

research or for other socially beneficial activities. This is even more relevant because they are usually financed by public funds which directly compete with other beneficial goods, such as social security or healthcare. This is justified by the potentially large public benefit that may arise out of the research that is facilitated by the infrastructure. However, measuring or accurately predicting these benefits is extremely difficult. A lack of use of the infrastructure by the scientific community or a lack of research outcomes thus constitute a waste of resources which is a political but also an ethical concern.

The above issues are not trivial and contribute to the wider discourse on what can be referred to as *e-infrastructure ethics* in this paper which means a set of moral principles and standards that should govern the use of e-infrastructures. This paper proposes that one of the approaches that can be utilized to address these ethical issues is to develop a responsible/ethical innovation through a practical application of ethical thinking to the design of its products and services. This approach combines technical, ethical and social mechanisms to provide solutions based on reflexive actions. It is a proactive and preventive measure motivated not only by the imperative to consider research ethics questions but also by the need to responsibly govern the application of generated complex, big data and the necessity to embed relevant principles into the design. The latter should lead to the establishment of practical and responsible mechanisms/approaches informed by ethics and the law and designed to integrate core ethical principles and legal provisions into the e-infrastructure.

Figure 1. Ethical issues of different e-infrastructures.



The broad nature of the above e-infrastructural concerns is a reason to believe that a framework that embeds privacy or data protection into design is not sufficient to mitigate other social, economic, legal, philosophical and ethical concerns e-infrastructure raises. A responsible e-infrastructure technology requires a robust approach/mechanism that addresses fundamental concerns beyond privacy and data protection. There is a need to integrate relevant ethical principles into the development and application of e-infrastructures. This is important because as Simon (2016) observed, in building technologies, values are often unintentionally inscribed in them and in return they may promote or demote certain values. Even though e-infrastructure is not technology *per se,* as an innovation, it should facilitate responsible actions. Instead of unintentionally embedding undesirable values into the design, responsible e-infrastructure requires a conscious effort to intentionally embed desired values into the innovation in ways that can promote values desired by the relevant stakeholders who are or can be affected by the technology. While this sounds great, the practical questions on what principles and how to embed them into the design of e-infrastructures are unclear in literature or practice.

Drawing from the literature on Technology ethics and Design Thinking, this paper argues that the concept of Ethics-by-design (Mulvenna et al., 2017; Dignum et al., 2018) as a form of value-sensitive design (Friedman et al., 2006), provides a robust framework for ethical and legal considerations of risks in the development and application of research e-infrastructures. Even though this concept has been applied in robotics (Dodig Crnkovic and Çürüklü, 2012) and business (Moore, 2017) discourses, it is yet to be applied to the design of a research e-infrastructure. The authors draw extensively from their experiences of developing approaches of addressing data governance issues, data protection, research ethics compliance and dual use in a large-scale EU project.

4. ETHICS-BY-DESIGN AND TECHNOLOGY

The idea of ethics by design is not an entirely new concept and has been around for over 20 years. One of the earliest developers of the notion was Tonkinwise (2004); he based his ideas on the works of Scarry (1987), Latour (1992), Jelsma (2003) and Borgmann (1995). Tokinwise sees design as embodying ethics because it is the process for making the world friendlier to us and the effort to make it more caring towards us, and therefore more morally acceptable to us. He therefore maintains that 'forethought' is necessary to design things that make ethical outcomes easier or harder, and suggests that ethics by design/ in design makes it easier for people to be more ethical. Feister et al. (2016) also promote the concept of ethics in design because they agree with the notion that design is inherently tied to ethics. They maintain that integrating ethics in the micro-level everyday decisions and thinking about it throughout the engineering design process will encourage greater incorporation of ethical thinking into the entire design process. They therefore advocate for a human-centred model of design (HCD) which is situated around appreciation of users' knowledge, skills, and experiences and gives attention to all stakeholders that might be affected beyond the targeted user.

In applying similar principles to the design of autonomous AI agents (softbots/robots), Crnkovic and Çürüklü (2012) promote the idea that any intelligence acquired by such artefacts must come in conjunction with ethics through what they call 'ethical by design'. This takes into consideration the engineering ethics of designers, manufacturers, and maintenance services, the user's attitude, as well as those of the artefacts. In a similar vein, Dignum et al. (2018) have applied the concept of ethics by design to autonomous agents in an effort to explain how to

build ethically-aware agents. In this regard, they define ethics by design as "the methods, algorithms and tools needed to endow autonomous agents with the capability to reason about the ethical aspects of their decisions, and methods, tools and formalisms to guarantee that an agents behaviour remains within given moral bounds." They maintain that the moral, societal and legal values of autonomous agents must be taken into consideration in their design and that focus should be on ensuring thrust of AI systems rather than performance alone. They conclude that AI related developments must be designed to ensure societal good through three principles articulated as 'ART' – **A**ccountability (the system should be able to justify its decisions and actions to stakeholders); **R**esponsibility (of both the AI system and those interacting with it in accounting for decisions, diagnosing errors, or unexpected findings); **T**ransparency (in terms of explanation and clarity of algorithms and their results). Other concepts that they suggest can aid the integration of ethics by design include ethically aligned design, responsible research and innovation, and ethics of the design processes itself.

An attempt was made by Mulvenna et al (2017) to develop a manifesto of principles for Ethical by Design that would be all-encompassing and guide everyone developing or considering solutions regardless of the area, market or their own expertise. They firmly believe that design thinking can and should be 'ethical by design' such that it 'inherently supports the ethical development, selection, and use of products and services.' They therefore propose the following 12 principles as the manifesto for 'ethical by design' - the design should engender empathy for others, enable informed decision, offer alternative or customisation that respects people's right to choose how they wish to engage with a product, allow equitable access and which must be balanced by privacy and security, support progression of policy, and actively look for and challenge biases in the product or service. Other principles of the manifesto include a requirement to compliment differing needs, abilities, viewpoints and morals; support shared decision making and feedback, aim for sustainability, be realistic about what is possible, integrate planning for handling failure, and provide support throughout the lifespan of the product or service.

While agreeing that the above manifesto of principles for 'Ethical by Design' provides a pragmatic ethical framework for design processes, Lee et al. (2018) also noted that there are other ethical principles to be considered in design. Judging from the above identified ethical issues of e-infrastructures, some relevant principles will be to be proactive rather than reactive, do a lifecycle ethics, to respect human autonomy, non-instrumentalism of living beings, intellectual property rights and striving for inclusivity, diversity and environmental sustainability. But how can these be integrated into the design of e-infrastructures?

4.1. Ethics-by-design and e-infrastructures

As it has been noted earlier in this paper, e-infrastructure is not a technology but can be classified as an innovation. It is an innovation that uses a combination of technologies to foster complex relationships/collaborations. These emerging collaborations are outcomes of the design of the e-infrastructures and raise a number of ethical concerns. Shaping these relationships e-infrastructures foster at the design level is an imperative. It is evident that the design of e-infrastructure is fraught with the need to make ethically responsible choices that can shape the application of its tools, resources and services. There is a need for e-infrastructure developers to appreciate the responsibility they have in addressing ethical problems exacerbated by this innovation. Researchers (including commercial entities) are encouraged to

use the state-of-the-art services offered by e-infrastructures to develop cutting-edge results for civil society. In most cases, this involves a relationship between users and thousands of datasets from human subjects. The complex relationships within the e-infrastructure ecosystem present critical challenges and the integration and implementation of relevant principles will require the collaboration of stakeholders involved in e-infrastructures. Embedding these principles into the design of a responsible e-infrastructure where health data and supercomputing services are involved is a challenge facing e-infrastructure developers. Effective implementation of the ethical principles such as autonomy, proactivity, beneficence and non-maleficence will require collaborations between e-infrastructure developers, resource/service providers and the users. It will involve a mixture of technical, social and legal mechanisms that provide solutions to both identifiable and unanticipated problems. It requires ethics by design.

4.2. Ethics-by-design through rri for e-infrastructures

We propose that one possible way of building these principles into e-infrastructures is by applying the Responsible Research and Innovation (RRI) approach to the design. During a Thematic day at the 20th International conference on Principles and practice of multi-agent systems (PRIMA) in 2017, Juan Pavon suggested RRI as an ethics-by-design tool for multi-stakeholder intelligent systems (Dignum et al., 2018). His argument was that it is a tool that copes with issues that require cooperation. RRI is defined as a meta-responsibility framework that aligns the goals, purposes and processes of research and innovation to produce desirable outcomes (Stahl, 2013). E-infrastructures epitomize the convergence of research and technological innovations. RRI therefore, seems a good approach to achieve a responsible e-infrastructure that will positively support those who use it. The application of RRI into the building of this technology is a way of influencing e-research and e-science in ways that are socially acceptable, ethically responsible, legally compliant and environmentally sustainable.

Jirotka et al., (2017) articulated the importance of this approach in Information and Communications Technology (ICT). Recent works of some of the authors of this paper highlight how this inclusive and discursive mechanism can be applied in care robots (Stahl et al., 2019; Stahl and Coeckelbergh, 2016) and data governance in neuroscience (Fothergill et al., 2019). In Fothergill et al., (2019), a version of RRI based on Stilgoe et al., (2013) was used to unite and respond to large-scale neuroscience stakeholders' expectations by ensuring an ongoing, productive dialogue that produced beneficial results for research and innovation. This version endorsed by the UK Engineering and Physical sciences Research Council (EPSRC) adopts a reflective process of Anticipation, reflection, engagement and action (AREA) (Owen, 2014). In another publication, Jirotka et al., (2017) developed an extended version of this framework by adding what they called the 4Ps: *Process, product, purpose and people*. This was an attempt to provide practical guidance on key aspects of the research and innovation that shape the products or outcomes with central focus on purpose and the people who will be affected by the innovation. These applications of RRI highlight that this approach fosters a human-centred, collaborative and inclusive innovation that e-infrastructures require.

E-infrastructures create imbalances between those who have access and those who do not; those whose interests are addressed and those who lack representation; threaten the privacy rights of data subjects. A responsible e-infrastructure innovation should therefore consider the interests of many sections of the society, creating systems that will provide solutions with inclusive values. It requires meaningful engagement and reflection that can align the design with

societal values. The last thing an e-infrastructure developer wants is to design a system providing high quality research services that nobody uses or that raises new or exacerbates old ethical and social problems. RRI can equip e-infrastructure developers with the knowledge of relevant values and principles through reflective and inclusive activities and practical ways of putting action into practice. It can provide a way of ensuring that ethics is an essential component of the core functionality of e-infrastructure or is integral to the system without diminishing its scientific purposes. It is important to note that the effectiveness of this approach in the design of an e-infrastructure is yet to be tested. As members of the Ethics and society sub-project of the EU Human Brain project (HBP) (working towards the development of an international neuroscience e-infrastructure) we have an opportunity to test the impact of RRI as an ethics-by-design tool in the development and application of an e-infrastructure.

5. CONCLUSION

In this paper we have set out the ethical issues that are likely to arise due to the development and use of e-infrastructures. There are already a number of such infrastructures in operation and new ones are under development. It is reasonable to assume that researchers will continue to seek to benefit from new ways of generating and analysing data. The trend towards big science or big data science is therefore likely to continue. While the perception of what counts as 'big' is likely to evolve, if current trends continue, then the ability to generate data will continue to outpace the ability of individual researchers or institutions to collect and make sense of this data. E-infrastructures are a way of addressing this, providing means of storing, analysing, visualising and processing such data.

If this prediction turns out to be true, then we are only witnessing the start of the development towards more e-infrastructures. These may come in all sorts of shapes and sizes; they will have different funding models and different governance structures. What they have in common is that they may raise ethical concerns that go beyond what their initiators had foreseen. It is therefore important to start to think about these issues early. Some of them are easy to foresee, such as data protection concerns in cases of infrastructure dealing with patient data. Others may be less obvious but still important. We hope that this paper provides the basis for a more comprehensive discussion of these ethical issues. Established e-infrastructures may provide good practice guidance and examples that developing ones can learn from.

In this paper we have suggested that there are established ways of implementing ethical reflection in the development of technology, such as ethics by design or RRI. The next step is now to transfer these ideas from the level of individual technology projects to infrastructure projects. This is likely to be difficult for reasons of size and scale which can raise further questions. E-infrastructures are, by their very nature, open and can be utilised for different purposes. They, therefore, raise difficult questions about accountability and responsibility. It is, therefore, unlikely to be easy to simply 'implement' ethics in infrastructure. It will require dialogue, openness to critique and failure, an ability and willingness to experiment and learn. We hope that this paper can contribute to the process of critical self-reflection of people involved in e-infrastructure to ensure that ethical issues do not stand in the way of the immense benefits that e-infrastructures promise.

ACKNOWLEDGEMENT

This project/research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3).

REFERENCES

Aicardi, C., Bitsch, L., Bang Bådum, N., Datta, S., Farisco, M., Evers, K., Fothergill, B.T., Giordano, J., Harris, E., Jørgensen, M.L., Klüver, L., Mahfoud, T., Rainey, S., Riisgaard, K., Rose, N., Salles, A., Stahl, B., Ulnicane, I., (2018). Opinion on 'Responsible Dual Use' Political, Security, Intelligence and Military Research of Concern in Neuroscience and Neurotechnology.

Andronico, G., Ardizzone, V., Barbera, R., Becker, B., Bruno, R., Calanducci, A., Carvalho, D., Ciuffo, L., Fargetta, M., Giorgio, E., 2011. E-Infrastructures for e-science: a global view. J. Grid Comput. 9, 155–184.

Borgmann, A. (1995) The Depth of Design. In: Buchanan, R. and Margolin, V. (eds.) *Discovering Design*. Chicago: The University of Chicago Press, p. 283.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S., 2013. Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). Int. J. Epidemiol. 42, 992–1001. https://doi.org/10.1093/ije/dyt088

Crosswell, L.C., Thornton, J.M., 2012. ELIXIR: a distributed infrastructure for European biological data. Trends Biotechnol 30, 241–242.

D'Abramo, F., 2015. Biobank research, informed consent and society. Towards a new alliance? J Epidemiol Community Health 69, 1125–1128.

Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., Haim, G., Kließ, M.S., Lopez-Sanchez, M., 2018. Ethics by Design: necessity or curse?, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 60–66.

Dodig Crnkovic, G., Çürüklü, B., 2012. Robots: ethical by design. Ethics Inf. Technol. 14, 61–71. https://doi.org/10.1007/s10676-011-9278-2

e-IRG, n.d. e-IRG Whit paper.

Feister, M.K. et al. (2016) Integrating Ethical Considerations In Design. In: *STEM Intergration Through Engineering*. Ammerican Society for Engineering Education 123 Annual Conference & Exposition. New Orleans, LA USA.

Fothergill, B.T., Knight, W., Stahl, B.C., Ulnicane, I., 2019. Responsible Data Governance of Neuroscience Big Data. Front. Neuroinformatics 13, 28. https://doi.org/10.3389/fninf.2019.00028

Friedman, B., Kahn, P., Borning, A., 2006. Value Sensitive Design and Information Systems, in: Zhang, P., Galletta, D. (Eds.), Human-Computer Interaction in Management Information Systems: Foundations. M.E Sharpe, Inc, NY.

Gagliardi, F., Muscella, S., 2010. Cloud computing–data confidentiality and interoperability challenges, in: Cloud Computing. Springer, pp. 257–270.

Jelsma, J. (2003) Innovating for Sustainability: Involving Users, Politics and Technology. *Innovation: The European Journal of Social Science Research*, 16(2), pp. 103–116.

Jeffery, K.G., Bailo, D., 2014. EPOS: using metadata in geoscience, in: Research Conference on Metadata and Semantics Research. Springer, pp. 170–184.

Jirotka, M., Grimpe, B., Stahl, B., Eden, G., Hartswood, M., 2017. Responsible research and innovation in the digital age. Commun. ACM 60, 62–68.

Kania, E., 2020. Minds At War: China's Pursuit Of Military Advantage Through Cognitive Science And Biotechnology – Analysis. PRISM 8.

Latour, B. (1992) Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker, W.E., Law, J. and American Council of Learned Societies (eds.) *Shaping technology/building society: studies in sociotechnical change*. Cambridge, Mass: MIT Press, pp. 225–258.

Lawson, M., Lofstead, J., Lüttgau, J., 2019. With Great Power Comes Great Responsibility: Ethics in HPC. Presented at the The International Conference for High Performance Computing, Networking, Storage, and Analysis, Denver Colorado.

Lee, T. et al. (2018) Teaching Ethical Design in the Era of Autonomous and Intelligent Systems. In: *2018 World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC)*. 2018 World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC). Albuquerque, NM, USA: IEEE, pp. 1–4.

Moore, S.L., 2017. Ethics By Design: Strategic Thinking and Planning for Exemplary Performance, Responsible Results, and Societal Accountability. HRD Press, Amherst, Mass.

Mulvenna, M., Boger, J., Bond, R., 2017. Ethical by design: A manifesto, in: Proceedings of the European Conference on Cognitive Ergonomics 2017. pp. 51–54.

OECD, 2019. Recommendation on Responsible Innovation in Neurotechnology, https://www.oecd.org/science/recommendation-on-responsible-innovation-in-neurotechnology.htm

Owen, R., 2014. The UK Engineering and Physical Sciences Research Council's commitment to a framework for responsible innovation. J. Responsible Innov. 1, 113–117.

Pollock, N., Williams, R., 2010. E-infrastructures: How do we know and understand them? Strategic ethnography and the biography of artefacts. Comput. Support. Coop. Work CSCW 19, 521–556.

Rocher, L., Hendrickx, J.M., De Montjoye, Y.-A., 2019. Estimating the success of re-identifications in incomplete datasets using generative models. Nat. Commun. 10, 1–9.

Schroeder, R., 2007. e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production? Prometheus 25, 1–17.

Scarry, E. (1987) *The Body in Pain: The Making and Unmaking of the World*. Oxford University Press.

Simon, J., 2016. Value-Sensitive Design and Responsible Research and Innovation, in: The Ethics of Technology - Methods and Approaches. Rowman & Littlefield Publishers, London, pp. 219–236.

Stahl, B.C., 2013. Responsible research and innovation: The role of privacy in an emerging framework. Sci. Public Policy 40, 708–716.

Stahl, B.C., 2012. Morality, ethics, and reflection: a categorization of normative IS research. J. Assoc. Inf. Syst. 13, 1.

Stahl, B.C., Akintoye, S., Fothergill, B., Guerrero, M., Knight, W., Ulnicane, I., 2019. Beyond research ethics: dialogues in neuro-ICT research. Front. Hum. Neurosci. 13, 105.

Stahl, B.C., Coeckelbergh, M., 2016. Ethics of healthcare robotics: Towards responsible research and innovation. Robot. Auton. Syst. 86, 152–161.

Stilgoe, J., Owen, R., Macnaghten, P., 2013. Developing a framework for responsible innovation. Res. Policy 42, 1568–1580.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017) *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. New Jersey, USA: IEEE.

Tonkinwise, C., 2004. Ethics by design, or the ethos of things. Des. Philos. Pap. 2, 129–144.

van Rijn, A., Vandenbroucke, R., 2017. Guide to e-Infrastructure Requirements for European Research Infrastructures.

Versweyveld, L., 2019. Over 700 million euro of budget allocated to e-Infrastructure related projects in H2020 - close to 600 million euro to HPC related projects. E-IRG News Blog.

Viberg, J., Hansson, M.G., Langenskiöld, S., Segerdahl, P., 2014. Incidental findings: the time is not yet ripe for a policy for biobanks. Eur. J. Hum. Genet. 22, 437–441.

Wolf, S., Clayton, E.W., Lawrenz, F., 2018. The Past, Present, and Future of Informed Consent in Research and Translational Medicine. J. Law. Med. Ethics 46, 7–11.

Yang, F., Chien, A.A., 2016. Scaling Supercomputing with Stranded Power: Costs and Capabilities. Univ. Chic. Tech Rep.