Responsible Research and Innovation in the Digital Age

Marina Jirotka, Barbara Grimpe, Bernd Stahl, Grace Eden and Mark Hartswood

Introduction

At a time when increasingly potent technologies are being developed that have the potential to transform society, investigators in all fields, including ICT, are under growing pressure to consider and reflect on the motivations, purposes and possible consequences associated with their research. This pressure comes from the general public, civil society and government institutions. In parallel with these demands, there is a growing recognition that current ethics review procedures within ICT may not address broader concerns such as the potential societal consequences of innovation.

Instances of ICTs raising societal concerns abound. For example, alongside headline grabbing concerns that Artificial Intelligence (AI) may ultimately pose an existential threat to humankind, there are more prosaic, yet strongly felt, social transformations currently being wrought by AI technologies. For instance, AI is becoming an increasingly powerful protagonist in the story of how digital technologies are transforming the nature of work as more aspects are mediated digitally, including how work is allocated, assessed and rewarded. With these new forms of digital agency driving important aspects of labour markets, crucial questions arise as to whose interests they serve, and how to ensure accountability and transparency.

This is but one example of many debates around technological, product- or process-based innovations. Potential issues are wide-ranging and crucially, often emerge after technologies have been embedded into the mainstream.

.

There is a long history of ICT scholars and professionals trying to understand and address such issues. However, there are still numerous areas of concern. A novel concept - Responsible Research and Innovation (RRI) - has recently emerged in response to the challenges of designing innovations in a socially desirable and acceptable way. This approach may be useful for framing discussions about how to manage the introduction of future innovations in ICT. In this article, we discuss the origins of RRI, briefly consider relevant research from Computer Ethics and Human-Computer Interaction (HCI), and illustrate the need for a new approach to ICT research governance. Finally, we suggest ways in which the ICT community might draw upon a framework for RRI in ICT based on the findings of a recent interview study with the ICT community.

Ethics and Social Responsibility for ICT

Traditionally ICTs have been associated with the development of tools that possess discrete and transparent functionality meant to support specific tasks. However, today their 'diversity, scope, and complexity' have extended far beyond this, to becoming situated within the very fabric of our

daily lives.[17] Rather than being merely tools, the technologies now designed are arguably transforming and augmenting the world around us, where computer-generated information, objects and infrastructures 'coexist in the same space as the real world'.[1]

Debates about ethical issues in ICT are not new; researchers have been concerned with the practice of ethics in computing since the 1950s.[23] And with the emergence of Human-Computer Interaction (HCI) in the 1980s, researchers have attended to the design of usable interactions between people and computers where broader ethical and societal issues of application design and use have also been considered.[4]There are numerous ways in which ICT researchers have tried to address ethical questions, for example through participatory design,[13] ICT for development[10] and many others.

In addition to the approaches to ethics that come from within the ICT research and development communities, there is a rich array of complementary thought that similarly tries to address particular ethical issues. The field of computer ethics which draws on philosophy as well as computer science, information systems, sociology and many others has a rich history of reflecting on ethics of ICT.[6, 11]

Furthermore, professional bodies such as the ACM (https://www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct) IEEE (http://www.ieee.org/about/ethics.html) or BCS (http://www.bcs.org/upload/pdf/conduct.pdf) have developed codes and standards for professionals to adhere to for considering ethical issues. Whilst guidelines and standards are firmly in place, there has long been a debate as to the limits of these approaches. A key question becomes whether or not future ethical and societal challenges are likely to be amenable to being addressed in these ways.

All the above approaches to identifying and addressing ethical issues are valuable. What is currently lacking though is a way of combining them that will allow the broad range of stakeholders involved to systematically engage with goals, purposes, challenges, problems and solutions in research and innovation processes. This means that individual researchers, research institutions, professional bodies, research funders, industry, and civil society will need to collaborate more. In practice, that means to incorporate different kinds of knowledge, including that from citizens, to inform the goals, directions and trajectories of innovation in an inclusive way. This has been the case in some areas, for example privacy and data protection, where long-standing debates have led to regulation and legislation and to innovations in methods for design. However, in many areas of ICT this has not yet happened. In light of the societal importance of ICT, such a broader engagement may now be necessary. Other areas of research and innovation that have been more socially contested have a longer history of such engagement. We therefore propose to look at responsible research and innovation as a discourse that has arisen from these more contested fields and discuss whether and how it may be applied to ICT.

Responsible Research and Innovation (RRI)

RRI initiatives across policy, academia and legislation emerged over a decade ago.[5, 15] RRI began with an aim to identify and address uncertainties and risks associated with novel areas of research beginning with Nanotechnology[5] and moving to the environmental and health sciences including Geo-engineering[18] and Synthetic Biology.[21] The scope of RRI has since expanded to include Computer Science, Robotics, Informatics and ICT more generally.[8] RRI proposes a new process for

research and innovation governance. The aim is to ensure that science and innovation are undertaken in the public interest by incorporating methods for encouraging more democratic decision-making through greater inclusion of wider stakeholder communities that might be directly affected by the introduction of novel technologies.

In other words, RRI seeks to facilitate a more reflective and inclusive research and innovation process, from fundamental research through to application design. In each phase of the innovation process there may be certain responsibilities associated with activities that occur within them, particularly in relation to how decisions taken might impact upon society. The focus is on creating a new mode of practical research governance that would transform existing processes with a view to ensuring a greater acceptability and even desirability of novel research and innovation outcomes, whilst also identifying and managing potential risks and uncertainties. RRI requires a widening of scope from risk governance to the governance of innovation itself.[18]

There is a broad debate of the conceptual foundations of RRI and ways of implementing it in practice. Probably the most advanced framework for RRI currently in circulation is that proposed by Stilgoe et al.[18] who also provide a non-exhaustive list of a variety of possible RRI methods, tools and techniques such as, citizens' juries or moratoriums. This approach has been taken up in EU policy and research such as, the RRI Tools project (www.rri-tools.eu). It has also been adopted and adapted by the UK Engineering and Physical Science Research Council (EPSRC www.epsrc.ac.uk/research/framework/). The EPSRC's framework uses the acronym AREA to describe four key components of RRI: **Anticipate** possible outcomes of research and innovation, **Reflect** on motivations, processes and products, **Engage** with relevant stakeholders and **Act** accordingly to address issues revealed.

The ideas behind RRI and the AREA framework may be easy enough to understand, but they raise significant conceptual and practical questions. Fundamental problems include the fact that research and innovation do not follow linear and predictable patterns. Bunching together research and innovation blurs important boundaries and hides significant differences. Pluralistic democracies usually do not have a substantive consensus on what counts as acceptable and desirable. Stakeholder engagement can be misused for specific aims. The idea of RRI itself contains specific values and implementing it may engender power struggles.

Most participants in the RRI discourse are well aware of these issues.[14] It is thus important to understand that RRI is not an attempt to invent a new top-down way of governing research and innovation, but rather is a way of linking and embedding existing principles and activities with a view to broadening their reach and relevance. This means that RRI encompasses existing work such as participatory design, research ethics and professional codes and aims to ensure that they can develop synergies. This also includes building on extant research into corporate ICT governance. More precisely, RRI may be understood as a demand for multi-level ethics (systemic and institutional 'macro ethics' in addition to individualistic 'micro ethics'), the engagement of a broader variety of stakeholders and the inclusion of social, political and ethical issues in ICT governance.[7] It remains problematic though how these ideas can be put into practice-.

Embedding RRI in ICT

The challenges for embedding RRI into ICT innovation are extremely complex. First, we need to understand how ICT researchers and practitioners currently manage their professional

responsibilities as well as how they perceive the notion of RRI in order to assess how to move forward and 'fit' features of RRI to researchers' perceptions and expectations. A significant challenge lies in developing a set of practical actions within an RRI framework that may be adopted by the ICT community and how such an approach might be embedded and deployed within current organizational processes. In order to understand these issues, we conducted investigations into the ways that RRI concepts, tools and processes might be shaped to become a creative resource for innovation in ICT. Our work was part of the project 'Framework for Responsible Research and Innovation in ICT' (FRRIICT) funded by the EPSRC.

The ICT community landscape

We interviewed leading computer scientists, postdoctoral, researchers and PhD students as well as EPSRC portfolio managers and representatives of professional bodies in the UK.[3] The study provides the first extensive summary of current positions regarding the boundaries of professional responsibility and the identification of potential long-term societal consequences of ICTs. It is an important baseline giving us an opportunity to describe, understand and triangulate ICT researchers' and other stakeholders' issues and concerns across a variety of computer science domains including; mobile computing, artificial intelligence, photonics, and signal processing, to name a few.

Many researchers welcome enhancements to current governance processes such as, through the introduction of framing questions that help in reflecting on research outputs. Also, some embrace the further integration of social and ethical research into design and development. Apart from such perceived opportunities for RRI, many interviewees raised various concerns. Though many significant issues emerged, we outline five key concerns discussed by participants. Together these concerns raise problems that typically arise when integrating RRI into ICT. We therefore sought to relate these concerns to concepts and approaches that would allow specifying RRI in ICT.

The first recurring issue is the difficulty of predicting the potential uses of research outcomes. Some researchers say it may be inappropriate to attempt to predict future impacts in the context of ICT research because the uncertainties tend to be social rather than scientific, meaning technologies are socially shaped and not fixed. Researchers cite two unknown factors related to prediction. First in fundamental research, risks and uncertainties are identifiable only within the *contexts of their use.* Second, in application-oriented research, industry and user adaptation can change the *trajectory of ICTs* in unforeseen ways. This very open nature of ICT, its logical malleability,[12] interpretive flexibility[2] and the social production of technology make it even more difficult to predict outcomes of research and innovation than in other areas of science and technology research. We refer to these issues as related to the 'product' of ICT research and innovation.

A second issue emerging from the study points to the perceived differences between ascertaining risks and uncertainties in Computer Science to that in the Physical and Life Sciences. For example, researchers discussed what we refer to as the 'rhythm of ICT' where outputs may occur at a quicker pace than in the physical sciences. Software may be developed, released and 'go viral' potentially in the same day with little, if any, oversight, and can have far reaching effects on people's activities and societal structures. These issues relate to the 'process' of research and innovation.

A further distinguishing feature typical of ICT is what Johnson calls 'the problem of many hands'.[11]

This refers to the organizational and institutional reliance on a division of labour where most activities are split up between numerous different individuals. The problem will be increased beyond organisational boundaries by open source projects. Also, different disciplinary languages remain important, which makes interdisciplinary work important but hard to achieve in practice. Thus, ascribing accountability for eventual consequences is made difficult. These aspects point to the importance of considering 'people' in RRI in ICT.

A final issue concerns the notion of 'convergence'[9] where the increasingly pervasive nature of technologies in the age of the Internet, web 2.0 and pervasive computing, means that demarcating clear boundaries between systems, features and functionality becomes increasingly problematic. This means that it becomes increasingly difficult to discern the 'purpose' of ICT research and innovation.

In combination, these concerns pose significant challenges to RRI in ICT that may go beyond those in other fields. We therefore developed the 4 Ps outlined above (product, process, people and purpose) as well as other concepts and approaches to be explained next, to develop a framework for RRI that is specific to ICT.

Towards AREA Plus: 'Talking back' and specifying RRI with the voices of ICT researchers

The AREA acronym points to general points of interest of RRI, but more detail is needed for ICT research. The discussion so far has clearly shown that RRI in ICT cannot be realised in a prescriptive manner. The nuances of acceptability and desirability, competing interests and their embedding in social, economic and political structures mean that many aspects of ICTs will remain contested. RRI therefore cannot aim to establish overall definitions of what counts as responsible but needs to be understood as a contextual process that enables the development of sensitivities towards relevant issues and a willingness of various stakeholders to engage with one another, to become responsive to mutual needs.

Thus, we reconceptualise RRI for ICT as an ongoing cultural dialogue in which different voices from within the HCI community talk back to RRI proponents, in order to find ways of translating back and forth which forms of responsible ICT design and development might already be in place, in the making, or still to be developed. This approach is akin to the view put forward by Strand et al. (2015)[20] who developed a set of indicators for the European Commission that could be used for monitoring RRI across different disciplines, research themes and projects. Whilst proposing a comprehensive list of indicators, Strand et al. also suggest that *any* indicator set would ultimately need to be (re)developed in a given research or application context. Thus, our framework is, in a sense, self-critical by design: it is deliberately meant to be continuously questioned and adjusted.

We shall exemplify what such a dynamic and context-sensitive framework for responsible behaviour may include in the case of ICT. Given the lack of space, we focus on interviewees' comments on the difficulties of predicting the trajectories of ICT. While we regard this as an appropriate scepticism to be voiced in the overall RRI discourse, under 'anticipation' we also suggest different approaches such as, a collaborative quest for future solutions informed by experiences in the present. Crucially, this alternative view profits from existing ICT research. In other words, ICT researchers actually have a lot to *add* to the RRI discourse to make it more context-specific and useful. Reeves' (2012) analysis of 'envisioning' techniques is a case in point.[16] He makes clear that the social shaping of technologies is at the heart of computer science, not

external to it, as suggested by some interviewees. Visions, utopia, predictions, promises and hype have been produced for decades. Importantly though, much of this envisioning has been done rather unconsciously, thus shaping the trajectories of ICT in ways that shut down alternative paths. So there are implicit powers at play. Narratives, teleology and technological determinism proliferate, but are not sufficiently reflected.

In practical terms, the framework draws on such existing approaches to ICT and provides a variety of scaffolding questions. Each cell of the framework expands into deeper questions, suggesting literature, more detailed discussion and problematisation. For instance, after scanning the framework as a whole (figure 1) a researcher may want to consider to what extent impacts may be anticipated (figures 2 and 3). Various links provide questions for exploring different possible pathways, a more comprehensive line of reasoning and references.

	Process (rhythm of ICT)	**Product** (logical malleability & interpretive flexibility)	**Purpose** (convergence & pervasiveness)	**People** (problem of many hands)
Anticipate	Is the planned research methodology acceptable?	To what extent are we able to anticipate the final product, future uses and impacts? Will the products be socially desirable? How sustainable are the outcomes?	Why should this research be undertaken?	Have we included the right stakeholders?
Reflect	Which mechanisms are used to reflect on process? How could you do it differently?	How do you know what the consequences might be? What might be the potential use? What don't we know about? How can we ensure societal desirability? How could you do it differently?	Is the research controversial? How could you do it differently?	Who is affected? How could you do it differently?
Engage	How to engage a wide group of stakeholders?	What are viewpoints of a wide group of stakeholders?	Is the research agenda acceptable?	Who prioritises research? For whom is the research done?
Act	How can your research structure become flexible? What training is required? What infrastructure is required?	What needs to be done to ensure social desirability? What training is required? What infrastructure is required?	How do we ensure that the implied future is desirable? What training is required? What infrastructure is required?	Who matters? What training is required? What infrastructure is required?

Figure 1: The AREA Plus Framework

	Process (rhythm of ICT)	Product (logical malleability & interpretive flexibility)
Anticipate	Is the planned research methodology acceptable?	To what extent are we able to anticipate the final product, future uses and impacts? Will the products be socially desirable? How sustainable are the outcomes?

Figure 2: Selecting anticipation

TO WHAT EXTENT ARE WE ABLE TO ANTICIPATE THE FINAL PRODUCT, FUTURE USES AND IMPACTS?

The future cannot be predicted, but there is room for exploring different possible pathways. Also, researchers and other stakeholders can build on already existing formal and informal practices of anticipation in the ICT community.

EXPLORING DIFFERENT POSSIBLE PATHWAYS

- Who might be the intended audience(s) of the envisioned product?
- What is the context the envisioned product is meant to address? And what is the context in which this anticipation process itself is taking place?
- What issues of the *present* does the anticipation process target, or could target?
- What can we learn from earlier (historical) anticipation processes?
- In using a particular envisioning, what pathways might we be shutting down as possibilities, which endpoints might be excluded, which present issues are excluded?

(Scaffolding questions adopted and adapted from Reeves (2012))

ENVISIONING IN ICT

Reeves, S. Envisioning Ubiquitous Computing. In *Proceedings of the 30th Annual ACM Conference on Human Factors in Computing Systems* (Austin, Texas, USA, May 5-10). ACM Press, New York, 2012, 1573-1582.

It is hard to predict the trajectory of ICT innovations, including outcomes, future uses and impacts. However, ICT is a domain in which visions, utopia, predictions, promises and hype have been produced for decades. Importantly though, much of this envisioning has been done rather unconsciously, thus shaping the trajectories of ICT in ways that shut down alternative paths. There are implicit powers at play: narratives, teleology and technological determinism proliferate but are not sufficiently reflected.

Figure 3: Unpacking anticipation

The framework is meant to be adapted to the context that researchers and other stakeholders find themselves in. Thus, the idea is to productively 'open up', not 'close down' expert discourses.[19] At the same time, we do not question 'closure' *per se*. Any design and development process requires taking countless decisions, and realising these in soft and hardware solutions, at multiple points in time. However, closures may still leave room for diversity.[19]

In sum, certain forms of productive self-criticism already exist in ICT research and could be cultivated further under an extended AREA Plus framework. In this sense, EPSRC's original AREA principles are only a starting point for the reinvigoration and possibly extension of a much more

nuanced discourse with and within ICT research.

Future of the AREA Plus Framework

The framework that we have started to develop in the spirit explained and exemplified previously is not a panacea, and it cannot do miracles. Many of the questions of relevance are related to fundamentally opposing interests and socially and politically contested issues. Such conflicts will not disappear overnight. However, the evolving framework may allow individuals involved in them to better understand their own and others' positions and to contribute to better informed debate and higher quality policies and decisions.

In order to achieve this and maintain this progress, much remains to be done. The framework needs to be supported by substantive tools and specific guidance on particular topics, issues and technologies. We have developed a resource to provide these (www.responsible-innovation.org.uk) but this is only a starting point. Below we identify issues that are crucial to the successful further development and adoption of our framework.

Firstly, embedding RRI activities needs to be perceived by researchers as something that is *achievable*. As we explained, 'anticipation' becomes significantly less mysterious when realistically scoped and grounded in concrete practices, including specific envisioning techniques and questions. Implementing RRI is about finding ways to instantiate concrete achievable practices and not about unattainable ideals of 'perfect' foresight or 'risk-free' innovation. Also, RRI for ICT may require developing new initiatives that are likely to depend on more fine-grained case studies that go beyond the scope of this paper.

In addition an *integrated approach* is needed for successful adoption of the framework. RRI has to be sensitive to the relationships between researchers, practitioners and the hierarchies and organisational structures within which they are situated. Responsibilities need to be appropriately apportioned across the entire ecology of organisations that together deliver research and innovation.[8] Taking RRI seriously as a strategic concern would permit practices of anticipation, reflection, and engagement to occur in the formation of new research programmes by funding councils, and in the final stages of commercialisation at the academic / commercial interfaces. In between these poles it would recognise the role of funding councils, professional bodies and others in sustaining RRI practices within research teams by providing appropriate support, services and guidance. Thus, responsible behaviour becomes a collective, uncertain and unpredictable activity which is less about accountability and liability and more about care and responsiveness.[18]

There is evidence that these developments are under way. Awareness of RRI is starting to develop in academia and industry. There are many good reasons for this. Maybe the best one, and a good conclusion for this paper, is that RRI, while largely conceived as a risk management approach has a much more positive aspect to it. By incorporating active considerations of the future into design, engaging with stakeholders, reflecting on process, product and purpose and putting people at the centre of research and innovation, RRI may well provide inspiration and become a unique source of innovation and creativity.

Acknowledgements

We would like to thank the reviewers for their thoughtful comments. Part of the research underlying this article has been funded by the EPSRC.

References

1. Azuma, R. Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B. Recent advances in augmented reality. *IEEE Computer Graphics and Applications 21*, 6 (November 2001), 34-47.
2. Doherty, N. F., Coombs, C. R., and Loan-Clarke, J. A re-conceptualization of the interpretive flexibility of information technologies: redressing the balance between the social and the technical. *European Journal of Information Systems 15*, 6 (December 2006), 569-582.
3. Eden, G., Jirotka, M., Stahl, B.. Responsible research and innovation: Critical reflection into the potential social consequences of ICT. In *Proceedings of the Seventh IEEE International Conference on Research Challenges in Information Science* (Paris, France, May 29-31). IEEE, 2013, 1-12.
4. Ehn, P. Work-Oriented Design of Computer Artifacts. Lawrence Erlbaum Associates, 1990.
5. Fisher, E., and Rip, A. Responsible innovation: Multi-level dynamics and soft intervention practices. In *Responsible Innovation*, R. Owen, J. Bessant and M. Heintz, Eds. Wiley, 2013, 165-183.
6. Floridi, L. (Ed.). *The Cambridge Handbook of Information and Computer Ethics*. Cambridge University Press, 2010.
7. Gotterbarn, D. ICT governance and what to do about the toothless tiger(s): professional organizations and codes of ethics. *Australasian Journal of Information Systems 16*, 1 (November 2009), 165-184.
8. Grimpe, B., Hartswood, M., and Jirotka, M. Towards a closer dialogue between policy and practice: responsible design in HCI. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto, Canada, April 26-May 01). ACM Press, New York, 2014, 2965-2974.
9. Grunwald, A. Converging technologies: Visions, increased contingencies of the conditio humana, and search for orientation. *Futures 39*, 4 (May 2007), 380-392.
10. Heeks, R. ICT4D 2.0: The next phase of applying ICT for international development," in *Computer 41*, 6 (June 2008), 26-33.
11. Johnson, D.G. *Computer Ethics*, 3rd edition. Prentice Hall, 2012.
12. Moor, J. What is computer ethics? *Metaphilosophy 16*, 4 (October 1985), 266-275.
13. Muller, M.J., and Kuhn, S. Participatory design. *Communications of the ACM 36*, 6 (June 1993), 24-28.
14. Owen, R., Heintz, M., and Bessant, J., Eds. *Responsible Innovation*. Wiley, 2013.
15. Owen, R. Macnaghten, P. and Stilgoe, J. Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy 39*, 6 (December 2012), 751-760.
16. Reeves, S. Envisioning Ubiquitous Computing. In *Proceedings of the 30th Annual ACM Conference on Human Factors in Computing Systems* (Austin, Texas, USA, May 5-10). ACM Press, New York, 2012, 1573-1582.
17. Sellen, A. Rogers, Y. Harper, R. and Rodden, T. Reflecting human values in the digital age. *Communications of the ACM 52*, 3 (March 2009), 58-66.

18. Stilgoe, J., Owen, R., & Macnaghten, P. Developing a framework for responsible innovation. *Research Policy 42,* 9 (November 2013), 1568-1580.
19. Stirling, A. 'Opening up' and 'Closing down'. Power, participation, and pluralism in the Social Appraisal of Technology. *Science, Technology & Human Values 33*, 2 (March 2008), 262-294.
20. Strand, R. Spaapen, J. Bauer, M. Hogan, E. Revuelta, G. Stagl. S. Indicators for promoting and monitoring Responsible Research and Innovation. Publications Office of the European Union, 2015.
21. Tucker, J.B. and Zilinskas, R.A. The promises and perils of synthetic biology. *The New Atlantis 12* (spring 2006), 25-45.
22. Van den Hoven, J. Value-sensitive design and responsible innovation. In *Responsible Innovation*, R. Owen, J. Bessant and M. Heintz, Eds. Wiley, 2013, 75-83.
23. Wiener, N. *The Human Use of Human Beings*. Da Capo Press, 1954.