11

Responsible Research and Innovation in Information and Communication Technology: Identifying and Engaging with the Ethical Implications of ICTs

Bernd Stahl,[1] *Marina Jirotka,*[2] *and Grace Eden*[2]
[1]*Centre for Computing and Social Responsibility, De Montfort University, Leicester, LE1 9BH UK*
[2]*Oxford e-Research Centre, Oxford University, 7 Keble Road, Oxford OX1 3QG UK*

11.1 Introduction

The concept of "responsible (research and) innovation" (RRI) can refer to a number of separate but interlinked questions: who is responsible? what are they responsible for? why are they responsible? or what are the consequences of these responsibilities? are just some of the more obvious ones. This chapter will discuss some of the most pertinent of these questions with reference to information and communication technologies (ICT). It will do this by describing the context, background and findings of the European research project "Ethical Issues of Emerging ICT Applications" (ETICA) as well as the currently on-going UK research project on a "Framework for Responsible Research and Innovation in Information and Communication technology" (FRRIICT).

The chapter will start with a brief discussion of concepts of responsibility and responsible innovation that will allow the identification of important aspects that a responsible approach to ICT will require. It will then build a FRRIICT by discussing two different approaches to RRI, as represented by the two projects. This will lead to a

Responsible Innovation, First Edition. Edited by Richard Owen, Maggy Heintz and John Bessant.

discussion of further research, as well as policy requirements that need to be addressed in order for research and development in ICT to live up to the expectations of responsibility.

11.2 Conceptualizing Responsibility and Responsible Research and Innovation in ICT

One of the first definitions of RRI is offered by Von Schomberg in Chapter 3 and Von Schomberg (2011, p. 9) who suggests that it can be understood as "*a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)*." For the purposes of this chapter it is worthwhile briefly to unpack the ideas behind RRI in order to better understand how these ideas are related to ICT.

11.2.1 Responsibility as a Social Ascription

Such a conceptual unpacking requires a brief look at the concept of responsibility (see also Chapter 7). The etymological roots of the term come from responding or answering (Lewis, 1972), which is reflected in the French *responsabilité* (Etchegoyen, 1993), as well as in the German *Verantwortung* (Bayertz, 1995). This is a first important indication of the relational nature of the term. Responsibility can be understood as a social construct that establishes relationships between a set of different entities. Primary among them are the subject and the object. The subject is the entity that is held responsible. The object is that which the subject is responsible for. This relationship between a subject and an object is the core of the vast majority, if not all, responsibility ascriptions. Examples of this core responsibility relationship can be the responsibility of parents for their children, of politicians for their policies, of a criminal for her crime, or the responsibility of the storm for the damage it caused. This core relationship needs to be supplemented with other aspects for a responsibility ascription to be workable. Further ingredients are the norms according to which the subject is responsible for the object as well as the mechanism by which this relationship is established. Both norms and mechanisms differ for different types of responsibility, such as moral, legal or role responsibilities. In the case of legal responsibility the law is likely to provide the norm and a court, judge, or jury may attribute it. Professional responsibility requires professional rules and professional bodies to implement these. Depending on the type, there are different types of purpose of responsibility ascriptions. Responsibility can aim at retribution or revenge. For the purposes of this chapter on RRI in ICT we hold that responsibility is a prospective exercise (Chapter 7) that aims to achieve some desirable social state. Retrospective responsibility ascriptions (such as holding a copyright infringer legally responsible for a past infringement and sentencing them to a pay a fine) can also be understood to serve a forward-looking goal, such as re-socializing them into society or ensuring future compliance with copyright legislation.

Responsibility ascriptions are normally meant to have practical consequences. This requires consideration of the conditions for successful ascriptions. These conditions can refer to any of the individual components of an ascription (e.g., subject, object, norms),

or to the mechanism of defining and enforcing consequences or sanctions arising from a responsibility ascription. There is much literature on some of these conditions. A key topic, for example, is the question of who, or what, can count as the subject of responsibility. The traditional response, both in law and in ethics, has been that it should be the individual human being. Conditions that such individual human responsibility subjects are normally assumed to meet include a certain level of knowledge of the situation, freedom to act and influence outcomes, power to affect the object, and a set of personal characteristics such as rationality, intentionality, reason, or self-control. It is immediately obvious that many of these touch on significant philosophical problems, such as freedom of will or freedom of action. In addition, however, there has been significant debate about whether other entities can be considered responsibility subjects. Possible candidates are groups of individuals (French, 1992), corporations (Werhane, 1985), and also technical artifacts (Allen, Smit, and Wallach, 2005; Floridi and Sanders, 2004). Depending on who or what is considered a legitimate subject of responsibility ascription, the other components will vary.

It is easy to find examples of responsibility relationships that relate to ICTs. There are several big issues that are widely discussed and have led to significant regulation and legislation. The two primary examples here are privacy/data protection and intellectual property. The responsibilities of different actors in this area (for example, the responsibility of social network sites such as Facebook to safeguard users' privacy, or the ownership of biometric data by governments or companies) are highly contested. Other examples of responsibility in ICT refer to the role of software developers in assuring quality and functionality, the responsibility of companies for usability, or the responsibility of regulators for appropriate use of data.

These are currently contested issues in existing technologies, which can be exacerbated with regards to novel technologies that are currently being researched and developed. Looking at von Schomberg's earlier definition, one can now see that he touches on some aspects, such as possible subjects (innovators and societal actors), some of the aims to be achieved (sustainability, acceptability, desirability of products), but leaves others open (exact nature of the object, conditions, mechanisms of ascription). This is not surprising, partly because responsibility ascriptions are always an evolving network where novel circumstances require redefinitions and adjustments. The present chapter contributes to the process of such redefinitions.

11.2.2 Responsible Research and Innovation as Meta-Responsibility

As the previous section has shown, responsibility relationships that touch on ICT as subject, object, or in some other form are nothing new. There are already large amounts of definitions of responsibility in the area (Stahl, 2004). These may be collective, individual, professional, role, moral, legal, and other responsibilities. Many of these responsibilities are interlinked and mutually constitutive. A software engineer, for example, may have a professional responsibility to develop a system for a customer on time and to a specified quality. She will have legal responsibility to comply with existing regulation, for example, data protection or copyright law. She may feel a moral responsibility to ensure that users are treated fairly in the system and a responsibility toward her peer software engineers to ensure transparency of her code. She can have a role responsibility as team leader to ensure responsible behavior by her team members, which may put upon her a

responsibility to define responsibilities of others to meet shared goals. It may, therefore, be better to speak of a web of inextricably interlinking responsibilities.

Understanding this web of responsibilities is not trivial, as many of these responsibilities will take different forms and will have different degrees of visibility and external validity. The different responsibilities may be mutually supportive, but they may also be contradictory. For example, our software engineer has been asked to code a piece of software that may infringe a user's privacy. There is a dilemma here in terms of what her responsibility may be to bring this issue into discussion. What if senior-level managers are aware of the breach? The conflict occurs when someone needs to be held accountable for acting upon the implementation of such code. The negotiation and arbitration between the different types of responsibilities may be explicit, but they may also be tacit and implicit.

RRI as a relatively novel concept has to incorporate the fact that it enters a well-established playing field. Responsibility in research and innovation is nothing new and, as outlined above using the example of ICT, is already expressed in a web of responsibilities. It is unlikely that RRI will make a difference and affect science and technology research and innovation in a sustainable way if it is turned into one more responsibility ascription or relationship in an already well-populated field. Instead, we suggest, RRI should be conceptualized as a higher level responsibility or a meta-responsibility. This meta-responsibility is *a responsibility for responsibilities*. RRI can aim to align responsibilities, to ensure they move in a particular way. RRI can define socially desirable consequences that existing responsibilities can work toward and develop responsibility relationships that ensure that the achievement of such desired aims is possible.

In this chapter we explore this idea of RRI as a meta-responsibility in the field of ICT. Before going into detail it is helpful to have a brief look at the current state of discussion of RRI more broadly. This will allow clarification of which areas are in need of further understanding and thus which contributions to knowledge are required. On this basis we explore two pieces of research on RRI in ICT, which allow us to return to the content of RRI in ICT and beyond.

11.2.3 Responsible Research and Innovation: the Four "P"s

The reason why RRI has recently gained currency is due, to a large extent, to the real, or potential impact that a number of technical innovations can have, or are perceived to have on our lives. In the area of ICT this can, for example, relate to the ability to communicate in novel ways, for example, through social media, or the potential of large scale ubiquitous monitoring or surveillance through automatic sensor systems. Innovation can be understood as the process of making some type of potential phenomena useful for human lives through "creation of better or more effective products, processes, services, technologies, or ideas that are readily available to markets, governments, and society."[1] This raises the question of when such an innovation can be considered responsible. A look at current work on RRI in other fields shows that it is a concept that is meant to mediate the consequences of technical and other innovations on our individual and social lives. While innovation is associated with social benefit and economic growth, understanding its implications now takes on a new sense of urgency because technical and scientific

[1] See Wikipedia, available http://en.wikipedia.org/wiki/Innovation (accessed 10 July 2012).

innovations are happening on a global scale and at increasing speed. At the same time the social authorities who regulate innovation work or its evaluation are no longer clear (see Chapter 8 for further discussion). In a globalized world it is not obvious what is perceived to be good or on what grounds such normative evaluations can be made. This adds to the complexities of developing approaches to incorporating RRI concepts into current governance frameworks.

In this context of the globalization of research and the fragmentation of research governance, the questions loom large regarding what is considered to be desirable, desirable to whom, for what purposes, and how can we use technology to achieve it. Since we have no agreed substantive view of what is good, right, and desirable, RRI can be understood as an attempt to give a procedural answer to the question of how to deal with the uncertainties around innovation. One way of characterizing these processes is to use an alliteration that allows us to keep track of some of the core features of RRI in ICT, namely the four "p"s, which are: product, process, purpose and people. The purpose of using the four "p"s is to draw attention to the fact that, in addition to the widely recognized importance of both product and process of technical development, the purpose of the development needs to be considered and people involved in the innovation need to be incorporated in RRI (see Chapter 2 for further discussion of products and purposes and the importance of inclusive deliberation). In this chapter we will not be able to discuss all of these comprehensively, but will give an account of some of them in some more depth, in so far as they are relevant to ICT.

In the light of this complexity of responsibilities, their interlinking aspects and dimensions, the current chapter will draw a picture of RRI in ICT. Returning to the four "p"s, we start with a look at aspects of the product.

11.3 Building a Framework for RRI in ICT

Having briefly outlined some of the conceptual foundations of RRI (e.g., the concepts of responsibility in relation to ICT innovation, the idea of RRI as meta-responsibility and the four "p"s), we now provide an outline of two research projects which aim to explore the specifics of RRI in ICT. The first, the EU project ETICA explores ethical issues of emerging ICTs, whereas the second, the UK Engineering and Physical Sciences Research Council (EPSRC) project on a Framework for Responsible Research and Innovation looks at the current landscape of RRI within the UK ICT research community.

11.3.1 Product: ICTs and Their Ethical Implications

One way of approaching the question of RRI in ICT is to explore the ethical issues that future and emerging ICTs are likely to raise. This is what the ETICA project set out to do. It was funded under the European Union's Seventh Framework Programme and ran from April 2009 to May 2011. The idea of ETICA was to explore the ethics of emerging ICTs by following these four main aims:

1. Identify emerging ICTs.
2. Identify ethical issues likely to be raised by those ICTs.
3. Evaluate and rank these issues.
4. Provide recommendations on appropriate governance structures to address these.

This section will briefly highlight the main findings of each of the first three steps and will then turn to the question of governance below.[2]

11.3.1.1 Identification of Emerging ICTs

The ETICA project was predicated on the idea that, in order to facilitate a proactive approach to the ethics of novel ICTs, policy makers, researchers, research institutions, funders, and other actors in innovation processes need to gain a sound understanding of which technologies are likely to be relevant, and what their capabilities and constraints will be. While an exact knowledge of the future is elusive, as predictions cannot be certain, one can study current work in research and development to gain a sound understanding of emerging technologies. This follows the principles of technology foresight research, which aims to explore possible futures in order to allow the making of decisions that are likely to promote desirable states (Cuhls, 2003).

The methodology employed to identify emerging ICTs during the ETICA project was a structured discourse analysis of documents containing visions of future technologies. Two types of documents were analyzed: (i) high-level governmental and international policy and funding documents, and (ii) documents published by research institutions.

During data analysis more than 100 technologies, 70 application examples and 40 artifacts were identified. These were synthesized into the following list of emerging ICTs. Note that the term "emerging ICT" is used for any high-level socio-technical system that has the potential to significantly affect the way humans interact with the world. A good example of such as system would be "ambient intelligence" which can be understood as a vision of a generalized, embedded and ubiquitous technology deployment in the human environment to further a broad range of human purposes (Sadri, 2011). Such systems are deemed to be emerging if they are likely to be socially and economically relevant in the coming 10–15 years. The ETICA project highlights the following 11 ICTs.

- Affective Computing
  (use of computing to measure or express human emotions)
- Ambient Intelligence
  (ubiquitous and pervasive computing environment)
- Artificial Intelligence
  (representation of intelligence through artifacts)
- Bioelectronics
  (combination of biological materials or principles with electronics)
- Cloud Computing
  (remote shared computing services)
- Future Internet
  (novel technical infrastructure for networked services)
- Human-machine Symbiosis
  (direct combination of humans and artifacts)

[2] This chapter does not provide the space to discuss any of the issues discussed here in depth. However, all project deliverables and findings are available in full from the project web site (www.etica-project.eu). A more detailed account of the overall project can be found in Stahl (2011a).

- Neuroelectronics
  (link between computing and neurosciences)
- Quantum Computing
  (utilization of quantum effects for computing purposes)
- Robotics
  (embodied artificial agents, typically somewhat autonomous)
- Virtual/Augmented Reality
  (representation of reality through technical means).

For each of these technologies a detailed description was developed that indicated their definition, history, likely uses and constraints. For the purposes of this chapter it will be sufficient to give a brief overview of some of the shared features these technologies are expected to display. These include:

- Natural interaction
  (use of ICTs becomes similar to interaction with humans)
- Invisibility
  (as a result of miniaturization or embedding of artifacts)
- Direct link
  (either physical implant, direct contact (e.g., wearable ICT) or through sensors)
- Detailed understanding of the user
  (emerging ICTs will require a highly detailed model of the user to fulfill their tasks)
- Pervasiveness
  (ubiquitous embedding of ICTs in the human and natural environment)
- Autonomy
  (ability of ICT to act without direct user input)
- Power over the user
  (ability to structure the space of action of the user)
- Market driven
  (allocation decisions are reached through commercial exchange mechanisms).

For the purposes of the ETICA project, as well as this chapter, it is not crucial to know whether all of these aspects will materialize. Instead, the interesting question is whether these predicted futures allow us to come to a better understanding of what could be done now in order to make sure that the ethical and social consequences of the technologies are beneficial. Or, to put it differently, the question is whether we find ways of dealing responsibly with the technologies and their expected consequences.

11.3.1.2 Ethics of Emerging ICTs

The above list of shared features of the technologies gives an indication of possible ethical issues. During the ETICA project the consortium analyzed each of the technologies and explored the literature to see which ethical issues were predicted to be raised by them. Again, all this chapter can do here is to give a high-level overview of the approach and of those ethical consequences of the emerging technologies that were found in the literature.

In order to identify likely ethical issues of emerging ICTs, a literature analysis of the ICT ethics literature from 2003 onwards was undertaken. This started out with a novel bibliometric approach that mapped the proximity of different concepts in the ICT ethics

literature.[3] Using this bibliometric analysis as a starting point, a comprehensive analysis of the ICT ethics literature was undertaken for each technology.

The ethical analysis showed that there are numerous ethical issues that are discussed with regard to the technologies. The number and detail of these ethical issues varies greatly. This variation is caused by the differing levels of detail and length of time of discussion of the technologies. Ethical issues shared by several of the emerging ICTs can roughly be divided into issues that are already being researched and regulated, and those issues that are currently less tangible and further removed from current debates. The widely discussed ethical issues of emerging ICTs include *privacy, security, trust, liabilities, and digital divides*. Some of these were represented on the bibliometric map, others were not visible because this particular representation of the bibliometric data did not display them. Being widely discussed does not imply that they will remain unchanged. In fact, many of these issues will take new forms. Privacy, for example, is a well discussed issue that has led to a wealth of literature, and is even recognized as a human right (Humphreys, 2011). There are numerous cases in which technical systems and their affordances have created novel types of privacy issues. In the UK, for example, one could observe the government's attempt to collect genetic data in a national DNA database or to create a national ID system, both of which were met with strong opposition. Similar cases from the private sector have been discussed in depth, for example, with regards to Google's right to collect data for the Streetview application, the appropriateness of contextual advertisement, and many more. In the light of current technical developments, privacy is likely to become even more prominent due to new quantities of data, new ways of linking it, and maybe even new types of data (e.g., emotion-related data) which may raise qualitatively new issues. What these issues have in common, however, is that they are currently on the social and political agenda and are recognized as important and in need of further attention.

Less predictable ethical issues arising from the emerging ICTs tend to be centered on difficult conceptual issues, such as human identity, the relationship between humans and technologies, and relationships among individuals or groups. Individual identities may change due to the way we interact with technology. What we perceive to be normal may partly be a function of the affordances (Gibson, 1977; Norman, 1999) that technologies offer us and, therefore, is potentially subject to change. A good example of this is the difficult distinction between therapy and human enhancement (Bostrom and Sandberg, 2009; Coeckelbergh, 2011). Technical interventions that can be used to treat diseases can, in many cases, be used to improve human performance. Drugs that were developed to stop memory loss, for example, may help healthy university students improve their exam performance. Related issues can arise due to the way in which new technologies can change power relationships and traditional balances between individuals and groups, and this is the very area where prediction of future uses and likely outcomes may be impossible. Such changes can affect not only local cultures and collective self-views, but the ways in which societies might be organized. This does not have to be morally problematic, but it may well be in some respects. The scale of the change means that potential ethical issues need to be monitored closely.

[3] A detailed account of the methodology used in the ethical analysis can be found in Heersmink *et al*. (2011).

11.3.1.3 Evaluation of Emerging ICTs

The evaluation of emerging ICTs and their ethical issues was undertaken from four different perspectives:[4]

- **Law**:
  The analysis was based on the principles of human dignity, equality, and the rule of law. A review of 182 EU legal documents revealed that the legal implications of emerging technologies were not adequately reflected, that is, did not relate to the specific issues the technologies raise.
- **(Institutional) ethics**:
  The earlier ethical analysis was complemented by looking at opinions and publications of European and national ethics panels or review bodies, such as the European Group on Ethics in Science and New Technologies, the Nuffield Council on Bioethics, the German Ethics Council, to name but a few. Furthermore, the review covered the implied normative basis of technology ethics in the EU.
- **Gender**:
  A review of the gender and technology literature showed that, in the case of five technologies, such gender implications had already been raised in the literature. However, for the majority of the technologies no research on their gender implications had been undertaken.
- **Technology assessment**:
  This analysis asked how far developed are ICTs and what are their prospects of realization. The expected benefits and possible side effects were discussed, as well as the likelihood of controversy arising from the different technologies. The technology assessment perspective included the other assessments and aimed to provide a general insight into the consequences of the technologies.

The evaluation found that several of the technologies are so closely related that they should be treated in conjunction. Building on the criteria of the likelihood of coming into existence and raising ethical debate, the following ranking was suggested:

1. Ambient Intelligence
2. Augmented and Virtual Reality
3. Future Internet
4. Robotics, Artificial Intelligence and Affective Computing
5. Neuroelectronics, Bioelectronics, and Human-Machine Symbiosis
6. Cloud Computing
7. Quantum Computing.

This ranking allows the prioritization of activities and policies. Overall, the ETICA project addressed the product dimension of RRI in ICT by identifying those technologies which are likely to require being dealt with responsibly, and which possible ethical and social consequences they may have that need to be taken into account. While such an ethically sensitive foresight analysis is important to feed into the process of RRI, it

[4] A full discussion of these steps of evaluation, the methodologies used and the findings can be found in deliverable D3.2 "Evaluation Report" which is available from the deliverables section of the ETICA project web site at www.etica-project.eu.

is by no means enough to lead to better technologies or consequences. One further central aspect that needs to be understood is how those individuals who actually create the technologies – the engineers, researchers, and developers – see the technologies, their consequences and the responsibilities that they or others have in ensuring that the technologies are socially beneficial. In order to assess this question in more depth, the chapter now discusses the fundamental ideas and preliminary results of another project, which will highlight the aspect of people in RRI in ICT.

11.3.2 People: Landscape of ICT Ethics

In the UK, the "FRRIICT" project began in September 2011 and is conducting the first comprehensive investigation of ICT researchers' notions of "RRI" and the role this might play in the next decade of ICT research. Responsible research invites researchers, policy makers, and funders to reconsider the influences of near- and long-term research strategy on the shaping of funding priorities, taking into account the potential consequences to social, environmental, health, and behavioral well-being. The "RRI" programme creates an opportunity for reflection, where decisions about research goals are made not exclusively on the grounds of their technical or scientific attributes; so that, in addition to addressing technical grand challenges, RRI asks all stakeholders to consider the potential impacts, risks, and uncertainties of research outputs to wider society.

As part of the FRRIICT project, a landscape study is being conducted to understand current perceptions and attitudes toward RRI, using semi-structured interviews with researchers – those who seek funding – and with portfolio managers – those who manage funding calls, within the EPSRC ICT portfolio. The goal of the interviews is twofold; in the case of researchers we are interested in how professional responsibility is conceptualized within specialist domains, and how ethical issues might be identified, discussed, and resolved as research projects unfold. In the case of the portfolio managers we are interested in the processes and procedures that motivate the identification and direction of long-term funding strategy (e.g., horizon scanning and researcher-led workshops) which, in turn, influence the development of themed funding calls. In addition to these two groups, we are including other stakeholders, such as industry, professional organizations and charities, in our discussions around the potential future societal consequences of ICT research and how RRI may be embedded into current processes.

With a focus on what is commonly referred to as *upstream activities* (Fisher, Mahajan, and Mitcham, 2006), encapsulated in research policy and strategic direction, and *midstream activities* (where bench work and R&D activities take place: see Chapter 9), we can begin to create a picture of the current ICT research and innovation cycle with an aim of providing a set of recommendations and good practice that could be adopted by the EPSRC and the ICT research community. This may provide an opportunity for reshaping research policy and funding governance to include a new conception of responsibility that extends beyond conventional, although important, notions associated with ensuring "experimental reproducibility" when working in the laboratory and with following research process guidelines (such as gaining "informed consent" when involving human participants, or sufficiently protecting data). We present brief preliminary findings below and discuss foreseeable challenges to implementing RRI in actual practice, as well as the

potential for RRI to be implemented as a resource for creativity rather than a constraint for innovation, as underlined by Von Schomberg in Chapter 3.

In the last quarter of 2011 interviews were undertaken with 16 ICT researchers, ranging from professorial to post-doctoral positions, and across domains from cybersecurity, artificial intelligence, robotics, eHealth, natural language processing to human–computer interaction. Additionally, seven EPSRC portfolio managers who facilitate research funding in these and other areas, as diverse as theoretical computer science, visualization, and photonics, were interviewed in early 2012. In the interviews we aimed to find out how both funders and researchers frame the relationship between scientific research, professional responsibility and research outputs' potential impacts on society.

A noteworthy distinction has consistently appeared across both funder and researcher interviews that may underpin general perceptions of professional and social responsibility. For many researchers there is a correlation between the location of their research (within the spectrum of "fundamental" to "applied" (Calvert, 2006)) and how, or even *if* they should be concerned with the potential societal consequences of their research outputs. The proximity of research outputs to uptake and use by potential user communities may be considered as a commonsense understanding of the ways in which research may have a direct influence on society; the closer it is to being taken up by, or embedded in society, the greater the societal impact is likely to be. However, it can be argued that even blue skies research has societal consequences, even though these may materialize in the longer-term. The EPSRC funds at a Technology Readiness Level (TRL) of between 1 and 3, which is defined as the lowest level of "technology readiness." This is where basic principles are identified, technological concepts formulated, and experimental research conducted to validate those principles and concepts. At TRL 1–3, fundamental scientific research is conducted. However, many projects within the ICT portfolio fall within a "gray area" somewhere between fundamental and applied research. Most researchers we interviewed described their projects as a mix of fundamental and applied. The distinction has been articulated as that between the *prototype*, a working system that could, quite soon, be iterated into a fully-developed product, and the *demonstrator*, whose purpose is to show the utility of the science, with its potential commercialization many years into the future. In both cases fundamental research, which may focus on algorithm and infrastructure design, is conducted alongside software engineering and human interface development. The potential for research outputs having either long or short-term utilization leads to a series of often tacit presumptions for many of our interviewees about their professional responsibility, social obligations, and whether or not they should consider how their research outputs may lead to either positive or negative consequences within wider civil society.

Within the spectrum of "fundamental to applied," we find a further distinction is made by researchers, many of whom identify their own work as being either "generic" or "application-oriented." Both types of research address highly technical, scientific challenges through experimental design and intensive laboratory work. However, those who frame their work as generic, on the whole, conduct their research at arms-length to any potential user community; whilst those who describe their research as application-oriented work in close partnership with some designated user community for the duration of a research project. In this way, the potential *context of use* of research outputs is

drawn upon as a justification allowing researchers to distance themselves from any kind of professional responsibility related to societal consequences, risks, and uncertainties.[5]

> The research that we do doesn't inherently tell you how to use this research. So, we built a robot that can throw balls 6 meters compared to 1 meter from previous versions. So, there is no kind of moral value associated with that, it's more of a performance. We're focused on the performance. How you use that performance is not our part I think (R-05).

The view that fundamental research is somehow independent of the concerns of contemporary society is found mostly when research outputs contribute to the development of "enabling technologies," which can be described as technologies that could be adopted across a wide variety of domains (e.g., healthcare, military, industrial or domestic settings) and for any number of uses. In contrast, for those who conduct application-oriented research, the potential social consequences of their outputs might be more obvious to them.

> You have this machine in your house 24/7 then it will potentially collect a lot of data about you. Video data, audio data it tracks what you are doing and it monitors what you are doing. It tracks and monitors whose coming into the house, the phone calls you make and potentially other data. So the question is; how do people perceive that? Is that a problem? If yes, in what way? And how can we design systems to address these issues and to give people the feeling that you take them seriously (R-03).

Even though researchers differentiated between these two orientations (generic and application-oriented), they nevertheless stated that in all cases potential negative consequences, risks, and uncertainties should be obvious and unmistakable before any decisions to change the direction of an ongoing research path (or halt it completely) are made. For many, the most obvious issues center around personal data; its security, confidentiality, and personal privacy. Outside of the realm of personal data protection and human subject consent many researchers and funders struggle to identify potential negative consequences of ICT research. In research areas where devices are manufactured, only minimal concern has been expressed in relation to environmental factors. One researcher suggested that because they are not *directly* involved in the manufacturing process, those issues are of no concern to their research.

> I don't specify how things get made. I specify what I want on the chip and then the manufacturers are responsible for putting it there by whatever means they use (R-10).

On the whole, ICT researchers take a very strong stance that there should be no oversight of research strategy unless there is an obvious social consequence to projected outcomes. They also tend to share a suspicion of activities aimed at projecting likely outcomes of research and a resistance to the introduction of foresight methodologies for the purposes of predicting possible future consequences, risks, and uncertainties. This is an interesting response, as research funding is prioritized at present through foresight and horizon scanning techniques. However, this resistance could be attributed to researchers being

[5] The displayed text indicates quotes transcribed from participant interviews.

unfamiliar with RRI concepts, as the current focus of such exercises remains largely on the potential positive, transformative effects of technology.

To summarize, in fundamental research, societal consequences are framed as being identifiable only within their *contexts of use*, and in application-oriented research the possibilities of industry and user adaptation may change the *trajectory of a technology* in unforeseen ways.

> Any research can always have negative consequences. We take a lot of precautions in what we are doing but you can never fully predict whether there might be negative consequences or not. It would be completely unrealistic and dishonest if I said; 'oh, there will never be any negative consequences of my research'. You can't say that. Who knows how someone else might use our research. You can never predict (R-03).

Some of the researchers we interviewed have suggested that rather than asking the ICT research community to predict in advance the potential outcomes of their research they could be asked instead to "respond to changes consequentially" as they arise (see Chapter 2 for further discussion of iteration and responsiveness). Others have suggested allowing "sufficient discussion" around the possible positive and negative consequences of research within institutions and amongst colleagues so that academic stakeholders "felt there'd been a broad airing and discussion around what the challenges might be." These types of activities could be developed further within emerging RRI frameworks and may be seen as a "responsive" tailoring of RRI by the ICT researchers themselves, rather than the imposition of a top-down "governance" model. When considering researchers' responses that consequences, risks, and uncertainties need to be obvious, their general distrust of the use of foresight methodologies, and their cautious suggestions for improvements to current processes, we can conclude that there will be many challenges to implementing RRI in actual practice. A more approachable and acceptable RRI framework could facilitate discussion amongst the researchers themselves in the initial phases of research, in an informal and non-binding manner where Chatham House Rule[6] applies, ensuring free and open discussion. RRI also aims to facilitate a step forward "from public engagement to public dialogue" where researchers facilitate workshops where they can engage the public directly in order to understand their concerns and to iteratively feedback on research strategies. These two activities (ICT community reflection amongst colleagues, and engagement with the public in a dialog about the societal consequences of research outputs) appear, at the moment, to be difficult challenges to address with regard to researchers current commitments to the scientific and technical challenges of their domain. We need to take these concerns seriously if RRI is to be embedded into the research process.

We suggest that there are many challenges to developing policy and regulation in an arena where attempting to predict the potential uses of ICT research outputs is a tenuous one, and where there may *not* be obvious risks or harms, as in other areas of research such as geoengineering or nanotechnology. Indeed, there may be significant differences between ascertaining risks and uncertainties in computer science, and those in the physical and life sciences (e.g., nanotechnology, geoengineering and synthetic biology).

[6] When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed.

For example, software and hardware may be released without adherence to health, safety, and other risk assessment procedures, excluding data protection and informed consent discussed previously. In contrast the potential consequences to society might be more obvious in the physical and life sciences because researchers conduct experiments with materials or manipulate the properties of matter, whether they are biological, botanical, or synthetic.

Even so, the ICT research community needs to begin to consider ways in which reflective governance mechanisms and public consultations can extend greater influence upon the direction of long-term funding strategy, such that decisions about research goals are inclusive of the concerns of wider society and take into account their influences on society, environment, health, and behavior. Policy in this area will also need to be harmonized with research strategy on a global level, if this is feasible, perhaps through the newly formed Global Research Council (GRC),[7] in order to mitigate any risks to the UK as a contributor to worldwide scientific outcomes. This could provide an opportunity for GRC member states to remain ahead of the technology innovation curve and act as shapers and facilitators of innovation rather than run the risk of RRI being considered a constrainer of research. While it is open to debate whether such a global strategy would be feasible, it is easy to see that it would be desirable with respect to setting minimum standards and spreading costs equally. The recent funding call for RRI projects by the European Commission (2011) posits that one expected impact is that the "enlarged international network on RRI will help Europe advocate its normative model for RRI and will foster convergence of regional innovation systems at a global level."

11.3.3 Process: Governance of RRI in ICT

Having now discussed both the ETICA and FRRIICT projects, which address different but complementary aspects, one can ask what steps are necessary in order to ensure that new ICTs are being built responsibly. We discuss this question by first going through the recommendations that arose from the ETICA project, and then looking at further research that is required in order to ensure such an RRI approach to ICT.

Current ways of dealing with the ethics of technology, whether they are embedded in European research funding (e.g., FP7) procedures such as the ethics review, or whether they are used outside of European funding, tend to be based on biomedical ethics. Biomedical ethics developed from the Nuremberg trials as a reaction to the Nazi's and others' atrocities committed in the name of scientific research. Formulated in the Declaration of Helsinki (World Medical Association, 2008), these principles continue to be developed. They have proven to be successful in many respects and they can guide action with regard to established ethical issues and legal requirements. However, the applicability of bio-medical ethics to ICT research can be questioned. It is not *a priori* clear that the aim of the technical development is justified. Bio-medical ethics is based on informed consent to research processes and tends to neglect the research product. Particularly within the EU Framework Programme, this approach has led to a tick-box attitude that sees ethics as a legal compliance issue, divorced from application context and broader societal discourses. This position contravenes much ethical thinking and ethical traditions. RRI will need to move beyond this status quo. The ETICA recommendations

[7] www.globalresearchcouncil.org, www.nsf.gov/news/news_videos.jsp?cntn_id=124178&media_id=72475&org=NSF.

on governance and approaches therefore include recommendations to policy makers to provide a political framework conducive to a proactive engagement with ethics and to ICT researchers, organizations, and the civil society concerned with the question of how to affect social reality.

11.3.3.1 Recommendations for Policy Makers

Policy makers have an important role in creating the regulatory framework and the infrastructure to allow ethics to be considered in ICT. If emerging ICTs are to be developed in a responsible manner that allows the identifying and addressing of the social and ethical problems outlined above, then a framework and infrastructure for the development of responsibility needs to be provided. Such a framework should cover at least the following three main areas of policy activity:

- ***Provide a regulatory framework which will support Ethical Impact Assessment for ICTs:*** This recommendation addresses the institutional framework that will be required to recognize responsibilities, and develop mechanisms for discharging them. The idea of an "Ethical Impact Assessment for ICTs" was chosen because it can draw on precedent from areas of the environment, privacy, or equality. Researchers are often familiar at least with some types of prospective assessment, such as risk assessment, and may be willing to engage with ethics on this level. This may explain why one attempt to implement RRI relied on principles of risk management (Owen and Goldberg, 2010). A framework for ethical impact assessment in ICT could provide incentives to engage with relevant issues. It could, thereby, encourage discourses that may lead to the development of specific responsibilities.
- ***Establish an ICT Ethics Observatory:*** While the first set of recommendations aimed at providing a procedural framework for identifying and addressing ethical issues in ICT, this set of recommendations aims to provide the content required for an Ethical Impact Assessment. The work undertaken by the ETICA project, for example, provides important pointers toward possible ethical issues to be considered. Individuals involved in technical development are often not experts in these matters. A shared repository of ethics-related theories, practices, methodologies, and so on is a necessary condition of the development of widely shared good practice. One of the tasks of the FRRIICT project will be to develop a first prototype of such an observatory.
- ***Establish a forum for stakeholder involvement:*** This final recommendation for policy makers points to the necessity of institutionalizing important discourses that allow civil society and other stakeholders to engage on a content level with the policy – as well as technical – community. Such a forum is required to ensure that responsible innovation covers not only specific technical interests and perspectives but is allowed to reflect broader societal concerns.

11.3.3.2 Recommendations for Industry, Researchers, and Civil Society Organizations

Industry, researchers, and other individuals or organizations should adhere to the following recommendations in order to be proactive and allow innovation to be socially responsible. If the institutional framework, background, repository, and societal discourses are present, then the conditions will be favorable for the incorporation of ethics and reflexivity into technical work and application usage.

- ***Incorporate ethics into ICT research and development:*** This recommendation aims to ensure that ethical reflexivity is realized within technical work. Furthermore, it aims to sensitize stakeholders to the difficulties of discharging their responsibilities.
- ***Facilitate ethical reflexivity in ICT projects and practice:*** This set of suggestions aims to ensure that the different stakeholders realize that ethics is not a predetermined and fixed structure. Ethical issues are context-dependent and need to be interpreted in the particular situation. Interpretive flexibility (Doherty, Coombs, and Loan-Clarke, 2006) of technology renders it desirable not only for the participants in a technology development project to engage collectively in the initial definition of ethical issues that are to be considered, but also to review this initial definition continuously and engage with stakeholders involved in other stages of the technology development process.

11.4 Critical Reflections

The policy recommendations arising from the ETICA project as outlined above arose from a specific and technology-centered perspective, and will need to be interpreted in the light of the findings of the FRRIICT project. A core question is that of the reflective governance mechanisms in which RRI for ICT will develop. ETICA recommended a regulatory framework to promote an ethical impact assessment. This was motivated by the need to develop some kind of incentive for ICT researchers and practitioners to engage with responsibility issues. The concept of impact assessment was chosen because researchers tend to be familiar with other types of impact assessments, such as environment, risk, or equality impact assessments. Such impact assessments are usually compulsory but may take different forms. The idea was that by requiring such an ethical impact assessment, ICT researchers would need to engage with the issues but would not be forced to do so in a particular way.

While this idea is, in principle, still valid, it raises questions. Owen and Goldberg (2010) used mechanisms of risk assessment to encourage researchers to engage with RRI with mixed success (see Chapter 2 for a critique). Furthermore, the findings of the FRRIICT landscape study suggest that researchers may not react positively to what they perceive as external pressure to deal with ethics and responsibility. In addition, there is a danger that a compulsory ethical impact assessment may turn into a further box-ticking exercise. An initial attempt to formulate the principles of such an assessment (Wright, 2011) suggests that this might indeed be the way it could be perceived.

11.4.1 The Meta-Responsibilities of RRI

This brings us back to the earlier suggestion that RRI could be viewed as a meta-responsibility, as a responsibility for the definition and realization of responsibilities. Having explored the ethical issues of emerging ICTs, as well as the perception of stakeholders in the UK ICT field, we can now offer a more precise definition and content of RRI in ICT.

RRI is a social construct or ascription that defines entities and relationships between them in such a way that the outcomes of research and innovation processes lead to socially desirable consequences. Of course, socially desirable consequences may be very different for different stakeholders and so the diversity of viewpoints needs to

be taken into account and negotiation made possible in whatever framework we suggest. RRI contributes to existing vibrant networks of existing responsibility relationships and ensures that these become compatible and synergetic. The activities of RRI will encompass the following aspects:

- Description of subjects of responsibility in ICT (e.g., researchers, professional bodies, funders, policy makers)
- Attribution of objects to subjects (e.g., sustainability, usability, privacy, transparency)
- Facilitation of discourses on norms, values, and ethical foundations
- Design of sanction and incentive structures
- Reflection on preconditions for successful responsibility relationships.

All of these points are already incorporated in existing responsibility ascription. The meta-aspect of RRI is that they are usually done on a local or *ad hoc* level and are often in conflict, or even contradictory. RRI will require a higher level discussion of what diverse groups within different societies want to achieve through their technologies, and then consider how to identify and negotiate those aims and how they might be achieved.

The last point – the reflection on the preconditions for successful responsibility relationships – may be the most difficult one. The question of when responsibility may have desired effects is surrounded by conceptual and empirical uncertainties. It is rather safe to say, however, that one of these conditions will be to ensure that researchers, developers, and practitioners in ICT understand the problems surrounding ethics and responsibility and accept that it is in their interest, not only as ICT professionals, but also as citizens, to engage with these questions. This leads to a clear need to embed such questions into ICT education from primary to tertiary levels (Stahl, 2011b). An understanding of ethics and responsibility is a necessary (albeit not sufficient) condition for researchers to take such issues seriously (see Chapter 2 for further discussion on education, training, and capacity). Such an understanding can lead to an intrinsic motivation to deal with the ethics of ICT. This intrinsic motivation will likely have to be combined with an extrinsic motivation, as represented by a reflective governance framework. The relationship between these two drivers for engaging with RRI will be subject to empirical review and policy discussion.

11.4.2 Further Research

This chapter has shown that work on RRI in ICT is still in its infancy. Despite a long tradition of exploring ethical issues of ICT going back to the very beginnings of the technology (Wiener, 1954), and despite a rich field of research in computer and information ethics (Floridi, 2010; Himma and Tavani, 2008; van den Hoven and Weckert, 2008), we are far from having generally acceptable and accepted ways of dealing with these issues. This prior work can sensitize researchers, practitioners, and policy makers, but it gives little guidance on how to *practically* deal with issues when they arise.

RRI as a meta-responsibility does not have to dwell on the details of individual responsibility relationships (e.g., what is the responsibility of the software engineer), but it will need to develop a framework that will allow the definition of individual responsibilities and the relationship between different instances of responsibility.

The ICT Ethics Observatory recommended by the ETICA project, which is now being developed as a prototype in the FRRIICT project under the title of Observatory for RRI in ICT, can be understood as a mechanism of facilitating coherent and consistent responsibilities across the ICT field. This means that it will need to reflect the conceptual uncertainties and the contested nature of many ethical issues. At the same time it will have to provide substantive and procedural guidance to the different stakeholder groups. In addition, its success will depend on the willingness of the ICT community to make use of it and contribute to it.

There are additional considerations of preconditions for RRI in ICT that need to be taken into consideration. One important aspect of this is the socio-economic environment in which ICTs are developed and used. The majority of ICT research expenditure (and therefore its subsequent outcomes) is spent in private companies. These are normally driven by a profit motive and will disregard RRI if it appears to contravene this profit interest. There is thus a question about the relationship of RRI and profits that needs to be explored and that points toward an incorporation of ideas of business ethics into RRI. And one should admit that there is at least the potential for a conflict of interest. But RRI does not have to oppose profits and one can make a strong and evidence-supported argument that early attention to ethical and social aspects is in the interest of companies. However, one should also be honest enough that RRI may oppose some activities or products that companies would find beneficial.

Another aspect of RRI in ICT that the FRRIICT project has started to touch upon, but that will need closer attention, is the role that policy makers, and in particular research funders play as responsibility subjects in their own right. Much of the attention of the ETICA and FRRIICT project was focused on the technologies and the researchers who develop them. It should be clear, however, that in many cases the ethical issues of a technology are largely shaped by funding calls and funding decisions. Research funders may sometimes be in a better position to understand and evaluate the social impact of a technology than individual researchers or research groups. Funders not only provide the resources but often shape specific calls and have an overview of related research, which can allow them to develop a more detailed understanding of the social context of the research and its possible consequences. This raises the question of how the responsibilities of funders can be expressed and how they might be related to other responsibilities, a key question for the meta-responsibility of RRI.

A final point worth mentioning is the overall democratic legitimation of publicly funded research, and the role that civil society can and should play in RRI for ICT. Von Schomberg's opening definition of RRI prominently underlined the role of societal actors, but leaves open who these actors are and how they can be incorporated into ICT research. This is reflected by the ETICA recommendation to establish a forum for stakeholder involvement. While this idea of societal feedback is probably plausible for publicly funded research in democratic societies, it is by no means clear how it could be implemented. It is even less clear how it could be evaluated (Rowe and Frewer, 2000). And finally, there will be costs associated with it which would need to be justified by empirically confirmed benefits of such engagement.

All of this shows that much work is to be done in order to develop a practical and beneficial overall FRRIICT. Such a framework will need to draw upon the manifold activities in other areas of research and innovation. The present chapter has provided an insight into some current activities and demonstrated that RRI is best conceptualized as

a meta-level activity that influences existing and novel responsibility relationships. We believe that this understanding of RRI can help further the goal of achieving desirable and desired results within ICT, as well as in other fields and disciplines.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no 230318. The authors acknowledge the contribution of the consortium members to the research which was summarized in parts of this chapter.

This research benefited from the activities undertaken in the project "Framework for Responsible Research and Innovation in ICT," EPSRC reference EP/J000019/1.

References

Allen, C., Smit, I., and Wallach, W. (2005). Artificial morality: top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology,* **7**(3), 149–155. doi: 10.1007/s10676-006-0004-4

Bayertz, K. (ed.) (1995). *Verantwortung: Prinzip oder Problem?* Wissenschaftliche Buchgesellschaft.

Bostrom, N., and Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. *Science and Engineering Ethics,* **15**(3), 311–341.

Calvert, J. (2006). What's special about basic research? *Science, Technology & Human Values,* **31**(2), 199–220.

Coeckelbergh, M. (2011). Human development or human enhancement? A methodological reflection on capabilities and the evaluation of information technologies. *Ethics and Information Technology,* **13**(2), 81–92. doi: 10.1007/s10676-010-9231-9

Cuhls, K. (2003). From forecasting to foresight processes – new participative foresight activities in Germany. *Journal of Forecasting,* **22**(2–3), 93–111. doi: 10.1002/for.848

Doherty, N. F., Coombs, C. R., and Loan-Clarke, J. (2006). 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), 569–582.

Etchegoyen, A. (1993). *Le Temps des Responsables*. Julliard.

European Commission (2011) Work Programme 2012, Capacities, Part 5, Science in Society, Brussels, http://ec.europa.eu/research/participants/portal/page/capacities?callIdentifier=FP7-SCIENCE-IN-SOCIETY-2012-1(accessed 12 October 2012).

Fisher, E., Mahajan, R. L., and Mitcham, C. (2006). Midstream modulation of technology: governance from within. *Bulletin of Science, Technology and Society,* **26**(6), 485–496. doi: 10.1177/0270467606295402

Floridi, L. (ed.) (2010). *The Cambridge Handbook of Information and Computer Ethics*. Cambridge University Press.

Floridi, L., and Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines,* **14**(3), 349–379. doi: 10.1023/B:MIND.0000035461.63578.9d

French, P. A. (1992). *Responsibility Matters*. University Press of Kansas.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw and J. D. Bransford (eds.), *Perceiving, Acting and Knowing* pp. 67–82. Hillsdale: Lawrence Erlbaum Associates.

Heersmink, R., van den Hoven, J., van Eck, N., and van den Berg, J. (2011). Bibliometric mapping of computer and information ethics. *Ethics and Information Technology,* 13(3), 241–249.

Himma, K. E., and Tavani, H. T. (2008). *The Handbook of Information and Computer Ethics*. Hoboken, NJ: John Wiley & Sons, Ltd.

van den Hoven, J. and Weckert, J. (eds.) (2008) *Information Technology and Moral Philosophy* (1st edn). Cambridge University Press.

Humphreys, S. (2011) Navigating the Dataverse: Privacy, Technology, Human Rights: Discussion Paper, International Council on Human Rights Policy, Geneva, Switzerland, www.ichrp.org/en/zoom-in/new_report_dataverse_privacy (accessed 07 July 2011).

Lewis, H. D. (1972). The non-moral notion of collective responsibility. In P. A. French (ed.), *Individual and Collective Responsibility: Massacre at My Lai*. pp. 116–144. Cambridge, MA: Schenkman Publishing Co.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions,* **6**(3), 38–43. doi: 10.1145/301153.301168

Owen, R., and Goldberg, N. (2010). Responsible innovation: a pilot study with the U.K. engineering and physical sciences research council. *Risk Analysis: An International Journal,* **30**(11), 1699–1707. doi: 10.1111/j.1539-6924.2010.01517.x

Rowe, G., and Frewer, L. J. (2000). Public participation methods: a framework for evaluation. *Science, Technology & Human Values,* **25**(1), 3–29. doi: 10.1177/016224390002500101

Sadri, F. (2011). Ambient intelligence: a survey. *ACM Computing Surveys,* **43**(4), 36.1–36.66. doi: 10.1145/1978802.1978815

Stahl, B. (2004) *Responsible Management of Information Systems*. Hershey, PA: IGI Publishing.

Stahl, B. C. (2011a). IT for a better future: how to integrate ethics, politics and innovation. *Journal of Information, Communication and Ethics in Society,* **9**(3), 140–156. doi: 10.1108/14779961111167630

Stahl, B. C. (2011b). Teaching ethical reflexivity in information systems: how to equip students to deal with moral and ethical issues of emerging information and communication technologies. *Journal of Information Systems Education,* **22**(3), 253–260.

Von Schomberg, R. (ed.) (2011). *Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields*. Publication Office of the European Union, Luxembourg, http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf (accessed 01 November 2011).

Werhane, P. H. (1985). *Persons, Rights, and Corporations*. Prentice Hall.

Wiener, N. (1954). *The Human Use of Human Beings*. New York: Doubleday.

World Medical Association (2008) Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. http://www.wma.net/en/30publications/10policies/b3/17c.pdf (accessed 23 August 2012).

Wright, D. (2011). A framework for the ethical impact assessment of information technology. *Ethics and Information Technology,* **13**(3), 199–226. doi: 10.1007/s10676-010-9242-6