Assessing responsible innovation training

Stahl, Bernd carsten; Aicardi, Christine; Brooks, Laurence; Craigon, Peter j.; Cunden, Mayen; Burton, Saheli datta; Heaver, Martin de; Saille, Stevienna de; Dolby, Serena; Dowthwaite, Liz; Eke, Damian; Hughes, Stephen; Keene, Paul; Kuh, Vivienne; Portillo, Virginia; Shanley, Danielle; Smallman, Melanie; Smith, Michael; Stilgoe, Jack; Ulnicane, Inga

DOI:
10.1016/j.jrt.2023.100063

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Stahl, BC, Aicardi, C, Brooks, L, Craigon, PJ, Cunden, M, Burton, SD, Heaver, MD, Saille, SD, Dolby, S, Dowthwaite, L, Eke, D, Hughes, S, Keene, P, Kuh, V, Portillo, V, Shanley, D, Smallman, M, Smith, M, Stilgoe, J, Ulnicane, I, Wagner, C & Webb, H 2023, 'Assessing responsible innovation training', *Journal of Responsible Technology*, vol. 16, 100063. https://doi.org/10.1016/j.jrt.2023.100063

Journal of Responsible Technology 16 (2023) 100063

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Assessing responsible innovation training

ARTICLE INFO

Keywords
Responsible innovation
Doctoral training
Education
Assessment

ABSTRACT

There is broad agreement that one important aspect of responsible innovation (RI) is to provide training on its principles and practices to current and future researchers and innovators, notably including doctoral students. Much less agreement can be observed concerning the question of what this training should consist of, how it should be delivered and how it could be assessed. The increasing institutional embedding of RI leads to calls for the alignment of RI training with training in other subjects. One can therefore observe a push towards the official assessment of RI training, for example in the recent call for proposals for centres for doctoral training by UK Research and Innovation. This editorial article takes its point of departure from the recognition that the RI community will need to react to the call for assessment of RI training. It provides an overview of the background and open questions around RI training and assessment as a background of examples of RI training assessment at doctoral level. There is unlikely to be one right way of assessing RI training across institutions and disciplines, but we expect that the examples provided in this article can help RI scholars and practitioners orient their training and its assessment in ways that are academically viable as well as supportive of the overall aims of RI.

1. Introduction

What kind of skills, knowledge, attitudes and competencies do researchers need in the context of responsible innovation (RI)? Should they know the definition of RI and its theoretical underpinnings? Should they be required to have a relatively detailed understanding of RI which often implies familiarity with matters such as methodologies of foresight or stakeholder engagement that can be far removed from the researchers' core scientific expertise? Or should they be willing and able to continuously reflect on and address social and ethical aspects of their own research, as it develops and raises new potential social and political issues, concerns and challenges?

Wherever one stands on these questions, it is widely accepted that it is important for researchers to be exposed to RI throughout their research training and career development. The integration of RI into researcher training has therefore been recognised as a key condition of RI success. There have been a significant number of RI research and support activities that aim to develop RI training material, curricula etc with a view to supporting this training. And while such training can take place across the entire research career, starting at undergraduate or even secondary school level, and continuing to senior research management positions, the focus in many cases has been on RI training for PhD students (Gerrits et al., 2021; Hesjedal et al., 2020; Tokalić et al., 2021).

RI training at PhD level can therefore draw on a rich array of material and support structures. It has increasingly been institutionalised, not least in the UK where the current round of Centres for Doctoral Training (CDT) have to respond to a funding requirement that stipulates that RI must form part of the training activities as well as the CDT structure (Ten Holter et al., 2022). While such a broad integration of RI into doctoral training is thus well established, there is little guidance as to how it should be implemented, a question that is largely left to the individual CDTs. This openness in terms of content is likely appropriate in light of the very different research questions, methods and possible outcomes of research across disciplines and institutions.

The openness of RI training at PhD level may be welcome on a content level, but it raises questions on other levels. Is RI training across disciplines, institutions, cultures and societies comparable and equivalent and should it be? Which level of RI engagement is appropriate for doctoral students in different disciplines? What consequences does and should it have? By integrating RI training into the rather formalised structures of university doctoral training, these and similar questions become unavoidable. One crucial part of the answer to these types of questions in other types of higher education is provided by assessment. Assessments are meant to provide evidence that specific learning outcomes have been achieved. This is a crucial component of the vast majority of current higher education provision. It is thus reasonable to ask how RI training should and can be assessed. This question forms the core research question of this article and has recently gained additional prominence when, in preparation for the next round of CDT funding the UK funding councils responsible for funding CDTs in engineering, physical science and artificial intelligence announced that in the next round CDTs should provide "[…] a formal, assessable programme of taught coursework, which should develop and enhance, for example, […] responsible innovation […]" (EPSRC, 2022)

The answer to the question how RI training can be assessed is clearly of practical importance to researchers who want to respond to this call and, more broadly, to scholars who aim to integrate RI training in broader research training structures. It is, however, also of conceptual

https://doi.org/10.1016/j.jrt.2023.100063

Available online 10 June 2023

interest to all participants of the RI discourse, given that it forces a reflection on the link between RI and RI training, the nature, options and limitations of such training and the way it can be integrated into formal research training structures.

This editorial article therefore brings together a number of authors who have been involved in RI training and aims to start and support a discussion within the RI community about RI training and in particular about its assessment.

The article proceeds as follows. It starts with a conceptual overview of RI, research training, in particular on a doctoral level and the challenges and open questions regarding options for RI training and assessment. The bulk of the article then consists of examples of RI training and assessment across various institutions and disciplines. This is followed by a discussion section that aims to tease out key lessons learned and insights that are of practical interest to individuals designing RI training as well as of theoretical relevance to the RI research community and beyond.

2. Responsible innovation training and the challenges of assessment

The main purpose of this article is an exchange of practical experience of RI training assessment. However, such practical experience is typically developed against a theoretical background and can usually only be appreciated and critically reflected upon when relevant theoretical perspectives are available. We therefore offer a brief overview of some aspects of theoretical interest with relevance to RI training assessment. The purpose of this overview is not to provide a comprehensive or systematic review of the relevant literature but to highlight key questions that can drive RI training assessment and that need to be answered – implicitly or explicitly – in any RI teaching and assessment regime.

2.1. Responsible innovation

The concept of responsible innovation or responsible research and innovation (RRI) has been well-discussed but at the same time remains under development. One of the most prominent early descriptions of the term provided by Stilgoe et al., and Macnaghten (2013) highlights four key components, namely anticipation, reflexivity, inclusion and responsiveness. This corresponds with von Schomberg's (2013) view that RI can be understood as a process of increasing mutual responsiveness between societal actors and innovators that is meant to lead to increased acceptability, sustainability and societal desirability of innovation processes and outputs. We will ignore the conceptual niceties of the distinction between responsible innovation, which has been more prominent in the UK (Owen, 2014), and the European discourse on responsible research and innovation which focuses on the so-called pillars of RRI (European Commission, 2013). While we agree that this distinction is worthy of deeper understanding (Owen et al., 2021), it will be sufficient for our purposes to see the overarching intention, which the EU's Rome declaration (2014) summarised as aligning research and innovation to the values, needs and expectations of society. It is important to see that RI, while it is a relatively novel term that has been in use since the early 2010s builds on older discourses that similarly cover the relationship between science, research and innovation with society, such as science and technology studies, governance of emerging technologies, innovation studies, technology assessment, philosophy of technology and others (see e.g. Randles et al., 2022; Shanley, 2021). The concept of RI thus covers and includes activities that significantly predate it.

For the purposes of this article that focuses on RI training and its assessment, it is more important to understand the content and expected outcomes of RI than the details of the conceptual debate. Several of the activities that fall under the umbrella of RI result from the definitions indicated earlier. The initial focus on anticipation, reflexivity, inclusion and responsiveness has been transformed by the UK's Engineering and Physical Science Research Council into its RI strategy that uses the AREA framework which stands for anticipation, reflection, engagement and action. The EU's pillars partly overlap with this framework, but they primarily serve to highlight the policy areas that the European Commission held to constitute RI. These are public engagement, gender equality, science education, ethics, open access and governance (European Commission, 2012).

Put differently, these ideas imply that research is done responsibly, if it incorporates anticipation, reflection, engagement, responsiveness, ethics, (gender) equality, science education, open access and appropriate governance mechanisms. One could argue that responsible innovation should furthermore cover other established points of individual and social responsibility that are not normally listed in the RI discourse but that nevertheless are important in terms of aligning research and innovation with societal needs and expectations. Examples could be the realisation of health and safety rules, or the prevention of misuse and dual use. Put this way, the description of RI raises a number of questions that need to be reflected on in the context of RI training. These include:

- Who is responsible for what?
  The list of components of RI indicates that it refers to activities and processes that are typically allocated to various individuals or organisational functions. Some of these may sit primarily with individual researchers, groups or projects, while others sit with institutions. Questions of equality, diversity and inclusion of research teams, for example, are heavily driven by organisational EDI policies, embedded in HR guidance and based on legal requirements. Open access to data and outputs are subject to funding rules and organisational policies that are often overseen by libraries.
- Is RI a cognitive or an empirical concept?
  The introduction to RI seems to suggest that its aim is an empirically measurable social state, i.e., the alignment of research and innovation with societal needs and expectations. The cognitive aspect of it that can be subject to RI training can refer to mechanisms that promise to achieve this outcome. There is, however, no guarantee that any of these activities will empirically achieve the desired outcome and there is no way of ensuring that the mechanisms that are taught are comprehensive.
- How much RI is enough?
  This question also has a cognitive and a practical component. The practical component refers to the question of how much RI activity should be undertaken for it to be deemed sufficient. Similarly, with regards to the cognitive side and training, the question is how much does a researcher need to know about RI to be deemed to be capable of discharging their responsibility?
- How much RI training and assessment can in practice be achieved within the budgetary constraints imposed on RI?
  Following from the general question of how much RI is enough, this question points to the resource constraints that will drive the ability to deliver RI training and thus influence what can and cannot be included in the training.
- Can RI be measured, and should it be measured? What counts as the success of RI? Would measuring and operationalizing RI and its success favour certain, more instrumental and mechanistic approaches to RI, namely, counting the number of women on committees as success, while disregarding reflection and reflexivity that cannot be easily operationalized and measured?
- How do we measure the success of RI?
  Implementing RI requires resources that could be used for research or non-scientific purposes. The question of how much RI is enough thus calls for ways of measuring the outcomes of RI activities to assess whether the desired state of affairs has been achieved.
- When do we measure the success of RI, and does it correlate to the success of RI training?
  RI implementation is an evolving process, and it is challenging to assess (or measure) long term impacts of RI practice. Also, should those measures be focused on research practice, project outputs or

both? RI training on postgraduate students and researchers themselves and how best to assess that?

- How to compare the success of RI training between different disciplines (and institutes) and even within the same discipline?

  The content of CDT RI modules and RI training provision (in house vs. external) differs widely within and between institutions. A uniform mandate to assess RI can be read to imply that RI training should be commensurate across disciplines.
- What is the value and (to whom) to assess RI practice?

  There are various stakeholders who might have an interest in the assessment of RI including the question whether a particular researcher can prove their proficiency. This includes individual research students, the training institution, the funding body who mandates the training but also potential future employers or professional bodies who can define a minimal level of expectations.
- How can the outcomes of the assessments be used to shape RI governance?

  This question refers to the fact that RI training is not (only) done as an end in itself, but that it is meant to pave the way for RI implementation in research projects and programmes. From this perspective, the key to successful RI is to ensure that it can unfold in RI governance in practice, which then arguably should form part of the assessment.
- How to make the assessment meaningful and useful (and to whom) and prevent to be taken as a box-ticking exercise?

  This question points to a larger set of issues around assessment in training and education more broadly. While assessment is entrenched from primary school onwards, it is typically portrayed as a means of ascertaining that learning outcomes are met. At the same time, meaningful assessment that can shed a light on the success in achieving learning outcomes is often difficult to realise. The stereotypical 'tick-box exercise' refers to assessment that fails to achieve such broader aims and renders assessment a hoop for learners to jump through with little added value.
- How can RI training help resolve frictions between RI components?

  Anticipation and reflection can lead to questioning, for instance, the desirability and ethicality of uncritical compliance with open access demands; or who defines which societal values research and innovation should align with and privilege. More generally, RI cannot promote anticipation and reflection for other research agendas and projects, while failing to apply these same principles to itself. This complicates the possibility of a stable teaching core for RI.

We do not suggest that these questions have clear and uncontroversial answers or that these are the only questions one could ask. What this list of questions highlights is that the practice of RI training itself is not settled and therefore the assessment of such training raises numerous questions. However, while these questions remain largely open, it is the case that any RI training regime will need to provide at least implicit responses to some of them to inform the structure and content of the training which then drives assessment.

2.2. Doctoral education and assessment

This article focuses on assessment of training at the doctoral level, but it may well be applicable to other types of RI training. Furthermore, the article focuses on doctoral training in the UK, given that it was triggered by the call for RI assessment in UK CDTs. The national characteristics of the UK's doctoral training landscape may not allow for a direct translation into the national doctoral training schemes in other countries. UK universities, for example, rely heavily on the doctoral dissertation as the main component of a doctoral journey which differs from the US model that has a more explicit emphasis on taught aspects of a doctoral degree.

UK universities are independent legal entities, unlike many continental European universities which are state bodies. The vast majority of them have the legal form of charities, which means they are not profit oriented. They nevertheless need to balance their budgets and can produce financial surpluses. The costs of doctoral studies are covered by a mixture of tuition fees and government-funded subsidies. The CDTs that require RI assessment are primarily funded by UKRI (www.ukri.org) which is a state-funded body that distributes the majority of project-specific research funding.

While UK universities are autonomous, they are overseen by various government bodies and rely heavily on peer review through external examiners and similar mechanisms to uphold academic standards. The body tasked with defining and overseeing quality standards across the entire UK teaching landscape is the Quality Assurance Agency (QAA). The QAA sets expectations for academic standards for all levels of higher education. For doctoral degrees, for example, this includes the characteristics that successful PhD students have to have demonstrated, the abilities they have acquired and their transferable skills (QAA, 2014). More detailed guidance is provided in the Characteristic Statement on Doctoral Degrees (QAA, 2020). This includes their context, forms and key features as well as, of importance to this article, the doctoral outcomes and assessment.

The QAA document states that "assessment is at the heart of doctoral degree standards" (QAA, 2020, p. 15). It continues to stress that the candidate's achievements are tested through the final doctoral assessment which includes a review of written materials as well as an oral examination. Other types of assessments, such as module assessment or formal reviews are seen as predominantly progress milestones that lead to the final assessment.

The assessment of RI thus raises many questions in its own right. At the same time it happens in an institutional framework which displays national, disciplinary, organisational and other idiosyncrasies. It thus stands to reason that there is not just one way of assessing RI but that the way to respond to a mandate to assess is to draw on past experience and good practice. We therefore describe a set of examples of RI assessment in the next section.

3. Examples of RI assessment

This section consists of detailed examples of how RI assessment or aspects of RI assessment could be realised. These examples illustrate diverse approaches to teaching and assessing RI at various UK universities. The first example comes from a company, while the rest are from academics at different career stages and from a range of scientific disciplines.

3.1. ORBIT RRI - Foundation in RRI workshop (Serena Dolby, Paul Keene, Martin de Heaver)

ORBIT-RRI ltd (www.orbit-rri.org) is a spin-out company from the University of Oxford and De Montfort University in the UK that has the task of promoting a culture of RI. Training CDT students is among its core activities. The Foundations workshop is a one-day event providing an overview and introduction to RI. It has been delivered to approximately 2000 doctoral students and supervisors at 28 CDTs.

3.1.1. Audience and content of training

The workshop targets students, researchers, and anyone working in research and development with an interest in Responsible Research and Innovation (RRI). No prior knowledge of RRI is assumed, making the workshop accessible to everyone interested in learning about RRI.

ORBIT has consistently prioritised the practical application of RRI, emphasising not only definitions but also strategies for implementation. The Foundation in RRI workshop aims to furnish attendees with a comprehensive introduction to RRI and its relevance, as well as practical guidelines for ensuring responsible conduct. This includes an overview of RRI's importance and historical context, analysis of stakeholders and agency, and examination of the relationship of RRI to research ethics and integrity. Participants are also instructed in utilising the RRI framework, including tools for determining the intensity of RRI implementation and maturity modelling.

The participants are presented with case studies and suggestions for reflection aimed at facilitating the implementation of RRI in a given scenario. The cases are chosen to represent a range of interests and discipline areas, some of which are more closely linked to the nature of the group than others. Following group discussions focused on the challenges they face, participants utilise their previous knowledge and additional research to collectively develop a presentation detailing their proposed approach for tackling the specific RRI challenges presented in the scenario. Finally, participants present their solutions as a group to the trainers and their peers, who provide constructive feedback, highlighting areas of strength, potential areas for improvement, and answering any inquiries raised by the collective. In the context of the CDTs, the students are then given the opportunity to utilise a bespoke online tool to evaluate their own research project in terms of various measure relating to RRI, and further to create an action plan, which includes a set of questions for them to address regarding the RRI aspects of their project.

3.1.2. *Intended outcome of the training and the assessment*

The workshop's learning objectives are to ensure that participants gain an understanding of the background and adoption of RRI, knowledge of the AREA Framework as adopted by EPSRC in the UK, the EU Pillars of RRI, and familiarity with critiques and challenges for RRI. The assessment of these objectives involves participants completing a short survey before and after the workshop. The pre- and post-workshop surveys consist of five questions, providing participants with the opportunity to rate their knowledge of RRI and confidence levels of applying it in their work.

The survey consists of 5 questions and we ask them to rate their answer on a scale of strongly disagree, disagree, neutral, agree and strongly agree (see Fig. 1):

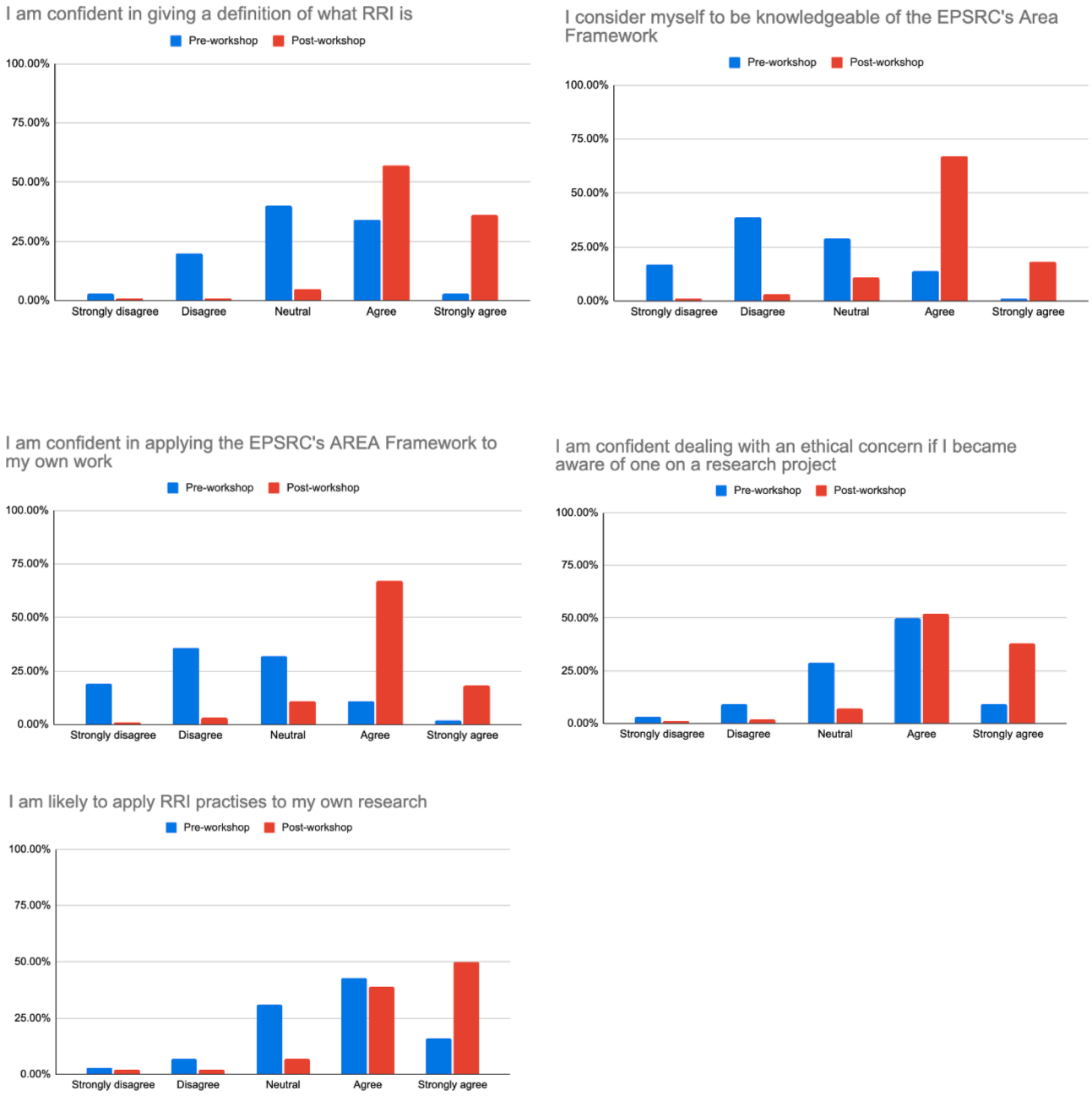

Fig. 1. (a-e): Responses to ORBIT assessment survey (blue (left)=pre, red (right)=post-workshop); figures based on 270 pre-workshop responses and 149 post-workshop responses, collected between 2019 and 2023.

- I am confident in giving a definition of what RRI is.
- I consider myself to be knowledgeable of the EPSRC's Area Framework
- I am confident in applying the EPSRC's AREA Framework to my own work.
- I am confident dealing with an ethical concern if I became aware of one on a research project.
- I am likely to apply RRI practises to my own research.

3.1.3. Critical reflection

The above figures demonstrate that participants feel they have developed an understanding of RI and that they are more capable of implementing it in their research. The advantage of this assessment regime is that it is low-cost and easy to administer. Disadvantages include the self-reported nature of the assessment which is not validated by any more objective measures. The assessment furthermore is very coarse, i.e. does not allow a distinction between different aspects of RI that may be more or less well understood.

To improve the evaluation process, it may be possible to conduct an additional assessment during a follow-up with the participants at a future interval, such as 6-12 months after initial implementation of RRI practices in their personal projects. The follow-up discussion may take place in person, involving both the participant and their supervisor, addressing their actions and potential opportunities for further improvement. However, logistical challenges may arise when attempting to conduct follow-up discussions, presenting potential barriers to this approach.

Anecdotal reflection also shows that within any one group, some participants struggle more with the concepts than others. This may come from prior familiarity with the terms and concepts as well as how much they see it relating to their particular area of research. However, once past the more background elements and into the case analysis part, almost all participants begin to see how it can be applied to the board spectrum of research and the importance that RRI has for society as a whole.

3.2. RI assessment in two existing CDTs (Christine Aicardi)

This example details the case of two Centres for Doctoral Training that were funded from 2019 by the EPSRC and UKRI respectively for the training of 5 cohorts of PhD students. Both included provisions for the delivery of R(R)I training by a named senior research fellow in Science and Technology Studies, contracted for 10% FTE by each CDT.

3.2.1. Context

It is outside the scope of the present example to detail the content of the RI training dispensed and the many questions, issues and challenges encountered during this particular experience of embedding RI in CDTs, but a minimum of context is necessary to set the scene for the decision to include a form of RI assessment and for the shape eventually taken by this assessment. In the round of calls for CDTs that would onboard PhD students cohorts over 2019-2024, the EPSRC and UKRI outlined the necessity to embed RI in the training of PhD students, but the high-level guidelines left much to the interpretation of the bidding teams and universities' research offices, who in the present case grasped the overarching practical role of RI but had a limited understanding of its cognitive role and conceptual aspects. As a result, the management teams of the two CDTs were very willing to support their RI lead but were of limited help when it came to developing the RI training that would be provided. For the RI lead in charge, this meant shouldering the accountability that may come in time for the RI training dispensed as part of the evaluation of the CDTs. One of the CDTs, which management team was building up from the experience of a pre-existing CDT, was especially concerned with anticipating evaluation in the development of training programmes. This was the general context that led to designing a RI assessment for the RI training that would be dispensed in both CDTs.

3.2.2. RI assessment design: taking inspiration from 'impact'

Two key principles and a key limitation framed the design of the RI assessment by the RI lead, in agreement with the management of both CDTs:

- First principle: RI is an ongoing process throughout the lifecycle of a research project, and the assessment should espouse this process.
- Second principle: The CDTs are accompanying postgraduate students on their journey to becoming fully fledged researchers, be it in the public or the private sector. The RI assessment should thus be designed in a way that places students at the centre, as Early Career Researchers on the brink of a career either in industry or in academia, and help them acquire skills that will be useful to their future occupations no matter what.
- Limitation: It was implicit that any RI assessment would concern the students only and not their supervisors. This was limiting the scope and value of said assessment and its capacity for opening up uncomfortable questions at CDT level.

The angle taken as a result of this framing was to go back to the general definition of research impact provided in advance of the REF2021, which stated that "impact is defined as an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia." We can all agree that the way impact is defined for the REF is highly debatable. But in the context of designing a RI assessment for CDT students, this particular definition of societal, environmental and economic impact maps rather well with RI's lofty goals of making positive contributions to society while minimising negative ones. Hence the idea of shaping the RI assessment for the two CDTs concerned as an exercise, for the students, in embedding RI as 'impact' into the lifecycle of their PhD projects and having a kind of mock impact case study to annex to their PhD dissertations by the end of their doctoral studies. This, at least, should provide them with skills they can put to use in their future careers, be it in academia or else. It also made sense to the students as something that may not be enjoyable but could be useful to them – and having the buy-in from the first concerned was paramount in view of the second principle above.

3.2.3. RI assessment in practice

This is how it works in practice. At the end of their 1st year, once the students have attended a mandatory 'Introduction to RRI' short course, they put in practice what they have learned as they each work on a 'RRI self-assessment and action plan' report (3 pages max). They submit these first draft reports to the RI lead, who provides them with feedback on RI-relevant aspects of their research and discusses potential actions with them during one-to-one tutoring sessions. This cycle is repeated once a year during their remaining years at the CDTs. It forces the students to come back to their RRI self-assessment and action plan, and revise it in view of how their research has evolved over the year, how they themselves may have changed in their outlook, which RI-friendly activities they have engaged in and which new interests they may have developed.

The RI report in its final state, including the RI-linked activities undertaken by the student throughout their doctoral project and further actions that may be warranted in the future but that they did not have the means to engage in within the scope of their PhD project, will be appended to their PhD dissertation. This whole concept of RI assessment has been well received by the one CDT's advisory board that it was presented to. It was not planned to mark the RI report at all as it is not in

the spirit of the exercise, but it is a step that the EPSRC in particular would like us to take. We have settled for now on the halfway plan to have either the CDT director or a panel from the CDT management team giving a qualitative evaluation of the report. It will not be an element of the viva, though. This would be highly problematic on many levels that will not be detailed here, but an important element is that the students are highly dependent on the willingness and interest of their supervisors, and on the institutional resources they are provided with, for developing RI approaches in their projects.

3.2.4. Reflecting on an ongoing experience

The first cohorts of the two CDTs are only now in their final year, and many have been granted extensions because the pandemic has disrupted their doctoral studies, so there is no knowing how well this form of RI assessment works when brought to its conclusion. But so far, although the RRI reports are inequal, the RI lead is overall pleasantly surprised by the overall quality of the reports. Several of the students appear to have found the exercise much more interesting than they expected at first and have done an excellent reflective job. They have engaged in good conversations with the RI lead and been grateful for new perspectives that they were not getting otherwise. Some of them are also soliciting the RI lead for advice on RI-friendly activities they engage in (such as developing interventions to promote the importance of decolonising computer science research within their Department), which is taken as a positive sign that they find value in RI perspectives.

The RI (self-)assessment was initially meant to go through a round of group presentations, discussion and feedback, but the COVID-19 pandemic came in the way. So far, the current format appears to be working rather well, but we are giving more thoughts to how we could re-introduce such group activities in the RI assessment design for the upcoming CDT bids.

A downside of this mode of RI assessment is that it is time consuming for the RI lead, so that it only works for reasonably sized CDT cohorts. Also, it works well when the CDTs address research topics that the RI lead is otherwise quite familiar with through their own research, which is the case here. But someone providing a generic RRI training with no particular expertise in the research domain covered by the CDTs would be hard put to provide a satisfactory level of relevant feedback to students. With the two CDTs under consideration in this example, the RI lead provides (1) the Introduction to RRI training for 1st years, and (2) advanced thematic workshops for 2nd and 3rd years on topics relevant to the CDTs domain areas as well as tutoring for the RI assessment. Going forward into the new round of CDT bids, we consider that these two tasks could be dissociated, as the former does not require the domain-specific understanding required by the latter.

I will add one last point regarding the demand by UKRI and EPSRC for the assessment of taught programmes such as RI in the present CDT calls. The current CDTs were the first that included a mandatory RI component. The experience of the RI lead in charge of the two CDTs in this example, is that it has taken the past four years to get to the point where it is now, with the upcoming CDT bids, that their institution and new bidding CDT teams take their expertise a lot more seriously and are keen to get some RI input into their CDT applications. It may be that this is, in part, a benefit of the requirement for RI assessment, that CDT applicants feel they will be the ones assessed eventually. This raises anew the question of what and who is assessed, and also the question of the positionality of the RI lead in the CDTs – my positionality – as well as the positionality of us, RRI specialists, as a group, in this context. Ironically, this new requirement for RI assessment may actually raise our status in the upcoming CDT bids, which most of us would probably welcome. But this means we have stakes in the whole endeavour, that we need to take a good hard look at and be open about.

3.3. UKRI Horizon Centre for Doctoral Training (Peter Craigon and Virginia Portillo)

The Horizon Centre for Doctoral Training (CDT) is focused on multidisciplinary work in the digital economy at the intersection of personal data and digital products, services and experiences. The multidisciplinary nature and breadth of the digital economy attracts a wide variety of students in terms of background and research disciplines (e.g., psychology, engineering, politics, sociology, art, health, computer science). This is reflected in the different approaches and topics of PhD projects they undertake. Each student project is supported by an industrial partner (coordinated by Horizon) to embed the research in a real-world setting. The partner, along with a multidisciplinary supervision team, work with the student throughout their PhD programme, bringing a broad range of perspectives to guide their work. Such a varied range of students, disciplines and potential approaches means that research training and methodology need to be applicable to the wide range of projects and disciplines represented.

Such breadth and variety are accommodated in the CDT's approach to training, which includes Responsible Innovation (RI), delivered alongside the students' research work throughout the four years of their programme. The training is mainly front-loaded in the first year, this aims to support and shape the development of their research project proposals. The focus of this training is a set of bespoke in-house modules (n. 12 – 160 credits at level 4) intended to expose the students to a broad range of academic perspectives, research approaches, methodologies and skills to conduct their PhD research and beyond. Within the first year, there is a specific module dedicated to RI training which includes a number of sessions (i.e., 2h to full day) over a 7-month span. These are delivered in-house by academics and researchers with experience in putting RI into practice within their own research projects. Over the course of the module students are required to submit three iterations of their developing research proposal in increasing detail, resulting in a final proposal which is the subject of their first-year review and the basis of their PhD research project (which they start in year 2). Within their proposal the students are asked to provide a statement of how RI applies to their work and how they will incorporate it in their proposed PhD activities. This statement is intended to put particular focus on RI but it is hoped that the activities will be embedded through the design and undertaking of the project rather than considered as a separate element. The different iterations of the proposal are reviewed by a range of researchers and academics that form part of the Horizon CDT network, including the student's supervisor(s). The intention is for feedback to help improve and inform the development of the final PhD project proposal. In addition to that assessment process, peer review sessions have supported the development of students' research proposals, and ultimately their PhD project. In short, students hear from a number of distinct voices that bring with them a range of perspectives on RI.

To support the development of the RI statement and the PhD proposal, several RI training activities are delivered. Many of the activities, for example using card-based tools, have been developed as part of RI research within the Horizon DER Institute, to help researchers engage with RI and embed it in their work. So, using those tools with the CDT students is intended to be mutually beneficial. At the start of the module activities are focused on RI in general terms, progressively shifting their focus onto the students individual PhD research topics as they progress through the activities. A summary of the RI module activities include:

- Introduction to Responsible (Research and) Innovation: explaining the main principles, frameworks and questions that RI asks of researchers, with real world examples and interactive reflective exercises.

- Interactive group sessions: for example using card- based tools and running Ethical Hackathons (see Section 3.4). Sessions using cards, for example the Moral-IT cards (Urquhart & Craigon, 2021) and RI prompts and practice cards (Greenhalgh et al., 2022; Portillo, Greenhalgh, Craigon, & Holter, 2023), have provided the students with the opportunity to apply RI principles and approaches to technology scenarios, identify RI challenges and reflect accordingly, and to design and plan RI activities to maximise benefits and minimise risks as part of their own projects.
- RI "clinics" (consultation sessions): two RI researchers provide individual RI support and feedback to students who are preparing their final PhD research proposals. Students share in advance their draft RI statement and feedback is tailored to each project proposal. This approach has been recently introduced in the Horizon CDT programme recognising the variety, complexity and multidisciplinarity of the projects undertaken via the CDT.

The assessment of RI training is necessarily shaped by the nature of the CDT, its varied students' research interests and projects. Furthermore, the fit within the wider requirements of proposal development and assessment of a PhD within a university setting, and the perspectives of those reviewing proposals and work from a wide range of disciplines. This is reflected in the objectives of the RI module which focus on equipping students with a broad appreciation of RI in research and the ability to apply and communicate its relevance and significance to their research with the aid of appropriate frameworks and tools.

3.3.1. Critical reflections

3.3.1.1. What is 'good' RI?. Reflecting on the delivery of the CDT RI training, the notion of assessment raises questions. If, for example, RI is understood through the AREA 4Ps framework (Jirotka et al., 2017), then what constitutes 'good' (or appropriate) anticipation or reflection, for example, and how can students be 'trained' to do this? Such issues are contextual, difficult to define and may only be realised through an ongoing reflexive approach throughout research, which we hope to begin to support and instill through the RI activities delivered. The variety of projects and approaches of the CDT show how RI needs to be carefully situated and considered in each individual context, as what is 'good' for one may well not be in another with 'training' flexibly taking this into account. This is supported by individual one to one support and discussion with and between the students to provide alternative perspectives and help to apply the frameworks and principles of RI to their specific context and its attendant nuances, approach, and challenges.

3.3.1.2. Who are we to decide?. Part of the value of RI is found in encouraging and enabling researchers to consider their work from a wider societal point of view. The perspective of RI (e.g., societally desirable, ethically acceptable, sustainable (Von Schomberg, 2011) is a valuable challenge to the location of expertise within academia. It is not for academics within universities to decide what is societally desirable for example, this is up to 'society'. For instance, to address social desirability requires a mechanism to access and incorporate the view(s) of 'society' into the assessment of RI activities, the training provision and the research they inform. If RI training is assessed within universities, then this may reinforce a division between 'science' and 'society' (Owen et al., 2012; Shanley, 2021) which RRI practices have been trying to overcome. How to approach this is a significant ongoing challenge.

3.3.2. Conclusions

RI is a dynamic, evolving and contextual process so assessing RI practice is challenging not only within a multidisciplinary CDT but due to the nature of RI and research processes themselves. The Horizon CDT recognises this in its RI training programme. A tangible approach to assess RI training is to evaluate elements of RI identified by the student within their project (e.g., challenges, risks, inclusion of stakeholders), and RI activities planned and put into practice during their PhD. To make it an embedded practice, this could be assessed not only at the start of the project (PhD project proposal) but throughout their PhD work (e.g., as part of each annual review). For example, the maturity of their RI learning process and practice could be assessed by incorporating RI reflexive content into the thesis, for example in each research chapter. This not only would help students and assessors to realise how they have (or not) articulated RI within their projects and identify facilitators and barriers within their project from the start to the end of their PhD, but will hopefully plant an 'RI seed' for their future careers.

Assessment of RI training should ultimately benefit the recipients (i.e., students, researchers) and help professionals to deliver high standards in line with society's desires and needs. Involving students, researchers, professionals but also external voices in designing how this could be achieved would help build RI best practice.

3.4. The RI focused ethical hackathon (Helena Webb and Liz Dowthwaite)

3.4.1. Hackathons and the RI focused ethical hackathon

Hackathon competitions are generally very familiar to students in computer science and engineering and are often included in university curricula for software engineering (Gama et al., 2018). In a 'traditional' hackathon, participants work on a programming challenge in teams over a short period of time (such as 48 hours) (Komssi et al., 2015). Lecture-style information delivery sessions are included to set the challenge and introduce/revise key skill areas but much of the event is dedicated to giving teams time to work together to tackle the challenge, which typically requires students to create some kind of functioning software tool/system. Teams present their tool/system at the end of the event and a panel of judges awards prizes to the best ones. The judging criteria generally relate to software skills and quality, so hackathons provide a fun way for students to consolidate their learning by applying it in practice.

In recent years, this traditional hackathon model has been broadened in various ways. For instance, green hackathon events (Zapico et al., 2013)set challenges to create sustainable software and/or hardware solutions, and data-focused hackathons (or 'datathons') challenge data scientists and others to build platforms for the effective analysis of large datasets (Aboab et al., 2016). Hackathons are also becoming increasingly multidisciplinary, with an emphasis on bringing together participants from different disciplinary perspectives to encourage creative approaches to problem-solving and promote peer learning.

The broadening of the traditional hackathon has also seen the development of the RI-focused ethical hackathon. The original idea for this model arose from the FRRICT project on responsible innovation in ICT and it was then further developed by members of the Human Centred Computing theme at the University of Oxford (Webb et al., 2019). It was trialled in a collaborative project involving the Universities of Nottingham, Oxford and Edinburgh (Patel et al., 2019) and is now an embedded part of the RI module for the Horizon CDT (section 3.3). As with the traditional model, the RI focused ethical hackathon sets a design challenge. However, rather than focusing purely on the technical features of design, the challenge requires interdisciplinary teams to consider how a new tool, product or system might address ethical and societal issues. The remainder of this case study describes and reflects on the use of the RI focused ethical hackathon as a learning mechanism within the Horizon CDT.

3.4.2. RI focused ethical hackathons in the Horizon CDT

In the RI focused ethical hackathon, the challenges set require teams to apply an interdisciplinary perspective in order to identify pathways towards a responsible innovation. As an example, a challenge we commonly set in the CDT is to *create a new, responsible social networking platform.* Students are encouraged to consider how mechanisms to

address problems such as misinformation, hate speech and excessive data collection for algorithmic personalisation could be embedded within the affordances of the platform itself. In addition, they are encouraged to consider further responsibility mechanisms, such as community practice guidelines, platform terms and conditions etc., as well as to plan for the financial sustainability of the platform. In their presentations to the judges, teams describe the overall structure of the platform, with mock ups to show its interface etc. and also describe and justify the responsibility mechanisms they have chosen. To help prepare the students for the task, the opening sessions of the event are dedicated to information delivery and structured group work. First, to help participants understand how responsibility can be embedded in real world research and innovation, we discuss how we have incorporated RI into our own projects and the issues that we have encountered. Following this we provide a series of case studies for participants to discuss in small groups. These case studies cover media coverage of current innovations that are contentious or controversial in some way and are chosen to be relevant to the overall challenge topic/domain. (For instance, when running the responsible social network challenge mentioned above, we use case studies on concerns over misinformation and excessive personalisation on online platforms, and algorithm controversies including the COMPAS recidivism prediction score (Angwin et al., 2016), and the moderation algorithm used to provide A-level results in England and Wales in summer 2020 (Smith, 2020). We ask the groups to anticipate the different potential impacts of the innovation, identify who might be affected by it and how, discuss who is responsible for negative impacts, and suggest various ways these impacts could be avoided or mitigated. Groups focus on different case studies and report back on their responses to the wider cohort to encourage further debate.

The challenge task is set to be deliberately interdisciplinary and therefore teams are composed of students from backgrounds such as computer science, engineering, law, social science, journalism, arts, business, and management. To fully embrace the challenge, participants need to share their own disciplinary expertise and consider how it can be combined to create an innovative and responsible solution. Quality of groupwork and interdisciplinarity are included in the judging criteria alongside factors such as the quality of the overall design and the extent to which ethical and societal issues have been identified and addressed.

3.4.3. Reflections

On the first few occasions that we ran the RI focused ethical hackathon we collected student feedback after the event. We combined this with our own observations to iterate and improve the model. Now that the event is a regular component of the Horizon CDT module, we note that the following are important to ensuring the success of the event:

- The RI focused ethical hackathon can form an excellent addition to RI training and assessment. It can be seen as an instance of RI in action, as it requires students to identify the responsibility issues arising from an innovation and work collaboratively to consider different ways to address them. Students can also be encouraged to bring in explicit RI activities and terminology – for instance to draw on the AREA framework to reflect on the innovation, identify stakeholders etc.
- Although information-giving sessions are included, the RI focused ethical hackathon works best as a supplement to formal RI training. It can be scheduled to take place after students have undertaken other sessions on the principles of RI, its frameworks etc. and be delivered as an opportunity to put these into practice in a fun way.
- Learning outcomes are maximised when the teams are a mix of participants from technical and non-technical backgrounds, and the challenge requires interdisciplinary thinking. Removing the expectation for any programming to be conducted is also valuable since coding tends to be a solitary activity that does not encourage collaborative working.
- Since students often find it difficult to know how to translate RI principles into practice, the case study discussions provide a very useful structured activity to try this out. In addition, teams benefit from regular check-ins with the hackathon facilitators to discuss progress. During these check-ins the facilitators can provide teams with prompts regarding further responsibility issues to focus on, discuss the feasibility of proposed solutions and point to useful resources.
- Prizes are given to the teams judged to be the best at meeting the challenge. However, the real assessment involved in the RI focused ethical hackathon relates to how and how well the students have engaged with the responsibility issues across the event. The facilitators use the check-in sessions and presentations to observe this and then provide feedback to students at the end of the event.

3.4.4. Conclusions

The RI focused ethical hackathon can form a valuable supplement to formal training in RI. It provides students with the opportunity to apply what they have learnt to a real-world case study and gain first-hand experience of the creative problem-solving spaces that can be opened up through interdisciplinary working. The structure of the model allows for necessary information-giving and feedback, balanced with periods of self-organised work. In combination with the competitive element, this typically makes the experience of the RI focused ethical hackathon fun as well as educational.

3.5. Using LEGO® Serious Play® (LSP) for MSc and PhD training (Stevienna de Saille)

LSP is a facilitation method originally used for product development within the LEGO Corporation. It uses a bespoke set of bricks and a simple guided methodology to help individuals and groups explore complex topics and/or generate creative solutions to shared problems (Rasmussen, 2006), and is increasingly used in academic teaching and research (James, 2013). In this section I discuss how LSP can be used to help introduce the concepts underpinning RRI, such as social shaping of technology, stakeholder and public engagement, and ethics and values in design, alongside different RRI frameworks, aimed at allowing students to visualise and discuss what this might look like in practice in their own research. Each workshop, although covering similar ground, may be tailored to the specific discipline of the CDT, the number of students, the time available, and the extent to which the CDT would like to combine theoretical elements. These can be interwoven as extremely short lectures of 5-10 minutes, followed by using the bricks for deeper exploration of individual student projects. An example protocol used as part of a CDT for biochemistry is discussed below. Similar protocols have been used with students in CDTs and summer schools in universities in the UK and abroad, and for disciplines such as law, economics, engineering, architecture and medicine. Students work together in small groups of 4-8, each with a bespoke set of 57 bricks, one of several sets developed by LEGO for this purpose.

3.5.1. Content

When used in conjunction with other CDT training such exercises (see Table 1 for the module overview) can provide a more personalised engagement with RI, especially when coupled with a longer introduction to how technology and society co-create each other, a more in-depth investigation of local innovation systems, definitions of stakeholders and different forms of engagement, etc. The LSP is then used to help the students explore the relevance of these to their own particular project, ending with a plan for active response. In terms of timing overall, such workshops are especially beneficial in the early months of a PhD, when students are still working out their research strategy, but are appropriate at any stage of research.

Table 1
Module overview.

Mode of delivery	Content
Short lecture	Introduction to the social shaping of technology and ethics & values as guides to decision making
Warm-up exercises	A standard series of skills-building exercises to familiarise participants with using the bricks to build metaphorical models to answer challenge questions, ending with building a device for travelling over snow.
Short lecture	Models of innovation, intro to EPSRC and EU frameworks
Build 1	Build a model to tell a story about your PhD research
Whole-room discussion	Exploring ethical values and motivations mentioned ('I want to help, discover, create etc), consider potential beneficiaries
Build 2	Build a model that helps you explore what risks (beyond the technical) you might foresee – do they affect the same people, different people, how might the overall goal affect society?
Build 3 (optional, if workshop is sufficiently long)	Adding to this model, who might you want to engage with early on to help you understand the potential outcomes? (involves an extra set of minifigures and other bricks for each table)
Whole-room discussion	Reflection on the benefits and risks, potential stakeholders and publics whose knowledge would be beneficial to their projects
Build 4	Using the same models, think about the BIGGEST risk you anticipated. Now modify the model to reduce that risk & tell us what you did and why.
Short lecture/summary	Recap of what they have done and how it fits into an RI framework – anticipating what the project offers, reflecting on unevenness of benefit and risk to different genders or social groups, engaging with other people, ideas, ethical frameworks, etc which might be relevant to mitigating those risks, considering how to incorporate that added knowledge into the research.

3.5.2. Assessment

To date, such workshops have not been part of the assessment process and it is uncertain whether traditional assessment would provide significant benefits to the students. However, it is possible to assess – both during the session and subsequently – the extent to which the workshop has helped them to consider new approaches and new ideas. This might be done best through developing an informal questionnaire which can both examine new knowledge developed and provide a reflective tool they can refer to throughout their project. This might include questions which ask them to describe their responses to the build challenges, and to relate this to the chosen RI framework.

3.6. Inclusion of RI discussion in project proposals (Vivienne Kuh)

3.6.1. Background

This example discusses the provision of Responsible Innovation training to ~100 PhD students per year across 8 Centres for Doctoral Training funded by EPSRC and UKRI delivered by a Lecturer in Responsible Innovation at a higher education institution. These CDTs focus on a wide variety of disciplines from lab-based chemistry to mechanical engineering & quantum computing.

Historically, much RI/RRI work has been undertaken by a social scientist/team of social scientists, embedded in a research team or field, wherein they aim to develop understanding of the disciplinary context, the relevant literature, and the socio-cultural context, both internal and external, within which the researchers work. This knowledge gathering is then used as the basis to stimulate reflection and activity in the research team to explore the RI/RRI implications of their work.

In this case, working across 8 disparate disciplines, this level of immersion was not practical or achievable. Rather than positioning themselves as expert authority on the social & ethical dilemmas presented by a particular field of research, the RI lecturer frames the programme as a process of co-creation, aiming to empower PhD students, and where interested (for there has been no scope for mandating their involvement), senior researchers in the CDT community, to map, examine and push the boundaries of socio-ethical reflection in relation to their research interest. The number of CDTs covered by the programme has also provided a logistical challenge in terms of the "embedding" of RI required by EPSRC. In response to this a novel RI Facilitators programme has been developed, where PhD students from the CDTs are trained and supported by the RI lecturer to help facilitate workshops, and research, develop and deliver their own workshops, rooted in the needs and disciplinary contexts of their CDT. This programme aims to further embed RI in CDTs' working practice and culture through peer-to-peer learning, as well as priming research leaders of the future with the skills to facilitate conversations with their peers about RI.

3.6.2. Content

The RI programme equips PhD students to:

- Understand the relevant frameworks for RI, why research funders might think it a necessary part of their training, and what it might mean for them as PhD students
- Develop skills and understanding in anticipation, reflection and engagement
- Map out the potential impacts of their research project, and broader field

First-year students in the CDTs attend a core programme of training: Introduction to Responsible Innovation, Introduction to Public Engagement, Responsible Innovation & Anticipation and Mapping Social & Ethical Issues. A programme of optional Extension Workshops is offered that allows students to dig in to "cross-cutting" themes and issues arising out of the RI programme. Extension workshops for this year's programme include Reflexivity for Scientists and Engineers, Giving Nature a Voice in our Research & Artefacts from the Future - a series of practical workshops in which students work with an artist to develop speculative objects for a museum in the far future. Throughout their PhD process, students can book 1-1 sessions with the RI Lecturer to discuss how they might explore RI in their PhD process and can access support from the Public Engagement team to plan and deliver engagement activities.

3.6.3. Assessment

Assessment has not been part of this training for all but one of the CDTs, for two key reasons. Firstly, the nature of the training is not didactic, but co-creative, thus traditional forms of assessment wherein the student's success or otherwise in understanding and synthesising the information imparted by the lecturer would not be applicable. Furthermore, the numbers of students enrolled in the programme of RI training across the 8 CDTs (approximately 100 students a year) would have made it logistically impossible for the RI lecturer to provide adequate marking and feedback in a timely fashion.

One CDT has found a way to include RI as a part of their formal assessment structure. In this CDT, students undertake a project module in their first year and use this experience to write a "Project Proposal" for their PhD which then forms the summative assessment for the module. As part of this proposal, they are required to discuss "how Responsible Innovation must be considered in the planned research". The students receive the following provocations to guide their discussion on RI:

- Although it's really early days, think about the possible future impacts of your PhD research project. What do you hope it could lead to and what might change in the world in the next 30 years as a result of your research, however small? (you might look at past research

breakthroughs in your area or similar areas and how they have effected change in the world).
- Think about the ways your research could become entangled (ethically, politically, socially or environmentally) and who (or what) could be affected. How plausible are the scenarios you are imagining?
- Based on some of the ideas discussed in the responsible innovation sessions, describe one practical thing you could do next year to understand and explore these entanglements better, and how you would feed this back into your research.

The RI Lecturer provides feedback on this aspect of the project proposal, shares reading related to the issues and ideas they have identified and suggests how they might explore them further, as well as extending an invitation to book 1-1 time. In this sense, this part of the assessment that focuses on RI, is both summative and formative, providing a space to consolidate what they have learned and discovered thus far, as well as pointing them towards further work they could do.

3.6.4. Critical reflection

This form of assessment works well in this limited context. One of the challenges of the RI programme not being formally assessed is that this status can have an impact on students' motivation to attend and engage fully, when they are overwhelmed by a large number of assessed modules in their first year. For students in the CDT in this case study, RI is presented as integral to the development of their project proposal, and one would hope that having experience in this at a crucial stage in their early development as researchers, they will embed these considerations in to how they conceive and write project proposals in future.

However, as indicated previously, were all the CDTs enrolled in the programme to decide on this form of assessment, the amount of feedback that would then be required of the RI Lecturer would not be achievable. In future programmes, in an imaginary where the teaching and assessment of RI is more generously resourced, it would be interesting to explore a more substantial "Impact Plan" as implemented by colleagues elsewhere in this paper. A small number of PhD students involved in previous RRI training programmes within EC & EPSRC funded projects at the institution have included a chapter in their thesis on the RRI implications of their projects, and these have been well received by examiners. Whilst the RI Lecturer cannot mandate that this happens, and there are many complexities involved in the positionality of the RI Lecturer in relation to the CDT students, not being their supervisor, nor even in their field of research, these students' experiences are shared as part of the introductory workshop and support is offered should students wish to take this route.

3.7. Assessment at the UCL RRI Hub - (in alphabetical order: Saheli Datta Burton, Stephen Hughes, Melanie Smallman, Jack Stilgoe)

The RRI Hub at UCL ('the Hub' hereafter) based at the Department of Science and Technology Studies (UCL-STS) provides a diverse range of RRI-focused trainings and workshops intended for doctoral students (including those in EPSRC CDTs) as well as professionals both within and beyond university settings. Since its setup in 2019, the Hub has expanded its offerings considerably with its current CDT training portfolio alone spanning eleven of UCL's EPSRC funded CDTs including those jointly run with (and involving CDT students at) the Universities of Cambridge (Center for Electronic & Photonic Systems), Cardiff (Compound Semiconductor Manufacturing), Loughborough (Energy Resilience and the Built Environment), Manchester (BioDesign) and Imperial College London (BioDesign). This training is offered at three levels, aimed at the three years of PhD study; basic (level 1) for first year PhDs, intermediate (Level 2) for second years and advanced (Level 3) for third years. Beyond this CDT portfolio, the Hub's executive education offer not only trains interested professionals of varying seniority across diverse sectors (including transport, health, banking, and construction) but also (non-CDT based) doctoral cohorts at British universities (including at UCL).

An inevitable result of this diversity of offerings is substantial pedagogic cross-pollination across the schemes. We outline, below, the three approaches to assessments currently in use in CDT trainings at the Hub.

3.7.1. Assessment type 1 (summative): 1500 to 2000-word report

Assessment type 1 (see Fig. 2) is aimed at students across the engineering disciplines and is one of the assessed (marked with feedback) components of a standard credit-bearing module comprising:

- 10 lectures at the taught Undergraduate or Postgraduate levels
- or a shorter format of 5 lectures aimed at third year doctoral students (CDT Level 3; starting end of May 2023 in the current 2022-23 academic year)

In this assessment type, students are expected to draw on various concepts, theories and class discussions of real-world examples related to RRI in particular and STS more broadly that are taught in class to present an analysis of how these apply (or may be applied) to their own research (for doctoral students) or areas of interest (for undergraduate or masters students). This presentation is formally assessed (i.e., marked and tutor feedback provided) and is expected to reflect the extent to which students understand and implement the 'Anticipation' and 'Reflection' components of the EPSRC AREA framework.

Likewise, feedback on the presentation from other students in the class – as a proxy for engagement with diverse stakeholder in real-world settings – is an important component of students understanding and implementation of the Stakeholder Mapping and Engagement component of the AREA framework. Finally, students are expected to integrate this feedback into the 'Individual Report' (see Fig. 2) and consider meaningful and doable action items to show how their research project (or area of research interest) follows socially responsive trajectories of technology development.

In summary, a key learning outcome of this assessment is to go beyond familiarising students with the conceptual aspects of RRI to considering the real-world challenges and processes of implementing RRI in practice.

3.7.2. Assessment type 2 (formative): Foresight reports

Assessment type 2 is aimed at doctoral (including CDT) students and professionals across multiple disciplines (hereafter 'learners'). It is designed to assess a learner's ability to provide 'foresight' on RRI considerations applied to their own area of work or research.

Having completed 3 weeks of the course, learners should be able to think about responsible innovation to a sufficient level of depth and nuance. Across three weeks, they critically reflected on the role that values play in science and technology, they considered the power of science and the ethical and social responsibilities that this confers on researchers, and they acknowledged the ambiguity and uncertainty involved in anticipating the intended and unintended impacts of science and the difficulty in identifying who ought to be responsible.

These insights have equipped learners with the tools required to anticipate and reflect on the impact and value of their work within a context of uncertainty and change. The foresight assessment offers them an opportunity to apply this knowledge to their own research. A foresight report is a document which anticipates and identifies a range of plausible future possibilities that are linked to research or innovation. It involves reflecting on the values that are embedded in those futures, evaluating them, and recommending how the research can respond to these different scenarios.

The learners are asked to write a 1000-word foresight report based on their current work. We recommend that they structure the report as follows:

Introduction - an overview of your research/innovation and what the reader can expect in your foresight report.

Fig. 2. Assessment type 1 (January – April 2023).

Anticipation - identification of key possible and plausible futures that may play out as a result of your work. Possible and plausible futures are the range of future scenarios that you believe your research or innovation may bring about. These scenarios are typically socio-technical, meaning that they involve social and technical (or scientific) elements. Your anticipation should clearly describe how the possible futures may arise from current research or innovation.

Reflection - evaluation of the values that underpin the scenarios you have identified, above. This will involve identifying the various personal, interpersonal, and/or contextual (social) values that are relevant to the scenarios and how they might be ordered and prioritised.

Response and summary - identify what action you can take in the present to ensure that your work is aligned with public needs and avoids public harms. This will involve thinking about the future scenarios that your research might bring about, and the various values that are involved, and considering what actions you need to take in the present to prepare.

Learners are assessed using the assessment criteria, below (see Fig. 3).

3.7.3. Assessment type 3 (formative): Asynchronous component with synchronous sessions

Assessment type 3 is aimed primarily at doctoral (including CDT) students across multiple disciplines and across various years (levels) of PhD study, although it has so far remained confined to the engineering and life sciences disciplines. The asynchronous component provides a basic (Level 1) introduction to RRI and its salience for students through a discussion of the various societal implications of controversial real-world applications, cases and scenarios. This introduces students to RRI rather than assessing their understanding, which is the focus of the synchronous session that follows.

The 'live' session uses group work specifically role playing in groups (as part of a light-touch case study method of teaching). The aim is to engage students with salience of anticipation, the complexity of diverse stakeholder perspectives and the necessity of public engagement as a key mechanism for (re)building public trust in their field of emerging research, but also science and expertise more broadly.

Responsible Innovation Foundations – Foresight Report Assessment Criteria

UCL

Criteria Title	Very Poor	Poor	Satisfactory	Good	Very Good
Anticipation	There is no potential research impact and corresponding future identified. The anticipation section may be incoherent and chaotic. An incoherent and chaotic section has no structure and involves numerous errors in reasoning.	Potential research impacts and the futures they could bring about are not clearly identified. Either the description of the research is unclear, the outline of the possible impacts is unclear, or the links between the two are unclear.	Potential research impacts and the futures they could bring about are clearly identified. A clearly identified future will make a direct link between the research and its future impacts. It will not try to predict a singular future impact but explore several possible impacts and perspectives.	Potential research impacts and the futures they could bring about are clearly identified and plausible. A possible future that is plausible is one that provides enough detail and supporting evidence or argument to be credible and justified.	Potential research impacts and the futures they might bring about are clearly identified, plausible, and compelling. A compelling future is novel and insightful. It is thought-provoking and inspiring. It may also be difficult to refute. The report includes intended as well as unintended impacts.
Reflection	The reflection does not identify values or potential bias at all. The reflection section may be incoherent and chaotic. An incoherent and chaotic section has no structure and involves numerous errors in reasoning.	The reflection does not clearly or accurately identify values or the potential for bias.	The reflection clearly and accurately identifies values and the potential for bias. Clearly and accurately identified values and biases will be specific, appropriate, and relevant.	The reflection clearly and accurately identifies values and the potential for bias. It examines the historical context of values. An examination of historical context considers how values have been shaped by research and innovation in the past and how it may change in the future.	The reflection clearly and accurately identifies values and the potential for bias. It examines the historical context of values. Values are critically evaluated with sophistication and nuance. A sophisticated and nuanced evaluation is complex and sensitive to detail. It may demonstrate a capacity to work within contexts of contradiction and uncertainty.
Response	The response does not clearly identify any actions that can be taken at all. The response section may be incoherent and chaotic. An incoherent and chaotic section has no structure and involves numerous errors in reasoning.	The response does not clearly identify appropriate actions that can be taken in the present.	The response clearly identifies appropriate actions that can be taken in the present. Clearly identified and appropriate actions are specific and detailed and likely to resolve the issues identified in the Anticipation and Reflection sections.	The response clearly identifies appropriate actions that can be taken in the present. It evaluates who is responsible. An evaluation of who is responsible considers whose values are embedded in the research in the first place and whose values should be embedded in the response.	The response clearly identifies appropriate actions that can be taken in the present. It evaluates who is responsible. The response is inventive and creative. An inventive and creative response offers new perspectives on the issues raised, offering novel, interdisciplinary, or unexpected solutions to the issues identified.

Fig. 3. Assessment criteria for assessment type 2 (November – December 2022).

3.7.4. Concluding reflections

One piece of frequent recurring feedback that we received from students that did not necessarily come across in assessments is that they found the content to be "thought-provoking". This came in form of feedback and evaluation about the courses. One avenue for further inquiry might be to think about how" thought provocation" might be evaluated. This leads to a second point that we struggled with when designing assessments and linking them to learning outcomes. One of our core aims has always been to find a way of bringing students to care about responsibility in relation to their work. This is a difficult thing to learn or assess. Some students seem to already arrive at our classes caring about ethical issues while others approach responsibility as another hurdle in the way of their PhD work. Care is something that has been discussed a lot in responsible innovation literature (Adam & Groves, 2011), but we are still struggling to find a way to systematise it. Perhaps, of course, care should not be systematically taught so much as cultivated. And, of course, caring cultures require careful thought and consideration in relation to issues of power and the distribution of ethical labour (Davies & Horst, 2015; Politi & Grinbaum, 2020). The fact that the majority of the learners to whom we teach responsibility also happen to be the ones with the least amount of power to actually change things is not lost on us. Nor is the fact that responsible innovation is most effective when it is enacted systematically (Spruit et al., 2016) which problematises the idea of individual assessments in the first place. As a result of these considerations, we feel that robust discussion continues to be the best way to facilitate and encourage careful reflection on responsibility in a way that can empower while maintaining perspective on broader structures of power.

At the same time, we recognise the concomitant need for those with greater power in the research and innovation ecosystem (than doctoral researchers) to co-enact systemic changes. A meaningful mechanism for doing this would be for those with relatively greater power such as PhD supervisors, principal (or co) research investigators, CDT directors, faculty heads and so on to embed meaningful RI as an integral part of core PhD studies. So far, RRI training has largely remained on the periphery of PhD curricula - a mandatory checkbox exercise - rather than a key aim of doctoral research systems. However, our experience of teaching and assessing learner's understanding or implementation of RI in practice emphasize the salience of the latter. For instance, PhD learners with supervisors closely involved in enacting *RI in practice* (e.g., through ongoing stakeholder engagement especially in the earlier stages of research) revealed not only a keener interest in RI during training but also often related how they were emulating their supervisor in integrating RI in their own research. First, this suggests that through their own practice of RI, senior actors in the research ecosystem often serve as compelling and much-needed examples of how RI might look like in practice (in the laboratory etc). This inevitably makes it easier for early career researchers to emulate RI in practice from early on. Second, it further suggests that embedding a culture of RI as an integral part of 'core' PhD training, rather than as just another checkbox exercise, is how RI can play a meaningful role in effecting systemic change.

3.8. Values-based planning as a mechanism to develop, support and assess RRI (Christian Wagner, Michael Smith)

3.8.1. Assessing RI as part of a stakeholder-values framework

Examples of RI assessment in this paper have explored a variety of systematic approaches to assessing RI at the doctoral level. Complementing these practical examples is a process being developed at the University of Nottingham which revolves around a stakeholder-values based planning framework and assessment tool for natural resource management (Wallace et al., 2016). We include this approach in the paper here as it highlights the possibility of overtly training RI as a situated process which is subjective to the specific societal culture of its stakeholders.

3.8.2. Process overview

Within the planning and in particular, the environmental management literature, values-based planning has been shown to provide an efficient mechanism to gather often disparate and conflicting views of multiple stakeholder groups and transparently leverage the resulting information to drive planning decisions (Wallace et al., 2022). How stakeholders value a natural system is directly affected by the state of key system elements (e.g., wildlife, lakes). The state of an element is measured by element properties (e.g., number of species, quality of the lake's water) and the state of any system element is altered by managing system processes (e.g., predation rates on wildlife, filtration of water entering the lake). Benefits of a values-based framework, in particular such as introduced by Wallace et al, include:

- The acknowledgement and mapping of conflicting stakeholder views
- Recognition of uncertainty and vagueness in stakeholder priorities – as well as in traditionally quantitative information on the state of key system elements (i.e., species diversity, water depth), processes (i.e., water salinisation or fire intensity) and end-state values (i.e., level of recreational satisfaction),
- Overt and transparent tracing of management decisions to stakeholder values priorities and associated improved explanation of management actions to stakeholders, and
- More efficient and effective stakeholder engagement, reducing stakeholder fatigue and disconnect between managers, planners and stakeholders (Smith et al., 2016).

The core steps of values-based planning are (Wallace et al., 2022):

1 Elicitation and joint definition of values as end-states of human well-being (e.g. the need to live in a suitably benign environment, fulfilment of aesthetics enjoyment, having a meaningful occupation, social, philosophical, and/or spiritual fulfilment) and enduring beliefs concerning the preferred ethical properties of human behaviour (e.g. Equity, across all stakeholders or treating wildlife in a humane manner).
2 Elicitation of the importance of each value to stakeholders or stakeholder groups within the context of the management system of interest
3 Stakeholder assisted identification, definition, and prioritisation of the system elements to be managed. For example, in a natural resource management context, this may comprise sets of biotic and/or abiotic elements including meaningful groupings of local wildlife (e.g., the waterbirds or the small mammals).
4 Elicitation of the properties used to define the state of the elements and processes which can be subject to management. For example, the abundance, health or diversity of specific biotic elements or the integrity of a dam wall.
5 Mapping of the function between different possible element states and stakeholder valuation of the element based on stakeholder elicitation and/or literature. For example, the more species diverse (a property) a grassland (an element), the more aesthetically pleasing (a value) humans may perceive it.
6 Establish the current state of the properties of the elements (with uncertainty) and expected current value to the stakeholders.
7 Assessment of the risk that each element is exposed to (e.g., with current management, will the element remain in a state that is sufficiently valued by the stakeholders over the management period).
8 Identification and evaluation of each possible management action based on its expected impact on element state (e.g., expected change in abundance of a given species) and associated valuation response by stakeholders in order to choose management actions which are most likely to best deliver against stakeholders' priority values.

3.8.3. Why and how values-based planning may support RRI?

The nature of what responsible research and innovation is, is

subjective, driven by local cultural and social norms. This poses an obvious challenge to the fostering—and assessment of RRI: how do we support and assess something without a clear, a priori definition?

It is here that values-based planning may provide valuable mechanisms. First, the elicitation and importance rating of values with stakeholders (Step 1) in the context of RRI, provides a direct approach to identify and engage the set of stakeholders in relation to the research questions, helping to establishing how the research might support or undermine the extent to which the stakeholders might value the management system at the end of the management period. Step 2, above, provides a mechanism to capture discord and variation between stakeholders in respect to the importance of individual values.

Overall, the values-elicitation framework provides a goal-setting exercise from an RRI perspective, allowing an assessment of how the research and its methodologies might affect the state of key system elements, from which expectations around the expected stakeholder 'value' response can be anticipated.. By mapping (Step 5) the way stakeholders might value different elements in response to the adoption of specific research methodologies and resources, researchers and broader research stakeholders can chart how, to what degree, and when priority values will be realised – and for which stakeholder(s).

This in turn, provides a potential pathway for a direct means of evaluating and assessing RRI, based on how well given research will affect ongoing management that is delivering or has delivered the priority values of stakeholders.

3.8.4. Summary

Values-based planning approaches are designed to support planning at the complex confluence of often uncertain and conflicting stakeholder priorities and management, e.g., by linking the management of element state to stakeholder wellbeing via realisation of end-state values. The mechanisms developed for values-based planning can provide a toolset for the support, development, engagement with, and assessment of RRI in particular because they encompass an initial values-elicitation stage which overtly lays out the priorities for the given planning or research process. This can help address the key challenge of defining what 'responsible' RRI is within a local (stakeholder) context, and how it may be achieved, providing vital anchoring for the fostering and subsequent assessment of RRI. At the student training level, this stakeholder-values process can provide students with a tool to elucidate and assess RRI throughout their own careers and domains of working, while also providing a direct means of assessing the degree to which their own doctoral research delivers the respective end-state and behavioural values.

4. Discussion and conclusion

The examples collected in the previous section show a broad range of approaches to RI across disciplines and institutions. They demonstrate the richness of the RI discourse which underpins the way RI is taught to PhD students and others. They also show the breadth of approaches to RI assessment in doctoral education which ranges from very light-touch self-reported measures to formative and summative assessments. The integration of the assessment into the broader doctoral training structure also varies widely. We hope that these examples can provide inspiration to individuals and institutions who are tasked with the development of curricula for doctoral training and the integration of assessment.

The examples furthermore demonstrate that many of the issues and questions we listed in the overview remain unresolved and can create challenges for RI training and assessment. Our examples show that there remains a tension between the cognitive component of RI which can be taught and assessed in an abstract way and the desired change of social practice which is a long-term goal and more difficult to assess.

It is not the purpose of this editorial article to provide one model of RI training or assessment, but to offer an array of options that can inspire ideas and indicate good practice. We hope that the examples listed in the previous section have achieved this aim. At the same time they show that there are a number of fundamental and structural issues that need to be kept in mind when considering the future of RI training and assessment.

Key among these considerations is the role of RI in the broader research and research training landscape. Many of the RI-related policies and requirements are driven by research policy and implemented by research funders. The attention that these funders pay to RI and the terminology they use to express their aims changes over time. In the UK, for example, commitment by funding bodies to RI remains strong, as evidenced by the requirement to include it into the next round of CDTs. The EU, on the other hand, which for a long time was the main proponent and funder of RRI puts much less emphasis on it in the current research framework programme Horizon Europe. Such research policy and funding principles have important consequences across the research ecosystem, as they set priorities and incentive structures. This directly impacts questions such as those around the availability of resources for RI. We have seen that some of the RI assessment approaches are labour intensive and only sustainable, if significant resources are dedicated to them.

Probably even more important than the allocation of resources to RI is the status it has in the research and innovation ecosystem. Some of the authors of our examples have pointed to the limits of their influence in shaping and realising RI training and assessment. A higher level of recognition of RI can motivate more senior researchers to engage with it which can have positive motivating effects on early career researchers including PhD students. More senior involvement can also pave the way for a more organic integration of RI, avoiding the perception that it is a specialist skill that can be tested through one-off simple assessments. We have seen that there are questions around power and its relative differential between doctoral researchers, supervisors, investigators, programme directors and other parts of higher education structures are important influencing factors that shape the content of RI training and assessment.

RI trainings and assessment of doctoral researchers can therefore be seen as one aspect of the broader integration of RI at an institutional level which, in turn, is one aspect of the integration of RI into national and international research and innovation ecosystems. This higher-level RI integration is clearly of crucial importance for the practice of RI by individual researchers. It therefore stands to reason that assessing this integration of RI on the institutional and ecosystem levels is a task that is at least as important as the assessment of individual RI training. The assessment of RI training for doctoral students will likely be one component of such higher-level assessment, and it will be a necessary but not a sufficient condition for RI to have an impact. We therefore hope that this editorial not only provides inspiration for the assessment of RI training for doctoral researchers, but also contributes to the broader discussion of the integration of RI in the research and innovation ecosystem and fundamental political changes that would require.

Declaration of Competing Interest

No conflict of interest of any of the authors.

Acknowledgments

The Horizon Centre for Doctoral Training is supported by: UKRI Grant No. EP/L015463/1 (EPSRC Horizon CDT in My Life in Data), UKRI Grant No. EP/S023305/1 (EPSRC Horizon CDT in Creating Our Lives in Data). This work was supported by the UKRI grants: EP/T022493/1 (Horizon: Trusted Data-Driven Products) and EP/V00784X/1 (TAS Hub). Christine Aicardi's work is supported by UK Research and Innovation [grant number EP/S023356/1 UKRI CDT in Safe and Trusted Artificial Intelligence] and by the Engineering and Physical Sciences Research Council [grant number EP/S022104/1 EPSRC CDT in Smart

Medical Imaging]. Vivienne Kuh's work is supported by the in-kind contribution of University of Bristol to the following EPSRC CDT grants: EP/S021728/1 CDT in Composites Science, Engineering and Manufacturing, EP/S021795/1 CDT in in Future Autonomous Robotic Systems (FARSCOPE-TU: Towards Ubiquity), EP/S023593/1 CDT in Aerosol Science, EP/S024107/1 CDT in Technology Enhanced Chemical Synthesis, EP/S022937 CDT in Interactive Artificial Intelligence, EP/S023275/1 CDT in Future Innovation in Non-Destructive evaluation (FIND), EP/S023607/1 CDT in Quantum Engineering EP/S023569/1 CDT in Computational Statistics and Data Science: COMPASS. The UCL RRI Hub is supported by the following: UKRI EPSRC Responsible AI for Long-term Trustworthy Autonomous Systems (RAILS): Integrating Responsible AI and Socio-legal Governance (EPSRC Grant Ref: EP/W011344/1); EPSRC & SFI CDT in Energy Resilience and the Built Environment with the University of Loughborough (EPSRC Grant Ref: EP/S021671/1); EPSRC & SFI Centre for Doctoral Training in Transformative Pharmaceutical Technologies (EPSRC Grant Ref: EP/S023054/1); EPSRC CDT in AI-enabled Healthcare Systems (EPSRC Grant Ref: EP/S021612/1); EPSRC CDT in Geometry and Number Theory at the Interface (known as 'The London School of Geometry and Number Theory') (EPSRC Grant Ref: EP/S021590/1); EPSRC CDT in Bioprocess Engineering Leadership (EPSRC Grant Ref: EP/S021868/1); EPSRC CDT in BioDesign Engineering with the University of Manchester and Imperial College (EPSRC Grant Ref: EP/S022856/1); EPSRC CDT in Delivering Quantum Technologies (EPSRC Grant Ref: EP/S021582/1); EPSRC CDT in Cybersecurity (EPSRC Grant Ref: EP/S022503/1); EPSRC CDT in Intelligent, Integrated Imaging In Healthcare (EPSRC Grant Ref: EP/S021930/1); EPSRC CDT in Compound Semiconductor Manufacture with the University of Cardiff (EPSRC Grant Ref: EPS024441/1); EPSRC CDT in Foundational Artificial Intelligence (EPSRC Grant Ref: EP/S021566/1); EPSRC CDT in Connected Electronic and Photonic Systems with the University of Cambridge (EPSRC Grant Ref: EP/S022139/1). The work undertaken by Bernd Stahl, Inga Ulnicane, Damian Eke and Paul Keene has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement no. 945539 (HBP SGA3). We would like to thank Adrian Hazzard for his input on the Horizon CDT case study.

References

Aboab, J., Celi, L. A., Charlton, P., Feng, M., Ghassemi, M., Marshall, D. C., Mayaud, L., Naumann, T., McCague, N., Paik, K. E., Pollard, T. J., Resche-Rigon, M., Salciccioli, J. D., & Stone, D. J. (2016). A "datathon" model to support cross-disciplinary collaboration. *Science Translational Medicine, 8*, 333ps8. https://doi.org/10.1126/scitranslmed.aad9072. -333ps8.

Adam, B., & Groves, C. (2011). Futures tended: care and future-oriented responsibility. *Bulletin of Science Technology & Society, 31*, 17–27. https://doi.org/10.1177/0270467610391237

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica*. URL https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 4.30.23).

Davies, S. R., & Horst, M. (2015). Crafting the group: Care in research management. *Soc Stud Sci, 45*, 371–393. https://doi.org/10.1177/0306312715585820

EPSRC, (2022). EPSRC centres for doctoral training [WWW Document]. URL https://www.ukri.org/opportunity/epsrc-centres-for-doctoral-training/ (accessed 5.2.23).

European Commission. (2013). *Options for strengthening responsible research and innovation (Report of the expert group on the state of art in europe on responsible research and innovation)*. Luxembourg: Publications Office of the European Union.

European Commission. (2012). *Responsible Research and Innovation - Europe's ability to respond to societal challenges*. Brussels: European Commission, Publications Office.

Gama, K., Alencar Gonçalves, B., & Alessio, P. (2018). Hackathons in the formal learning process. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE 2018* (pp. 248–253). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3197091.3197138.

Gerrits, E. M., Bredenoord, A. L., & van Mil, M. H. W. (2021). Educating for responsible research practice in biomedical sciences. *Science & Education*. https://doi.org/10.1007/s11191-021-00295-y

Greenhalgh, C., Craigon, P., Portillo, V., Dowthwaite, L., Perez, E., Webb, H., Wagner, H. G., & Stahl, B. C. (2022). Responsible Research and Innovation (RRI) prompts and practice cards (version 2.1, August 2022). 10.17639/nott.7243.

Hesjedal, M. B., Åm, H., Sørensen, K. H., & Strand, R. (2020). Transforming scientists' understanding of science–society relations. stimulating double-loop learning when teaching RRI. *Science and Engineering Ethics, 26*, 1633–1653. https://doi.org/10.1007/s11948-020-00208-2

James, A. R. (2013). Lego Serious Play: A three-dimensional approach to learning development. *Journal of Learning Development in Higher Education*. https://doi.org/10.47408/jldhe.v0i6.208

Jirotka, M., Grimpe, B., Stahl, B., Hartswood, M., & Eden, G. (2017). Responsible research and innovation in the digital age. *Communications of the ACM, 60*, 62–68. https://doi.org/10.1145/3064940

Komssi, M., Pichlis, D., Raatikainen, M., Kindström, K., & Järvinen, J. (2015). What are hackathons for? *IEEE Software, 32*, 60–67. https://doi.org/10.1109/MS.2014.78

Owen, R. (2014). The UK engineering and physical sciences research council's commitment to a framework for responsible innovation. *Journal of Responsible Innovation, 1*, 113–117. https://doi.org/10.1080/23299460.2014.882065

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy, 39*, 751–760.

Owen, R., von Schomberg, R., & Macnaghten, P. (2021). An unfinished journey? Reflections on a decade of responsible research and innovation. *Journal of Responsible Innovation, 8*, 217–233. https://doi.org/10.1080/23299460.2021.1948789

Patel, M., Webb, H., Jirotka, M., Davoust, A., Gales, R., Rovatsos, M., & Koene, A. (2019). Harnessing interdisciplinarity to promote the ethical design of AI systems, in: ECIAIR 2019 European Conference on the Impact of Artificial Intelligence and Robotics, Oxford: UK. pp. 246–253.

Politi, V., & Grinbaum, A. (2020). The distribution of ethical labor in the scientific community. *Journal of Responsible Innovation, 7*, 263–279. https://doi.org/10.1080/23299460.2020.1724357

Portillo V., Greenhalgh C., Craigon P.J., Holter C.T., 2023. Responsible Research and Innovation (RRI) prompts and practice cards: A tool to support responsible practice. In *First International Symposium on Trustworthy Autonomous Systems (TAS '23)*, July 11-12, 2023, Edinburgh, United Kingdom. ACM, New York, NY, USA, 7 Pages. https://doi.org/10.1145/3597512.3599721.

QAA. (2020). *Characteristics Statement: Doctoral Degree*. Quality Assurance Agency.

QAA. (2014). *The Frameworks for Higher Education Qualifications of UK Degree-Awarding Bodies*. Quality Assurance Agency.

Randles, S., Tancoigne, E., & Joly, P.-B. (2022). Two tribes or more? The historical emergence of discourse coalitions of responsible research and innovation (rri) and Responsible Research and Innovation (RRI). *Journal of Responsible Innovation, 9*, 248–274. https://doi.org/10.1080/23299460.2022.2061306

Rasmussen, R. (2006). When you build in the world, you build in your mind. *Design Management Review, 17*, 56–63.

Rome Declaration, (2014). Rome declaration on responsible research and innovation in Europe.

Shanley, D. (2021). Imagining the future through revisiting the past: The value of history in thinking about R(R)I's possible future(s). *Journal of Responsible Innovation, 8*, 234–253. https://doi.org/10.1080/23299460.2021.1882748

Smith, H. (2020). Algorithmic bias: Should students pay the price? *AI & Soc, 35*, 1077–1078. https://doi.org/10.1007/s00146-020-01054-3

Smith, M. J., Wagner, C., Wallace, K. J., Pourabdollah, A., & Lewis, L. (2016). The contribution of nature to people: Applying concepts of values and properties to rate the management importance of natural elements. *Journal of Environmental Management, 175*, 76–86. https://doi.org/10.1016/j.jenvman.2016.02.007

Spruit, S. L., Hoople, G. D., & Rolfe, D. A. (2016). Just a cog in the machine? The individual responsibility of researchers in nanotechnology is a duty to collectivize. *Science and Engineering Ethics, 22*, 871–887. https://doi.org/10.1007/s11948-015-9718-1

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*, 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Ten Holter, C., Stahl, B., & Jirotka, M. (2022). Built in, not bolted on: Responsible innovation in UK CDTs - the new normal? Learning Organization.

Tokalić, R., Buljan, I., Mejlgaard, N., Carrió, M., Lang, A., Revuelta, G., Marušić, A., & HEIRRI Consortium. (2021). Responsible research and innovation training programs: Implementation and evaluation of the HEIRRI project. *Forensic Sciences Research*, 1–11. https://doi.org/10.1080/20961790.2021.1970319

Urquhart, L. D., & Craigon, P. J. (2021). The Moral-IT Deck: A tool for ethics by design. *Journal of Responsible Innovation, 81*, 94–126. https://doi.org/10.1080/23299460.2021.1880112

Von Schomberg, R. (2013). A vision of responsible research and innovation. In R. Owen, M. Heintz, & J. Bessant (Eds.), *Responsible innovation* (pp. 51–74). Wiley.

Von Schomberg, R. (2011). *Towards responsible research and innovation in the information and communication technologies and security technologies fields*. Luxembourg: Publication Office of the European Union.

Wallace, K. J., Wagner, C., Pannell, D. J., Kim, M. K., & Rogers, A. A. (2022). Tackling communication and analytical problems in environmental planning: Expert assessment of key definitions and their relationships. *Journal of Environmental Management, 317*, Article 115352. https://doi.org/10.1016/j.jenvman.2022.115352

Wallace, K. J., Wagner, C., & Smith, M. J. (2016). Eliciting human values for conservation planning and decisions: A global issue. *Journal of Environmental Management, 170*, 160–168. https://doi.org/10.1016/j.jenvman.2015.12.036

Webb, H., Jirotka, M., Inglesant, P., & Patel, M. (2019). Human centred computing approaches to embed responsible innovation in HCI.

Zapico, J. L., Pargman, D., Ebner, H., & Eriksson, E. (2013). Hacking sustainability : Broadening participation through Green Hackathons. In *Presented at the Fourth International Symposium on End-User Development*. Denmark: IT University of Copenhagen. June 10-13, 2013.

Bernd Carsten Stahl[a,*], Christine Aicardi[b], Laurence Brooks[c], Peter J. Craigon[a,d], Mayen Cunden[e], Saheli Datta Burton[f], Martin De Heaver[g], Stevienna De Saille[h], Serena Dolby[g], Liz Dowthwaite[a,d], Damian Eke[i], Stephen Hughes[f], Paul Keene[g], Vivienne Kuh[j], Virginia Portillo[a,d], Danielle Shanley[k], Melanie Smallman[f], Michael Smith[l], Jack Stilgoe[f], Inga Ulnicane[m], Christian Wagner[n], Helena Webb[a,d]

[a] *School of Computer Science, University of Nottingham, UK*
[b] *Dept. of Informatics and School of Biomedical Engineering & Imaging Sciences, King's College London, UK*
[c] *Information School, University of Sheffield, Sheffield, UK*
[d] *Horizon Digital Economy Research, School of Computer Science, University of Nottingham, UK*
[e] *Independent Researcher*
[f] *Dept of Science and Technology Studies, University College London*
[g] *ORBIT RRI limited, Leicester, UK*
[h] *Dept of Sociological Studies, University of Sheffield, Sheffield, UK*
[i] *Centre for Computing and Social Responsibility, De Montfort University, Leicester*
[j] *School of Chemistry, University of Bristol, UK*
[k] *Department of Philosophy, Faculty of Arts and Social Sciences, Maastricht University, The Netherlands*
[l] *Biodiversity and Conservation Science, Department of Biodiversity, Conservation and Attractions, Wildlife Research Centre, Western, Australia*
[m] *De Montfort University, UK*
[n] *Lab for Uncertainty in Data and Decision Making (LUCID), School of Computer Science, University of Nottingham, UK*

[*] Corresponding author.
E-mail address: bernd.stahl@nottingham.ac.uk (B.C. Stahl).