

Mapping Cross-Disciplinary Perspectives on Responsible Conduct of Research:

A Delphi Study

Stage 1 Registered Report proposal

Sarahanne M. Field, Jackie Thompson, Tom van Drimmelen, Jennifer Ferrar, Bart Penders,

Sarah de Rijcke, and Marcus R. Munafò

Abstract

Responsible conduct of research (RCR) is ~~generally agreed to be a laudable goal. It promotes high imperative for the quality and trustworthiness of research practices, which should lead to more credible findings, and instill confidence in the~~ and the health of the whole research community system. However, it is as yet unclear to what extent RCR differs across disciplines. Currently, many approaches to research and training in RCR are either ~~generalised across all disciplines, at a high level presented as universally applicable~~ (e.g., international frameworks on research integrity) or at the other extreme, discipline-specific. Relying on the expertise and knowledge of a carefully selected multidisciplinary panel of RCR scholars and practitioners, this Delphi study aims to expand the current (underspecified) frameworks of RCR to develop a more diverse and comprehensive concept of what constitutes RCR across disciplines, ~~along with. We will conduct a mapping that captures this updated understanding. The modified reactive Delphi process will begin with participants refining, in which panellists progressively refine their judgement of the importance of individual dimensions of RCR to their respective discipline, starting from~~ a provisional list of dimensions ~~of RCR collated derived~~ from previous literature and interviews, ~~then will proceed with several rounds of rating the importance of each dimension to particular disciplines.~~ [After completion of the study, we will report the details of participant numbers, rounds of Delphi, and a summary of results here.]

CRedit statement: [to be completed in final submission]

The research enterprise is an impressive, powerful engine, capable of generating huge global change as well as catastrophe, as it deals with uncertain facts, disputed values, and high stakes in the context of a need for urgent decisions (see Funtowicz & Ravetz, 2018). When scientific research becomes misaligned with the needs and expectations of society, at best, it *becomes irrelevant*, failing to adequately meet society's needs, and falling short of the promises it makes. At worst, it *becomes unethical*, putting the lives of people and the environment at risk.

Responsible conduct of research (RCR) promotes the production of robust and reliable knowledge, can be used to foster a healthy research culture, and ultimately will lead to the public's trust in the research process, its output, and its implementation. In short, conducting research activity responsibly is crucial to the health of the knowledge generation enterprise going forward. In practice, this is achieved by upholding or improving crucial principles such as transparency, accountability, and integrity in how research is conceived, carried out and communicated. However, these broad principles may manifest as different concrete behaviours, across different disciplines. Currently, however, we know little about how much what constitutes 'responsible conduct of research' should differ across disciplines.

Over time, a growing awareness of the heavy responsibility the scientific enterprise carries has led to the development of frameworks which underpin individual and institutional codes of research conduct. Almost 30 years ago, 'ELSA' (Ethical, Legal and Social Aspects of emerging science and technology) was first introduced as a framework for research development and funding, to reorient scientific research practices to make them more effective and – crucially – more ethical and self-aware. Due to its origins, state Zwart, Landeweerd and van Rooij (2014), ELSA applied largely to disciplines such as genomics and other life sciences, bioethics, science and technology studies, technology assessment, philosophy of science, and science communication. Since then, the framework has been built

upon and developed, making way for what is known as the Responsible Research and Innovation (RRI) framework, which is more widely applicable and more focused on addressing larger-scale socio-economic challenges than its predecessor ELSA ~~was~~.

As with ELSA before it, RRI and, derived from it, responsible conduct of research (RCR)¹ is institutionalised as an instrument of policy, rather than being a discipline in and of itself (Tallacchini, 2009). It refers to a methodological attitude to be applied to research conduct: a strategy to change how research is conducted in practice, regarding its responsibilities to society. Von Schomberg (2013) points out that there is no agreed-upon definition of what RCR is. ~~Rather, he invites readers to consider what RCR as a top-down signifier might denote, in relation to the disciplines and research processes with which it engages. Despite this declaration, on page 9 of a 2012 article~~ However, Von Schomberg does propose a somewhat concrete definition of RR(I), describing it as “a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society).” ~~Narrowing the scope of the)~~ (Von Schomberg 2012, p. 9). ~~Though we do not intend to propose a formal definition for the current of RCR – as exploring what constitutes RCR is the aim of this study, we consider responsible conduct of research (RCR) to be a topic – we would contend that requires the synthesis of many disparate aspects. While~~ it is distinct from concepts such as research integrity (RI) ~~or responsible research and innovation (or RRI), it is, although~~ no doubt closely related. We argue that RCR casts a broader net than the typical definitions of research integrity, that is, promotion of confidence and trust in research and the research process. ~~This~~

¹ Although the RRI framework involves both research and innovation, our focus is primarily on the research conduct aspect of the framework and will henceforth refer to RCR, the responsible conduct of research.

~~broader remit of~~ On the other hand, RCR includes dimensions that overlap with those of RRI, such as the responsibility obligation of research has for honest to deal honestly and transparent dealings transparently with citizens and society. However, while conceptualisations of RRI typically include impacts of technological innovations and research output on society, RCR concerns the subset of dimensions or responsibilities relating to the activities involved in conducting research.

~~Underpinned by a recent scoping review conducted by our research team (Field et al., 2024), in alignment with the arguments of many scholars and organisations gone before, we contend that RCR will promote the co-production of robust and reliable knowledge, can be used to foster a healthy research culture, and ultimately will lead to the public's trust in the research process, its output, and its implementation. In a previous~~ In short, conducting research activity responsibly is crucial to the health of the knowledge generation enterprise going forward.

~~In that same~~ scoping review study, with the view of eventually developing our own RCR framework, we made steps toward determining how RCR might vary across disciplines, and what elements are shared. We found that although some dimensions of RCR are viewed as more applicable to some disciplines than others (e.g., anticipation and transparency), others were applicable across many fields (e.g., integrity). We concluded that ~~communities of practice built around future RCR framework~~ efforts promoting RCR might benefit from updating and diversifying existing frameworks such that they allow room to accommodate different epistemological traditions and are sensitive to emerging fields which operate at the forefront of scientific advancement (such as genomics), or which are especially prone to ethical dilemmas (such as artificial intelligence) or both (such as nanotechnology).

~~This proposal concerns a~~ In this study that builds, we aim to build upon the foundation laid previous literature and guidelines to construct an updated collection of

dimensions of RCR and a mapping of how important they are to different research disciplines, including those traditionally underserved by our scoping review. Namely, we plan to conduct a existing frameworks. Using the Delphi study which, with the help of technique of structuring communication, we will enlist a multidisciplinary expert panel, will to help us pin down the most salient and crucial elements of a new mapping of RCR that descriptively lays out the key differences betweenmap the manifestations of RCR in different disciplines. The Delphi process will help us identify core tenets of RCR that go beyond the disciplines central to previous ~~RR~~RCR frameworks, along with more niche, discipline-specific elements. ~~We aim~~We consider the Delphi technique to develop be particularly suited for this goal, as it accommodates the inclusion of a framework, using these RCR dimensions, broad range of participants, and is likely to yield results that reflects the needs of the academic community, reflect the whole group's opinions (rather than capturing just the perspectives of the most vocal participants). It also ensures relative anonymity of responses, which tends to allow participants to disagree with the explicit intention of representing as many disciplines as possible. one another more freely.

We aim for our mapping to fill the gap between the two extremes that existing conceptualisations of RCR tend to fall under: either high-level frameworks designed to be universally applicable across all disciplines (e.g., the Singapore Statement on Research Integrity, the Australian Code for Responsible Conduct of Research, ~~or~~ the All European Academies European Code of Conduct for Research Integrity²), or prescriptive guides tailored to the practical instruction of researchers within a specific discipline or field (e.g., RCR training designed for members of a university department as part of a degree or continuing professional development, or mandated by funders such as the National Institutes

² The Singapore Statement can be found here: <https://www.wcrif.org/guidance/singapore-statement> and ALLEA's code of conduct here: <https://allea.org/code-of-conduct/>

of Health, or guidance from discipline-specific learned societies such as the Society for Improvement of Psychological Science). ~~The findings of this Delphi study will also be relevant to subsequent research on the dimensions of RCR. We note that while our goals involve the development of a novel RCR framework, the current registered report focuses only on the Delphi study. The outcome of the Delphi study will be the basis of the development of the later framework, which will be central to a separate, future article and details on how the framework will be developed are thus outside of the scope of this registered report.~~

Note that we do not share any explicit hypotheses at this juncture, where one might expect them. This is because the proposed study has the goal of developing our list of RCR dimensions, rather than testing any specific theory, and is exploratory and descriptive as a result. Providing that our methodology is sound, and the Delphi carried out as planned, valid and reliable results could take many forms.

~~Lastly~~Finally, it is also important to consider the role of this study as it forms a component of a larger multi-year project, which aims to broadly develop a diverse understanding of how RCR is conceptualised and applied across different research disciplines. The conceptual mapping we will have co-produced with the help of RCR experts during this Delphi study will form a scaffold for interaction with communities of practice in the project's latter half, helping them to contextualise where particular disciplines and practices sit in relation to others in the overall ecosystem of RCR. Its broad remit will also help to spotlight the perspectives of disciplines that have been more peripheral in discussions and evaluative frameworks on RCR so far. Thus, we aim for our mapping to stimulate a more nuanced understanding of cross-disciplinary conceptions of RCR within the communities that work to embed practices *in situ*. While our output can be used more broadly to assist other interested entities (such as individuals or research groups) in conceptualising and applying

RCR principles for their own needs, that is a secondary purpose. Therefore in our aim to generate a tool that presents a wide perspective on the RCR sphere, we have cast our epistemic net broadly.

Method

Ethics approval for this proposal was granted by the University of Bristol's School of Psychological Science Research Ethics Committee (reference number 12071), in the United Kingdom (October 14th, 2022) and by the University of Leiden, the Netherlands (January 19th, 2023). **The Present Study**

~~To try to develop a list of RCR dimensions that are practically applicable and relevant to a wide range of scientific disciplines without the explicit involvement of members of the scientific community would be remiss. In this study, we aim to build upon previous literature and guidelines, crystallising distributed community knowledge into an updated collection of dimensions of RCR and a mapping of how they apply to different research disciplines, including those traditionally underserved by existing frameworks. To achieve these goals, we plan to assemble a disciplinarily diverse Delphi panel of RCR experts, a selected group of individuals with experience in RCR frameworks, scholarship and/or practice. A pre-constructed reference document, comprising a proposed list of RCR dimensions, will be sent to our Delphi panel. The panel will be invited to suggest additions to the list, and later asked to judge each RCR dimension in the revised list on its importance to RCR within the panellist's discipline of expertise.~~

The Delphi technique is a method of facilitating and structuring group communication processes (Linstone & Turoff, 2002, Linstone & Turoff, 2011). The

Formatted: Font color: Auto

procedure aims to get a refined set of participant evaluations on a set of ‘target items’ (in our case, RCR dimensions) and is known for its ‘iterative’ nature. Participating panellists provide judgements (depending on the specific approach, these might be quantitative, qualitative or both), which are collected by researchers, summarised, and sent back to participants for subsequent judgement rounds. This evaluation-summary process is repeated for multiple rounds and ends when researchers are content with the resulting item list, or when some other stopping criterion (such as a pre-registered number of rounds) is reached.

~~We have chosen the Delphi approach as it accommodates the inclusion of a broad range of participants, and is likely to yield results that reflect the whole group’s opinions (rather than capturing just the perspectives of the most vocal participants). It also ensures relative anonymity of responses, which tends to allow participants to disagree with one another more freely.~~

We plan to use a so-called ‘reactive’ Delphi method, in which panellists respond to an existing pre-constructed reference document ([available on https://osf.io/irf47](https://osf.io/irf47)) instead of creating one themselves (Salkind, 2007, p. 243). However, in order to make the best use of their expertise, our Delphi panel will be encouraged to build upon the initial reference list with their own diverging ideas, suggesting missing RCR dimensions. In later rounds, the panellists will be asked to judge each RCR dimension in the revised list on its importance to RCR within the panellist’s discipline of expertise. As such, our Delphi study is both based on existing literature and perspectives (which are captured in the reference document), as well as being highly exploratory in that it seeks to chart out expert perspectives on a broader, multi-disciplinary RCR not yet represented in frameworks and guidelines.

Formatted: Font color: Auto

Formatted: Font color: Auto

While most Delphi approaches are consensus-based, meaning that they aim to converge on a selection of important elements of a reference document (Diamond et al., 2014; von der Gracht, 2012), our approach aims to map and refine the existing breadth of

perspective on various dimensions of RCR. It is important to note that though this does constitute a departure from the typically practiced Delphi, this actually is a reversion to how the Delphi process was originally intended. Multiple authors have commented that the value of a Delphi study lies exactly in mapping the distributions of opinions instead of generating consensus (Scheibe, 2002; Linstone and Turoff, 2011).

~~Our Delphi study is both based on existing literature and perspectives (which are captured in the reference document), as well as being highly exploratory in that it seeks to chart out expert perspectives on a broader, multi-disciplinary RCR not yet represented in frameworks and guidelines.~~

~~Method~~

~~Ethics approval for this proposal was granted by the University of Bristol's School of Psychological Science Research Ethics Committee (reference number 12071), in the United Kingdom (October 14th, 2022) and by the University of Leiden, the Netherlands (January 19th, 2023).~~

Initial Reference Document: A Proposed List of RCR Dimensions

Before constructing the reference document, i.e., a proposed list of RCR dimensions, the authors conducted a scoping review of the existing RCR literature (Field et al., 2024) and interviews with RCR scholars and practitioners (see the interview guide at <https://osf.io/xv98y>). The interviews consisted of 10 one-on-one interviews and one focus group which included two moderators and eight participants. The articles included in the scoping review and interview transcripts were subject to a thematic analysis, conducted by SMF, which involved coding topically salient sections of text and combining related codes into themes (see the appropriate appendix for a detailed description of this process at <https://osf.io/jrf47>). This analysis generated a series of overarching themes which reflected

Formatted: Font color: Auto

salient dimensions of RCR from the literature on the topic. These dimensions are the core of the initial reference list, to which dimensions taken from existing, older RCR frameworks were added (i.e., the Singapore Statement on Research Integrity, and the Australian Code for Responsible Conduct of Research).

It is important to note that the choice of dimensions, and their respective definitions, are not intended to represent an authoritative list of dimensions of RCR, nor the only way to carve up these concepts. Rather, they were designed to maximise the information gain from the Delphi process, by covering a broad range of concepts and reducing redundancy. For example, the dimension “integrity” carries many different connotations and facets, and many of these are already covered by other proposed dimensions gleaned from the research literature and interviews, such as “rigour,” “transparency,” and others. Therefore, we defined the dimension of “integrity” to cover a more constrained facet not already mentioned, concerning the possession of and adherence to strong moral principles. We also aimed to avoid vague or overly broad definitions, as these would not allow us to know which aspect of a multi-faceted dimension the panellists were responding to. The purpose of this study is not to come up with a consensus definition of any of these concepts; instead, the definitions are intended to make sure the concept space is covered adequately, and that participants are clear about the concepts they are rating. This caveat is included in the instructions to participants.

Similarly to the Singapore Statement, this initial list of RCR dimensions contains both core principles, or ‘metaresponsibilities’ (which we defined in Field et al., 2024, as somewhat more diffusely defined principles that guide RCR practice and relate to or influence many other dimensions of RR), as well as more concrete, practical, and specific responsibilities. We will not make distinctions between these different ‘levels’ of dimensions in the reference document, however. We do not wish to prime or sensitise participants, influencing them to weigh items more or less heavily than they would have done in isolation.

While some frameworks such as the Singapore Statement do make these distinctions (and we expect that most participants will be at least aware of these), much of the literature, including articles by some participants, does not. The relevant scholarship on the topic is far from concluding that there even *are* different levels of applicability or concreteness of responsibilities. Thus, we argue that it makes the most sense to keep the structure of the initial list simple and allow the participants to judge dimensions as naturally as possible. Any structure we impose on the output list (i.e., the document resulting from the Delphi process) will occur as a part of later work on producing a novel RCR framework, and further discussion of it is outside the scope of this study.

Participant Sample

For our research aims, we are looking for a diverse – in terms of academic discipline – expert panel. As a guide for this disciplinary diversity, we use the 2021 Research Excellence Framework (REF) Units of Assessment (UoA). There are 34 UoA listed (Table 1), which in our estimation provides a cross-section of research disciplines sufficiently granular to provide a diverse participant pool (for which a categorisation with only 6 or 7 elements would not be sufficient), without having so many categories that we would have difficulty finding a participant from each category (some lists contain hundreds of small discipline categories). Potential participants are *not* required to be a practicing scholar within the UoA they have expertise on, although many may be; others may be an expert by dint of studying practices in that field through a different paradigm. For instance, a medical anthropologist may be recruited to be an expert in medicine rather than anthropology for the purposes of our study.

Panel	Unit of Assessment	
A	1	Clinical Medicine
	2	Public Health, Health Services and Primary Care

Formatted Table

	3	Allied Health Professions, Dentistry, Nursing and Pharmacy
	4	Psychology, Psychiatry and Neuroscience
	5	Biological Sciences
	6	Agriculture, Food and Veterinary Sciences
B	7	Earth / Environmental Sciences
	8	Chemistry
	9	Physics
	10	Mathematical Sciences
	11	Computer Science and Informatics
	12	Engineering
C	13	Architecture, Built Environment and Planning
	14	Geography and Environmental Studies
	15	Archaeology
	16	Economics and Econometrics
	17	Business and Management Studies
	18	Law
	19	Politics and International Studies
	20	Social Work and Social Policy
	21	Sociology
	22	Anthropology and Development Studies
	23	Education
	24	Sport and Exercise Sciences, Leisure and Tourism
D	25	Area Studies
	26	Modern Languages and Linguistics
	27	English Language and Literature
	28	History
	29	Classics
	30	Philosophy
	31	Theology and Religious Studies
	32	Art and Design: History, Practice and Theory
	33	Music, Drama, Dance, Performing Arts, Film and Screen Studies
	34	Communication, Cultural and Media Studies, Library, and Information Management

Table 1. Research Excellence Framework 2021 Units of Assessment (retrieved from:

<https://www.ref.ac.uk/panels/units-of-assessment/><https://2021.ref.ac.uk/panels/units-of-assessment/index.html>)

From these disciplines, we plan to recruit individuals who are well-versed in the theory and/or practice of RCR.³ These people may have published on the topic, developed frameworks and codes based on RCR, may be involved in hands-on RCR training or

³ Note that while many researchers interested and involved in open/transparent research and reform science topics have shown interest in being involved in our RCR research including this Delphi panel, we are selective when it comes to the individuals we will invite to participate. This is because although open and reform research topics are adjacent and relevant to RR, we are considering RCR in terms of its formal frameworks (for instance in terms of the framework officially used by the European Commission's Framework Programs) and require input from individuals with expertise in this specific RR(I) framework and its use in research and policy.

community activity, or support researchers and research teams in RCR-related research projects and institutions. For our recruitment strategy, we have operationalised the above concepts into the following inclusion criteria: a participant must have (co)authored at least one peer reviewed article (in articles of more than two co-authors, their position in the author list must indicate leadership in the project in terms of its content; i.e., being in first or second author position, or being corresponding author) including ~~the following keywords: “RCR”, “RRI”, “responsible research and innovation” “research integrity” or “responsible research”~~ (as per our previously published scoping review, see Field et al., 2024) any relevant keywords (see below for a list) AND/OR include one or more of these keywords in their personal institutional webpage AND/OR have taught RCR/RRI to researchers, AND/OR have been involved in a project focusing on RCR/RRI (such as the European Commission’s NewHoRRizon, MoRRI or SUPER MoRRI projects: <https://newhorizon.eu>; ~~<https://super-morri.eu/morri-2014-2018/>~~; <https://super-morri.eu/morri-2014-2018/>) or a job that involves RCR/RRI (e.g., a research integrity officer), AND/OR, finally, have been part of a RCR/RRI network or working group (such as the RRING network: <https://ring.eu>, or the UKRI: <https://www.ukri.org>). Relevant keywords include “RCR/responsible conduct of research”, “RRI/responsible research and innovation”, “research integrity”, “responsible research”. We will also consider using keywords pertaining to particular areas of RCR (such as “open research”, “open science”, “research ethics”, “public engagement”, or any of the other aspects identified in our initial reference document) as ways to identify potential participants, but for these more particular terms we would require that, upon consultation, the potential participant provides additional justification that they have expertise to answer the Delphi on RCR more generally.

While these operationalisations do not ensure that all participants on the list will be RCR experts, we consider them a valid proxy for the purpose of this study. If people lead

research projects on RRI or RCR, list these topics as being their areas of expertise, or train and support others in RCR/RRI activities, they will likely have a sufficiently deep knowledge of the currently established tenets of RCR/RRI and some of the existing scholarly literature, and perhaps the application of RCR/RRI principles in policy to engage meaningfully with the stimulus provided as part of the Delphi process⁴. Additionally, in the contact emails, we will be clear about the kinds of participants we are looking for for this study, and that while our recruitment methods are reasonable, we might have made errors in ~~judgment~~judgement. Should the participants we approach feel that they are not suitable for purposes of participating in our study, they may also exclude themselves on these grounds.

We also note that despite our goal of developing output that is diverse in terms of the scientific disciplines that are represented in it, this output will represent only a limited selection of countries, regions and cultures. While the broader project within which this study is situated concerns RCR in the UK and regions of Europe, and a Euro-centric approach is appropriate to those ends, we emphasise that our findings will be produced with the input of a largely Western participant sample. We discuss the impact of this on our findings further in the limitations section in the discussion.

Panel Size

~~Ideally, we would recruit two to three members of each of the 34 UoA's, in order to maximise the chance that the outcomes of the study are in fact disciplinary differences instead of personal differences. General interest in our project so far would indicate that we should expect a higher than average response rate. We received enthusiastic responses to our~~

⁴ We recognise that while assembling a panel of RCR experts is appropriate for the aims of this Delphi study, and for the wider aims of the project the study is part of, this approach risks leading to a somewhat homogenous set of dimensions based predominantly upon pre-existing frameworks and models. We recognise that other strategies may lead to a more substantially different or transformative framework in comparison with what already exists.

recruitment calls for general involvement in our interviews, the Delphi and any other potential empirical studies relating to our RRPractice project⁵, which went out on social media and via various relevant networks in both the UK and Europe in mid-2022. Despite this, it is largely agreed (e.g., see Murphy et al., 1998, or Keeney, Hasson & McKenna, 2006) that the overall response rates of Delphi panels tend to be low, and attrition rates high (retention rates vary widely from study to study, however they decrease sharply as a function of the number of rounds in most studies).

Recommendations and empirical studies on Delphi methods vary on desirable sample size. For instance, 20-30 participants seems to be a sufficient panel based on Melander (2018; note that this is more than typical consensus Delphi panels tend to require), while Turoff (2002) suggests that between 10 and 50 panellists is sufficient for a dissensus Delphi. Choosing a minimum number necessarily contains an arbitrary factor, as well as a pragmatic one. Considering our goal of disciplinary diversity in the expert panel, we elected to use a minimum on the higher end of the average that these two sources suggest. As such, we have set the minimum panel size at the start of the process to be 30 panellists. The ideal panel size in a Delphi study is largely dependent on the purpose of the particular study. We are using this Delphi study to map the disciplinary differences regarding RCR. As such, it is important to reasonably minimise the chance that differences in judgements between panellists stem from their personal convictions rather than their disciplinary backgrounds. One way to do so is to ensure we have a sufficient panel size, to balance out differing opinions. However, we will also mitigate this risk of noisy data by the fact that we recruit experts, whose views should already take into account their familiarity with the range of differing opinions in their field. Additionally, we explicitly ask participants to answer based on their knowledge of the

⁵For purposes of clarity, what we refer to as the RRPractice project is a large research project that subsumes the current, proposed study. More information on RRPractice is available here: <https://www.cwts.nl/projects/current-projects/rrpractice>

field, rather than their personal beliefs. Finally, since the primary purpose of our study is to investigate diversity between disciplines in general, rather than making inferences about specific disciplines, we do not consider a minimum number of panellists per discipline to be necessary. Of course, the expected increase in reliability and representativeness of an increased panel size must be weighed against the feasibility of the study, both in terms of difficulty of recruitment and retention, and in terms of time and resources required to analyse (qualitative) data and draft the feedback reports. As such, we have set the minimum panel size at the start of the process to be 30 panellists.

To avoid disciplinary bias in our sample, i.e., where a disproportionate amount of experts would have a background in a specific discipline, we also decided to include two measures of diversity. First, we will require a minimum amount of disciplines present/represented in our sample before we start the Delphi study. As such, the starting panel must represent a minimum of 15 disciplines, with no more than ~~3 participants from one single discipline~~ three participants from one single discipline. Second, we will require a minimum amount of three participants from each of the four “panels” of the REF UoA structure that we used as our guide for disciplinary diversity (the four panels roughly correspond to: health and life sciences, natural sciences, social sciences, and arts and humanities).

We will initially approach three persons from each UoA, aiming for well above our minimum sample size. We will start the Delphi process when either 1. at least two persons from each UoA have agreed to take part, or 2. after three weeks of recruitment have elapsed, as long as our minimum panel size and disciplinary diversity requirements are met. If this is not the case we will continue recruiting until the minimum numbers are met.

Should enough participants drop from the study such that the total N drops below 15 or the amount of disciplines represented drops below 10, we will resume recruitment until

these minimum thresholds are met. In this case we will still maintain a maximum of three participants per discipline. While this is not ideal methodologically speaking, our goal of having the input of a diverse and large enough panel is more important than having continuity across rounds. Any attrition and replacement will be thoroughly and transparently documented in the final manuscript.

Recruitment Strategy

The recruitment strategy for this list will be largely centered on the articles included in the scoping review that preceded the Delphi study (Field et al., 2024). First and second authors of included articles will be searched whether they satisfied the inclusion criteria. Where a specific discipline is missing potential participants, we will carry out google searches with combinations of the following keywords: “RCR”, “RRI”, “responsible research and innovation” “research integrity” and “responsible research” and the names of the specific UoA’s, until we have a list of eligible participants per discipline. Should more recruitment be necessary, we will then actively recruit people through our own networks, including those identified in a previous, more general call for participants for the wider project through professional contacts of the authors, social media, and local university networks.

Potential recruits will be emailed to ask whether they are willing to participate as experts in our Delphi panel. We will send consent forms, participant information statements and study information (once again, via email) to those people that consent to participation in the Delphi. In addition, we will also ask for recommendations of other participants who fulfill the above-mentioned criteria, from those who declined to participate. Potential participants pointed out by declining participants will be vetted as to whether they meet the inclusion criteria before they are invited to participate.

In order to acknowledge the panellists' effort in this research project, we will offer authorship to each participant who participated in at least three rounds (for these purposes, the initial modification phase is also counted as a round) and is willing to contribute to the drafting of the final manuscript. We will also offer to list all panellists as contributors in a statement in the final manuscript, if they wish. Because we are recruiting experts, and this is necessarily a small and select population, we will allow participants to skip rounds if need be, although this will be discouraged.

Delphi procedure

The goal of our Delphi process is to establish an inclusive list of various dimensions of RCR, and to estimate how important these dimensions of RCR are to the represented research disciplines, including as broad a range of perspectives on this as possible. To do so, the Delphi process will take place over two phases, the latter of which is divided into multiple, iterative rounds. Figure 1 summarises the process, including the development of the initial reference document by the authors (the blue box in Figure 1.). This method is based on an earlier published Delphi study, in which SMF was an author (Pittelkow et al., 2023). Pittelkow and colleagues successfully used a modified reactive Delphi method to establish a checklist for communicating the rationale behind conducting replication studies, and the protocol described and followed in Pittelkow and colleagues' paper provides a kind of proof of concept for the methodology we set out in this proposal.

The process will start with an initial modification phase, *Phase 1* (the green box in Figure 1), in which participants can suggest additions to the proposed list of dimensions. Here the goal is to broaden the scope of the initial list of dimensions, capturing the various disciplinary perspectives of the panel. After the research team incorporates these suggestions and updates the dimension list, the second phase, *Phase 2* (the red box in Figure 1) will involve multiple rounds in which participants rate the importance of these dimensions to

RCR in their discipline. In this phase, the goal is to probe which items are more broadly appreciated by the sample (i.e., which might be universally valuable in RCR practice), versus which might be more discipline-specific. In this way, the list of RCR dimensions can be ‘weighted’ by importance across disciplines.⁶

⁶ Note that the weighting information captured in Phase 2, comparing importance of dimensions across disciplines, will help the research team structure the RCR cross-disciplinary mapping we aim to develop in a further stage in the broader project.

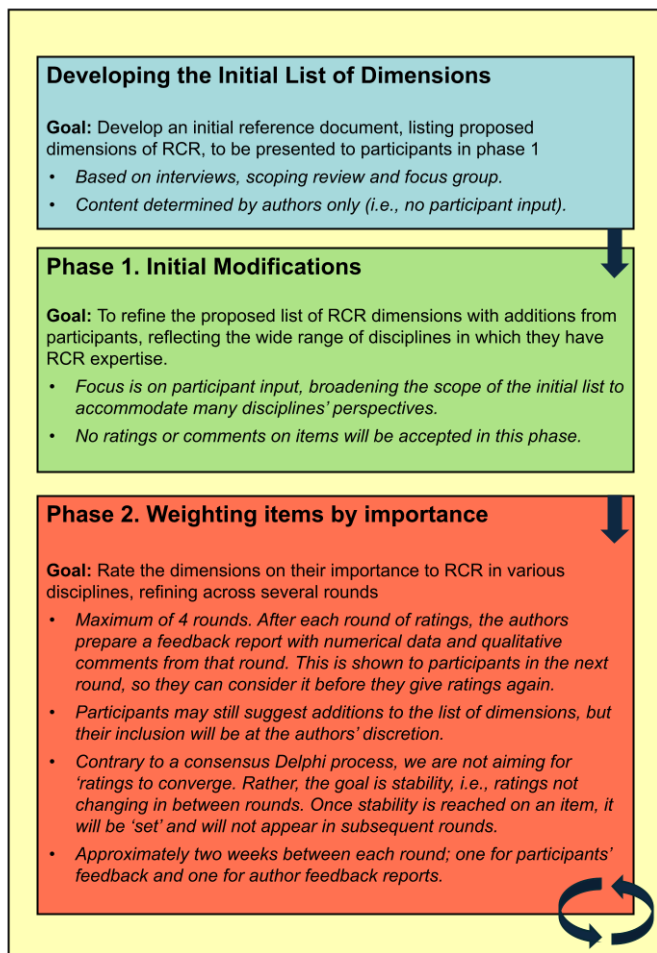


Figure 1. Flowchart showing the phases featuring in this modified Delphi protocol.

Each phase or round of this Delphi will be hosted asynchronously in an online environment that panellists can access through a link provided in an email. When panellists follow this link, they will arrive at a Qualtrics survey containing the Delphi questionnaire. Each questionnaire will start by asking which discipline a participant identifies with for the

purposes of the Delphi, which they will have confirmed with the researchers beforehand during recruitment. Prior to Phase 1 of the Delphi, we also ask for two participant demographics, namely their geographical region, and years of expertise in their discipline; this is merely to understand the overall makeup of our panel, and will be reported separately to any analysis of the main data.

In Phase 1, we will present the experts with the full list of the proposed dimensions, and ask whether they deem any dimension missing from the list. They will be encouraged to add any dimensions they think should feature in the list. Panellists can answer this question in a textbox with unlimited characters. After Phase 1, the research team will refine the list (see “Feedback reports and analysis plans” section below for details) and then dimensions (including their names and definitions) will be set for all future rounds of the Delphi process. However, if, from analysis of the qualitative data in Phase 2, it is overwhelmingly clear that two or more groups of panellists are making different interpretations of a particular dimension, the research team may have to change or split that specific dimension for future rounds. All changes to the dimension list will be noted in the results, and reported in detail in the supplemental materials.

In Phase 2, which will involve up to four rounds, we will ask for ratings of importance for each dimension. The survey will present each proposed dimension consecutively, and will ask the participants how important they consider the dimension of RCR to their specific discipline on a 7-point rating scale ranging from “extremely unimportant” to “extremely important”. In addition, the participants will be encouraged to motivate their answers in a textbox with unlimited characters, though a motivation is not required to move to the following dimension. Note that, other than a brief explanation as part of the initial survey instructions, we will not attempt to define to the participants in any great detail what “important” means. Although in general it is advisable to be as precise as possible

in elicitation of survey measurements, we believe that in this case trying to prescribe particular aspects of the concept of “importance” would counterproductively narrow our measurement, when an intuitive, broad understanding of the word may more closely capture the essence of what we wish to measure, i.e., the sense that a dimension “matters” in that discipline. In round 1 of Phase 2 (i.e., the first round of ratings), only the name and definition of each dimension will be presented to participants. In all subsequent rounds, a feedback report (see section below) summarising data from the previous round will also be presented for each dimension. In Round 3 and 4, participants will only be presented with dimensions that have not reached stability in the prior rounds. We will conclude Phase 2 after a maximum of four Delphi rounds. Melander’s review suggests between two and three rounds is the average for a consensus Delphi; however, because we expect a particular diversity of disciplines and perspectives in our Delphi, we will allow up to four rounds if needed.

~~Note that in Phase 2, participants will not be asked to add new dimensions. Allowing the addition of new items after Phase 1 is complete will needlessly complicate the process and cause it to take longer than might be reasonable, risking higher attrition and discontinuity. Should participants still wish to add items, despite this, it may be indicated in the response boxes provided, and will be considered by the research team. In extreme cases (where many participants suggest the addition of the same kinds of items), new additions will be considered though we believe this to be unlikely. What is and is not required of participants for each phase will be clearly communicated to them.~~

In all stages of the Delphi, panellists are asked to answer based on their expert understanding of the view from their discipline as a whole, not simply their own personal opinions. Ideally, they focus on a single discipline in this exercise, but some respondents (such as those that are more senior or more embedded in the RCR/RRI sphere, or work across multiple fields) might have more interdisciplinary input. We will ask them to answer

in relation to their primary field of expertise (that is, the one we recruited them for), however we will also ask that they add any additional insights from other fields they're familiar with into an open text box at the end of the survey.

For each round, as well as the initial modification phase, panellists will have one week to two weeks to fill out the questionnaire. All rounds will have one week to two weeks for the research team to complete that round's interim analysis, after which the following round will commence. These weeks will be staggered and run from Wednesday to Tuesday of the consecutive week, to accommodate for panellists' potential leave weeks.

Feedback reports and analysis plans

After Phase 1, during which participants suggest additions or changes to the dimensions in the initial proposed list, the research team will screen these suggestions based on whether they are 1) sensible, i.e., relevant to RCR, coherent, and factually correct, and 2) different enough from existing dimensions to add value to the list. If suggested additions are merely narrower versions of existing dimensions, we may simply add the suggestion as an example to the definition of the applicable dimension. At least two researchers will screen all the participant-suggested additions and changes, recording their recommendations individually, then these will be compared across the team and any discrepancies will be resolved by group discussion. The team will then amend the proposed dimension list with all changes or additions agreed by the team to be both sensible and adding value. All suggested additions and changes will be collated, along with the decisions the research team makes on them, and recorded in supplemental materials.

Note that in Phase 2, participants will not be asked to add *new* dimensions. Should participants still wish to add dimensions, despite this, they may indicate so in the response boxes provided. These suggestions will be screened by the research team on the basis of the

same two points as after Phase 1, as outlined above. However, the barrier for including new dimensions during Phase 2 will be much higher, as addition of new items at this point in the process substantially complicates the process, risking higher attrition and discontinuity. As such, new proposed dimensions must contain a highly compelling argument that it is sensible and non-redundant. What is and is not required of participants for each phase will be clearly communicated to them.

After each round of the rating phase, we will analyse the panellists' answers and draft a feedback report. The feedback reports will consist of the descriptive statistics of the aggregated ratings (median and interquartile range, IQR) and an analysis of the qualitative feedback per dimension (as was done in Pittelkow et al., 2023). All qualitative feedback – the panellists' input – will be organised per item and placed in an appendix to the feedback report, accessible to participants by clicking a web link. This feedback will be anonymised if necessary, but otherwise left untouched. Specifically, this means that the feedback reports will not contain names of persons, institutions, or geographical locations that may be present in the qualitative data. In addition, if we determine that a particular section of text in the qualitative data – such as a description of an institution or location – allows the panellist to be identified, we will delete this as well. All deletions will be clearly marked.

Qualitative feedback will be bundled per dimension and analyzed by the research team using a simple form of thematic analysis. Though we encourage the participants to be as detailed in their responses as time allows them to, we do not expect the qualitative data to be highly complex, as the participants will be answering a specific question. This is why we will use a “small q” or “coding reliability” version of thematic analysis, focusing on structured codebooks, independent coders, and consensus between coders (Braun & Clarke, 2023; Finlay, 2021; Kidder & Fine, 1987). For each round of the Delphi study, two researchers on the team will independently open code the data, after which they will meet to establish a final

codebook with which the data will be coded. The result of this analysis will be summarised and given to the experts in the feedback report for the subsequent round, along with an appendix with the 'raw' qualitative data. The key to our analysis method is transparency, as we will provide access to the 'raw' qualitative data, as well as the coding books with which they were analysed. In addition, since our analyses will be distributed to all participants/experts before every new round, along with their own data, interested participants/experts are welcome to, and invited to, check the analysis as the Delphi process continues.

In subsequent rounds, the participants will be presented with the feedback report, and will be asked again to score the list of dimensions. This will continue until we reach stability on all dimensions, though for the sake of feasibility we will conduct no more than four rounds of ratings.

Since we selected our participants in this study on the basis of their individual expertise – which surpasses our knowledge of their disciplines – we will not perform any stringent quality checks on the content of the data, quantitative or qualitative. Data quality should be aided by the fact that participants are encouraged to write down free-text justifications to every question, which should prompt them to answer questions thoughtfully. If we notice suspicious patterns in the data, however (such as all items answered with the same choice) we will contact participants individually to check that they meant these.

Stability

An important aspect of the Delphi process is the concept of stability, or when we consider the answers to be similar enough between two or more subsequent rounds that we can consider the answer "settled" or "definite". Since panel responses can vary greatly between rounds - as per the explicit aim of the Delphi process - it is important to assess

whether the panel's response on any given dimension is still developing or whether it can be considered settled. In fact, different authors have argued that assessing the level of consensus in a Delphi study is meaningless without having assessed stability of responses, since the response may not be an accurate reflection of the conclusive ~~judgment~~[judgement](#) of the panel (Dajani et al., 1979, Scheibe et al., 2002). To reduce participant burden, where stability is reached, the item will be considered 'set', and not feature in subsequent rounds.

We use a simple metric for stability: for each dimension, we will take the absolute value of the change in ratings for each participant. If the mean of these absolute-value change scores is less than the equivalent of 1 point on our rating scale (i.e., 16.66% of the total breadth of the rating scale), we will consider the dimension stable. This tracks closely with the recommended cutoff for stability of 15% difference recommended by Scheibe et al. (2002), which is based on an empirical estimation of the random change between rounds. However, we will also temper this quantitative stability judgement with qualitative analysis: if the qualitative data contain novel arguments for the importance or unimportance of a dimension that we have reason to believe may sway the panel substantially in the following round, we will not consider the dimension to be stable.

Reflexivity/positionality

Finally, we wish to be transparent about the contributions that we, as individuals and as a team, approach the subject of RCR. This allows the reader to evaluate our decisions with personal context in mind; given the flexibility that exists in our design; especially as the dimensions are developed between rounds, ~~and as the framework is built.~~ Our team comes from a background in science [and technology](#) studies, metascience and research integrity, and we have previously published on the topics of epistemic diversity, responsibility and quality (e.g., [Field & Derksen, 2021](#); [Muller & de Rijke, 2017](#); [Penders, de Rijke & Holbrook, 2020](#); [Penders & Goven, 2010](#); [Valkenburg, Dix, Tijdink and de Rijke, 2021](#); [Van](#)

Drimmelen et al. 2023). We have also previously published the scoping review on which this Delphi study directly builds. As a result, we are aware of the literature on RCR and adjacent topics. Unavoidably, our decisions will be rooted in this knowledge, from the initial choices we have made to develop this study, to the choices we will collectively make as we co-construct an RCR framework with the input of our experts. In our supplemental materials on OSF (<https://osf.io/prvds>), This means that the experts that we invite to participate in the panel may be familiar with our work. This contains a risk of biasing the study, since their decision of whether or not to participate may depend on their estimation of our work. However, the direction of this potential bias is uncertain. In addition, this risk is mitigated greatly by the fact that the study population is large and diverse, and its majority will have had no professional or personal interaction with any of the researchers. In our supplemental materials on OSF (<https://osf.io/prvds>), we provide individual statements about our positions in relation to the present study, structured using orienting questions proposed by Barry et al. (1999) and Olmos-Vega et al. (2023), to further highlight our link to the research our group is conducting.

Data Sharing

We believe in the importance of data sharing, both from the perspective of accountability, as well as the potential re-use of our data. As such, we will share all data and analysis, guided by the TOP Guidelines. We will do this by making the feedback reports openly available in a suitable repository. These feedback reports will include the pseudonymised ‘raw’ data, both quantitative and qualitative, along with our analyses of this data, subject to any redactions by study participants for their privacy. Final versions of all study materials will be uploaded to this page before the study starts.

Expected results

Results

Our results will be structured as follows: first, as an overview of the main findings, we will provide a table containing all final dimensions from Phase 2 of the Delphi process, noting which dimensions reached stability, and outlining the final variances and interquartile ranges, as a proxy for consensus, and the median rated importance, as an indication of importance. ~~It~~ (see Table X, below). This table contains data from the entire panel, across all represented disciplines. On our 1-7 importance rating scale, we will interpret median ratings corresponding to numerical values lower than 3 as low importance, between 3 and 5 as moderate importance, and greater than 5 as high importance. These straightforwardly map onto the verbal labels of the rating scale.

However, we will not declare any *a priori* thresholds for interpreting levels of consensus in the data. While it is imperative to state specific thresholds for interpreting data in a Delphi study, if that study aims to attain consensus (Grant et al., 2018; Williams & Webb, 1994). ~~However,~~ our goal is different, ~~as we~~. We aim to refine and map the existing perspectives on dimensions of RCR, rather than determine a consensus of which dimensions are most important. Indeed, multiple authors have argued that the emphasis on a binary consensus-nonconsensus divide is a crude way of employing the Delphi technique. For example, Scheibe and colleagues note that “considering that there is a strong natural tendency in the Delphi for opinion to centralize, resistance in the form of unconsensual distributions should be viewed with special interest” (2002, p. 271), see also (Linstone & Turoff, 2011).

As such, we do not propose any confirmatory analyses for declaring ‘consensus’ for any given dimension, but will instead present the quantitative data descriptively. We are interested in ~~(and expect we might realistically find)~~ at least three categories of response distributions: either a strong peak around a single point (i.e., universal agreement), a

Formatted: Font: Not Italic

relatively flat distribution across all points, or a multi-modal distribution with two or more distinct peaks. We do not propose to differentiate these statistically, given the relatively small sample sizes and exploratory nature of this work, but instead will present our descriptive interpretations along with visualisations (e.g., histograms) of the complete importance rating data for each dimension, so readers can visually assess the shape of distributions to infer the ‘universality’ of responses across the panel—(see Figure(s) X, Y, Z, below).

In addition to the ~~main table containing the median importance ratings across the entire panel~~ primary results above, we will also provide, ~~for context, a table providing a complete view of the Delphi results broken down by discipline, by-reporting a separate table with the median rated importance of each discipline category per dimension, and the number of respondents per discipline category.~~ Since our sample will include ~~no more than three only a small number of~~ panellists per discipline, ~~and likely less than that category,~~ we ~~will~~ must be cautious in making any inferences on the basis of these analyses, ~~though we do consider them worthwhile sharing. We will find it useful (for purposes of summarising and interpreting our findings) to pre-specify some simple labels to categorise these measurements of importance. Our categories of importance are easy to delineate: on our 1-7 scale, we will interpret median ratings corresponding to numerical values lower than 3 as low importance, between 3 and 5 as moderate importance, and greater than 5 as high importance. These straightforwardly map onto the verbal labels of the rating scale and advise readers from making any strong inferences either.~~

To contextualise all the abovementioned quantitative findings, we will also provide an analysis of the qualitative data from throughout the Delphi process for all Phase 2 dimensions. ~~All of our results will be discussed in a discussion section.~~

References

- Barry, C. A., Britten, N., Barber, N., Bradley, C., & Stevenson, F. (1999). Using Reflexivity to Optimize Teamwork in Qualitative Research. *Qualitative Health Research*, 9(1), 26-44. <https://doi.org/10.1177/104973299129121677>
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401-409. <https://doi.org/https://doi.org/10.1016/j.jclinepi.2013.12.002>
- Dajani, J. S., Sincoff, M. Z., & Talley, W. K. (1979). Stability and agreement criteria for the termination of Delphi studies. *Technological Forecasting and Social Change*, 13(1), 83-90. [https://doi.org/https://doi.org/10.1016/0040-1625\(79\)90007-6](https://doi.org/https://doi.org/10.1016/0040-1625(79)90007-6)
- Field, S. M., & Derksen, M. (2020). *Experimenter as automaton; experimenter as human: exploring the position of the researcher in scientific research*. *European Journal for Philosophy of Science*, 11(1), 11. <https://doi.org/10.1007/s13194-020-00324-7>
- ~~Field, S. M.,~~ Thompson, J., de Rijcke, S., Penders, B., & Munafo, M. (2024). Exploring the Dimensions of Responsible Research Systems and Cultures: A Scoping Review. Preprint: <https://osf.io/6z4mx/>
- Funtowicz, S., & Ravetz, J. (2018). Post-normal science. In *Companion to environmental studies* (pp. 443-447). Routledge.
- Grant, S., Booth, M., & Khodyakov, D. (2018). Lack of preregistered analysis plans allows unacceptable data mining for and selective reporting of consensus in Delphi studies.

Journal of Clinical Epidemiology, 99, 96-105.

<https://doi.org/https://doi.org/10.1016/j.jclinepi.2018.03.007>

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing*, 53(2), 205-212.

Linstone, H. A., & Turoff, M. (2002). *The Delphi Method: Techniques and Applications*. Addison-Wesley.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*, 78(9), 1712-1719.
<https://doi.org/https://doi.org/10.1016/j.techfore.2010.09.011>

Melander, L. (2018). Scenario development in transport studies: Methodological considerations and reflections on Delphi studies. *Futures*, 96, 68-78.

Müller, R., & de Rijcke, S. (2017). Thinking with Indicators. Exploring the Epistemic Impacts of Academic Performance Indicators in the Life Sciences. *Research Evaluation*, 26(4), 361-361. <https://doi.org/10.1093/reseval/rvx033>

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment (Winchester, England)*, 2(3), i-88.

National Health and Medical Research Council. (2018). Australian Code for Responsible Conduct of Research. <http://www.nhmrc.gov.au/guidelines/publications/r41>

Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2023). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher*, 45(3), 241-251. <https://doi.org/10.1080/0142159X.2022.2057287>

Penders, B., de Rijcke, S., & Holbrook, J. B. (2020). Science's moral economy of repair: Replication and the circulation of reference. *Accountability in Research*, 27(2), 107-113. <https://doi.org/10.1080/08989621.2020.1720659>

~~Penders, B., & Goven, J. (2010). Nutrigenomics and the stewardship of scientific promises. *Biotechnology Journal*, 5(9), 909-912.~~

~~<https://doi.org/https://doi.org/10.1002/biot.201000069>~~

Pittelkow, M. M., Field, S. M., Isager, P. M., van't Veer, A. E., Anderson, T., Cole, S. N., ... & Van Ravenzwaaij, D. (2023). The process of replication target selection in psychology: what to consider?. *Royal Society Open Science*, 10(2), 210586.

Salkind, N. (2007). Encyclopedia of Measurement and Statistics. In. SAGE Publications, Inc. <https://doi.org/10.4135/9781412952644>

Scheibe, M., Skutsch, M., & Schofer, J. (2002). Experiments in Delphi Methodology. In H. A. Linstone & M. Turoff (Eds.), *The Delphi Method: Techniques and Applications* (pp. 257-281). Addison-Wesley Publishing Company.

Tallacchini, M. (2009). Governing by values. EU ethics: soft tool, hard effects. *Minerva*, 47(3), 281.

Turoff, M. (2002). The Policy Delphi. In H.A. Linstone and M. Turoff (eds.). *The Delphi Method: Techniques and Applications*, Reading MA: Addison-Wesley, pp. 80-96.

~~Valkenburg, G., Dix, G., Tjeldink, J., & de Rijke, S. (2021). Expanding Research Integrity: A Cultural Practice Perspective. *Science and Engineering Ethics*, 27(1), 10.~~

~~<https://doi.org/10.1007/s11948-021-00291-z>~~

~~Van Drimmelen, T., Slagboom, M. N., Reis, R., Bouter, L. M., & Van der Steen, J. T. (2023). Decisions, Decisions, Decisions: an ethnographic study of researcher discretion in practice. *MetaArXiv*. <https://doi.org/https://10.31222/osf.io/7dh3t>~~

~~Von Schomberg, R. (2011). Towards responsible research and innovation in the information and communication technologies and security technologies fields. Available at SSRN 2436399.~~

Von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden*, 39-61.

Formatted: Font color: Custom Color(RGB(34,34,34)), Highlight

- Von Schomberg, R. (2013). A vision of responsible research and innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society*, 51-74.
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401-409.
<https://doi.org/https://doi.org/10.1016/j.jclinepi.2013.12.002>
- Grant, S., Booth, M., & Khodyakov, D. (2018). Lack of preregistered analysis plans allows unacceptable data mining for and selective reporting of consensus in Delphi studies. *Journal of Clinical Epidemiology*, 99, 96-105.
<https://doi.org/https://doi.org/10.1016/j.jclinepi.2018.03.007>
- Salkind, N. (2007). *Encyclopedia of Measurement and Statistics*. In. SAGE Publications, Inc.
<https://doi.org/10.4135/9781412952644>
- Williams, P. L., & Webb, C. (1994). The Delphi technique: a methodological discussion. *Journal of Advanced Nursing*, 19(1), 180-186.
<https://doi.org/https://doi.org/10.1111/j.1365-2648.1994.tb01066.x>
- Zwart, H., Landeweerd, L., & Van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the European research funding arena from 'ELSA' to 'RRI'. *Life sciences, society and policy*, 10, 1-19.