

TRUSTparency

Deliverable 3.1

Report on the end-user level of consensus for the collected interventions

Author: *Daniele Fanelli*

Reviewers: *Panagiotis Kavouras, Kalle Videnoja*

Project title: Increasing reproducibility through the co-creation of interventions that support a transparent and trustworthy ecosystem

Project acronym: TRUSTparency

Grant Agreement no.: 101188172

Lead contractor for this deliverable: Science of Science Ltd (SoS)

Deliverable factsheet:

Project Number:	101188172
Project Acronym:	TRUSTparenCcy
Project Title:	Increasing reproducibility through the co-creation of interventions that support a transparent and trustworthy research ecosystem
Title of Deliverable:	Deliverable 3.1 Report on the end-user level of consensus for the collected interventions
Work Package:	WP3
Due date of deliverable	M18 - 31/7/2026
Actual date of submission	30/7/2026
Author:	Daniele Fanelli
Reviewers:	Panagiotis Kavouras, Kalle Videnoja

ABSTRACT:	The scope of this deliverable is to assess the community's possible concerns or doubts about specific reproducibility interventions that will be included in TRUSTparenCcy's toolbox. This is line with TRUSTparenCcy's mission to ensure that reproducibility policies and interventions are deployed in ways that minimise their costs and the risk of unintended consequences. This deliverable fills some gaps in our knowledge, since the discourse on reproducibility interventions and policies tends to focus on their assumed benefits, while the potential implementation costs or downsides are rarely discussed. We invited 565 members of national reproducibility networks to share and discuss their concerns for the effectiveness or implementation of reproducibility interventions on the comCensus platform, using a list of 41 intervention types grouped in five categories. Circa 16% (N=87) responded, sharing 397 posts and reactions, sharing diverse concerns relevant to every intervention in the list. The participants' concerns map well onto challenges in terms of resources, receptivity and relevance suggested by our analysis of the literature and developed into the scheme of inclusivity challenges and epistemological disparities, reported in D1.3. Overall, these results validate TRUSTparenCcy's assumption that the interventions and policies to improve reproducibility present potential challenges and unintended consequences that risk being overlooked by mainstream metascience, and support the project's tenet that policy-making needs to be informed by bottom-up, anonymous feedback collected from the research community.
Keyword List:	Reproducibility, disparities, inequality, inclusivity, consensus, comCensus

Consortium

	Role	Name	Short Name	Country
1.	Coordinator	University of Oslo	UiO	NO
2.	Partner	Aarhus University	AU	DK
3.	Partner	Science of Science Ltd.	SoS	GB
4.	Partner	Amsterdam UMC	AUMC	NL
5.	Partner	University of Latvia	UoL	LV
6.	Partner	National Technical University of Athens	NTUA	EL
7.	Partner	Finnish National Board on Research Integrity	TENK	FI
8.	Partner	Aristotle University of Thessaloniki	AuTh	EL
9.	Partner	University of Santa Catarina	UFSC	BR
10.	Partner	National University	NU	PH
11.	Partner	Kwame Nkrumah University of Science and Technology	KNUST	GH
12.	Partner	Lviv Polytechnic National University	LPNU	UA
13.	Associated Partner	Open Science NL	OSNL	NL

Revision History

Version	Date	Revised by	Reason
0.1	18.01.2026	Daniele Fanelli	Draft plan outlining scope and methodology
0.2	15.07.2026	Daniele Fanelli	First complete draft of the report
0.3	17.07.2026	Panagiotis Kavouras	Comments on the first draft
0.4	19.07.2026	Daniele Fanelli	Second complete draft of the report
0.5	24.07.2026	Kalle Videnoja, Panagiotis Kavouras	Comments on the second draft
1.0	27.07.2026	Daniele Fanelli	Final version addressing reviewer's comments

Table of contents

1. Introduction	5
1.1 Scope	5
1.2 Community	6
2. Methods.....	7
2.1 Question set up	7
2.2 Participant invitation	8
2.3 Analysis	9
2.4 Transparency and reproducibility	9
3. Results.....	9
3.1 Descriptive statistics	9
3.2 Concerns by category	10
4. Differences by group	14
4.1 Gender differences.....	14
4.2 Career stage differences	15
4.3 RRR breakdown	16
5. Conclusions	23
6. Appendix.....	24
6.1 Line-numbered list of interventions	24
6.2 List of contributors	24

1. Introduction

The TRUSTparenCy project aims to empower institutions involved in the research process to develop their own reproducibility policies and guidelines - where “reproducibility” is the extent to which design, implementation, analysis, and reporting of a study enable a third party to repeat the study and assess its findings. Its main output, the Reproducibility Promotion Plan (RPP), is a tool that supports institutions in implementing and evaluating interventions, that is, concrete actions to embed reproducibility practices in everyday work and to monitor their effectiveness.

A core assumption of TRUSTparenCy, building on the work of iRISE, TIER2 and OSIRIS, is that reproducibility does not apply equally to all types of research, and that the cost-effectiveness, barriers, and unintended consequences of any given intervention can vary substantially across fields, institutions and demographics. The “toolbox” produced by TRUSTparenCy therefore aims to offer not only a list of interventions but also guidelines that alert policy-makers (that is, leaders and decision makers within research institutions and organisations) to the factors they should consider before deploying any of them.

Work Package 3 is dedicated to the co-development of the first draft versions of these interventions, of the training materials and of the RPP. Within WP3, Task 3.1 has the function of collecting and reporting the end-user community’s views on the interventions mapped in Work Package 1, so that this empirical evidence can feed back into Tasks 3.2, 3.3 and 3.4. The present deliverable reports on the first stream of this work.

1.1 Scope

Deliverable 1.1 reports a list of interventions, which were subsequently categorised and classified into 5 overarching categories, 18 subcategories and 41 distinct intervention types. These will form one of the three core elements of TRUSTparenCy’s outputs and they will facilitate the pilot testing of the RPPs by the pilot testing by institutions.

In line with TRUSTparenCy’s mission to ensure that reproducibility policies and interventions are deployed in ways that minimise their costs and the risk of unintended consequences, the scope of this deliverable is to assess the community’s possible concerns or doubts about specific reproducibility interventions that will be included in TRUSTparenCy’s toolbox.

This focus is central to TRUSTparenCy’s mission. The discourse on reproducibility interventions and policies tends to focus on their assumed benefits, occasionally acknowledging that their effectiveness may be unequal across fields. The potential costs or downsides, conversely, are rarely discussed. Part of the reason is that researchers may lack the opportunity to express contrarian views, and part is that the topic of reproducibility has acquired an unfortunately moralising dimension, such that many researchers might not feel comfortable expressing open criticisms.

As articulated in Deliverable 1.3, there are at least nine distinct dimensions that represent potential challenges, barriers, hidden costs and unintended consequences resulting from the implementation of reproducibility interventions. These nine dimensions are grouped under three higher-level themes:

- a) **Resources**, comprising the dimensions of (1) *funds*, (2) *time*, and (3) *infrastructure*;
- b) **Receptivity**, comprising the dimensions of (4) *knowledge*, (5) *incentives*, and (6) *regulations*;
- c) **Relevance**, comprising the dimensions of (7) *phenomena stability*, (8) *methods homogeneity*, and (9) *philosophical compatibility*.

These dimensions may unequally affect researchers across fields, countries and demographics. Our purpose is to help the community express and discuss any such concerns openly and safely. The resulting information will guide the pilot tests, will feed back into the RPP (where each intervention will be scored along some or all of the nine dimensions identified in D1.3), and may be submitted for to be published as a peer-reviewed article.

1.2 Community

The community of end users invited to provide feedback comprises two main components. The first consists of members of national Reproducibility Networks (RNs) across 32 countries. We collected N=565 email addresses of members who were either occupying leadership positions within these networks or had indicated themselves as local contacts. Table 1 shows the distribution of these contacts across countries and the type of role they hold within their respective network.

The second component consists of the members of the TRUSTparenCy Reproducibility Community, hosted on LinkedIn (<https://www.linkedin.com/groups/10088652/>), which brings together other stakeholders and interested parties beyond researchers, which currently includes N=101 members.

In addition, the following members of TRUSTparenCy also contributed their views: Daniele Fanelli, Mads P. Sørensen, Panagiotis Kavouras. These TRUSTparenCy members responded like any other participant, and had therefore an equal (not greater) opportunity to shape the results than all other participants.

Table 1. Members of national Reproducibility Networks contacted for this consultation, by country and type of role.

Country	Leadership / Governance	Local contact / Representative	Total
Australia	9	1	10
Belgium	0	5	5
Benin	0	1	1
Brazil	7	32	39
Burkina Faso	0	1	1
Canada	0	47	47
Croatia	6	9	15
Denmark	9	2	11
Egypt	0	1	1
Ethiopia	0	3	3
Finland	5	0	5
France	2	6	8
Gambia	0	1	1
Georgia	1	0	1
Germany	28	50	78
Ghana	0	1	1
Italy	7	56	63
Kenya	0	3	3
Malawi	0	1	1
Netherlands	23	0	23
Nigeria	0	9	9

Country	Leadership / Governance	Local contact / Representative	Total
Norway	13	1	14
Portugal	5	2	7
Rwanda	0	1	1
Slovakia	0	24	24
Somalia	0	1	1
Sudan	0	2	2
Switzerland	33	9	42
Tanzania	0	1	1
Tunisia	0	1	1
United Kingdom	0	145	145
Zimbabwe	0	1	1
Total	148	417	565

2. Methods

2.1 Question set up

As planned in the project, we will use the comCensus platform, which was specifically adapted to TRUSTparenacy's needs as reported in Deliverable 2.1.

The key benefit of comCensus is that it collects validated responses and yet does not require registration or collect meta-data, making the data anonymous at the source. Moreover, it supports very large numbers of participants and, unlike a survey, it facilitates the exchange of viewpoints and allows participants to discuss at any level of detail the matter at hand (e.g. discuss the merits of individual reproducibility interventions). This methodology brings to the consultation some of the depth that would normally be associated with a focus group or a structured meeting, while preserving the scale and the openness of an anonymous survey, making this exercise substantively as well as statistically meaningful.

Further advantages of comCensus include the following: the entire conversation can be made public, and in any case is retained as a citable record of the discussion; participants can disclose their name without breaking the anonymity of their responses, which allows them to be transparently credited as contributors in any publication resulting from the exercise; and the platform integrates graphical and AI tools designed to help analyse and synthesise the rich qualitative and quantitative data generated by the discussion. In particular, the platform offers interactive graphics that allow to search and filter the posts exchanged between participants, and it implements particular algorithms that allow a dedicated AI to distil and interweave the views expressed, taking into account the level of consensus for each position.

Following the comCensus methodology, participants were invited to respond to a single question, referring to a separate document that contains the full list of interventions (as compiled by the scoping review of Task 1.1), structured across different categories and sub-categories (as developed by Task 1.2). Different possible formulations of the question were examined and tested with consortium members, both during and following the 2nd TRUSTparenacy General Assembly that took place in March 2026.

The question found to be most suited for the scope is the following:

“In which category of reproducibility interventions do you have the greatest reservations about cost-effectiveness or practical implementation?”

The options to choose from consisted in five descriptive categories, each comprising a subset of reproducibility intervention types: Open Policies, Planning and Training, Open Data, Methodology and Reporting, Tools and Validation. These categories were deemed better suited to the purpose than the taxonomy developed in Task 1.2, because they are more directly descriptive of different types of practices.

The question is accompanied by the following subtext:

“We have compiled 41 types of reproducibility intervention, classified in 5 broad categories. These may form the basis of reproducibility policies in your field or institution.

Please:

- *Vote for the category you have the greatest reservations.*
- *Explain your choice, by citing specific interventions that concern you.*
- *Mention how your field or personal background shapes your views.*

Responses are anonymous. If you would like to be acknowledged in the report, you can optionally share your details.”

A document associated with the question and visible to participants displayed the full list of types of interventions within each category, with a line number, for easy reference. The same list is included in the Appendix to this report for reference. In the text that follows, we shall refer to specific interventions by line number, preceded by the acronym ILN (Intervention Line Number).

After answering the question by articulating their answer, participants could choose to disclose up to five demographic characteristics: Gender, (Man, Woman, other), Career stage (Student, Early, Mid, Established, Retired), Country Income Level (Low, Lower-mid, Upper-mid, High), Type of institution (Public research performing, Private research performing, Research Funding, Research Publishing, other), and Research Type (Mathematics & Statistics, Natural Sciences (Theoretical), Natural Sciences (Empirical), Engineering and Technology (Theoretical), Engineering and Technology (Empirical), Medical (Basic), Medical (Clinical), Social Sciences (Quantitative), Social Sciences (Qualitative), Humanities (Sources), Humanities (Creative), Other).

Participants could also choose to share their name, surname, and affiliation, potentially including a link to a personal or professional page of their choice. As mentioned above their names cannot be associated with their responses whatsoever. The names and affiliations of respondents were displayed in a page associated with the discussion page, whereas participants who had chosen not to disclose their identity appeared on the page as “NA”.

2.2 Participant invitation

To encourage higher response rates, each candidate participant was sent an invitation email that communicated the personal nature of the invitation, addressing the candidate by name and mentioning the specific country of the reproducibility network of which the candidate was a member.

Invitations were sent out in batches of 20 between the 18th and the 22nd of June 2026, then a single follow up email was sent on June 30th. On July 3rd, an invitation was also posted on the LinkedIn channel of the TRUSTparenCy Reproducibility Community.

Of the 565 emails sent, 11 were returned with an error, leaving a sample of 554 potential participants. However, circa 10% of email invitations received an automatic response due to the recipient being on annual leave. This suggests that the actual sample of participants that was reached by the invitation whilst the question was open for responses is likely smaller, possibly N=500 or less.

2.3 Analysis

Descriptive statistics and text summaries were generated directly from the data, using the comCensus interface, powered by the LLM grok-4.20, with temperature set at 0 to ensure maximum accuracy and consistency.

Descriptive graphs were generated including all available responses.

Text were generated recursively using tools available on the platform. In all cases, the AI was prompted to generate an initial draft, which was subsequently reviewed and edited by Daniele Fanelli.

In particular, the per-category summaries were generated with the prompt (square brackets contain editorial notes) : “Summarise the concerns and reservations (and any counter-arguments made in reaction to them) for interventions falling in the category [Category name as it appears in the text] ([shorter category name used for voting], listed from parx to pary [where x and y refer to starting and ending line in the relevant portion of the intervention list text]). Cite specific examples and include any relevant references cited.” The Cross-category concerns were initially drafted with the following prompt: “Summarise the general concerns and reservations expressed, which cut across multiple categories.”

The gender and career differences used, respectively, the prompts: “How do the responses of women differ from those of men?”, and “How do the responses differ across career levels, from early to established?”. Finally, the RRR breakdown data was generated by including in the prompt the part of the text of Deliverable 1.2 that describes the three themes and nine dimensions, as follows: “This is a 3-thematic, 9-dimensional framework of possible obstacles to the adoption or effectiveness of reproducibility interventions: [text from D1.2] For each of the 41 interventions, produce a table that gives: the intervention name, a comma-separated list of the dimensions that are relevant to that intervention based on how it was discussed, and a summary of the concerns about that intervention that are relevant to those dimensions.”.

In each case, the results were compiled into a draft text that was then edited on the platform itself, before being exported to a Word file for further editing and formatting. Each claim made by the comCensus AI includes links to the one or more posts that generated it, This allows to verify directly each claim in the text against what participants actually said, and make any other necessary amendments.

Therefore, it is important to emphasize that the texts included in this report are not purely auto-generated summaries. The texts were generated with a distinctive algorithm designed to distil the level of consensus, and they were subsequently verified for their accuracy and edited by a human researcher - the author of this report.

2.4 Transparency and reproducibility

The comCensus page containing the list of participants and the full record of the discussion is available at: https://comcensus.org/public/publicMain.php?q=TRUSTparenCy_interventions

The pseudonyms associated with each claim are retained in the text, so that the reader can independently verify the accuracy and correctness of the synthesis offered in this document.

3. Results

3.1 Descriptive statistics

In total, N=87 participants responded to the invitation, and contributed a total of N=326 posts and reactions. Thus, the response rate was at least 16% and, given that at least 10% of invitees were on annual leave, the actual response rate is likely higher. The relative majority of participants self-declared to be Man (N=46, 53%), Mid-career (N=34, 39%), from a High-Income country (N=40, 46%), from a Public research performing organisation (N=65, 75%), and doing Quantitative Social Science (N=38, 44%).

All participants except 13 opted for sharing their name and affiliations, in order to be credited for this report. Their names are included, with sincerest appreciation, in the Appendix to this report.

The most frequently voted category was Open Policies (33%), but votes were spread over all five categories (Figure 1).

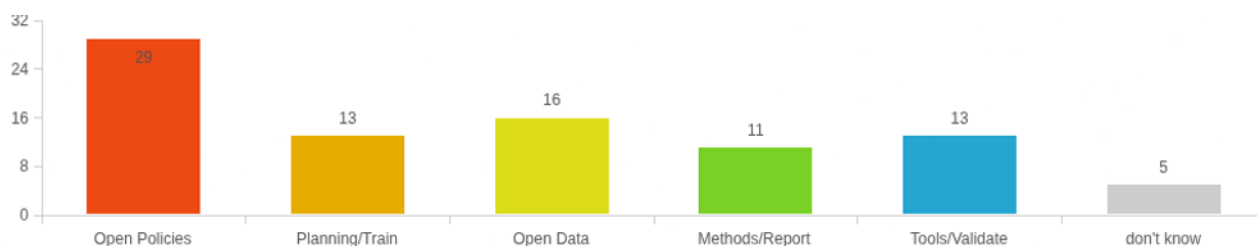


Figure 1: Distribution of votes across the five categories. Numbers above the bar are vote counts.

3.2 Concerns by category

Open Policies

Several participants viewed Open Policies as prone to becoming bureaucratic tick-box exercises that depress creativity and produce minimal compliance rather than meaningful behaviour change, especially when mandatory (Confident Dolphin, Bright Alligator, Philosophical Quelea). This concern drew support in replies that highlighted how institutional or journal mandates on items such as data-sharing statements (ILN 3), completeness of reporting (ILN 2), or registration (ILN 4) quickly lose utility once researchers focus only on appearing compliant (Bright Alligator, Confident Dolphin).

Many argued a persistent gap exists between policy creation and action, with interventions like funder policy on data management (ILN 8), government policy on registration (ILN 9), or research institution support services (ILN 10) proving time-consuming to negotiate and approve yet often resulting in “yet another policy” that few read or operationalise (Enthusiastic Earthworm, Philosophical Quelea, Uplifting Possum). Opinions split on whether this renders the whole category ineffective or merely signals that policies must be paired with training, incentives, and enforcement—which are frequently absent (Optimistic Crab, Philosophical Quelea, Focused Donkey).

Enforcement and infrastructure deficits were widely cited as rendering journal policies on data sharing (ILN 3), registration (ILN 4), or outcome-agnostic acceptance (ILN 7) largely performative; authors promise sharing “on reasonable request” but rarely deliver, while verification demands resources and oversight the current publishing system lacks (Lovely Ant, Ethical Urraca, Blissful Ovenbird). Some saw optional elements such as open science badges (ILN 5) or results-free peer-review (ILN 6) as easily circumvented without repercussions, producing only superficial uptake (Ethical Urraca, Unassuming Gorilla).

Equity concerns featured prominently: mandating openness without accompanying funding penalises smaller labs, early-career researchers, and those in lower-income settings through article processing charges or added workload, particularly around journal policies on reporting or data sharing (Courageous Urraca, Empathetic Xantus, Prosperous Quoll). Relatedly, participants noted that policies entangle multiple stakeholders influenced by capitalist incentives, so journals have little motivation to invest in verification or quality control unless it yields profit, limiting practical rollout of badges, registered reports, or outcome-agnostic acceptance (Uplifted Cheetah, Blissful Ovenbird).

Specific to results-free peer-review (ILN 6) and outcome-agnostic acceptance (ILN 7), some worried about added administrative burden, reviewers’ inability to separate methodological quality from findings, and insufficient evidence that gains in reproducibility justify the complexity (Happy Skylark); others countered that emerging meta-research demonstrates improvements in bias control, quality, and citations while system-level waste decreases, making the objection overstated (Focused Donkey). Government policy on

registration (ILN 9) elicited split views: some feared politicians lack domain expertise and could impose ill-fitting rules (Prosperous Chickaree), while others pointed to successful clinical-trial registration as evidence that such policies can reduce publication bias when properly scoped (Humorous Newfoundland).

Journal policies were additionally criticised for rigidity that ignores disciplinary nuance (e.g., badges favouring “mainstream” quantitative open science), fostering performative compliance and unintended hierarchies (ILN 5, Unassuming Gorilla, Lively Hippopotamus); replies were mixed, with some arguing policies can accommodate variation if co-created and others maintaining they remain too blunt (Flexible Tayra, Graceful Oyster). Overall the discussion treated these reservations as prevailing but not absolute, stressing that policies alone achieve little without aligned incentives, infrastructure, and monitoring (Optimistic Crab, Candid Rattlesnake, Focused Donkey).

Planning & Training

Several participants argued that Planning & Training interventions, while necessary, face serious practical barriers because shifting delivery from grassroots efforts to institutions tends to prioritise easily measured outputs like trainee completion rates over genuine behaviour change or skills development, often defaulting to ineffective lecture-style formats without hands-on practice (Candid Rattlesnake, Bright Alligator). This view drew broad agreement, with particular reservations expressed about ILN 18 (research institution provides training) and ILN 17 (researcher trained on study protocol), where institutional involvement was seen as inviting tick-box compliance that depresses creativity and produces minimal real-world uptake (Confident Dolphin, Confident Dolphin, Bright Alligator).

Opinions were split on whether such training can ever be cost-effective without accompanying mandates or rewards: some held that researchers already show high motivation (e.g., via departmental surveys) but lack the harder-to-teach skills for interventions like ILN 13 (preregistration), 15 (statistical analysis plan), or 14 (optimisation of study design), so cheap knowledge-focused approaches will continue to be favoured over resource-intensive skills training (Candid Rattlesnake, Lively Rhinoceros, Yummy Shark), while others countered that training rarely changes behaviour at all in “publish or perish” cultures where path-of-least-resistance incentives dominate and evidence from implementation research shows limited impact (Inspiring Partridge, Gentle Damsel fish).

A smaller set of posts raised unresolved concerns that one-size-fits-all training (e.g., ILN 12 on formulating the RQ or par16 on standardisation of experimental protocol) risks excluding non-standard approaches or assuming all researchers possess both the skill and will to engage, with advanced competencies like detailed preregistration or R programming potentially taking years and reproducing inequalities in under-resourced or Global South settings where job insecurity and heavy teaching loads make additional requirements burdensome rather than beneficial (Unreal Mouse, Sincere Wasp, Mediator Otter). These points remained with no clear community consensus emerging beyond agreement that institutional role models (ILN 19) and tailored field-specific delivery could mitigate some risks but would still require structural reforms to incentives and culture (Fair Kookaburra, Involved Quoll).

Open Data

Several participants viewed open data interventions as imposing high infrastructural and financial costs that institutions cannot easily sustain, particularly for data repositories (ILN 21), data storage solutions (ILN 23), and direct data deposition (ILN 24) amid ever-growing dataset sizes that outpace institutional capacity and may instead require national-level investment (Sympathetic Tarantula, Imaginative Rabbit). This concern drew some agreement as a field-dependent tradeoff rather than a universal flaw (Bright Alligator).

Opinions were split on whether mandatory elements such as long-term conservation, anonymisation, metadata requirements, and standardized data preparation procedures (ILN 28) create excessive burdens that divert researcher time from core science, with some calling these necessary costs for progress while others saw them as overlooked downsides that produce bureaucratic tick-box exercises instead of genuine utility (Imaginative Rabbit, Trusting Lionfish, Confident Dolphin).

Researchers broadly agreed that interventions aimed at researchers sharing data (ILN 25) and post-publication data requests (ILN 27) suffer from lagging training on FAIR practices, leading to superficial compliance such as vague “data available on reasonable request” statements that fail to deliver reusable outputs; this was tied to absent incentives, insufficient time allocation in grants, and a lack of mandatory training by data stewards as part of professional development (Thriving Caribou, Stunning Seal, Talented Reindeer).

Disciplinary diversity emerged as a contested point: some argued that data quality control (ILN 29), standardized file system structures (ILN 22), and data management tools (ILN 26) cannot be applied uniformly because qualitative or sensitive data are hard to anonymise or share without creating inequities that favour quantitative fields, while others countered that these costs are still justified or can be mitigated through flexible national infrastructure (Sympathetic Tarantula, Cooperative Damsel, Pioneering Caracal).

Several participants held that open data lacks practical quality assurance in practice, with repositories often accepting unstructured files or datasets missing codebooks, turning ILN 29 into the most difficult unmet need absent dedicated teams (Fit Tegu, Idealistic Newfoundland, Mediator Impala).

A supported view held that the entire category is a “moving target” because there is little consensus on repositories, deposit standards, long-term curation responsibility, or funding models, compounded by rapidly evolving funder mandates and the need for stronger enforcement mechanisms to move beyond performative uptake (Talented Reindeer, Affable Kangaroo).

Data safety, GDPR risks, and data sovereignty concerns (especially for Indigenous or sensitive materials) were raised as unresolved barriers to researcher sharing (ILN 25) that policies currently handle poorly (Victorious Oyster, Mediator Impala, Warm Tarsier).

Methods & Reporting

Several participants expressed reservations about standardization interventions in this category potentially becoming overly prescriptive or bureaucratic, risking a “big brother” perception among colleagues if not presented supportively (Amazing Reindeer, Confident Dolphin). One researcher argued that items such as a software tool that documents data retrieval procedures (ILN 33), a tool to standardize data analysis procedures (ILN 36), researcher uses common method for data analysis (ILN 35), or researcher uses standard method for data acquisition (ILN 31) could impose unrealistic rigid standards, gatekeeping, or workflow complications in fields like psychology where software options and analytic pipelines vary widely, though reporting checklists (ILN 38) and recommended documentation frameworks were viewed more favourably if framed as flexible best-practice starting points rather than mandates (Attractive Fly).

Others contended that practical implementation of standardized methods and reporting requires broad community agreement on standards, which is difficult to achieve given differences in culture, power dynamics, and traditions across research groups (Keen Chicken). One participant noted these interventions would be difficult to implement consistently across disciplines, institutions, and regions because methodological approaches differ and reporting guidelines remain concentrated in biomedicine rather than universal (Humorous Newfoundland).

Concerns were also raised that data-collection reporting protocols and standardized documentation frameworks (ILN 32, 34, 37) for complex real-world scenarios (such as multi-source online data pipelines with gaps, non-random missingness, or rapidly changing collection methods) risk being built on overly simple experimental templates, creating false confidence in validity while generating high ongoing costs for protocol updates and institutional commitment (Focused Donkey). One researcher highlighted that these interventions often overlook or fail to fit research in methodological development, where the classical hypothesis-acquisition-testing workflow does not apply (Prosperous Groundhog).

Reservations were voiced that forcing common methods or standardized tools for data analysis (ILN 35, 36, 37), real-time procedure documentation, or reporting checklists (ILN 38) could institutionalize lowest-common-denominator approaches, penalize innovative econometric or experimental designs, create

administrative burdens that disproportionately affect junior researchers, and encourage tick-box compliance without deepening critical thought (Humble Echidna, Amiable Stoat). One participant suggested such standardization constrains the creativity of the data-analytic process itself (Playful Copepod).

A further concern specific to psychological interventions was that inadequate reporting limits clinical replicability in practice and in replication trials, producing research waste even when checklists or standardized documentation frameworks are nominally used (Persistent Capybara). These points largely stood without broad contestation or replies that altered their reception.

Tools & Validation

Several participants argued that wide variation in researchers' skills and knowledge across fields renders Tools & Validation interventions ineffective or lacking in credibility without deeper training (Perfect Frogmouth, Bright Alligator, Confident Dolphin). This view drew broad agreement as a core practical barrier.

Many emphasised that researchers should retain freedom to select tools relevant to their specific research question, given vast differences across disciplines from arts and humanities to life sciences and mathematics (Enough Guanaco, Bright Alligator, Candid Rattlesnake, Lively Rhinoceros), a point that garnered substantial support.

Concerns focused on specific interventions such as statistical tools to estimate power (ILN 40), bias analysis (ILN 41), multiverse analysis (ILN 42), and effect-size filters (ILN 43), which several described as costly to scale in complex samples like multilingual clinical or developmental studies where recruitment is slow and conditions vary (Adventurous Gazelle, Happy Xolmis). In fields like chemosensory science, power estimation and replication (ILN 47) were called often infeasible for rare populations, with high analyst burden for bias analyses yielding marginal gains that rarely justify the cost (Clean Sandgrouse, Uplifting Possum); however this drew opposition from those stressing power's role in controlling false positives and cherry-picking, especially in underpowered social science (Gentle Damselish, Shining Seal, Self-assured Sambar).

Peer-reviewer checks of reporting (ILN 45) and discrepancy review (ILN 46) were highlighted for hidden costs, as they add hours to already strained reviewers at a crisis point and risk becoming unpaid labour without publisher investment (Flexible Tayra, Happy Skylark, Fit Tegu), a concern that received broad support. Replication (ILN 47) itself was seen as rare in practice, with under 2% of submissions being replications because researchers see little incentive to invest the time (Flexible Tayra).

Some raised that these interventions require outside effort from non-primary researchers and carry risks of direct conflict that could discourage uptake (Civil Chinchilla, Happy Skylark), though this remained mixed without clear consensus. A minority view that power analysis is a "necessary nightmare" that trades robustness for accessibility (Fair Dugong) was opposed by most replies, which affirmed its necessity (Gentle Damselish, Shining Seal, Incredible Sambar).

Cross-category concerns

One of the most common concerns expressed across intervention categories is that interventions often lead to superficial compliance or tick-box exercises when mandates outpace training, incentives, enforcement, and infrastructure, undermining their intended benefits (Philosophical Quelea, Optimistic Crab, Blissful Ovenbird). This concern drew broad agreement in threads emphasising the policy-action gap and risks of performative adoption.

The discussion also expressed widespread reservations that varying skill levels, disciplinary differences, and contextual factors (such as qualitative versus quantitative work or resource availability in smaller labs) make uniform tools, validation approaches, policies, or training ineffective without greater flexibility and bottom-up tailoring, as rigid application can reduce credibility or penalise certain fields (Perfect Frogmouth, Enough Guanaco, Candid Rattlesnake, Candid Rattlesnake).

Several participants contended that hidden resource demands cut across categories (including reviewer time for validation and discrepancy checks, sustained skills training over knowledge-focused lectures, infrastructure for data or complex documentation, and opportunity costs for early-career or under-funded researchers) frequently exceed available support and may not justify marginal gains in all contexts (Flexible Tayra, Candid Rattlesnake).

Opinions were split on top-down policies versus community-driven or incentive-based alternatives: some saw proactive policies as essential to counter perverse incentives and embed values (Candid Rattlesnake), while others countered that they invite bureaucracy, unintended consequences, or inequity without aligned assessment criteria (Bright Alligator, Confident Dolphin, Confident Dolphin).

4. Differences by group

Across the demographic variables collected, two main patterns emerged.

4.1 Gender differences

Women more frequently raised reservations about the Planning & Training and the Tools & Validation category than men did (Figure 1), often citing wide variation in researchers' skills and knowledge that can render such interventions ineffective or lacking in credibility (Perfect Frogmouth, Graceful Oyster, Happy Xolmis), a concern that drew substantial agreement from both men and women (Unassuming Gorilla, Courageous Urraca, Incredible Sambar). Women also highlighted practical infeasibility and high analyst burden in specific fields such as chemosensory science or plant pathology, where power analysis or replication can demand unfeasible sample sizes or yield highly variable outcomes that require adjusted expectations around replicability (Clean Sandgrouse, Happy Xolmis, Fair Dugong), points that met opposition from some who countered that power analysis remains essential for controlling false positives even if difficult (Gentle Damselish, Shining Seal) while others expressed partial agreement (Unassuming Gorilla).

Votes by category, partitioned by gender

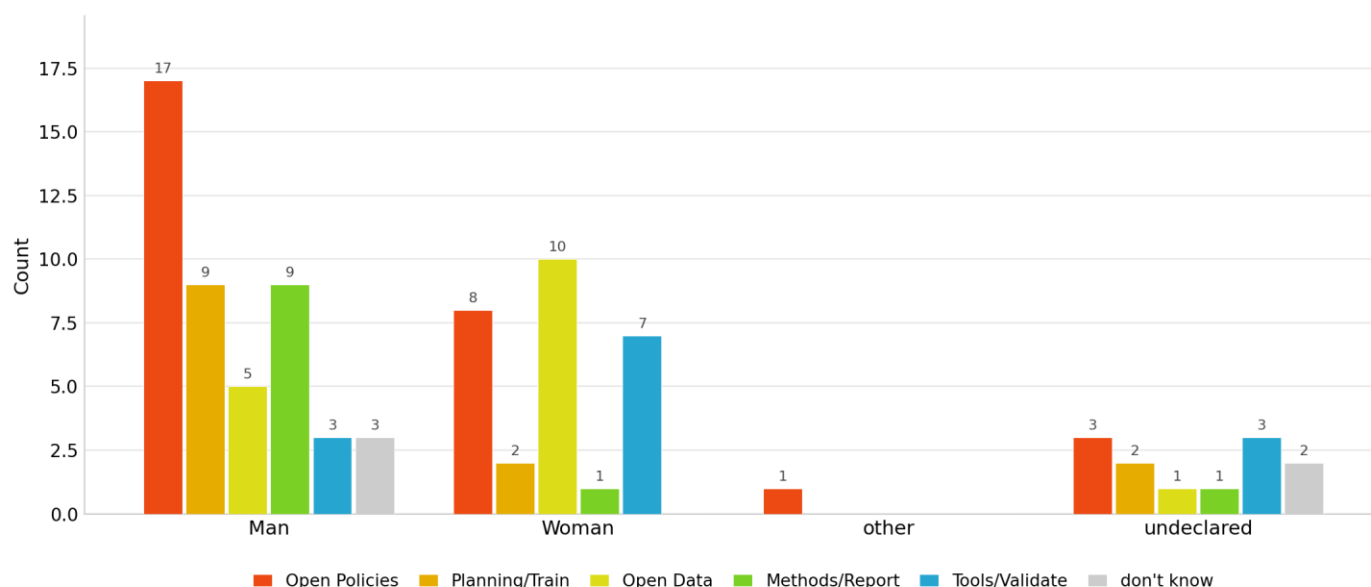


Figure 2: Distribution of votes by gender. Colours reflect intervention category (Red: Open Policies; orange: Planning & Training; Yellow: Open Data; Green: Methods & Reporting; Blue: Tools & Validation; Grey: don't know).

In contrast, men more often selected and expressed reservations about Open Policies, arguing that such interventions risk becoming brute-force measures, box-ticking exercises, or sources of bureaucratic burden that depress research creativity without delivering intended benefits (Bright Alligator, Confident Dolphin, Uplifting Possum, Lovely Ant), views that received broad agreement including in replies that reinforced how mandatory policies lead to minimal compliance rather than genuine uptake (Confident Dolphin, Bright Alligator). Men additionally stressed that government or journal policies can lack nuance, invite unintended consequences such as sovereignty risks in governmental data or performative open-science washing, and prove hard to enforce without dedicated resources (Prosperous Chickaree, Trusting Lionfish, Ethical Urraca), concerns that were contested by others who noted policies can still provide helpful leverage and norms when paired with flexibility (Focused Donkey, Flexible Tayra).

Some women raised related but distinct equity objections that mandates without funding can penalise smaller labs or create performative badges that ignore methodological diversity (Courageous Urraca, Unassuming Gorilla), a split that produced mixed reception (Keen Chicken, Flexible Tayra, Graceful Oyster).

Women were also more visible in voicing that training investments tied to tools or planning categories can be worthwhile once embedded in lab culture yet remain hampered by resource constraints or misalignment with precarious careers, particularly from Global South or early-career perspectives (Sincere Wasp, Mediator Otter, Philosophical Quelea), positions that stayed largely unrated or drew partial agreement without clear community consensus. Men more readily framed Open Data or Methods & Reporting reservations around infrastructure costs, disciplinary mismatches, or risks of constraining innovation in analysis pipelines (Imaginative Rabbit, Sympathetic Tarantula, Tidy Llama), points that elicited mixed or opposed replies depending on the thread.

4.2 Career stage differences

A second pattern suggests a tendency for early career researchers to vote proportionally more for Tools/Validation, Open data and Open policies, whereas the topics of Planning/Training and Methods/Reporting were mainly discussed among Mid-career and Established researchers (Figure 3). Early career researchers tended to emphasize equity, financial burdens, and limited personal control in Open Policies and Tools & Validation, while established researchers focused more on systemic gaps like enforcement, institutional lag, and whether policies produce real behavior change.

Votes by category, partitioned by career stage

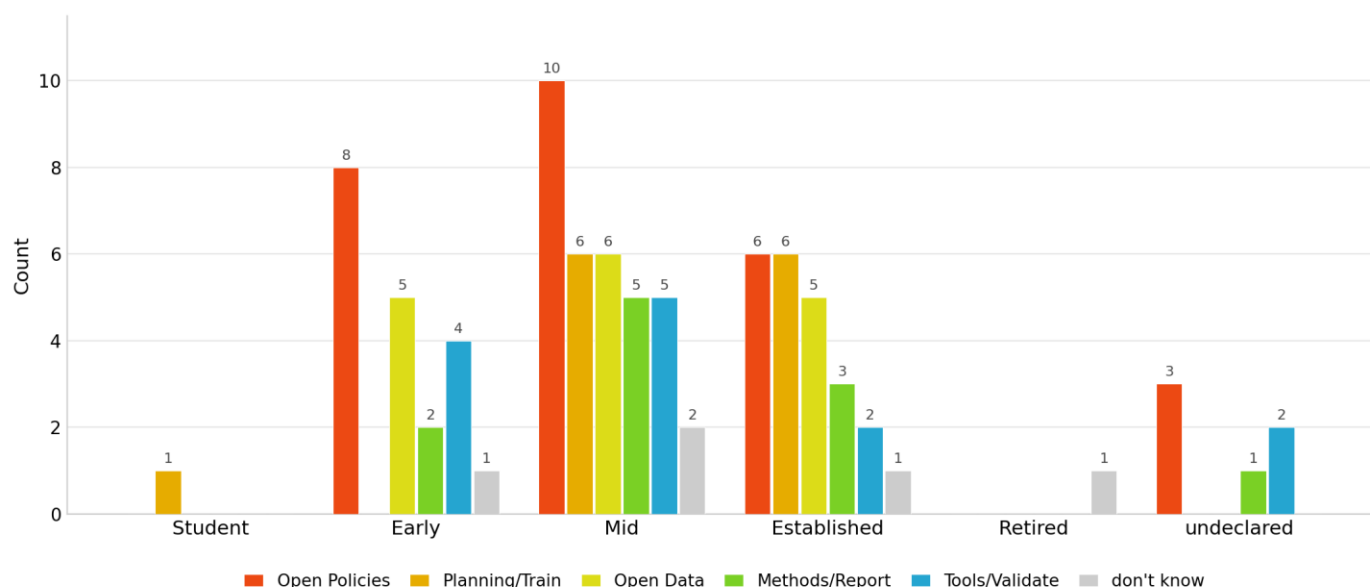


Figure 3: Distribution of votes by career stage. Colours reflect intervention category (Red: Open Policies; orange: Planning & Training; Yellow: Open Data; Green: Methods & Reporting; Blue: Tools & Validation; Grey: don't know).

Early-career participants often framed Open Policies as risking inequity because mandates for data sharing or open access (via APCs or compliance costs) penalize smaller labs, early-career investigators, and lower-resource settings without accompanying infrastructure or funding (Courageous Urraca, Lovely Ant, Ethical Urraca, Graceful Oyster). These points received mixed or supported reception, with some replies concurring on feasibility barriers while others countered that openness can still be achieved at multiple levels or that individual researcher responsibility remains key (Keen Chicken, Flexible Tayra).

Early-career voices also raised reservations about Tools & Validation (e.g., power analysis as a “necessary nightmare” that trades robustness for accessibility, or replication infeasible in hard-to-recruit populations), but these were largely opposed by broader community replies that stressed power’s importance for controlling false positives and cherry-picking (Fair Dugong, Clean Sandgrouse), with established researchers particularly likely to disagree and argue such tools remain essential even if adjusted for feasibility (Gentle Damselish, Shining Seal, Uplifting Possum).

Mid-career contributors, who supplied several of the most-supported posts, converged on skill variation and freedom-to-choose arguments against heavy reliance on Tools & Validation or top-down Open Policies. They argued that wide differences in researcher abilities render validation tools ineffective or lacking credibility unless skills training is deepened, and that communities should develop tools bottom-up rather than have them imposed (Perfect Frogmouth, Enough Guanaco, Bright Alligator, Candid Rattlesnake). These posts drew broad agreement, including from both early- and established-career respondents. On Planning & Training, mid-career posts stressed that institutionalizing training risks box-ticking, lecture-style delivery, and metrics focused on completion rather than behavior change; they advocated delivery by active researchers rather than administrators (Candid Rattlesnake, Bright Alligator, Confident Dolphin, Confident Dolphin). This perspective was supported, with replies noting that motivation is already high but skill translation is the harder, under-resourced piece.

Established researchers more frequently highlighted implementation gaps, “open-science-washing,” and the disconnect between policy and measurable behavior change, often viewing policies as cheap to write but costly or ineffective without monitoring. They described policies as easily circumvented or performative unless paired with verification, incentives, and cultural shifts, and warned that institutions lag because they “care less about science” than individual researchers (Uplifting Possum, Philosophical Quelea, Blissful Ovenbird, Shining Seal, Honest Ant). These points received mixed-to-supported framing; some replies agreed on the need for enforcement and monitoring while others noted that policies still create useful leverage and norms even if imperfect (Focused Donkey, Flexible Tayra). Established participants were also more likely to contest or partially agree on Open Data costs and disciplinary mismatches, sometimes arguing that training and infrastructure lags are the real bottleneck rather than the policies themselves (Imaginative Rabbit, Prosperous Groundhog).

Across all levels, Open Policies attracted the greatest volume of reservations, but the framing shifted: early-career posts stressed immediate personal and equity costs, mid-career posts emphasized unintended consequences and loss of flexibility, and established posts focused on long-term ineffectiveness without enforcement. Planning & Training and Tools & Validation reservations were more concentrated among mid-career voices and generally received stronger community support when they highlighted training quality and skill gaps rather than outright rejection. Purely oppositional stances (e.g., dismissing power analysis or replication outright) were more common among early-career posters but were contested or rejected by established and mid-career respondents who saw these tools as non-negotiable for credible science. Unrated posts from all levels added nuance around disciplinary specificity and Global-South/resource-constrained contexts but did not shift the prevailing supported views.

4.3 RRR breakdown

To gain further insights into the nature of the concerns and challenges perceived by participants, each of the 41 listed types of interventions was examined in terms of the Resources-Receptivity, Relevance (RRR) themes and across the 9-dimensional system that categorises inclusivity challenges and epistemological disparities developed in Deliverable 1.3. As shown in Table 1, all types of interventions were discussed in

terms of one or more themes and two or more dimensions, and all of the nine dimensions of the RRR scheme were relevant to at least some of the interventions.

Table 1: RRR-analysis of concerns expressed by participants for each intervention. The RRR scheme comprises a total of three themes and nine dimensions: Resources (funds, time, infrastructure); Receptivity (knowledge, incentives, regulations); Relevance (phenomena stability, methods homogeneity, philosophical compatibility). See introduction and D1.3 for further explanations.

Category	Intervention name	RRR dimensions	Summary
Open Policies	Journal policy on completeness of reporting	Receptivity: Incentives Resources: Time Relevance: Methods homogeneity	Several participants viewed journal policies as prone to box-ticking that fails to improve actual practice and consumes researcher time without delivering proportional gains (Confident Dolphin, Bright Alligator, Philosophical Quelea). This concern drew broad agreement when linked to open-policies interventions generally. Others noted that without monitoring or enforcement such policies remain largely performative (Lovely Ant, Ethical Urraca), a point that received support in replies.
	Journal policy on data sharing	Resources: Funds, Infrastructure Receptivity: Regulations Relevance: Philosophical compatibility	Community discussion highlighted that data-sharing mandates can impose infrastructure and compliance costs that fall unevenly on smaller labs or fields with sensitive data, potentially exacerbating inequities (Courageous Urraca, Imaginative Rabbit). Several argued these policies risk becoming tick-box exercises or “open-science washing” unless paired with enforcement and resources (Uplifting Possum, Philosophical Quelea). Views on whether policies themselves or only their implementation are the problem were contested (Focused Donkey, Flexible Tayra).
	Journal policy on registration	Receptivity: Incentives Resources: Time Relevance: Methods homogeneity	Participants noted that registration policies can slow publication in competitive fields, creating incentive misalignments, and that governmental or journal versions may lack nuance for different disciplines (Prosperous Chickaree, Bright Alligator). Some saw value in clinical-trial registration while others worried about bureaucratic burden and unintended restriction of exploratory work (Humorous Newfoundland, Snazzy Xolmis); opinions remained split.
	Journal awards open science badges	Receptivity: Incentives Relevance: Methods homogeneity Philosophical compatibility	Badges were described as low-cost signals that nonetheless risk rewarding only “mainstream” open-science practices and disadvantaging qualitative or non-standard methods (Unassuming Gorilla, Lively Hippopotamus). Several participants viewed them as performative when verification is absent (Blissful Ovenbird,

Category	Intervention name	RRR dimensions	Summary
			Graceful Oyster); the discussion showed mixed support with calls for greater flexibility.
	Results-free peer-review	Resources: Time Receptivity: Knowledge Incentives	Reservations centred on increased editorial workload, reviewer training demands, and uncertainty whether reviewers can meaningfully evaluate without results (Happy Skylark, Bright Alligator). These concerns were contested; several replies cited emerging evidence of improved quality and argued system-level administrative savings outweigh initial costs (Focused Donkey, Flexible Tayra), producing an open disagreement.
	Outcome-agnostic acceptance by journal	Receptivity: Incentives Resources: Time Relevance: Methods homogeneity	Discussed mainly as part of registered-report models, this intervention was seen by some as requiring workflow changes that journals may resist and that could disadvantage fields with high phenotypic variability (Candid Rattlesnake, Charismatic Cat). Support was mixed; a minority viewed it as essential to combat publication bias while others worried about practical uptake (Humorous Newfoundland).
	Funder policy on data management	Resources: Funds, Infrastructure Receptivity: Regulations	Funder policies were criticised for imposing costs and infrastructure demands without always supplying corresponding support, especially in lower-resource settings (Imaginative Rabbit, Courageous Urraca). Several participants noted tension with regulations such as GDPR or data-sovereignty rules (Mindful Quelea, Victorious Oyster); the view drew partial agreement but also counter-arguments that such policies ultimately reduce long-term waste (Trusting Lionfish).
	Government policy on registration	Receptivity: Regulations, Incentives Relevance: Philosophical compatibility	Participants expressed concern that government mandates risk misalignment with scientific needs because politicians lack domain expertise (Prosperous Chickaree, Bright Alligator). This was contested by those pointing to successful clinical-trial registration that reduces bias (Humorous Newfoundland, Stunning Seal); the exchange remained contested.
	Research institution provides support services	Resources: Infrastructure, Funds Receptivity: Knowledge	Institutional support services were seen as essential yet frequently under-resourced, leading to uneven availability across institutions and countries (Sympathetic Tarantula, Imaginative Rabbit). Several argued that without dedicated funding these services become another unfunded mandate (Philosophical Quelea).

Category	Intervention name	RRR dimensions	Summary
Planning & Training	Formulate the RQ	Receptivity: Knowledge Relevance: Methods homogeneity Philosophical compatibility	Concern focused on one-size-fits-all interpretations that may not suit qualitative or exploratory work (Unreal Mouse, Lively Rhinoceros). The point received mixed reception; some saw clear research questions as universally beneficial while others warned against philosophical incompatibility with interpretivist approaches.
	Preregistration	Resources: Time Receptivity: Incentives Relevance: Methods homogeneity	Preregistration was described as adding burden that may be justified only in high-degrees-of-freedom fields; in others it can become performative or discourage creative designs (Charismatic Cat, Candid Rattlesnake). Views split between those emphasising its role in reducing QRPs and those stressing context-dependence (Prosperous Chickaree, Focused Donkey).
	Optimization of study design	Receptivity: Knowledge Resources: Time	Participants noted that design optimisation requires skills that are unevenly distributed and that training alone may not overcome incentive structures favouring speed (Candid Rattlesnake, Fair Kookaburra). The discussion showed broad support for the goal but reservations about practical uptake without cultural change.
	Statistical analysis plan	Receptivity: Knowledge Resources: Time Relevance: Methods homogeneity	Similar concerns to preregistration were raised; plans can become rigid and penalise innovative or field-specific approaches (Attractive Fly, Unreal Mouse). Several saw value when flexibly applied but warned against checklist-style compliance (Confident Dolphin).
	Standardization of experimental protocol	Relevance: Methods homogeneity, Phenomena stability Receptivity: Regulations	Standardisation was viewed as difficult in fields with high phenotypic variability or sensitive data, where protocols cannot be uniformly applied (Happy Xolmis, Mindful Quelea). The community largely agreed that disciplinary differences must be accommodated.
	Researcher trained on study protocol	Receptivity: Knowledge Resources: Time	Training was regarded as necessary but hard to scale effectively; lecture-style delivery often fails to produce behaviour change (Candid Rattlesnake, Bright Alligator). Several emphasised that training must be delivered by active researchers rather than administrators to be credible (Confident Dolphin, Graceful Oyster); this received strong support.
	Research institution provides training	Receptivity: Incentives, Knowledge	Institutional training risks becoming a tick-box exercise when delivered by non-research staff or without alignment to daily workflows (Confident

Category	Intervention name	RRR dimensions	Summary
		Resources: Funds	Dolphin, Bright Alligator). Participants broadly agreed that researcher-led, field-specific training is more effective but requires dedicated funding that institutions often withhold.
	Research institution provides role models	Receptivity: Knowledge, Incentives	Role models were mentioned infrequently; one participant suggested they could mitigate risks of institutional training but the point remained unrated and isolated (Candid Rattlesnake).
Open Data	Data repository	Resources: Infrastructure, Funds Receptivity: Regulations	Repositories were seen as essential yet costly to maintain, especially for large or sensitive datasets; many institutions lack adequate infrastructure (Sympathetic Tarantula, Imaginative Rabbit). Concerns about long-term curation and GDPR compliance were widely shared (Mediator Impala, Victorious Oyster).
	Standardized file system structure	Receptivity: Knowledge Resources: Time	Standardised structures were viewed as helpful but requiring training and time that researchers often lack; uptake depends on disciplinary norms (Attractive Fly, Skilled Possum). The discussion was mixed.
	Data storage solution	Resources: Infrastructure, Funds	Storage solutions were repeatedly flagged as increasingly expensive and unevenly available, particularly at teaching-oriented or lower-income institutions (Sympathetic Tarantula, Imaginative Rabbit). National-level infrastructure was proposed as a partial remedy but recognised as costly.
	Direct data deposition	Resources: Time Receptivity: Regulations Relevance: Phenomena stability	Direct deposition raises workload, anonymisation, and regulatory hurdles (especially for sensitive or fast-changing data); several participants noted it can penalise smaller labs (Courageous Urraca, Mindful Quelea). Views were mixed on whether benefits outweigh costs in all fields.
	Researchers share data	Receptivity: Incentives Resources: Time Relevance: Philosophical compatibility	Data sharing was described as time-intensive with weak incentives; many researchers default to “available on request” statements that are rarely honoured (Lovely Ant, Blissful Ovenbird). Concerns about free-riding and idea theft were also raised (Relaxed Junco). The community broadly supported the principle but noted practical and cultural barriers.
	Data management tool	Receptivity: Knowledge	Tools were criticised for steep learning curves, maintenance costs, and risk of creating new dependencies (Skilled Possum, Attractive Fly).

Category	Intervention name	RRR dimensions	Summary
		Resources: Infrastructure	Several called for modular, well-supported tools and more data stewardship rather than over-optimised IT solutions.
	Post-publication data requests	Receptivity: Incentives Resources: Time	Requests were seen as low-cost for journals but high-burden for researchers; compliance is patchy and enforcement rare (Lovely Ant). The discussion treated this as a weak substitute for proactive sharing.
	Standardized data preparation procedure	Receptivity: Knowledge Relevance: Methods homogeneity	Standard procedures risk being too generic for qualitative or complex multi-source data, potentially creating false confidence in reproducibility (Focused Donkey). Participants called for field-specific guidance.
	Data quality control	Resources: Time, Infrastructure Receptivity: Knowledge	Quality control was described as one of the most important yet labour-intensive aspects; few groups have dedicated teams (Fit Tegu, Idealistic Newfoundland). This view drew broad agreement.
Methods & Reporting	Researcher uses standard method for data acquisition	Relevance: Methods homogeneity, Phenomena stability, Philosophical compatibility	Standard methods were seen as inappropriate for method-development research or fields with high variability; rigid application can stifle innovation (Prosperous Groundhog, Unreal Mouse). The point received mixed reception.
	Researcher uses standardized documentation framework (data collection)	Resources: Time Receptivity: Knowledge	Documentation frameworks add administrative load that may not scale to complex, real-time, or multi-pipeline data collection (Focused Donkey, Attractive Fly).
	Software tool that documents data retrieval procedures	Receptivity: Knowledge Resources: Infrastructure	Concern centred on the unrealistic expectation of a single tool across diverse software ecosystems and the learning cost involved (Attractive Fly, Skilled Possum).
	Real-time procedure documentation	Resources: Time Relevance: Methods homogeneity	Real-time documentation was viewed as burdensome in fast-moving or creative experimental settings and potentially discouraging non-standard designs (Prosperous Groundhog, Attractive Fly).
	Researcher uses common method for data analysis	Relevance: Methods homogeneity Receptivity: Incentives	Mandating common methods risks locking in lowest-common-denominator approaches and penalising innovative techniques (Humble Echidna, Attractive Fly). This was contested by those who saw standardisation as protective against questionable practices.

Category	Intervention name	RRR dimensions	Summary
	Tool to standardize data analysis procedures	Receptivity: Knowledge Relevance: Methods homogeneity	Similar reservations as above; tools can create rigidity and new dependencies (Skilled Possum, Humble Echidna).
	Researcher uses standardized documentation framework (data analysis)	Resources: Time Receptivity: Knowledge	Documentation was seen as necessary but time-consuming; without incentives it risks becoming performative (Attractive Fly, Uplifting Possum).
	Use of reporting checklist	Receptivity: Incentives Resources: Time	Checklists were criticised for encouraging tick-box compliance that reduces critical thinking and burdens junior researchers (Humble Echidna, Amiable Stoat). Several countered that well-designed checklists can educate when paired with training (Blissful Ovenbird).
Tools & Validation	Statistical tool to estimate power	Receptivity: Knowledge Resources: Time Relevance: Methods homogeneity	Power analysis was called a “necessary nightmare” that is hard to implement robustly and may reduce accessibility (Fair Dugong, Clean Sandgrouse). This minority view was opposed by many who argued it is essential for controlling false positives and ultimately saves resources (Gentle Damselfish, Shining Seal, Self-assured Sambar, Incredible Sambar).
	Bias analysis method	Receptivity: Knowledge Resources: Time	Bias analysis was grouped with other validation tools seen as imposing high analyst burden whose marginal gains may not justify cost in some fields (Clean Sandgrouse). The concern was contested; several maintained such methods are valuable when feasible.
	Multiverse analysis	Receptivity: Knowledge Resources: Time Infrastructure	Multiverse approaches were viewed as computationally intensive and skill-demanding, limiting scalability in complex or multilingual samples (Adventurous Gazelle). Mixed views emerged on whether transparency can substitute when full multiverse analysis is impractical.
	Authors use effect-size filter	Relevance: Phenomena stability Receptivity: Incentives	Rigid effect-size filters were criticised for potentially discarding theoretically meaningful small effects in developmental or clinical research (Adventurous Gazelle, Happy Xolmis). The point received partial agreement.
	Open workflow platform	Receptivity: Knowledge Resources: Infrastructure	Platforms were seen as useful but requiring training and maintenance; risk of over-dependence on institutional IT was highlighted (Skilled Possum).

Category	Intervention name	RRR dimensions	Summary
	Peer-reviewer checks reporting	Resources: Time Receptivity: Incentives	Peer reviewers are already overburdened; adding detailed reporting checks was described as unsustainable without compensation or dedicated staff (Flexible Tayra, Fair Chicken). This drew broad support, with calls for journals to fund such labour.
	Discrepancy review	Resources: Time, Infrastructure	Similar workload concerns were raised; discrepancy review was seen as valuable but likely to remain volunteer labour or unfunded (Flexible Tayra). The community largely agreed on the practical barrier.
	Replication	Resources: Funds, Time, Infrastructure Relevance: Phenomena stability	Replication was flagged as especially costly or infeasible for rare populations, longitudinal studies, or fast-changing phenomena (Clean Sandgrouse, Happy Xolmis, Adventurous Gazelle). While the scientific value was acknowledged, low submission rates and funding difficulties were widely cited (Flexible Tayra).

5. Conclusions

Members of national reproducibility networks, when invited to share and discuss anonymously their concerns for the cost-effectiveness and feasibility of a list of 41 types of interventions, responded in high numbers and expressing numerous and diverse opinions. The concerns they expressed map well onto challenges in terms of resources, receptivity and relevance suggested by our analysis of the literature and developed into the RRR scheme of inclusivity challenges and epistemological disparities (D1.3). Given that respondents are a self-selected group of reproducibility advocates, these results likely under-estimate the extent and diversity of concerns that may be common among the broader research community.

Overall, these results validate TRUSTparenecy's assumption that the interventions and policies to improve reproducibility present potential challenges and unintended consequences that risk being overlooked by mainstream metascience, and support the project's tenet that policy-making needs to be informed by bottom-up, anonymous feedback collected from the research community. Our results also validate the scheme developed in D1.3, by showing how each of the nine dimensions captures distinctive and real concerns expressed by the community.

The transparent and efficient methodology implemented on the comCensus platform will allow these results, and further insights yet to be distilled from the discussions that are going to take place during the pilot testing phase, to directly inform the work and outputs of TRUSTparenecy.

6. Appendix

6.1 Line-numbered list of interventions

The following list matches the list of interventions and line numbers that were shown to participants. Specific interventions are cited in the text in terms of their Intervention Line Number (ILN).

- | | |
|--|--|
| 1. Open Policies
2. Journal policy on completeness of reporting
3. Journal policy on data sharing
4. Journal policy on registration
5. Journal awards open science badges
6. Results-free peer-review
7. Outcome-agnostic acceptance by journal
8. Funder policy on data management
9. Government policy on registration
10. Research institution provides support services

11. Planning & Training
12. Formulate the RQ
13. Preregistration
14. Optimization of study design
15. Statistical analysis plan
16. Standardization of experimental protocol
17. Researcher trained on study protocol
18. Research institution provides training
19. Research institution provides role models

20. Open Data
21. Data repository
22. Standardized file system structure
23. Data storage solution
24. Direct data deposition
25. Researchers share data | 26. Data management tool
27. Post-publication data requests
28. Standardized data preparation procedure
29. Data quality control

30. Methods & Reporting
31. Researcher uses standard method for data acquisition
32. Researcher uses standardized documentation framework (data collection)
33. Software tool that documents data retrieval procedures
34. Real-time procedure documentation
35. Researcher uses common method for data analysis
36. Tool to standardize data analysis procedures
37. Researcher uses standardized documentation framework (data analysis)
38. Use of reporting checklist

39. Tools & Validation
40. Statistical tool to estimate power
41. Bias analysis method
42. Multiverse analysis
43. Authors use effect-size filter
44. Open workflow platform
45. Peer-reviewer checks reporting
46. Discrepancy review
47. Replication |
|--|--|

6.2 List of contributors

Here below is the list of all participants who agreed to share their name.

Surname	Forename	Affiliation
ADAIME DE BORBA	Anderson	Mackenzie Presbyterian University
Amaral	Olavo	Associate Professor, Institute of Medical Biochemistry, Federal University of Rio de Janeiro
Arcara	Giorgio	na
Auer	Susann	Postdoc, Dept. Plant Physiology, Faculty of Biology
Ayala	Ana Patricia	Research services librarian, University of Toronto
Barbour	Ginny	EIC

Belyk	Michel	Edge Hill University
Bochynska	Agata	University of Oslo
Bokhove	Christian	na
Bouter	Lex	na
Buljan	Ivan	Faculty of Humanities and Social Sciences, University of Split
Cawthorn	William	na
Ceneviva	Ricardo	na
Centeno	Eduarda	Executive Director BRRN
Conte	Erica	Unity Health Toronto
Costa	Ana	Post-doc reseracher
De Luca	Giulia Raffaella	na
Draper-Rodi	Jerry	Health Sciences University
Evans	Thomas	Professor in Occupational Psychology and Open Scholarship, University of Greenwich, UK
Fanelli	Daniele	Heriot-Watt University and comCensus
Findor	Andrej	Comenius University in Bratislava
Frank	Maximilian	na
Giner-Sorolla	Roger	University of Kent
Glerean	Enrico	Staff scientist / Aalto University
Grassi	Massimo	Department of General Psychology, University of Padua, Italy
Guzman-Ramirez	Liz	Data Steward's Team Lead
Höffler	Jan H.	na
Ihle	Malika	LMU Open Science Center coordinator
Inguyesi	Christine	Research Scientist/ Kenya Agricultural and Livestock Research Organization
Jola	Corinne	na
Kavouras	Panagiotis	Centre for Medical Ethics, University of Oslo
Kačmár	Pavol	Department of Psychology, Pavol Jozef Šafárik University in Košice, Košice, Slovakia
Kurschilgen	Michael	na
Learmonth	Gemma	Lecturer in Psychology, University of Stirling
Lee	Clement	na

Liuzza	Marco Tullio	Associate Professor, Department of Developmental Psychology and Socialization
Lunic	Anita	Senior assistant, Department of Philosophy, Faculty of Humanities and Social Sciences, Split
Makel	Matthew	Professor, Werklund School of Education, University of Calgary
Manzoni	Gian Mauro	Professor of clinical psychology, Department of Brain and Behavioral Sciences. University of Pavia
Marinho	Ana Carolina	na
Marusic	Ana	na
Mazzuca	Claudia	Post Doc- Department of Dynamic and Clinical Psychology, Sapienza University of Rome
Miller-Dicks	Matt	na
Miniusi	Carlo	University of Trento
Montefinese	Maria	Tenure Track Researcher, Department of Developmental Psychology and Socialisation, Padua University
Nakagawa	Shinichi	Prof / Biological Sciences / University of Alberta
Nosáľová	Ester	Assistant professor / Department of Psychology / Pavol Jozef Safarik University in Kosice
Nüst	Daniel	TUD Dresden University of Technology
O'Connor	Daryl	Professor of Psychology, University of Leeds
Ocklenburg	Sebastian	na
PENNINGTON	CHARLOTTE	SENIOR LECTURER IN PSYCHOLOGY, ASTON UNIVERSITY, BIRMINGHAM, UK. EMAIL: c.pennington@aston.ac.uk
Pardo Pardo	Jordi	University of Ottawa
Parma	Valentina	Monell Chemical Senses Center
Penido	Carmen	Senior scientist at Fundação Oswaldo Cruz
Rinke	Eike Mark	na
SAGBOHAN	JOSEPH	Chef Service/Bibliothèque et Archives/EPAC-UAC
Santini	Francesco	na
Seghezzi	Silvia	Lecturer at Birkbeck university of London
Seibold	Heidi	Digital Research Academy
Shanks	David	Professor, Division of Psychology & Language Sciences, University College London

Smailes	David	na
Solomons	Joel	Teaching and Research Associate
Stewart	Suzanne	Associate Professor of Psychological Science, Division of Psychology, University of Chester
TRIKI	LEILA	na
Tressoldi	Patrizio	Studium Patavinum, Padova University - Italy
Ursić	Luka	Department of Research in Biomedicine and Health, University of Split School of Medicine
VIANELLO	MICHELANGELO	Professor, University of Padova
Vidovic	Steven	na
Vukovic	Miro	University of Split School of Medicine
Waltman	Ludo	Leiden University
Winchester	Catherine	na
Zloteanu	Mircea	na
Zogmaister	Cristina	Università degli Studi di Milano-Bicocca
Šrol	Jakub	Slovak Academy of Sciences

Author: Daniele Fanelli

Reviewers: Panagiotis Kavouras, Kalle Videnoja