Open research at Oxford survey

round 1: PGR (12 Jan - 1 Mar 2021)

Malika Ihle, Reproducible Research Oxford Coordinator

Broadly, open research (also called ‘open scholarship’ or, in some disciplines, ‘open science’) refers to efforts to ensure that different elements of the research process are transparent and accessible. The aim of this survey is to assess the views of Oxford researchers on **open research practices**, related **training needs and provision**, and **recruitment criteria**.

The survey was designed and administered by Reproducible Research Oxford. The first round was aimed at post-graduate research (PGR) students working in any field of research and based in any Divisions of the University (Humanities; Mathematical, Physical, and Life Sciences; Medical Sciences; Social Sciences; Department for Continuing Education).

Participation in the survey was voluntary and all responses were anonymous.

The survey was circulated to all Oxford PGR through both the Divisional and Collegial communication channels (newsletters and email to students from graduate training coordinators) to maximise the diversity of views and representativeness of the survey results, and to minimise the self-selection bias (where only people already invested in a subject chose to take part) as much as possible.

### Open research practices

Definitions of relevant terms vary across fields of research and contexts. Therefore, we provided the following standardised entries for respondents to refer to throughout the survey.

**Open Access publication** An article, book chapter, book, or other scholarly work that is released with unrestricted access (i.e. available to the public to view online, or download, without registration, payment, or approval). This includes all forms of open access, such as ‘green’ (i.e. the accepted version of the work, after peer review, is shared by the author(s) in a repository), and ‘gold’ (i.e. the version processed by the publisher is released openly on the publisher’s system upon publication).

**Repository** A service dedicated to facilitating the preservation and sharing of research content, including digital repositories (for digital collections, data, code, etc.) and physical repositories (for collections, archives, reagents, specimens, etc.). The service may be operated by universities, governments, or private organizations, and it can be domain-agnostic or discipline-specific. Access to such repositories may be open or controlled (e.g. requiring registration, payment, or approval). Personal web-pages or other ad hoc methods of storing and sharing research content are not included.

**Data** Any information underpinning a piece of research. The information can be qualitative (e.g. source, archive, representation of art, artefact) and/or quantitative (e.g. measurements, machine output, simulation results). The data can be raw (i.e. as collected), cleaned (i.e. corrected for errors), or prepared/processed (e.g. transformed into a different format for analysis, or anonymized for sharing).

**Code** Custom software developed by researchers specifically for the purpose of conducting a piece of research (e.g. a computer program to extract, clean, or analyse data, or to generate simulation results), or to build components of a research data infrastructure (e.g. a data repository).

**Materials** Any element of the research process that can be coded digitally or shared physically (e.g. protocols, survey questions, instructions, intervention materials, videos of the study procedure, specimens, reagents, samples, and other items used to collect data and/or conduct the research).

**Preprint** An article, book chapter, book, or other scholarly work that is deposited in a repository ahead of peer review. Equivalent terms used in some disciplines are ‘working papers’ and ‘unpublished manuscripts’.

**Preregistration** The practice by which researchers specify elements of the planned work in a dedicated registry before observing the outcomes of the work. Examples include description of the planned approach for a qualitative study, and the data analysis plan for a quantitative study (i.e. a ‘pre-analysis plan’, which can be submitted either before the start of data collection or for previously collected datasets, before the start of data analysis).

**Registered report** A journal article format in which research question(s) and methodology are peer-reviewed before the work is conducted. A submission may be provisionally accepted for publication following peer-review, and eventually published if the authors follow through with the methodology specified in the accepted version (deviations from the registered plan are allowed, but they must be explicitly justified and noted as such). Acceptance of the article to the journal is therefore independent of the results obtained.

### Sample sizes

A total of **499** PGR students (i.e. **7.1%** of the total number of students registered in January 2021) provided their consent to participating in the survey, as well as their affiliations and their role, which were the only mandatory questions participants needed to answer to be included in the analysis. Among these, **123** (**24.6%**) respondents dropped out of the survey at various points before the last question. Table 1 shows the number of respondents per Divisions (Humanities abbreviated by ‘Hum’; Mathematical, Physical, and Life Sciences by ‘MPLS’; Medical Sciences by ‘MSD’; Social Sciences by ‘SSD’; Department for Continuing Education by ‘ContEd’).  
Given the low sample size (**5**) for ContEd, we will not interpret the results for this department separately. However, because the representativeness of this sample is in the same range as for the other Divisions (ContEd: 4.3% of the total number of PGR students in that Department took the survey, SSD: 4.7%), we will still include it to the upcoming visual representations.

Table 1: Total number of respondents by Division

|  | ContEd | Hum | MPLS | MSD | SSD | Total |
| --- | --- | --- | --- | --- | --- | --- |
| Number of respondents | 5 | 75 | 210 | 134 | 75 | 499 |
| Census 2021 (unofficial) | 116 | 1119 | 2600 | 1577 | 1600 | 7012 |
| Representativeness in % | 4.3 | 6.7 | 8.1 | 8.5 | 4.7 | 7.1 |
| Number of dropouts | 1 | 18 | 56 | 26 | 23 | 123 |
| Dropouts in % | 20 | 24 | 26.7 | 19.4 | 30.7 | 24.6 |

An additional **40** staff members (i.e. academics, research staff or fellows, or research support staff) filled out the survey. Their responses will be merged to the views gathered in the second round of data collection, which will target this population specifically.

### Awareness of open research practices

A total of **440** PGR students responded to the question:  
“**Which of the following research practices are you aware of, and which do you have experience with?** ‘Aware only’ indicates that the practice is applicable to your discipline, but you do not have direct experience with it. ‘Accessing/using only’ refers to resources made available by others. ‘Practising myself’ relates to implementation of the practice in your own research (in addition to, or instead of, accessing/using resources made available by others).”

Overall, **Open Access** publishing was indicated to be the most practiced among PGR students and only **4.1%** of the respondents indicated that they were not aware of it or not sure whether this practice was applicable to their work (see Figure 1). However, there was a large proportion (**59.3%**) of respondents who were aware or accessing open access publications but who did not practice it themselves. Analyses of barriers to this practice (see details below) reveals that the lack of funding for gold open access may be one of the cause, therefore **we recommend that the Bodleian Library’s team keep raising awareness of the possibility of publishing open access following the green route (e.g. on ORA)**.

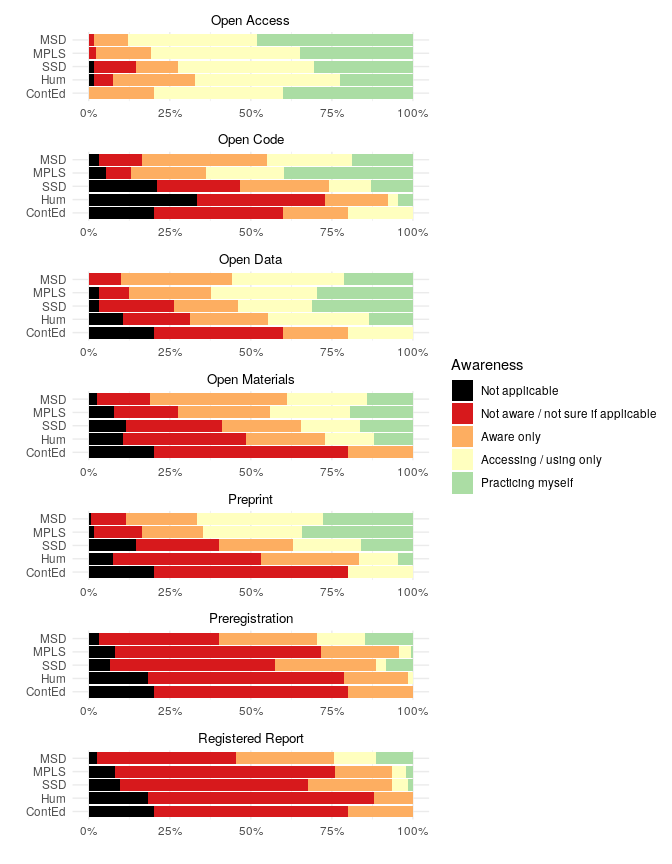


Figure 1: Distribution of responses for awareness of ORPs across Divisions. N = 440 PGR students. (ContEd: 5, Hum: 67, MPLS: 183, MSD: 123, SSD: 62).

**Open Data**, **Open Code**, **Open Materials**, and **Preprint** shared relatively similar level of practice (around **22.5%**), and unawareness (on average **18.5%**).

**Preregistration** and **Registered Reports** were the least practiced among PGR students (on average, by **4.9%** of the respondents), and the majority of PGR students (on average, **56.7%** of the respondents)) indicated not to be aware of these practices. When comparing Divisions, it appears that respondents from the Humanities and the SSD had overall less awareness of the open research practices proposed than respondents from the MSD and MPLS (on average **36.1%** vs **22.6%**, respectively).  
Another point worth noting: the two least known measures overall (namely **Preregistration** and **Registered Reports**) were the most practiced by MSD respondents and, in this Division, more than half of the respondents were at least aware of these measures (**57.2%** vs **25.1%** on average for all other Divisions). This can be explained by the fact that the registration of clinical trials (which is mandatory) is a very similar practice than Preregistrations and Registered Reports (which can most easily be applied to experimental fields).

## Overall effect of open research practices

A total of **439** PGR students responded to the question:  
“**In your opinion, what would be the overall effect of widespread adoption of the following practices in your field of research?**”

The vast majority of respondents (on average **65.7%**, ranging from **34.8%** to **89.7%**) found that the overall effect of a widespread adoption of open research practices would be beneficial, while very few found that it would be overall detrimental (on average **1.4%**, ranging from **0.5%** to **2.5%**). Lower levels of favorable opinions did not correspond to a higher level of unfavorable opinions but to a large proportion of respondent who indicated they were unsure about what the overall effect a widespread adoption of open research practices would be. This pattern of ‘uncertainty’ from respondents matches well the patterns of lower awareness shown for measures such as Preregistration and Registered Reports.

When looking at Divisions separately, the previous observations still prevails: respondents were largely thinking that the overall effect of widespread adoption would be beneficial while lower level of favourable opinions were only corresponding to higher level of uncertainty (whose patterns matched the patterns seen in the awareness question) rather than unfavorable opinions.  
**Preprint** was judged the most detrimental practice with **%** of the PGR students from the Humanities indicating that they think that the overall effect of its widespread adoption would be detrimental.

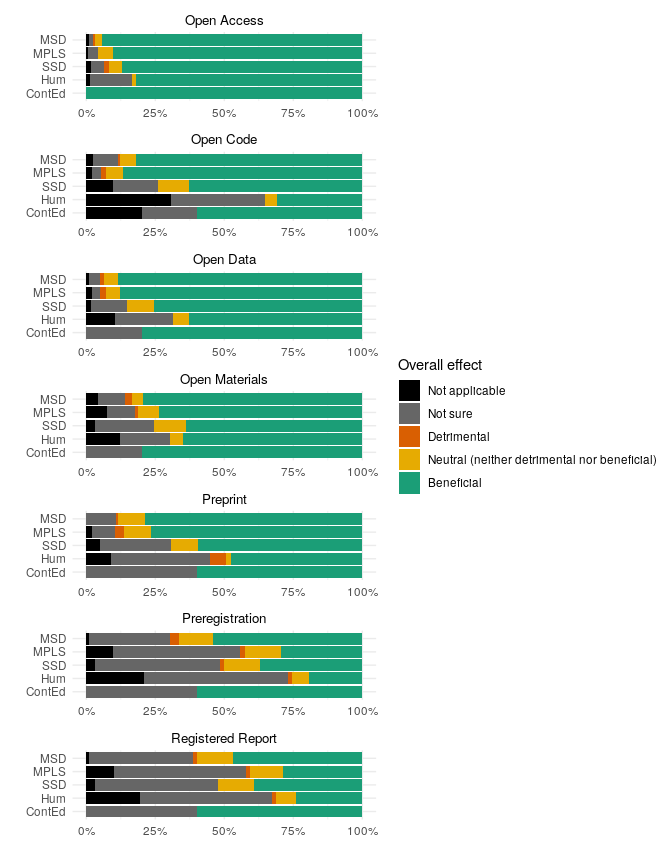


Figure 2: Distribution of responses across Divisions for the expected overall effect of widespread adoption of ORPs. N = 439 PGR students. (ContEd: 5, Hum: 67, MPLS: 183, MSD: 122, SSD: 62).

### Barriers to adoption of open research practices

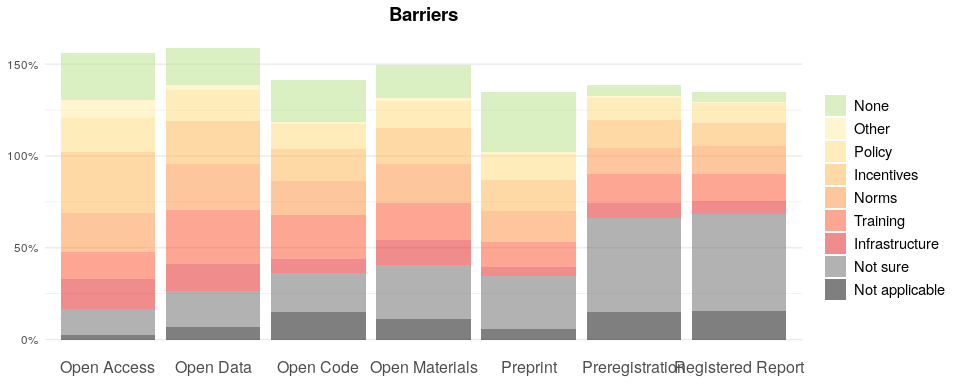
Different factors may contribute to the adoption of open research practices, including:  
- infrastructure, i.e. tools, resources, and/or services that enable implementation of the practice (e.g. repositories for digital or physical storage; registries for preregistration; publishing platforms or outlets) and of relevant community standards (e.g. guidelines relating to citation of data and metadata; ethical and legal guidelines for sharing sensitive data);  
- training, i.e the acquisition of knowledge and/or skills needed to implement the practice;  
- norms, i.e. widespread attitudes and behaviours that support and/or encourage the practice (e.g. the perception that the field is aware of the practice and in favour of implementation; interest from early-career researchers; support from supervisors, mentors);  
- incentives, i.e. resources and/or mechanisms that reward the practice (e.g. dedicated funding, awards; institutional recognition);  
- policy, i.e. a requirement to implement the practice by relevant stakeholders (e.g. institutions, funders, publishers).

Time may be perceived as an additional factor contributing to adoption of a practice. Rather, we consider availability of time, or lack thereof, as a consequence of the factors listed above. For example, a researcher may choose to invest the time required to adopt a practice depending on:  
- the availability of infrastructure or training, which facilitate implementation;  
- the existence of norms in the community, which mitigate the burden on the researcher;  
- the presence of incentives, or policy, which reward and/or mandate the practice in the researcher’s field or institution.

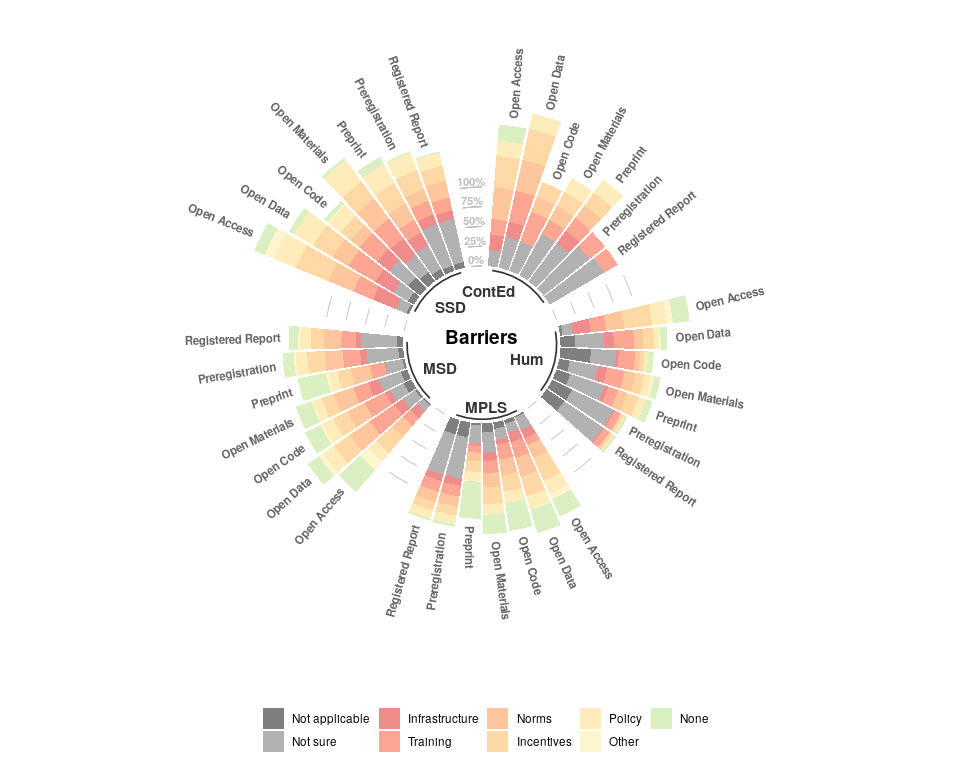
A total of **397** PGR students responded to the question:  
“**Do you face any barriers in adopting the following practices and, if so, what are they?** Tick all that apply and/or provide a brief description for any others.”

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ContEd | Hum | MPLS | MSD | SSD | Total |
| Number of respondents | 5 | 61 | 164 | 111 | 56 | 397 |

Because each respondent could select multiple answers, the sum of the percentages of respondents selecting each of the options extends beyond 100%. Each of the barriers can still be scrutinised individually in a similar way as before. For example, we can see that, overall, more than 25% of the respondents indicated a lack of incentives to practice **Open Access** publishing (**33.1%** to be exact), a lack of training to practice **Open Data** (**29.2%** to be exact), and no barriers to practice **Preprint** (**32.6%**). Overall, we can see that all barriers listed were thought to play some role in the adoption (or lack thereof) of open research practices.



While the lack of incentives to practice **Open Access** and lack of training to practice **Open Data** is shared across all Divisions, the ~25% of respondents thinking that there are no barriers to **Preprint** are mostly distributed among the MPLS and MSD Divisions where these practices are already more common.



#### Other barriers

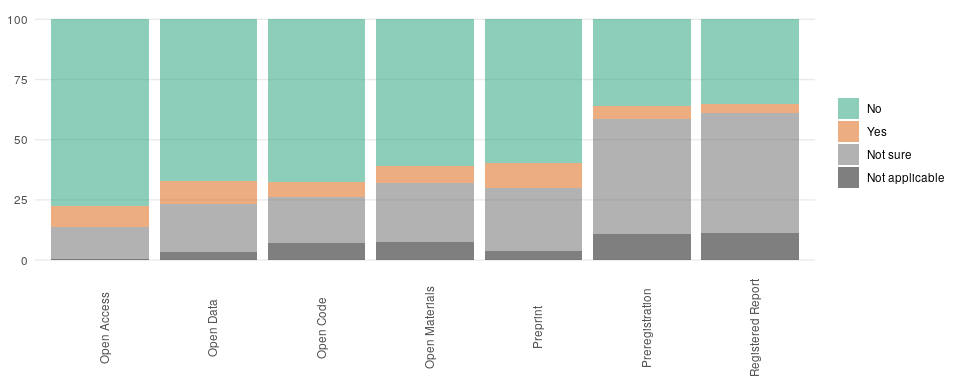
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Open Access | Open Data | Open Code | Open Materials | Preprint | Preregistration | Registered Report | Total |
| Financial cost | 35 | - | - | - | - | - | - | 35 |
| Ethical concerns | - | 4 | - | - | - | - | - | 4 |
| Fear of scooping | - | - | 1 | 1 | 1 | 1 | - | 4 |
| Resource not owned | 1 | 2 | - | - | - | - | - | 3 |
| Difficult resource management and lack of metadata standards | - | 1 | - | 1 | - | - | - | 2 |
| Resource not always digital | - | 1 | - | 1 | - | - | - | 2 |
| Lower quality | 1 | - | - | - | - | - | - | 1 |
| Time investment | - | - | 1 | - | - | - | - | 1 |
| Lack of peer review | - | - | - | - | 1 | - | - | 1 |
| Lack of funding for pilot studies | - | - | - | - | - | 1 | - | 1 |
| Not applicable to all disciplines | - | - | - | - | - | 1 | - | 1 |
| Impedes flexibility in protocols | - | - | - | - | - | - | 1 | 1 |

### Downsides to widespread adoption of open research practices

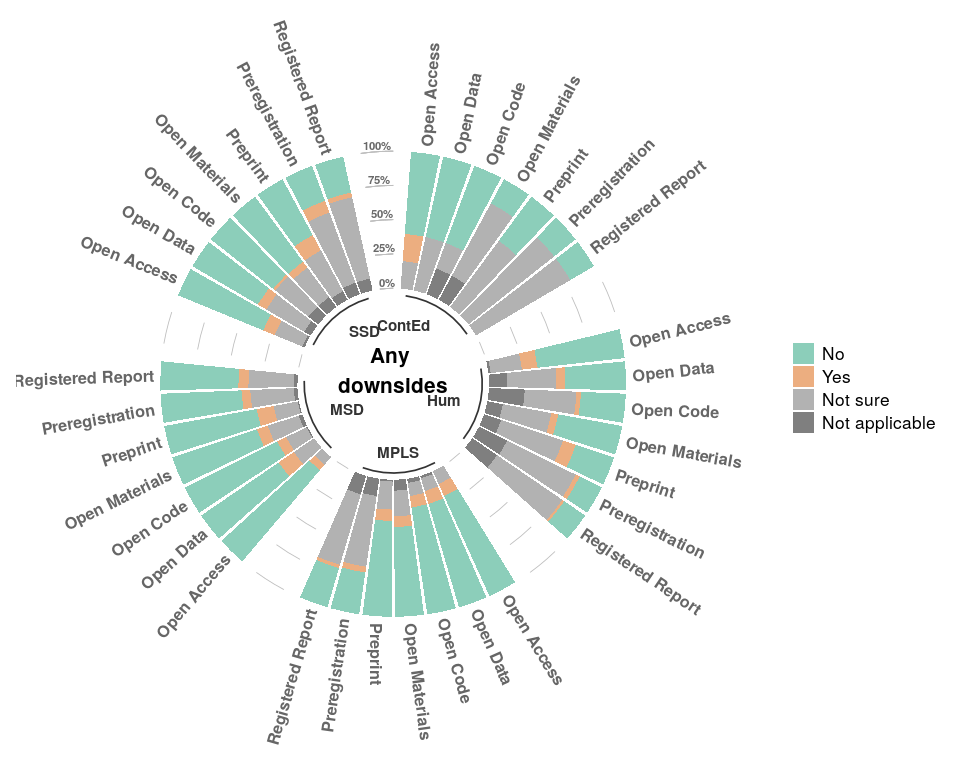
A total of **399** PGR students responded to the question:  
“**In your view, are there any downsides to widespread adoption of the following practices in your field of research?** If so, please provide a brief description.”

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ContEd | Hum | MPLS | MSD | SSD | Total |
| Number of respondents | 5 | 61 | 166 | 111 | 56 | 399 |

Overall, the vast majority of respondents (on average **57.7%**, ranging from **35.1%** to **77.5%**) found that the widespread adoption of open research practices would have no downsides, while very few found that it would (on average **7.3%**, ranging from **3.8%** to **10.3%**). Lower levels of favorable opinions, like previously, did not correspond to a higher level of unfavorable opinions but to a large proportion of respondent who indicated they were unsure about whether a widespread adoption of open research practices would lead to any downsides. T



In MSD, more than half (56.9% at minimum) of the respondents thought there were no downsides to any of the measures presented. The MSD (if we exclude ContEd) also presented the highest percentage of respondents (**13.5%** ) indicating downsides for an open research practice (namely, **Open Data**).



#### what downsides

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Open Access | Open Data | Open Code | Open Materials | Preprint | Preregistration | Registered Report | Total |
| Intellectual property concerns | 5 | 6 | 6 | 5 | - | - | 1 | 23 |
| Misleading due to lack of peer review | - | - | - | - | 23 | - | - | 23 |
| Ethical, safety, or security concerns | 1 | 13 | 2 | 4 | - | - | - | 20 |
| Financial cost | 18 | - | - | - | - | - | - | 18 |
| Time investment | - | 4 | 1 | 2 | 1 | 3 | 4 | 15 |
| Fear of scooping | - | 3 | - | 4 | 1 | 3 | 2 | 13 |
| No control over validity of reuse, misrepresentation, misuse | 1 | 3 | 3 | 2 | - | - | - | 9 |
| Lowers quality | 2 | 1 | 3 | 1 | 1 | - | - | 8 |
| Impedes flexibility in protocols | - | - | - | - | - | 4 | 2 | 6 |
| Impedes exploratory research | - | - | - | - | - | 3 | 2 | 5 |
| Prevents or complicates formal publishing | - | - | - | - | 3 | - | - | 3 |
| Loss of journal income | 2 | - | - | - | - | - | - | 2 |
| Interferes with blind peer reviewing | - | - | - | - | 2 | - | - | 2 |
| Not relevant for all fields | - | - | - | - | - | 2 | - | 2 |
| Fewer (prestigious) journal options | 1 | - | - | - | - | - | - | 1 |
| Challenges around continuity of ownership (e.g. for longitudinal dataset) | - | 1 | - | - | - | - | - | 1 |
| Resource not owned | - | - | - | 1 | - | - | - | 1 |
| Unusable if not from open source platforms | - | - | - | 1 | - | - | - | 1 |
| Needs corresponding increase in respect for null findings | - | - | - | - | - | 1 | - | 1 |
| Challenging timescale for ECR under short term contracts | - | - | - | - | - | - | 1 | 1 |

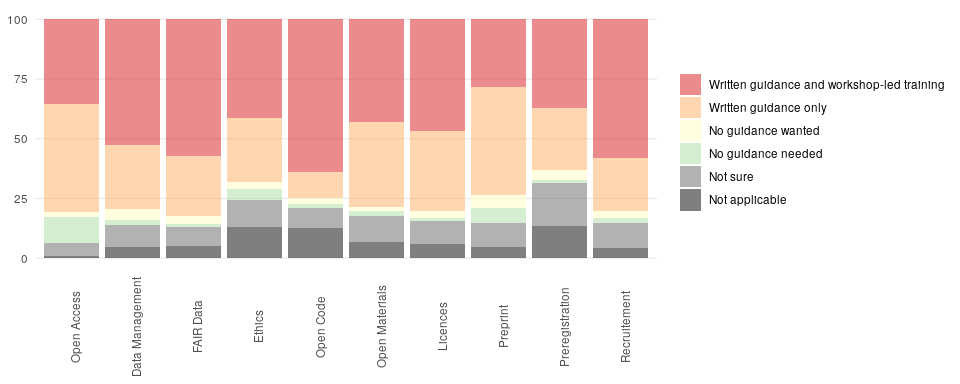
### Training needs

A total of **376** PGR students responded to the question:  
“**For which of the following topics do you think more guidance is necessary?** ‘No guidance needed’ indicates that sufficient guidance on the topic is available. ‘No guidance wanted’ indicates that you do not see the net benefit of engaging with the topic. Please list any additional topics in the empty boxes below.”

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ContEd | Hum | MPLS | MSD | SSD | Total |
| Number of respondents | 4 | 57 | 154 | 109 | 52 | 376 |

The topics listed were:  
How to publish open access articles, theses, or monographs [labelled **Open Access**]  
How to prepare data management plans [labelled **Data Management Plan**]  
How to prepare data and metadata for archiving, and possibly for sharing in a public repository in line with community standards [labelled **Open data**]  
How to prepare ethics applications that allow archiving of anonymised data in a public repository [lablled **Ethics**]  
How to write good quality code (including unit testing, version control, reproducible documentation) and share it publicly (e.g. selection of a repository that assigns a DOI) [labelled **Open Code**]  
How to prepare materials for sharing (e.g. in a digital repository that assigns a DOI, or in a physical repository) How to choose and apply licences for sharing resources (e.g. data, materials, and/or code) and navigate relevant legislation (including copyright law, privacy and GDPR law, commercial law, and institutional regulations) [labelled **Licences**]  
How to prepare preregistrations and/or registered reports (including experimental design, statistics, data simulation) [labelled **Preregistration**]  
Guidance on sharing preprints (e.g. selection of appropriate repository, consideration of publisher’s rights, ethical considerations) [labelled **Preprint**]  
How to assess job applications responsibly, fairly, and transparently [labelled ]

Overall, most respondents indicated they wanted some form of guidance (written guidance and/or worshop-led training) for each of the topics presented (on average **76.16%** of the respondents; ranging from 63.3% for Preregistration and Registered Reports to 82.4% for FAIR data.

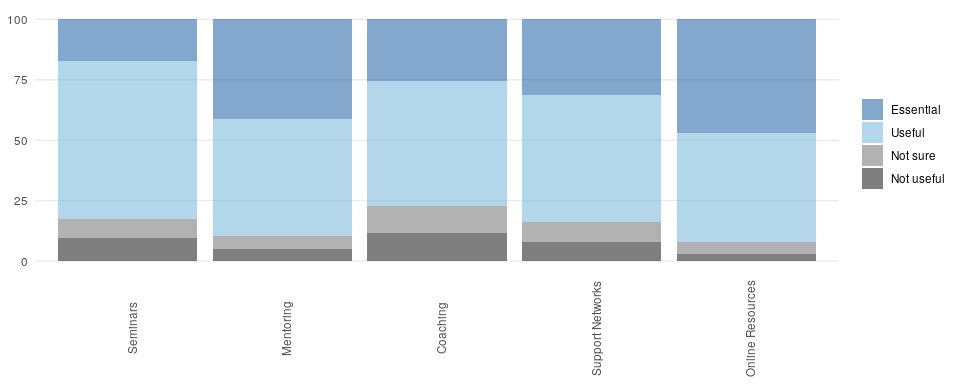


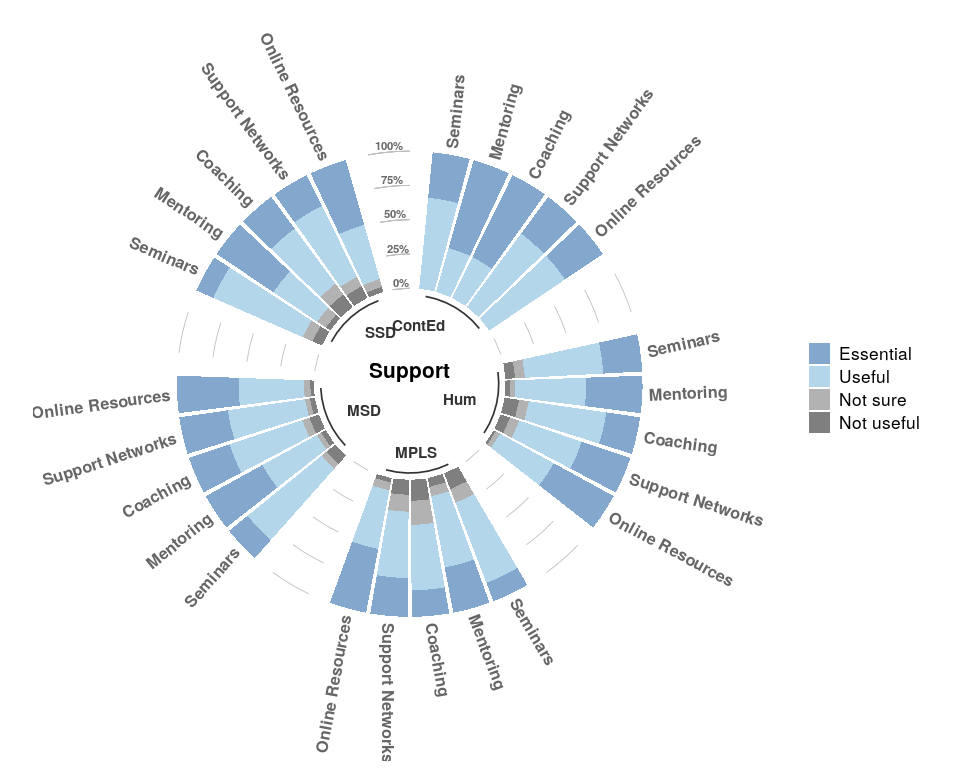


#### extra training requested

|  |  |
| --- | --- |
|  | Written guidance and workshop-led training |
| How to foster diversity and social justice | 3 |
| how to supervise/mentor/teach/collaborate | 3 |
| How to ensure that you have acquired employable skill | 1 |
| How to negotiate with publishers about book publications | 1 |

### Support



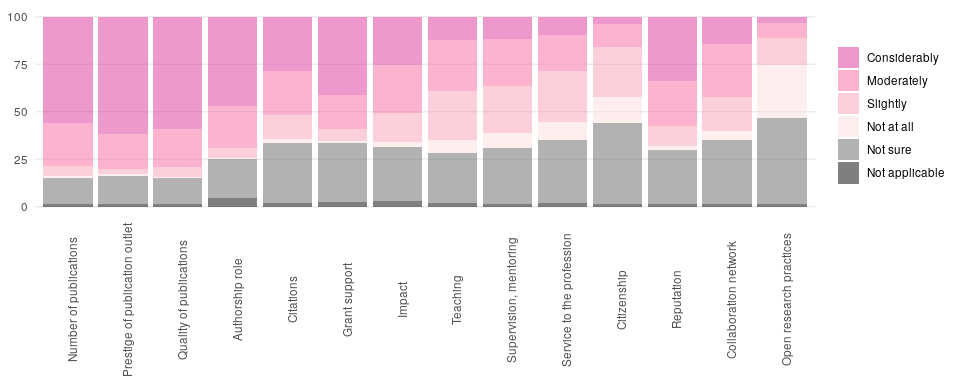


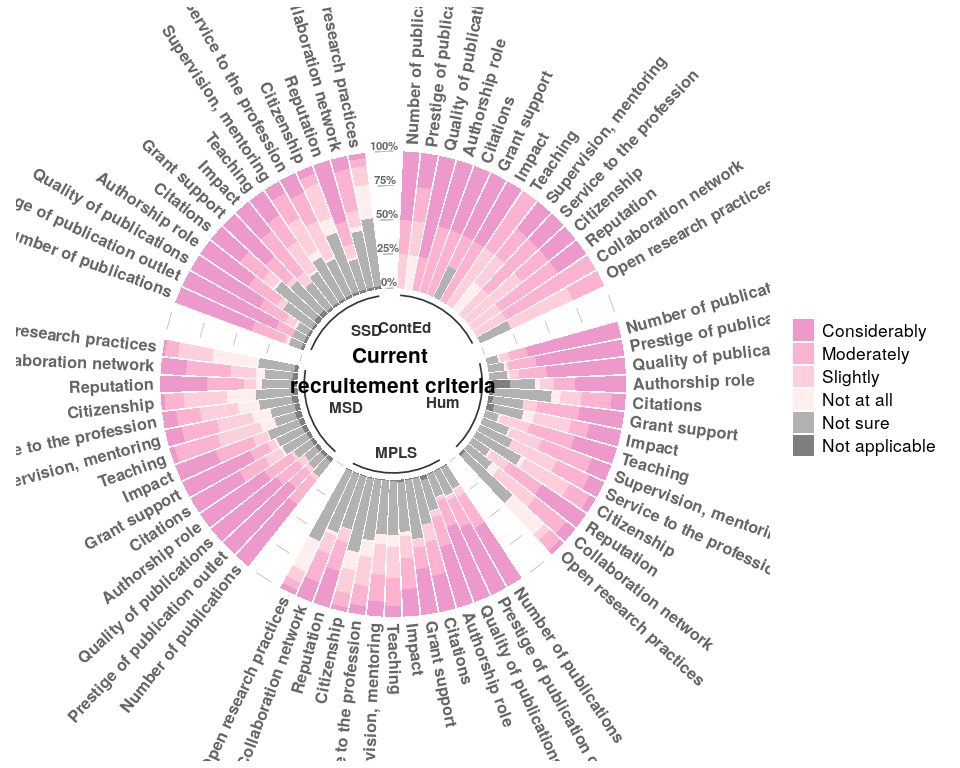
#### extra support requested

|  |  |  |  |
| --- | --- | --- | --- |
|  | Essential | Useful | NA\_ |
| University endorsement, recruitement criteria, and policies | 4 | 0 | 0 |
| Funding | 2 | 0 | 0 |
| Information on how to find funding for publishing in hybrid journal | 0 | 1 | 0 |
| Training on how to publish transparent research | 0 | 1 | 1 |

### Perceived current recruitment criteria

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ContEd | Hum | MPLS | MSD | SSD | Total |
| Number of respondents | 4 | 59 | 155 | 108 | 52 | 378 |



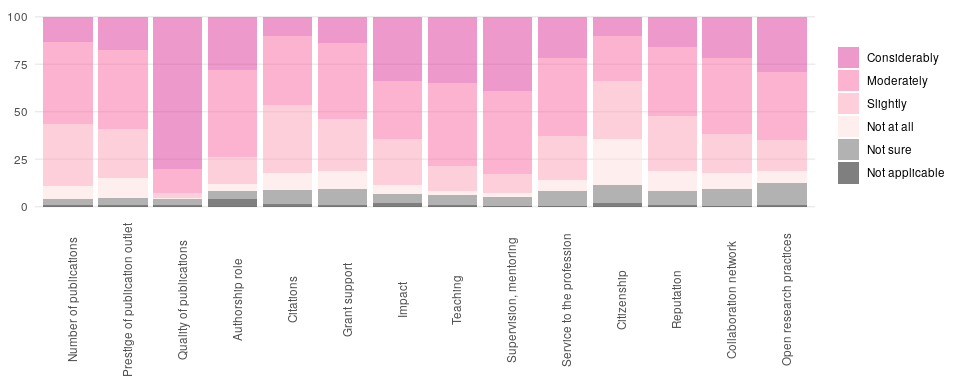


#### other perceived current recruitment criteria

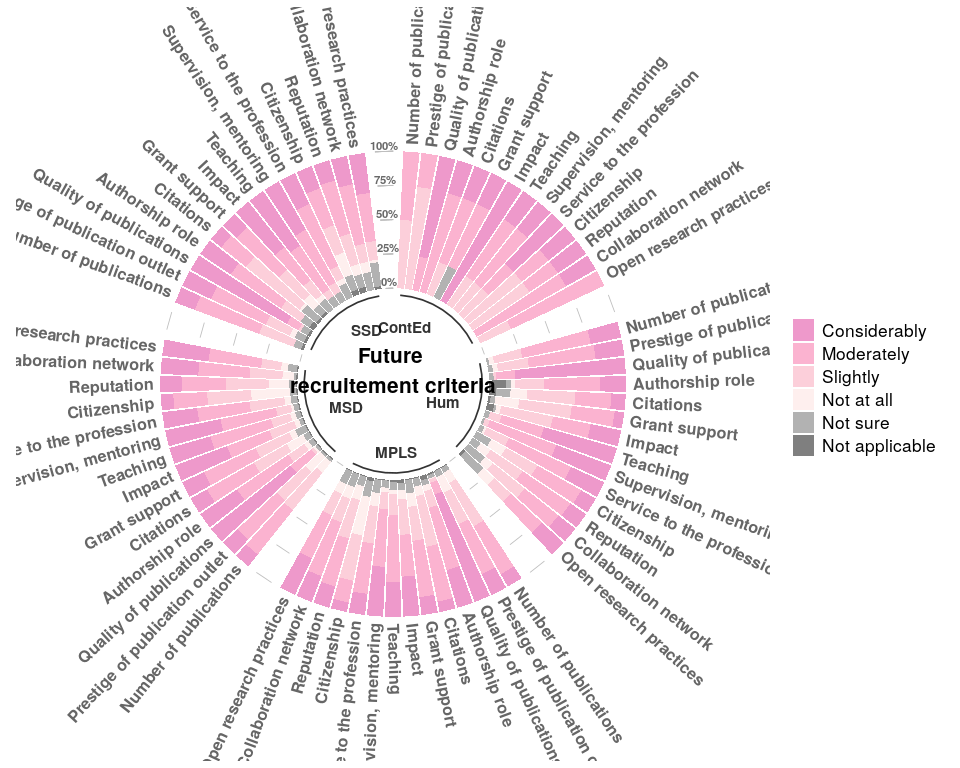
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Considerably | Moderately | Slightly | Not sure |
| Personal connections (e.g. academic references, people from the hiring department | 3 | 2 | 0 | 0 |
| Academic Background (not sure what is meant here) | 1 | 0 | 0 | 0 |
| Having expertise underserved in the department | 1 | 0 | 0 | 0 |
| Leading major fieldwork projects | 1 | 0 | 0 | 0 |
| Public engagement and social media presence | 1 | 0 | 1 | 0 |
| Diversity | 0 | 0 | 0 | 1 |

### Desired future recruitment criteria

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ContEd | Hum | MPLS | MSD | SSD | Total |
| Number of respondents | 4 | 57 | 155 | 109 | 52 | 377 |



teaching most desired criteria in Humanities not so elsewhere.



#### other desired future recruitment criteria

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Considerably | Slightly | Not at all | NA\_ |
| Diversity | 3 | 0 | 0 | 0 |
| Originality of research | 2 | 0 | 0 | 0 |
| Communication skills (explaining their research) | 1 | 0 | 0 | 0 |
| Leading major fieldwork projects | 1 | 0 | 0 | 0 |
| Medical impact | 1 | 0 | 0 | 0 |
| Availability of teaching in previous career | 0 | 0 | 0 | 1 |
| Personal connections (e.g. people from the hiring department) | 0 | 0 | 1 | 0 |
| Public engagement | 0 | 1 | 0 | 0 |