Improving Member Data

How to maximise the potential of your membership data

OPENING REMARKS

Sarah Cumbers, chief executive of the RSS

The RSS was very pleased to be supported by the Academy of Social Sciences (AcSS) and the Economic and Social Research Council to conduct this research on membership data.

In the course of this work, we have had conversations with many membership organisations that want to know more about their members – so that they can tailor events, or check that they are broadly representative of their communities – but who are unsure about how to do this. People's concerns cover a range of topics:

- Whether they should try and collect data routinely and store it in their database, or whether they should use surveys.
- What data they're allowed to collect.
- What types of questions they should ask to sensitively collect personal information.
- How they can encourage members to share their information.

This is something that we have been working through at the RSS as well. So this work was intended to make use of the expertise of our members to set out best practice across these areas – both so that we could reflect on it ourselves and improve our approach, and also to help other membership organisations.

This report is primarily written for membership organisations to help them think about how they collect, store and use their members' data.

We set out three core guiding principles: pragmatism, clarity and trustworthiness. These are general principles that can be broadly applied across different types of work with statistics and data, but they are especially important for membership organisations who often have limited resources and a requirement to maintain positive relationships with members.

Pragmatism means being realistic about what you will be able to do with your data. Probably the single most important recommendation throughout this report is to think about what you – realistically – will be able to do with the data that you collect and to tailor the questions that you ask to that end. Being realistic is important, as it means that you are more likely to be able to make use of the data. You don't need to know everything about your members – but it's important to be clear about your needs and collect information that you will use.

Running a data collection process effectively needs clarity. Clarity about your objectives and a clear understanding of your organisational abilities and the resources available to you. The right approach for you will depend on each of those things, since they will inform the questions you ask and the systems you adopt to answer them.

Trustworthiness is a core statistical principle – and this is especially true for membership organisations who need to maintain the support and engagement of their members. Transparency is the most effective way to build trust: this means communicating clearly with members about what you will use their data for and why you are asking for particular bits of information. Where you can involve members in the process of deciding what your organisation should use this data for, that is always helpful too.

This report sets out in detail how you can implement these principles in your data collection work. We hope that you find it valuable.

Acknowledgements

The RSS would like to express its gratitude and acknowledge the Academy of Social Sciences (AcSS) and the Economic and Social Research Council (ESRC) for funding this project and the RSS members who provided feedback and guidance on its contents.

CONTENTS

EXECUTIVE SUMMARY

This report offers practical guidance for how membership organisations can improve their member data. Although every organisation has unique circumstances, we outline good practices with respect to several key areas of common concern. These are based on common principles that all membership organisations should adopt:

- Be pragmatic
- Be clear
- Prioritise trust

Each of these principles helps guide thinking about which tools organisations should use to collect member data—a core topic addressed in Section 2 of this report. The "gold standard" approach is to use CRM systems integrated with a member-facing online portal to collect robust, foundational information about your members' characteristics. This will serve as the basis for conducting reliable surveys that paint a broad picture of your members' experiences and perspectives. In combination, a well-designed CRM and surveys will also facilitate reaching specific groups of members for qualitative research, which will give you a more nuanced understanding of them.

Thus, it is best to take an integrated approach to the tools that are available to you. Each one serves a complementary but distinct purpose, and although they can sometimes achieve the same thing, it is best to apply them where they are most suited for the task at hand.

Alongside the benefits of any particular approach to data collection, organisations need to consider which option is the most appropriate for them given their available resources. For instance, although well-designed CRM systems can save time and money in the long run compared to other options, they also require upfront investment and can be costly to modify once in place.

We offer three case studies for applying these tools, drawing from the Academy of Social Sciences, the Royal Geographical Society, and our own experience at the Royal Statistical Society with collecting member data. These cases illustrate practical lessons for how to collect better member data, including:

- The importance of collecting data at routine contact points, especially joining and renewal.
- The value of framing data collection as a chance for members to contribute to something more than just your own internal data collection, such as improving your organisation or your discipline.
- The potential for high response rates to diversity questions when they are asked at the right points and in the right way.

Although none of the illustrated cases are examples of perfect data collection, each touches on aspects of good practice that all membership organisations should consider adopting.

Recognising that some membership organisations struggle to justify collecting data, we highlight several hypothetical uses for improved member data, focusing especially on diversity-related data. However, apart from any particular use case, we emphasise that better data is a tool for building better relationships with your members and for gaining business-relevant insights about them. Some justifications for collecting diversity data could include:

- Establishing a baseline understanding of your membership.
- Enabling benchmarking of your data to external datasets.
- Understanding your membership pipeline.
- Evaluating Equity, Diversity, and Inclusion (EDI) initiatives.
- Identifying and engaging specific groups.

We also provide a set of diversity questions that organisations should consider using. For most questions, harmonised standards produced by the government serve as a solid baseline, but we have sought to address matters of specific relevance to membership organisations and have tailored our guidance accordingly. The precise questions that organisations ask will ultimately depend on their own circumstances and purposes—including the specific EDI initiatives they want to run—but good practice should ensure that members are respected in their identities and that you are gathering information that will be of practical use.

We also offer some general guidance on how data protection regulations relate to efforts to improve member data. Although we are not in a position to give legal advice, we have emphasised the importance of clearly justifying your collection of data, organising your data systems and record keeping, and the need to keep data secure.

Lastly, we offer some commentary on a few statistical tools or methods that have some potential for improving member data. However, none of these tools are entirely satisfactory substitutes for getting the basics right. In general, we recommend seeking expert statistical counsel, especially if use of these tools is being considered. For some membership organisations, this expertise could be found within their memberships.

Readers should remember that improving member data is a continuous process of learning. Keeping your members engaged in the process will help guide you on what to do, but you should also feel confident to make decisions that align with your strategic priorities and learn from them, recognising the need to avoid decision paralysis and take a pragmatic approach. This report will hopefully serve as a useful tool to that end.

1. INTRODUCTION

BACKGROUND

Membership organisations, professional bodies, and learned societies have a duty to serve their members while advancing their strategic objectives. A first step in this is understanding what an organisation's membership looks like – but many organisations do not have the level of information that they would like about their members' characteristics. This report is intended to help membership organisations design a framework for the collection of data about their members that works for them.

This report is based on the findings of several streams of work. Firstly, we conducted desk research to establish an understanding of common and best practices used in each topic covered by this report. The guidance provided in each section adapts the practices uncovered by this desk research to the contexts of membership organisations.

We also commissioned a telephone survey of our members that sought to answer questions that were of particular interest to this report but not covered in sufficient detail by prior work. Specifically, we were interested in knowing whether our members were in principle willing to share certain kinds of information with us, whether they supported our Equity, Diversity, and Inclusion (EDI) initiatives, how highly we should prioritise EDI in our strategic planning, and whether our communications had been effective with respect to previous EDI efforts.

More detail about the survey and its findings are available in Section 3. along with case studies drawn from semi-structured interviews with other membership organisations. Additionally, this report draws on conversations with experts from the RSS membership and external organisations who have experience working with data in a variety of contexts. The perspectives of these experts were helpful in shaping our guidance throughout.

Finally, we convened representatives of UK-based membership organisations to get a sense of their experiences. We asked them in small groups about the biggest challenges they have had with their member data, what they wish they could improve, and what their current systems and practices look like. It became clear that membership organisations (MOs)[1] frequently struggle to collect high-quality information about their members. Common issues include internal capacity constraints and financial costs, concerns about compliance with data protection regulations, and the feeling of being cross-pressured by different subsections of their memberships towards and against EDI initiatives. These are discussed here in turn.

Internal Capacity and Financial Costs

Collecting high-quality data can be resource intensive. Particularly among smaller organisations, it is not always clear where the staff capacity will come from to conduct a survey or perform analysis on more extensive data once it is collected. In other cases, an organisation might lack analytical expertise altogether.

For resource-constrained organisations, financial investment in improving their member data can feel unrealistic, especially when the concrete benefits may be unclear. In some cases, this issue is exacerbated by a changing environment in which memberships are declining and revenue is scarce.

This report sets out how an MO can improve member data without overburdening staff, in some cases even freeing up time with improved data systems. Many improvements start with using the right tools to collect data. We offer guidance on different data collection tools in Section 2 and case studies of their use in practice in Section 3.

GDPR

A common experience among many organisations is a lack of confidence in their ability to collect personal data on their members given the need to comply with data

---

[1] Throughout this report, we use 'membership organisations' or 'MOs' to succinctly refer collectively to membership organisations, professional bodies, and/or learned societies.

protection regulations. In our roundtables, this was one of the most commonly mentioned barriers, largely because organisations were unsure that they could justify collecting more data. Particularly for organisations that do not have the resources to comprehensively review data protection legislation, the easiest option is to take an over-abundantly risk-averse approach.

To some extent, navigating GDPR is a matter of training staff to understand it and build compliance into their ordinary ways of thinking about data collection, use, and storage. It is also about learning how to develop clear justifications for collecting data and a clear plan for how to use and secure it. Unfortunately, this too can take financial resources and staff time that not all organisations have readily available.

However, collecting necessary data can be done in a legal and ethical way, and confidence will be built over time as staff become accustomed to working within the GDPR framework. We offer guidance about improving member data in the context of GDPR in Section 6. Examples of possible justifications and use cases for collecting member data are given in Section 4.

Pushback from Members

Roughly a third of organisations represented at our roundtables indicated that collecting more extensive personal data about their members had resulted in significant pushback from parts of the membership, particularly when framed as part of EDI initiatives. The pushback sometimes comes from multiple angles at once: some members feel that EDI should not be prioritised at all, while others feel that it is of vital importance and that existing efforts are not sensitive or comprehensive enough.

In such cases, it is crucial that the MO has clearly defined and communicated reasons for collecting the data. If an organisation has followed the correct procedures and has defined appropriate justifications for their collection, and if they have appropriate safeguards in place to secure that data, then they should feel confident that they can collect it.

Although we recognise that MOs will have different needs, and while we recommend that they take individualised approaches to their member data, we have also heard that there is demand for a consistent, principles-driven set of diversity-related questions that can serve as a foundation for future collection initiatives. To help meet that demand, we have compiled guidance on asking specific diversity-related questions. This is presented in Section 5.

WHY IT MATTERS

For an MO, the importance of understanding your membership goes beyond advancing EDI objectives. For example, a lack of data can limit how effectively organisations respond to emerging issues within their memberships.

Consider an organisation that faces high attrition rates among members from marginalised backgrounds. Effective programming, engagement, and outreach efforts cannot be successfully implemented without high-quality data about where to target resources, and this can impact on revenue streams as well as compounding a lack of inclusion. In fact, an MO may not even be aware of risks and opportunities available to them if these disparities are only examined at the overall membership level; a concerningly high – and actionable – rate of attrition among a specific but small group might go unseen in an overall attrition rate that is only marginally concerning.

This point relates to a more general concern that the services offered by MOs have to adapt to meet the changing needs and wants of members. Many MOs are experiencing low rates of member engagement and declining membership numbers. To keep up, MOs need to better understand not only the characteristics of their members, but also members' behaviours when engaging with the organisation and their opinions of the services being offered. This requires improving their member data.

A MODEL FOR INSIGHT

Figure 01 (p. 08) shows a model of a 'typical' data-related insights project. It begins with a planning phase. This is where you will define the project's goals and specify how it will contribute to the overall strategic aims of your organisation. It is also where you will determine appropriate methods for achieving goals, allocate budget and capacity, engage stakeholders (including members and staff), and formalise a project plan.

Figure 01: A Model for Insights Projects

It is hard to understate how important the planning phase is. As well as making sure that you know how the project will run, this is also where you have a chance to terminate ill-conceived projects before they become expensive, and also where you can discover ways to make your project more efficient or better suited to your strategic priorities. You should carefully think through and formalise a plan with respect to:

1. What data you will collect, how, and why;
2. How you will use and analyse the data;
3. Anticipated or possible actions that will result from your analysis;
4. How you will communicate about the project and its results to members, and how you will involve them throughout its duration; and
5. Your plan for learning from this process and adjusting it going forward.

Critically, planning is the part of the process where many membership organisations get stuck, often because they lack confidence in one or more of the later phases. One of the aims of this report is to help MOs overcome planning paralysis and empower them to launch their projects—and do so confidently.

The subsequent phases are essentially about implementation of your plan, from data collection to analysis and eventually into actioning insights and formalising any learning from the project. In theory, the better your plan is, the more smoothly the rest of the process will go. These phases are not always cleanly separated, and you may find in practice that the boundaries are fuzzier than our illustration suggests. Phases may merge or loop into each other, rather than flowing cleanly from one phase to the next.

Figure 02 (p. 9) offers a set of questions that you should think of asking of all data-related initiatives. They touch on aspects of each phase of our model, and although they are not exhaustive, they will serve as a helpful launching pad for you to plan your own projects. They also reflect many of the principles that shape our guidance throughout this report. These include three core principles:

1. Pragmatism
    - Getting Things Wrong: organisations should not let the search for perfection prevent progress in improving member data.
    - Manage Expectations: take a realistic view of what a specific set of data can and cannot do; choose the right tool for the job, and keep expectations grounded.
2. Clarity
    - Clarity of Objectives: set clear and achievable goals and have a plan for how to action key insights.
    - Clarity of Needs: ensure there is a real need for data you collect and that existing data is insufficient before starting new projects.
    - Clarity of Abilities: understand the skills and resources available to your staff team and membership, and plan projects that account for them.
3. Trustworthiness
    - Communication: keep members informed at all stages of your projects, including how you will use, report, and secure their data.
    - Involve Members: wherever possible, bring members onto project teams, making use of their expertise.

Figure 02: Questions to ask before collecting member data

What are the questions we want to answer?

- Are there gaps in our understanding of our membership that create challenges for achieving our strategic objectives?
- Are there actions we want to take but don't because of a lack of data?
- How do we get from where we are to where we want to be? What role can data have in that transition?
- What are the things we know we don't know? Could data help us discover hidden gaps in our knowledge?

What will we do if we get answers to our questions?

- Does knowing the answers make a practical difference to our decisions or actions?
- If we get an answer we expect, what will we do? What about if we get an answer we don't expect?
- Do we have the resources to action any insights gleaned from better data? If not, could we leverage the insights to build capacity or gather resources, possibly by supporting external funding bids?

Do we already have what we need to answer them?

- Is the gap in our knowledge a genuine one? Would an answer be available by looking at other knowledge sources, both internal and external?
- Are we interested in confirming current assumptions that underpin our work? Do any of our current beliefs need additional evidence?

Can we expect to get a meaningful answer with the tools and resources available to us?

- What tools are we going to use? Are they fit for purpose?
- How many responses are needed to get useful answers? Can we achieve that?
- How much would it cost to get an answer? Think about monetary costs as well as staff capacity.
- Do we have the necessary skills in our staff team to collect the data? What about to analyse it? Can we bring in members to help?
- Are we able to outsource any part of the collection and analysis work externally?

After determining your objectives for member data, you should devote time to thinking about the tools that you will use for collecting it. Efforts to improve member data are largely built on these tools. Each have their own advantages and disadvantages, and this section sets those out.

You should view these tools as complementary, rather than discrete. Although they can sometimes be used towards the same ends, you can avoid duplicated effort by applying the right tool for the right job, and you will find that having a robust system for applying one tool will help you glean insights from the others more effectively.

CRM SYSTEMS – OVERVIEW

All MOs will already have some version of a member database, at least in so far as it is operationally necessary to have contact lists for all members of the organisation. Building a database of member profiles, by contrast, involves linking this minimal information with additional data, such as occupational or diversity-related data. This linked database is what defines what we refer to Customer Relationship Management systems (or just CRMs) in this report.

CRM systems hold data about members in a coherent and usually identifiable way on an internal system that can be accessed by staff. This data can include information about diversity characteristics, contact information, and other details collected by the organisation, often including the type of membership they have, their fields of interest, when they joined, and their sector of work. The specific details on the member profiles are customisable to the needs of the MO.

Most MOs will have a member portal on their website that can be integrated with their CRM. This can greatly expedite the process of collecting information from members and incorporating it into a member database.

CRM SYSTEMS – STRENGTHS

There are many advantages to developing member profiles via CRMs. Firstly, data can be collected routinely whenever members first join or renew their memberships. Collecting data at these points and storing it in a database ensures that the questions are posed to your entire membership at some point in their time with the MO. Joining and renewal are not the only touchpoints that you can leverage; virtually any instance of a member signing into your membership portal is an opportunity to get them to update their information.

Well-designed CRM systems can make data collection routine. The result is not just comprehensive data, but also potentially minimal time commitment from staff once the systems have been established and integrated into standard operating procedures.

Successful implementation of a CRM approach can produce membership-wide disclosure rates that far exceed what can reasonably be expected for surveys. The importance of this point is hard to understate, because having a nearly complete database of member profiles enables all of the following:

- Getting a reliable picture of the characteristics of the full membership.
- Tracking of trends in the membership over time and across your member pipeline.
- Analysis of the impact of initiatives to promote EDI.
- Reliable communication channels with members from specific backgrounds, including via targeted surveys or recruitment to focus groups and event or advice panels.

These are some of the core objectives of improving member data generally as they pertain to EDI, but the benefits extend to run-of-the-mill operational matters too, like being able to target members who have marked an interest in a particular subject on their profile. Put differently, a well-designed and well-utilised CRM elevates your relationships with members across the board.

Apart from the benefits CRMs offer for data collection, many now come with integrated analytical functionality, such as generating simple visualisations of the data. Especially for organisations that lack substantial analytical capacity on their internal staff teams, these tools can be very helpful for enabling basic insights. Even if you do have analytical capacity, you may find that using a CRM saves your project teams considerable time when reporting on your membership, as the more routine analysis can be easily streamlined.

In addition to these benefits, holding robust data about most of your members on a CRM improves your ability to conduct high-quality surveys, because you can directly check some assumptions about the representativeness of your respondents relative to your population (that is, your members). For now, the key point is that organisations should think about CRMs not as an isolated data collection tool, but as an interconnected enabler of more robust data practices in general.

CRM SYSTEMS – WEAKNESSES

Despite the important strengths of CRMs, they also have some weaknesses that should not be ignored.

CRMs can be hard to modify once established. Even relatively simple changes like adding a new field or modifying response options to a question can be expensive, depending on your CRM provider. Much of the challenge comes from the integration of CRMs with customer-facing member portals, since having multiple overlapping systems in that way creates technical complexity that comes with notable costs. Relative to surveys, CRMs are the less agile option, and it is worth the investment to get your CRM right. That investment will likely include:

- Staff time to plan the CRM's structure and contents.
- Financial resources to establish and maintain it.
- Staff time to learn how to use it.

Some of these costs are likely to come up front for organisations that do not have an existing CRM that can be modified. Although CRMs have the potential to save time and money in the long run compared to surveys, the initial costs should not be forgotten, especially because it is crucial to get things right so as to avoid the costs of later modifications.

Although data collection via CRM can eventually be made routine and nearly seamless, this does take time. If data has not historically been collected on a CRM, there will be a gap for those who joined the organisation before diversity forms were made a core part of the joining process. Although renewing members can be posed these questions again, an MO may find that those members are less inclined to fill out the forms, either because their interest has waned over time or because they make use of an automatic renewal process.

In such cases, it is worth trying to plug the gap by directly reaching out to members when they renew, or otherwise trying to reach these members through other avenues. Section 3 includes a case study of the RSS's CRM system that highlights the importance of reaching members at routine contact points.

Still, even assuming that disclosure rates at joining are very high (for instance, in excess of 90%), if disclosure rates at annual renewal are much lower (say, 20%), then it may take several years for the overall disclosure rate among the membership to reach very high levels if renewals are the only contact point at which you try to elicit disclosure.

Moreover, there is bound to be a portion of the membership that simply isn't willing to disclose their information, regardless of whether forms are distributed consistently at joining and renewal, or at any other contact points. If the characteristics of these non-responders differ systematically from those who do respond, then there will still be some amount of non-response bias in the figures produced by a CRM, even if the data is otherwise quite comprehensive. Non-response bias is not unique to this situation though—a more detailed discussion of non-response bias (as well as more technical aspects of surveys) is given in the Annex.

There is also a chance that members will be less likely to disclose information (especially if it is sensitive) if it will be associated with their name on a database. However, there are steps you can take to mitigate this issue, such as adopting strong security measures and communicating about them to members. Indeed, in our case study of our own CRM in Section 3, we present survey results that cautiously suggest that reluctance to disclose sensitive information on member profiles is less widespread than might be assumed.

CRM SYSTEMS – FURTHER GUIDANCE

Being pragmatic and clear about your objectives matters. It is unrealistic to expect 100% disclosure or response rates from any system: the goal should be to get to a sufficiently comprehensive database that can serve clearly defined purposes determined by your organisation.

When deciding what to include on your CRM, you should look for questions that are very commonly used for analysis and for which the number of members falling into various response groups are likely to be relatively large. You should also consider whether there are questions that can enable critical business-relevant insights or the ability to benchmark against relevant data sources. In general, your questions should also be informed by the types of EDI initiatives you want to run, as each initiative will require specific kinds of data.

Although there are data that are most likely to be useful to MOs, consider what is most appropriate in your case. Your thinking should be guided by the questions outlined in Figure 02 (p. 9), and the characteristics you target should be informed by the initiatives for which you want to use the data. More specific guidance about individual questions is covered in Section 5. However, we expect that a solid foundational set of information about members that will enable meaningful action will generally include:

- Sex and gender identity
- Ethnicity
- Age
- Country of residence
- Nationality.

More extensive information might also be helpful or necessary for your objectives. Accommodating the individual needs of members with disabilities at events or in the formatting of communications can be helped by asking particular questions about disability, for example.

There is a risk to storing personal data on a CRM, so you must establish a lawful basis for processing it. It will often be best to seek explicit and affirmative consent before processing members' personal data.[2] In addition, you can mitigate the risks by:

- Ensuring systems are appropriately secured with the support of professional IT specialists to safeguard the data from cybersecurity threats.
- Limiting access to the most sensitive data to staff whose role in the organisation requires them to see it. You should encrypt sensitive data both while it is at rest in your internal systems and when you are transferring it.

More detailed information about protecting personal data is provided in Section 6.

Whatever data you choose to collect on your CRM, collection should be routine. You could softly mandate members to return data collection forms when they join and/or renew, at a minimum—including 'prefer not to say' options for all categories of personal data. This ensures that all members are at least clearly posed the questions you want them to answer and have to deliberately choose whether they will provide the information or not.

Separately, think creatively about ways you can encourage members to engage with your digital platforms and use their member accounts. Although softly mandating the return of forms is one way to boost participation, it is somewhat crude and should be accompanied by positive incentives for engaging. One way to attempt this is to emphasise the value of the portal for networking opportunities. For instance, members that flag areas of academic or professional interest on their profiles can be better linked to other members who share those interests. You may be able to develop initiatives or groups for members with specific interests if they share that information, and communicating about this possibility may drive up engagement via your member portal.

---

[2] Other lawful bases exist besides consent, including legitimate interests. For what is called "special category data," you will also need to satisfy a condition listed under Article 9 of GDPR. For instance, you might use explicit consent or a not-for-profit exemption under certain circumstances. Further information is available in the Section 6, but it is important to note that the applicability of each lawful basis and Article 9 condition will vary depending on your individual circumstances.

You should offer as many positive incentives as possible to get members to engage organically, which will increase the number of members who choose to fill out or update their profiles voluntarily.

SURVEYS - OVERVIEW

Fundamentally, surveys seek to understand something about a population of interest (or simply 'population') by studying a subset of it. In the case of MOs, this will usually be their entire membership or subgroups within it, like all members who reside in the UK. Because of the impracticalities of getting responses from the entire membership, a survey looks to get responses from a subset of members—known as a sample—that can be used to estimate something about the population, like the prevalence of a certain characteristic, or the proportion of members that hold a certain opinion.

Once you have defined your population, your survey will proceed with the construction of a sample frame, which is the set of individuals from which a sample will be drawn. In survey research, the sample frame is not always the same as the population, because it may not be possible to identify or contact everyone in the population.

Membership organisations have a unique advantage here, namely that they will—in theory, assuming profiles are updated—have contact details for most or all of their members (that is, virtually their entire population). Because of this advantage, your sample frames can generally consist of all your contactable members, assuming that your population is, in fact, all or a subset of your members.

Once your sample frame is determined, you should choose a sampling method that will select members from your sample frame to be asked to participate in the survey. There are many sampling methods that you could use, each of which has its own advantages and disadvantages. For most MOs, the most common approach is likely to be reaching out to all members (or all members from specific groups, depending on your goals). A discussion of alternative sampling methods that you may wish to employ is presented in the Annex.

SURVEYS – STRENGTHS

Surveys can accomplish many different things depending on what questions an MO wants to answer. They can be distributed in a form that suits each MO's unique capabilities, such as by telephone, email, or via QR codes or even paper forms at events (although paper-based surveys are much harder to use and generally not a preferred option).

Importantly, they can also be run on either an ad-hoc or regular basis depending on the purposes of the research. For example, you may want to collect feedback about an event you recently ran and only need a one-off survey to get that information. Alternatively, you may want to conduct a survey on a regular basis, like a twice-annual pulse survey of your membership to understand how members feel about the organisation's direction. Making the survey regular will facilitate comparison of data over time, which is especially important if you intend to measure how the introduction of new EDI initiatives affects how members respond to it.

In general, the versatility of surveys is one of the reasons they are so widely used in research settings. You can tailor your surveys to answer specific questions facing your organisation and to collect reasonably large amounts of data in a relatively short period of time on virtually any topic.

Compared to CRMs, surveys offer a more agile approach to data collection as it is cheaper and easier to modify their content between survey waves. However, because comparability of data over time is an important part of attaining many objectives, you will want to be careful about changing the questions you include too often or too radically between projects.

SURVEYS – WEAKNESSES

Survey design can be technical. Your choice of sampling method needs to be informed by your research questions, and more complex sampling methods are not always straightforward to implement, especially without statistical or analytical expertise on the project team.

Survey responses can also be sensitive to the wording of questions, or even something seemingly mundane like the ordering of response options. The academic literature is full of examples of sources of error in surveys and debate about how to deal with them. In practice, most surveys rely on assumptions about those possible sources of error to underpin their findings. Some assumptions are better than others depending on the context and design of any given survey. A more detailed discussion focusing especially on non-response bias is given in the Annex.

Another challenge is that, to track the prevalence of diversity characteristics, surveys must be run at regular intervals (or at least frequently), which may be costly in terms of staff and member time and/or financial resources, depending on the mode, sampling strategy, and complexity of the survey. Moreover, a well-implemented CRM that has been integrated into joining and renewal processes can capture diversity information from a higher percentage of members than even regular surveys can.

Because of these limitations and complexities, the gold standard is to combine surveys with a robust CRM. You will be able to use the CRM to verify assumptions about the representativeness of your samples, calculate appropriate weights for your surveys if appropriate for ensuring representativeness, and also increase your ability to reach specific. You will also minimise the need to include the same diversity questions on every survey if you can link the responses to member profiles that already have that information.

SURVEYS – FURTHER GUIDANCE

As a starting point, we reiterate the importance of seeking statistical expertise when conducting surveys. Given the various complexities involved in survey design, we recommend that organisations seek analytical support for their projects if they do not have in-house expertise. Remember to leverage the experience of your members here—they can prove to be major assets when conducting surveys in particular.

Regarding sampling methods, your choices may be informed in part by the mode of the survey you are going to run. Assuming you have relatively comprehensive data on your CRM, then your ‘sample’ can be your entire sample frame if you distribute the survey via email, for instance. You would still need to think about the normal issues that can arise in surveys, like whether your estimates are precise enough to be practical or whether there is likely to be serious bias in the results. But since the administrative costs associated with emailing your members is low, getting this kind of reach is likely an appropriate choice. On the other hand, if you are running a telephone survey, then the resource burden per response will be much greater, and you will have to generate a more ‘traditional’ sample using other methods—see the Annex.

Think about ways to encourage responses from groups that may typically be less likely to respond, perhaps by sending out frequent and targeted reminder emails to non-respondents, offering the survey in multiple formats, posting links to the survey on your digital platforms, or asking some highly engaged members to promote the survey to their networks—especially if they are linked to communities that have a low propensity to respond. However, you should try to ensure that you are not then expanding participation beyond your intended sample.

If you have the resources to run follow-up surveys, you should consider targeting these at non-respondents to the primary survey. Although this will require concerted effort and creative thinking to get ordinarily non-responding members to engage with your surveys, this is one way that you can try to determine if there is substantial bias in your results based on patterns of non-response. It will also help you learn about the non-responding population in a way that can inform the design and analysis of future surveys.

You should consider weighting the responses according to characteristics that you believe have some relevance to the results when trying to estimate how your membership overall would respond. For example, you might weight the responses to ensure that they are proportionate to your population by sex if the responses differ widely by sex or if you have employed oversampling techniques, which involve intentionally sampling more members from a specific group to ensure they are adequately represented.

In the absence of a CRM, you could use national or industry comparator datasets to help determine appropriate weights to achieve (assumed) representativeness, but in most use cases it will be much better to use data from a CRM that captures characteristics for your full membership. However, weighting is not always appropriate for every situation, and we advise consulting a statistical expert if your weighting requirements are particularly complex.

Surveys work well as a way of getting a broad-brush idea of the views of your members and subgroups of them. Relatively simple survey designs can attain that level of insight. However, a majority of the membership organisations that participated in our roundtable indicated that low response rates had afflicted surveys they conducted in the past. While getting a high response rate is good, there are nuances to bear in mind, and we suggest that MOs should spend more time and attention understanding the precision of their survey results and possible biases that affect them. That discussion is somewhat more technical and is explored in the Annex.

There are some good practices that may boost response rates:

- Keep surveys short, simple, and focused on a clear topic. Cut extraneous questions and make sure you are using clear language, providing guidance for how to answer any questions that may be confusing.

- Use inclusive language when designing questions, especially if they are diversity-related. Members are more likely to engage if they can see themselves in the response options and feel that the language respects the nuances of their experience.

- Have clear, consistent, and frequent communications about the purposes of the survey and how it will contribute to your organisation’s objectives.

- Internally test your survey before distributing it to members to check that it is clear and easy to complete and captures the kind of information that you desire.

In general, it is good practice to run surveys that are not primarily focussed on collecting diversity data. Trying to collect diversity information alone does not typically inspire responses. Members will usually be more interested in participating when they can see a clear purpose for their involvement and when that purpose is something that they fundamentally care about. You should try to focus on things that are of tangible relevance to the members, and the diversity information can be a secondary purpose that can add nuance to the analysis. In Section 3, we include a case study of a survey conducted by the Royal Geographical Society that illustrates how to do this successfully.

Survey data can be anonymous, pseudonymous, or not anonymous. Anonymous surveys are generally going to be useful for understanding the perspectives of members when the subject is in some way sensitive or controversial, or otherwise when the members may be reluctant to express their opinions genuinely. Even when this is not the case, the response rate is likely to be better when respondents know the information cannot be linked to them directly.

At a baseline, you should always ensure that responses are anonymous in any outputs you produce. If individuals can be identified in your reporting, then this poses a risk to their privacy and will break their trust in your data practices.

Prior to reporting, full anonymisation of the data presents a challenge for analysts in so far as it prevents linkage of data between the responses and other information held about members on your CRM (or prior surveys, if you intend to collect data from the same individuals at multiple points in time). If you did not collect diversity data on the survey itself and have kept the data fully anonymised, then you will not be able to do any sort of demographic analysis on your results.

In practice, you should seriously consider pseudonymising response data rather than making it fully anonymous. Pseudonymisation involves removing any immediately identifying information like names or email addresses and replacing them with unique identifiers that can be linked to other parts of your data systems. The identifier could be your member IDs, which would somewhat reduce the sensitivity of the data if it were to be seen by external parties.

You should consider adding another layer of pseudonymisation beyond member IDs (which can be seen or known by staff throughout your organisation) to ensure that only the relevant analytical team can link the responses to member data. This could involve generating another set of unique IDs, perhaps called 'analytical IDs,' that correspond to each member ID. An extra security measure you could use is to store the data in access-controlled folders that require passwords.

If you do have a robust CRM with even minimal diversity-related information about members, you can use it to match diversity characteristics to responses via member IDs (with or without intermediate analytical IDs), email addresses, phone numbers, etc., thereby reducing the burden on respondents to answer repetitive diversity questions. However, you should ensure that you are clearly communicating to members how their responses will be used. If you intend to link their responses to personal information held on their member profiles, then members have a right to know. Communicating any pseudonymisation you will use for analytical purposes is good practice.

QUALITATIVE RESEARCH

Beyond CRMs and surveys, which intend to gather information about relatively broad sections of your membership, more qualitative methods like interviews or focus groups can generate data that can be used to understand something about your members' experiences in greater detail. One considerable benefit of that detail is its ability to reveal nuances that cannot be captured by other methods. However, the fundamental limitation to qualitative data is that it is expensive to collect and you will never reach a comprehensive view of your membership as a result.

For membership organisations, qualitative research is best placed as a complementary tool alongside the gold standard combination of surveys and CRMs. A survey might tell you that members from a certain marginalised group are relatively dissatisfied with the continuing professional development services that you offer, for example. But there is only so much information that you can gather via free-text explanations of that dissatisfaction. Qualitative research methods can let you zoom in on these perspectives where other tools could not.

The synergistic relationship between qualitative research and other tools runs the other way, too. You might conduct interviews with members of the LGBTQ+ community, asking for their views on how to design a new EDI programme for them. A very real downside to qualitative research is that it is not always clear that your findings will be externally generalisable beyond the group of people who participated, so you might follow up with a survey of LGBTQ+ members to see how they would feel about a proposal developed from the qualitative results.

Like with surveys, having a robust CRM will significantly improve your ability to conduct high-quality qualitative research. Recruiting members from specific groups for participation will be much easier if you know members' characteristics alongside their contact details. Additionally, being able to reach out to many members from a group will help you avoid having to rely on the same small set of highly engaged members to link you to their communities.

DATA COLLECTION TOOLS – SUMMARY

Figure 03 (p. 17) presents a table that summarises some of the key points from the above guidance on data collection tools. Use it as a reference guide when deciding between options for your unique circumstances.

In general, although developing a robust CRM can serve as a solid foundation for improving your member data, you should view each of the options available to you as complements rather than substitutes for one another. They each serve different purposes even though there is overlap. Use them in complementary ways to get the best results.

Figure 03: Summary Comparison of Data Collection Tools

Tool	CRM Databases	Surveys	Qualitative Research
What does it enable?	• Collecting the most comprehensive data about a majority of your members. • Tracking characteristics over time and identifying trends. • Gaining business-relevant insights about members, like attrition rates by group or building networks of interest groups. • Diversity monitoring for most characteristics, especially when they are less sensitive. • Targeted communications to and engagement with members from specific backgrounds.	• Measuring sentiment or opinions of a moderate number of members. • Collecting anonymous feedback about sensitive topics, or tracking the prevalence of the most sensitive characteristics. • Gaining more detailed or specific information than can be gathered via CRMs about elements of your organisation or its services. • Understanding and collating the perspectives of members about the strategic direction of your organisation.	• Understanding the nuances of member experiences; going into detail. • Examining how members actually engage with your website or other touch points in detail. • Developing a better understanding of groups within your membership, especially for small groups that can be hard to capture in surveys or where member perspectives do not align with easy assumptions.
What are its limitations?	• Lack of anonymity necessitates the use of more substantial security measures. • Can take time to build disclosure rates if forms have not been mandated in the past. • Often provide the least detail of the examined tools. • CRMs have up-front costs and can be expensive to modify once in place.	• Can be technical; analytical expertise is essential to enable better insights and avoid pitfalls. • Can be costly and often need to be done regularly for best results. • More detailed than CRMs, but less detailed than qualitative data.	• Can't reach a large number of members via qualitative research. • Although not unique to qualitative research, the most engaged members are likely to be the ones willing to participate in interviews or focus groups.
Key considerations	• A good CRM serves as a foundation for diversity monitoring, but also will enable better surveys and qualitative research. • Collection needs to be routine to get a good response rate. Gently require members to return forms (while letting them opt out of specific items) at common contact points. • Ensure access restrictions are in place so that the most sensitive information is only accessible to relevant staff.	• Seek out statistical expertise to support design, implementation, and analysis. Members can often be involved here. • Think clearly about your purposes and how best to design your survey before implementation, but also be careful to avoid organisational paralysis—good-and-done is better than perfect-but-never-started. • Although response rates are important, think in terms of the precision of your estimates and possible sources of bias.	• Use qualitative methods when you want a lot of detail and depth, rather than a broad understanding of something about your membership. • Consider where qualitative research can expand or fill gaps in your understanding of the member experience. Think about how it can lead to a more holistic view.

This section presents three case studies of different UK-based membership organisations that employ different tools for collecting member data. Each one illustrates elements of good practice or lessons for other organisations. In the case of our own case study of data collected on our CRM system, we have also included results from a telephone survey commissioned as part of this project.

Especially in combination with our general guidance earlier in this report about data collection tools, we hope that readers will glean valuable insights that they can apply in their own cases. Although none of the examples are perfect, a synthesis of their good practices can lead to a well-integrated system for robustly collecting member data.

CASE STUDY: ACADEMY OF SOCIAL SCIENCES

As part of their recruitment process, the Academy of Social Sciences (AcSS) ask nominees for fellowship of the society to complete a diversity monitoring form. It asks for the following pieces of information:

- Title, first name, and surname
- Email address
- Ethnicity
- Sex
- Gender identity
- Disability
- Type of workplace
- Area of the UK (England, Wales, Scotland, Northern Ireland)

To be considered for fellowship at the AcSS, individuals have to be nominated by two current fellows or one of their member learned societies. As part of this process, the nominators distribute the diversity monitoring form to the nominee, and they are instructed to return them via email to AcSS staff. For the vast majority of nominees, this level of contact is sufficient for them to complete and return the form. When the form is not returned, members of staff contact the nominee to request again that the form be completed and returned. There is also an option for nominees to complete and submit the forms online, although most nominees choose to return their responses via email.

Using this approach, the AcSS has managed to achieve completion rates that are consistently above 90% over the last four years, often approaching 100% in individual nomination rounds. These kinds of response rates are very difficult or impossible to get from surveys that are conducted later in a members' relationship with an organisation. With this level of completion, the AcSS get a comprehensive view of the diversity of nominees over time.

Nomination round	Diversity Monitoring Form Completion Rate (%)
Spring 25	91
Autumn 24	92
Spring 24	94
Autumn 23	98
Spring 23	92
Autumn 22	98
Spring 22	100
Autumn 21	98
Spring 21	100

Once the forms have been returned, AcSS staff manually transfer the data to a spreadsheet where it is anonymised. This spreadsheet is used to produce reports on the overall diversity trends among cohorts of nominated fellows, as well as to generate a picture of whether there are any clear patterns in what groups are being nominated by fellows from certain backgrounds or institutions. While this data can also be used post hoc to determine if there are trends in the types of nominees who are ultimately approved, the data is not used as part of the approval process and does not inform approval decisions in any way.

There are several elements of good practice that come through in the AcSS approach to diversity monitoring:

1. Their form is short and very easy to complete.
2. They have made data collection routine by diligently distributing the forms during the nomination process, when nominees are theoretically the most engaged they will ever be.
3. They take advantage of key aspects of the nomination process to boost response rates and enable useful insights. A key part of this is leveraging the existing personal contact that fellows have with the people they nominate; response rates are likely higher because of the personal touch that comes with receiving the forms from a known entity.
4. Collecting data on nominees rather than just new fellows means that a larger portion of the joining pipeline can be monitored, including of people who were considered for membership but ultimately turned down.

The main limitations of the approach outlined here generally relate to the fact that the AcSS do not produce a set of member profiles using the data. The lack of a database of information about individual members means it would be more challenging to do any follow-up EDI work that requires being able to identify or communicate directly with members from certain backgrounds. Additionally, this kind of diversity monitoring does not enable analysis of whether members from certain groups leave the organisation at different rates—a key insight that can be an indicator of the differential experience of members.

Relatedly, because it only tracks diversity in the nomination process, it cannot produce a picture of diversity within the existing membership at any given point in time. Lastly, because of the mode of collection, data has to be manually transferred to a spreadsheet that can be used for analysis, which places an additional burden on staff. Each of these issues could be overcome with the introduction of a CRM that integrates the diversity monitoring form.

Nevertheless, other membership organisations can take important lessons from the success that the AcSS has had in getting very high completion rates on their form. Especially for organisations that rely on a nomination process similar to theirs, the strong integration of existing fellows into the data collection process can be a major boon, as well as the collection of data on nominees rather than just accepted fellows. This is also a good example of data collection being routine, and other organisations would do well to incorporate that aspect of the AcSS approach.

CASE STUDY: ROYAL GEOGRAPHICAL SOCIETY

The Royal Geographical Society (RGS) used an external research consultancy agency called Membership Matters to help them design and implement their Navigating Geography Survey and conduct focus groups that would inform their recruitment strategy. The agency provided a comprehensive product from the design stage all the way to analysis and reporting, which took a lot of the burden off internal staff.

The survey was not targeted at understanding diversity, but did include questions about career stage, age group, country or region, gender, and ethnicity at the end. The bulk of the survey related to questions of strategic and operational importance for the Society, including but not limited to:

- What members valued about their membership.
- Reasons why former members left.
- What members would like the Society to prioritise, and how well they think the Society currently delivers on those priorities.
- Members' awareness of various services on offer.
- How members would feel about possible new services that could be offered.

The survey was distributed widely via email, with multiple follow-up emails to boost engagement from non-responders. It was also promoted on the Society's digital platforms (including their website, social media, and newsletters), and made use of aspirational, mission-based wording.

Rather than simply framing the survey as a 'member survey,' they framed it as an opportunity for individuals to contribute to the strategic future of the RGS, and thereby improve the field of geography. Phrases like 'Have your say', 'Contribute towards our mission', 'Help us have a significant impact', and others with a similar valence were common. Additionally, participation was incentivised by entering willing respondents into a £100 prize draw. The survey was distributed to more than 11,600 members, of which 41% completed it. The responses were kept anonymous and not linked to a CRM or other member database.

Recruitment for the focus groups was done through a question late in the survey that asked whether individuals would be willing to participate, as well as a couple of options indicating when they would be available. Participants were selected in part with the goal of getting a diversity of individuals in terms of age, location, etc.

The information gathered from the focus groups was used to validate the findings from the survey and create a set of four 'personas', which identified different types of individuals and their primary interests. Each persona was framed around core value propositions that aligned with various strategic priorities around which the Society could message and act. The personas had their own goals, pain points, and desires for the future of the Society. Using this information, RGS was able to produce an action plan that would help them improve their recruitment activities going forward.

There are several examples of good practice within the Society's approach to this survey:

1. The survey itself was not focussed specifically on diversity. Instead, it was designed to gain insight into what members value about their membership, why they may have left the organisation, what they would like to see the Society do differently, etc.
2. In the same vein, the survey was not framed as a 'member survey,' but as an opportunity for people (members and non-members) to contribute to a greater mission, namely influencing the direction of the Society and improving the field of geography. Framing the survey in this way likely improved response rates substantially.
3. The involvement of an external agency was important because of the ambitious scope and scale of the project, which included not just a survey but also qualitative research that was to be integrated with the survey. Their involvement reduced the overall capacity burden on staff while also enabling the Society to internalise learning from expert practitioners. It further shows the value of thinking about qualitative research as another tool that can complement other types of member data.
4. The insights gained from the survey and focus groups were highly oriented towards action. It allowed the Society to understand what members' priorities were and how well they were delivering on them, among many other mission-critical insights. It also helped the Society to overcome certain assumptions that they had made about what different categories of member wanted. This helped them to better present and communicate their services and improve member engagement.

It is important to keep in mind the intended objectives of the survey and qualitative research here. They were not intended to be a diversity monitoring exercise, nor were they strictly speaking an EDI-oriented project. Rather, the goal was to gain valuable insight that would inform a recruitment strategy that went beyond crude classifications on the basis of demography.

An EDI-focused project would likely involve more diversity oriented questions or a sampling strategy that prioritised collecting data from a representative sample based on specific characteristics. However, the approach taken was well-suited to the specific objectives of this project.

Thus, this is a good example of a well-executed survey that gathered valuable data from a high percentage of the Society's members. Other organisations should consider the value of having a clear strategy in place for their own surveys and the benefit that mission-based messaging, incentives, and active promotion can have on their work.

CASE STUDY: ROYAL STATISTICAL SOCIETY

The member-facing side of our digital platforms at the Royal Statistical Society is called 'MyRSS', which is a hub for members to manage their membership, subscribe to various member-led groups based on their interests, access an opt-in directory of RSS fellows, and otherwise access or engage with benefits of their membership.

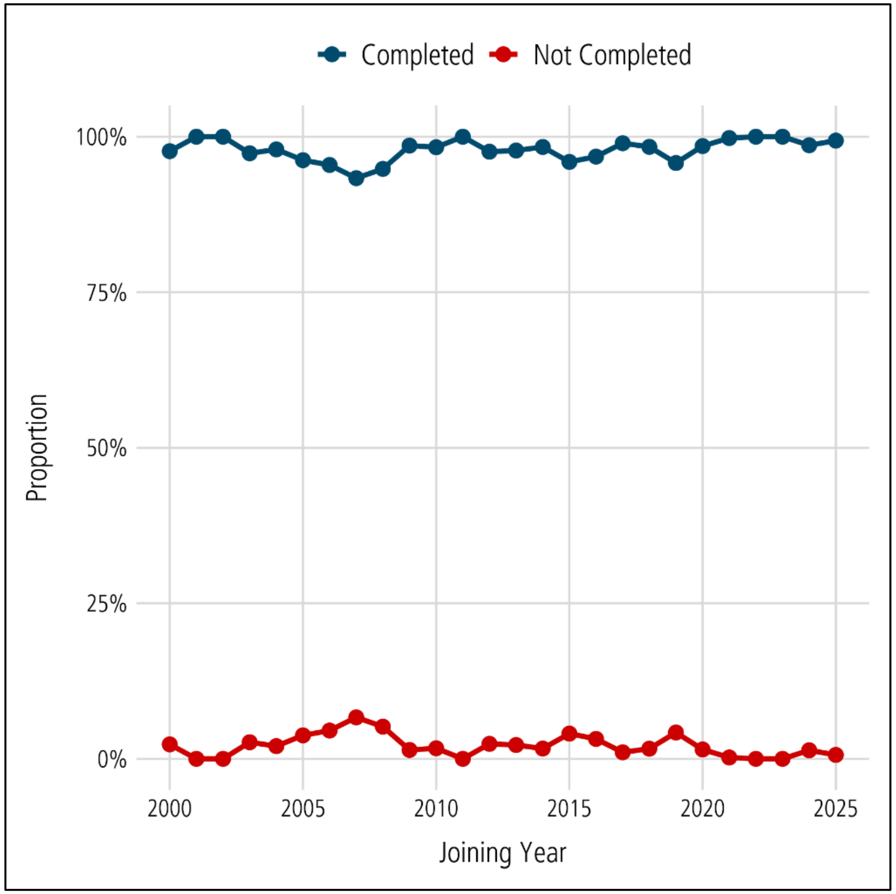
Through joining online, members create a profile that allows them to access MyRSS. Profiles have a section where members can voluntarily complete a set of diversity monitoring questions. Although this section has changed over the years and will be reviewed in the context of this report, the current questions ask about gender, ethnicity, disability, and religion. While members are directed to complete their profile in MyRSS after joining, and we occasionally send out mass communications asking members to complete our diversity monitoring questions, the vast majority of members do not do so.

Rather than relying solely on members to fill out their profiles, we also ask certain questions as part of our joining process using forms that are linked to our CRM. These questions, too, have changed over time; currently, the only diversity-adjacent characteristics we request on these forms are date of birth and country of residence. We can get an instructive picture of the importance of routine data collection for new members by looking at how completion rates for specific questions have changed with amendments to our join forms. That is, we can look at the proportion of members who have available information for a field on our CRM, and who are thus considered to have 'completed' that question:

1) Age, which has always been calculated using date of birth information that is generally required when joining the RSS.
2) Ethnicity, which has never been part of our join forms and is instead only part of diversity monitoring as captured on MyRSS.
3) Gender, which was previously on a join form until April 2020 and is now only included as part of member profiles voluntarily updated on MyRSS.

For age, the mandatory date of birth information results in very high completion rates throughout time, as would be expected. See Figure 04.

Figure 04: RSS Age Question (Date of Birth) Completion Rates



Despite being a required field on joining forms, we still see that the overall completion rate as reflected on our CRM for some years is not 100%. This could be due to irregularities in collection for some members, such as those that completed a paper form whose data may not have been fully copied into the CRM, or special subsets of our membership that are not required to complete a joining form, such as honorary members. Still, a mandatory question collected at a routine point for all members produces a completion rate that is very nearly 100%.

The reverse dynamic is visible when a question is neither mandatory nor included as part of our joining forms, as is visible in Figure 05 for our ethnicity question. Here, we see that completion rates are almost always below 25%. We also see a slight downward trend over time, with completion rates hovering around 15% in the most recent years.

This level of completion makes it challenging to use the data for most purposes. At this level, it is likely that a survey sent to all members directly would do a better job of capturing more responses to our ethnicity question and capture responses from a more representative set of members.

However, comparing completion rates for our ethnicity question with those for our gender question is illustrative of a more important point about routine data collection via CRM-integrated forms. Figure 06 (p. 23) shows how even non-mandatory fields can result in very high completion rates so long as they are asked of members when they join.

Prior to April 2020, our gender question regularly obtained completion rates of 100% (or close to it) in most years. This enabled us to have a practically complete understanding of our membership by gender. However, in 2020, we stopped asking for gender as part of our joining forms as we transitioned to a new CRM system, and there was a two-year gap before we started asking it as an optional diversity question in MyRSS. This was due to the time taken to finalise the questions and options and implement these in the CRM. The result has been that our gender completion rates per year now look much more like those for our ethnicity question than those for date of birth.

Figure 05: RSS Ethnicity Question Completion Rates

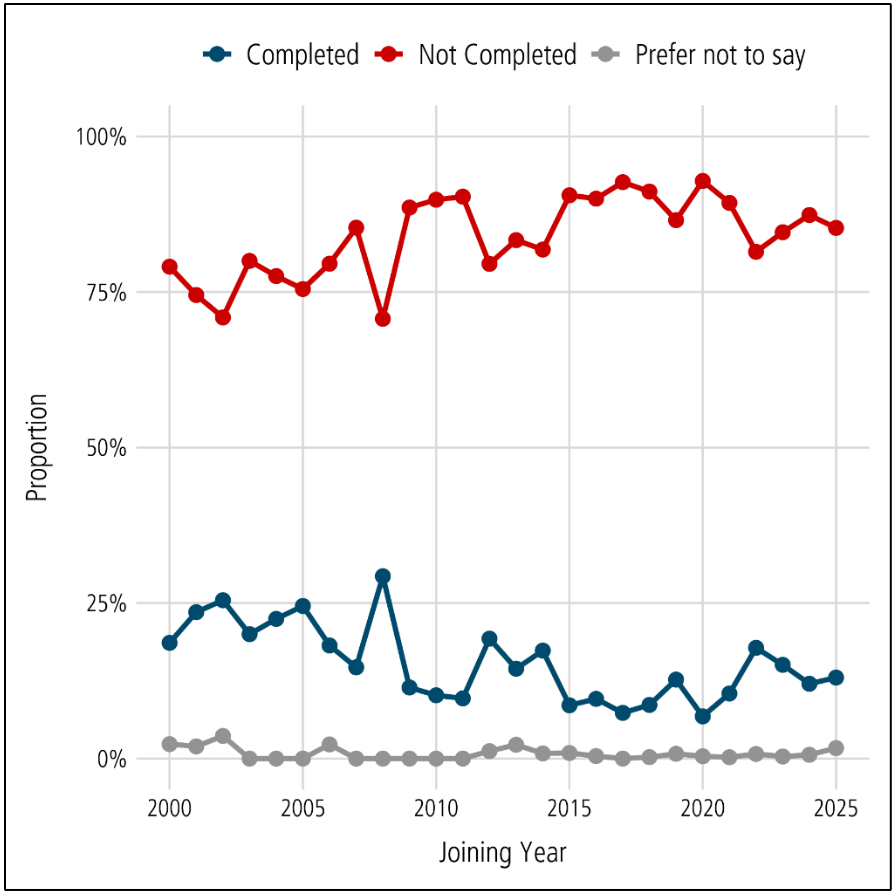
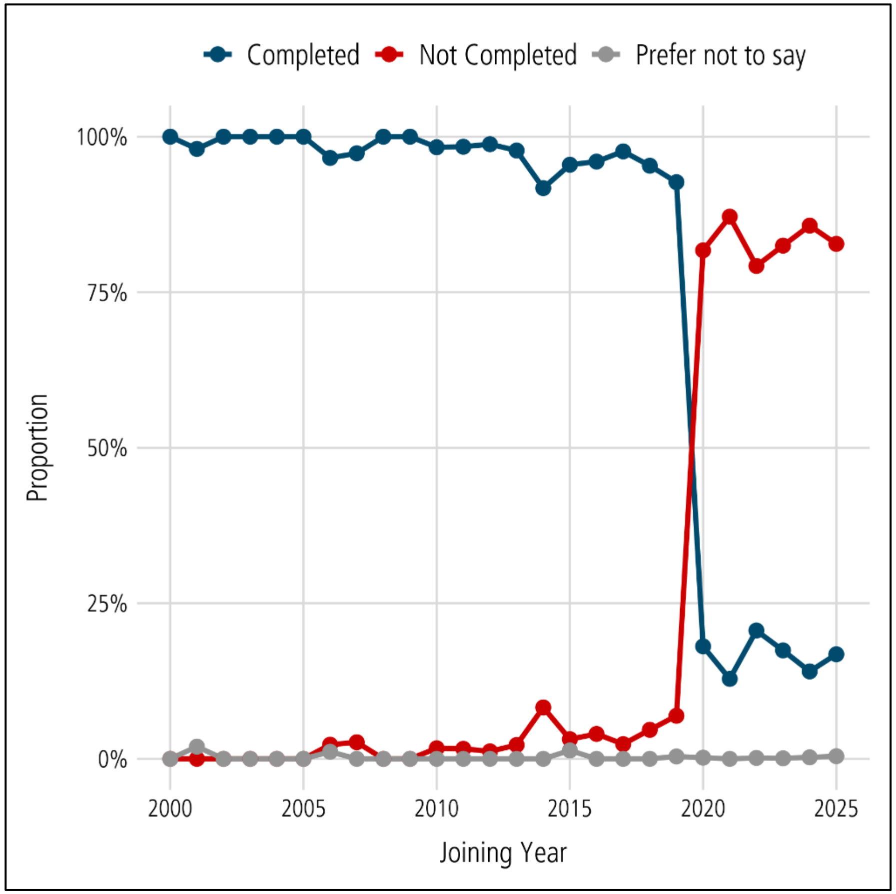


Figure 06: RSS Gender Question Completion Rates



It is possible that other questions would get lower completion rates than gender if they were included as part of our joining forms. This information is held on member profiles and therefore not strictly anonymous—as it will be visible to RSS staff—we might therefore expect that more sensitive questions could generate low completion rates if collected via our CRM.

To help us to get a better understanding of these points and assess the possibilities for collecting a wider range of data about our members, we commissioned Chord UK Ltd to conduct a telephone survey that ran from mid-March 2025 until the end of April 2025. Two-hundred members responded either by telephone or online via links sent through follow-up emails to members who did not respond to attempts to reach them by phone.[3]

The survey was targeted at current RSS members who had not completed the diversity monitoring section of their profiles on MyRSS as of 12 March 2025.[4] We were particularly interested in this group as they make up the majority of members and are likely to be less engaged than those with complete profiles, either because they have less time or interest in filling out their profiles, or because they have not engaged with our prior communications asking them to do so. Attempts to improve our systems would benefit from a better understanding of why members are not responding our previous attempts to get members to fill out diversity monitoring forms.

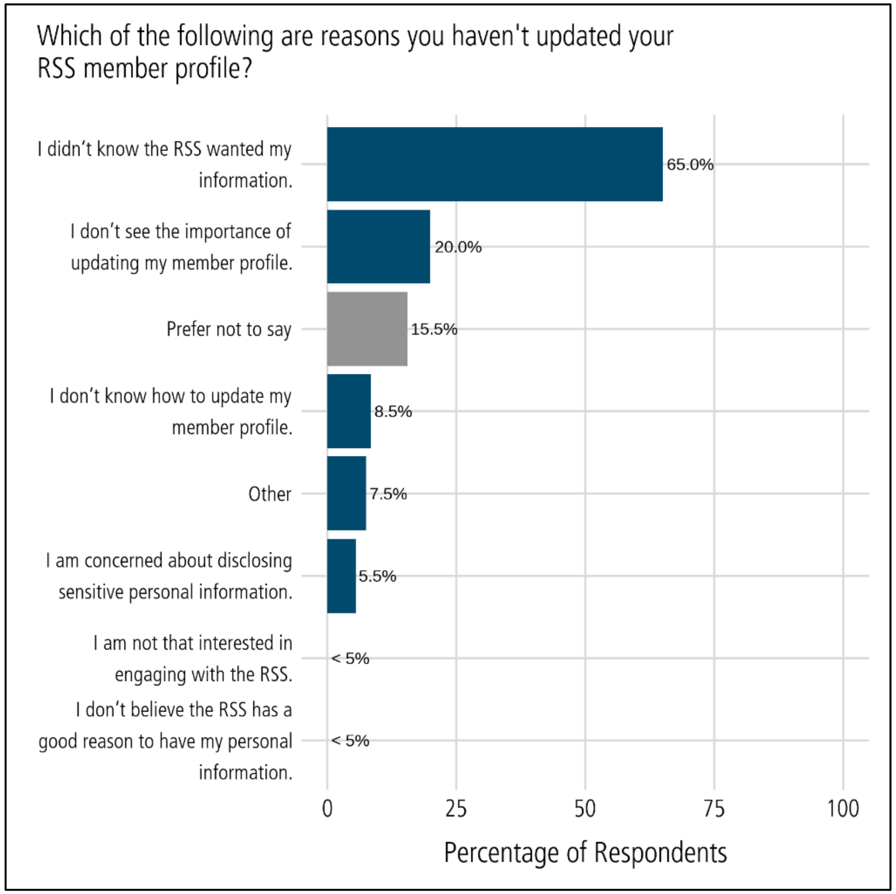
One of our questions attempted to get at that question directly, asking members to select from several reasons why they had not updated their profiles. Multiple options could be selected. The results are presented in Figure 07 (p.24). By far the most frequently cited reason was ‘I didn’t know the RSS wanted my information’ at 65%, followed by ‘I don’t see the importance of updating my member profile’ at 20%.

---

[3] Conservatively, this corresponds to a margin of error of +/- 6.9% for our binary-response survey items. While this is a wide margin, for many of our items it is sufficient for gaining practical insights.

[4] Because the survey was done primarily via telephone, we provided Chord with a contact list that was restricted to RSS members based in the UK. Thus, one caveat to our results is that they correspond to our UK-based member population.

Figure 07: Survey Question on Reasons for Incomplete Member Profiles



These responses cannot give us a complete view, but they are indicative of at least one major weakness of relying on voluntary and irregular member updates to their profiles. Specifically, communications encouraging completion do not necessarily reach their intended audience, because the message itself is not read by the members and/or because the rationale behind our need for updated profiles is not clearly and persuasively communicated.

Of course, some members may never feel it is important for the RSS to collect their information, regardless of our communications. However, the general sentiment expressed in response to other questions was that members' willingness to provide information often depended on the information being collected and what their intended purposes were.

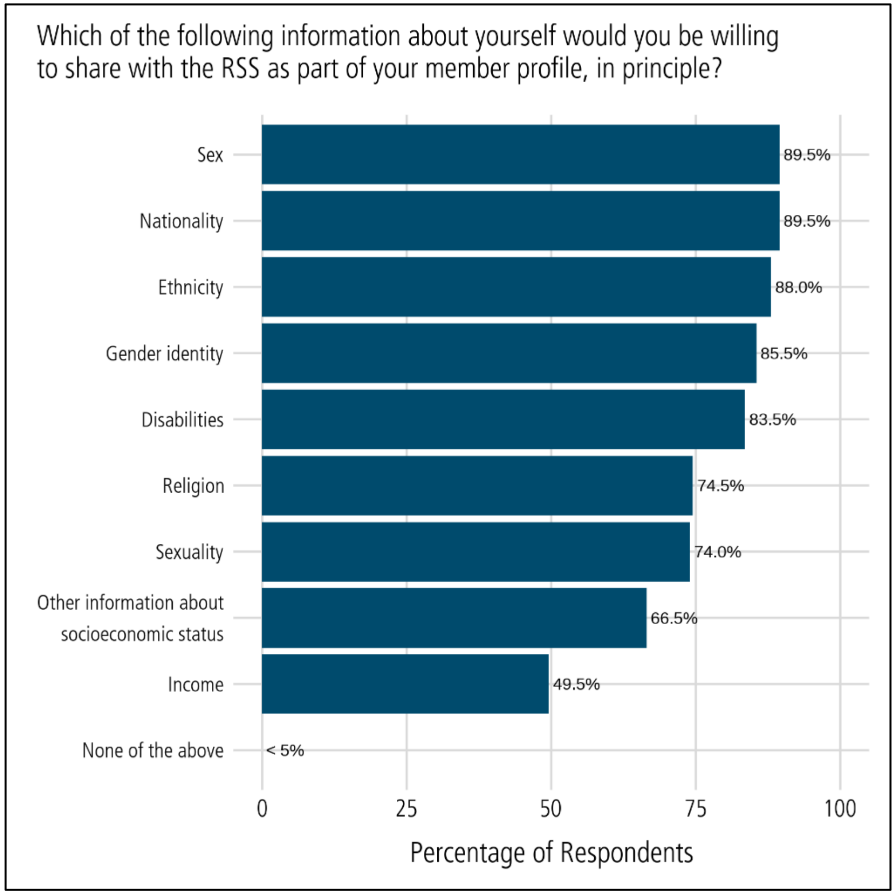
We asked a question that tried to get at the range of information that could in theory be held about members on MyRSS: 'Which of the following information about yourself would you be willing to share with the RSS as part of your member profile, in principle?' The results are presented in Figure 08 (p. 25).

We focused this question on member profiles because we are interested in understanding willingness to share these characteristics on a more sensitive medium that is less anonymous than surveys. It is important to gauge the ceiling for collection and storage of data via a member database given the numerous benefits such a database can yield for other data-related initiatives.

Overall, a large majority of respondents are willing to share most of the characteristics mentioned in principle. There is a noticeably lower willingness to share for religion, sexuality, and socioeconomic information. Just less than half of respondents indicated a willingness to share information about their income.

Respondents were invited to expand on their responses in an open-ended way, and the further detail emphasised some of the core concerns that people generally have around sharing their data – that they are too sensitive, too personal, or irrelevant to the RSS. Many indicated that the RSS would need to clearly communicate why it is collecting the data and how they would use it. A few mentioned the need for the data to be handled sensitively and not shared externally.

Figure 08: Survey question on characteristics members are willing to share



Some of the open-ended elaborations indicated the importance of the nuances of the questions being asked. One respondent emphasised the need to consider the government's harmonised standards when designing questions for these characteristics. Another mentioned the need for the data to reflect the complexity of people's lives, including other characteristics not mentioned in the question, like caring duties.

The results largely support the view that our formerly high completion rates for the gender question on our joining form could be achieved for other characteristics, too. Particularly for sex, gender identity, nationality, ethnicity, and disability, nearly all respondents were in principle willing to share this data. Asking for it when they join would likely get us close to comprehensive coverage. Religion and sexuality are somewhat more controversial, which is perhaps unsurprising given their deeply personal nature. Still, a sensitively posed question with clear communication could generate very high completion rates.

Importantly, we wanted to understand what could help incentivise members to update their information. Because any new, light-touch approaches to data collection for existing members will take time to boost completion rates, we need to be aware of the most effective tools for increasing engagement. We asked 'Which of the following would make you more likely to provide the RSS your personal information?' Results are presented in Figure 09 (p.26).

As expected, the most commonly selected option was 'Being asked for the information when I join or renew my membership.' Naturally, members are more likely to give their information at a usual and expected touch point. Also unsurprising is that clear communication about our uses for the information is a very commonly cited factor for boosting willingness to give information.

Just under half of respondents indicated that they would feel more likely to provide information if offered incentives with a monetary value. A similar percentage also identified more consistent reminders to update their profiles as a factor that could boost their willingness. Notably, only 5% of the respondents explicitly indicated that they would be unwilling to provide their information and that none of these factors would make a difference.

Figure 09: Survey question on factors to improve disclosure rates

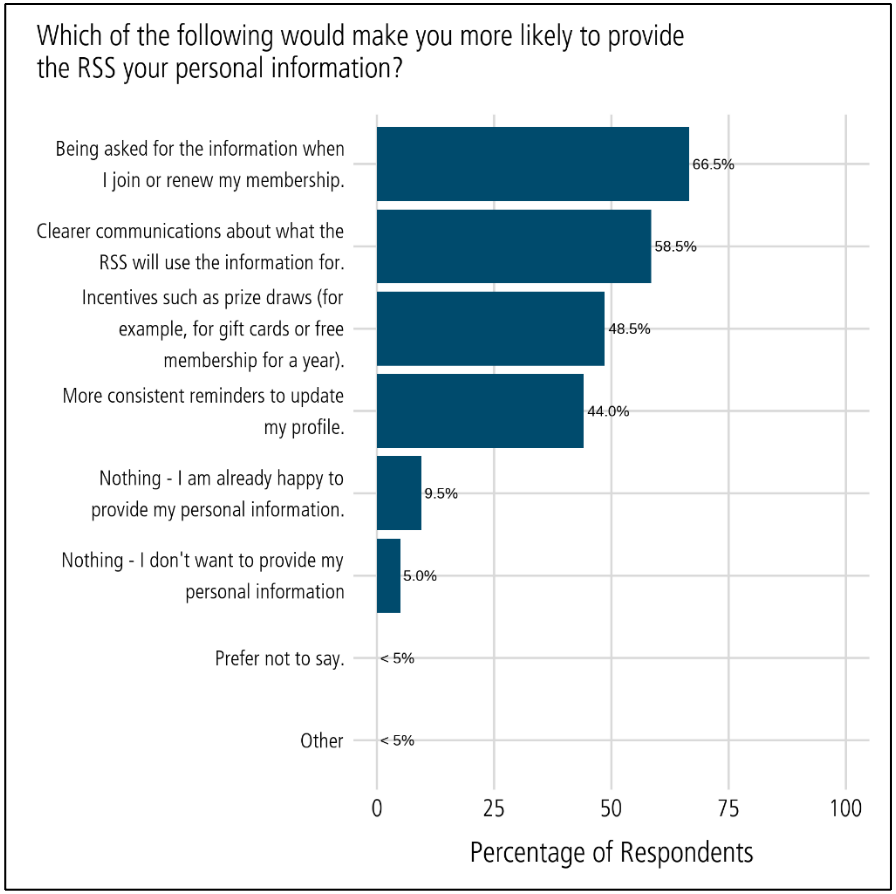
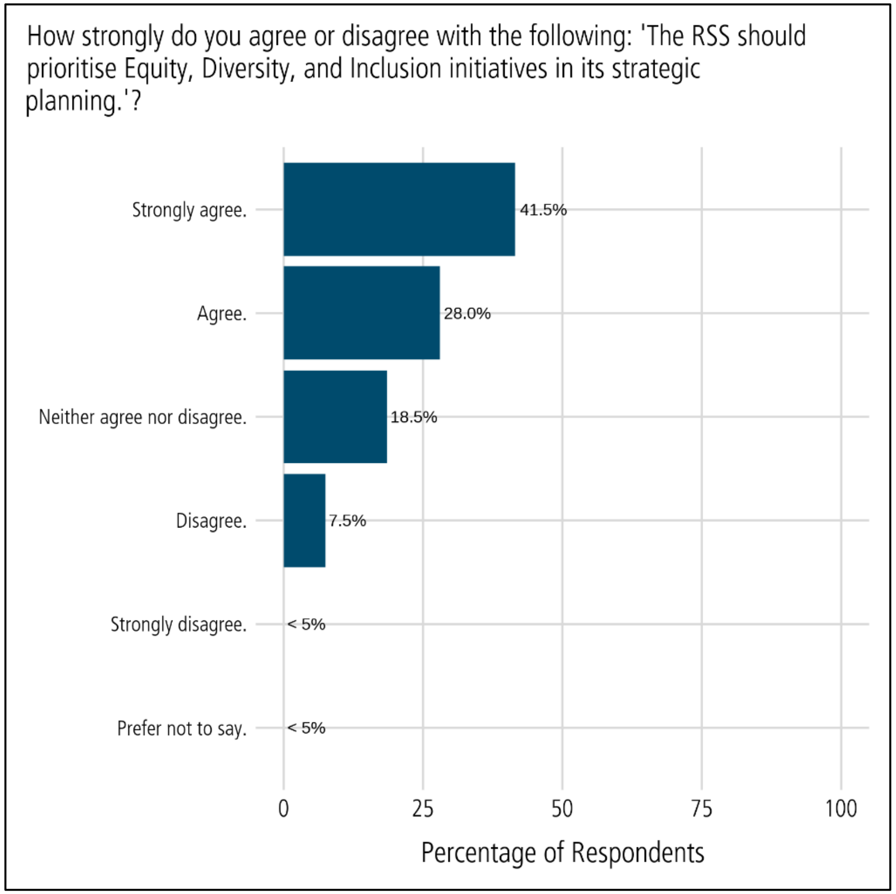


Figure 10: Survey question on prioritising EDI in strategic planning



Moving now to a fundamental question that underpins many of our reasons for being interested in the characteristics of our members, we asked respondents to indicate how strongly they agreed or disagreed with this statement: ‘The RSS should prioritise Equity, Diversity, and Inclusion initiatives in its strategic planning.’ The results for that question are displayed in Figure 10.

As a membership organisation, it is crucial that we make decisions that are reflective of the interests of our members. Therefore, knowing how our members want us to prioritise EDI is key for justifying any of our initiatives going forward, including any collection of member’s data that purports to support those initiatives.

Just under 70% of respondents agreed or strongly agreed that we should prioritise EDI initiatives in our strategic planning. Of those that did not agree, the majority did not express a clear opinion for or against the statement. However, it is important to note that the open-ended elaborations associated with this question revealed strong feelings on all sides of the spectrum, from those that felt EDI is of central value to the RSS and those that felt it is a distraction from our core mission. Some felt that, although EDI should be prioritised, it should not be the top priority in place of issues more immediately related to statistics.

We also asked about how well we are currently doing to communicate with members about our thinking around EDI and our initiatives in that area: 'How strongly do you agree or disagree with the following: "The RSS clearly communicates with me about its Equity, Diversity, and Inclusion initiatives."?' Those results are presented in Figure 11.

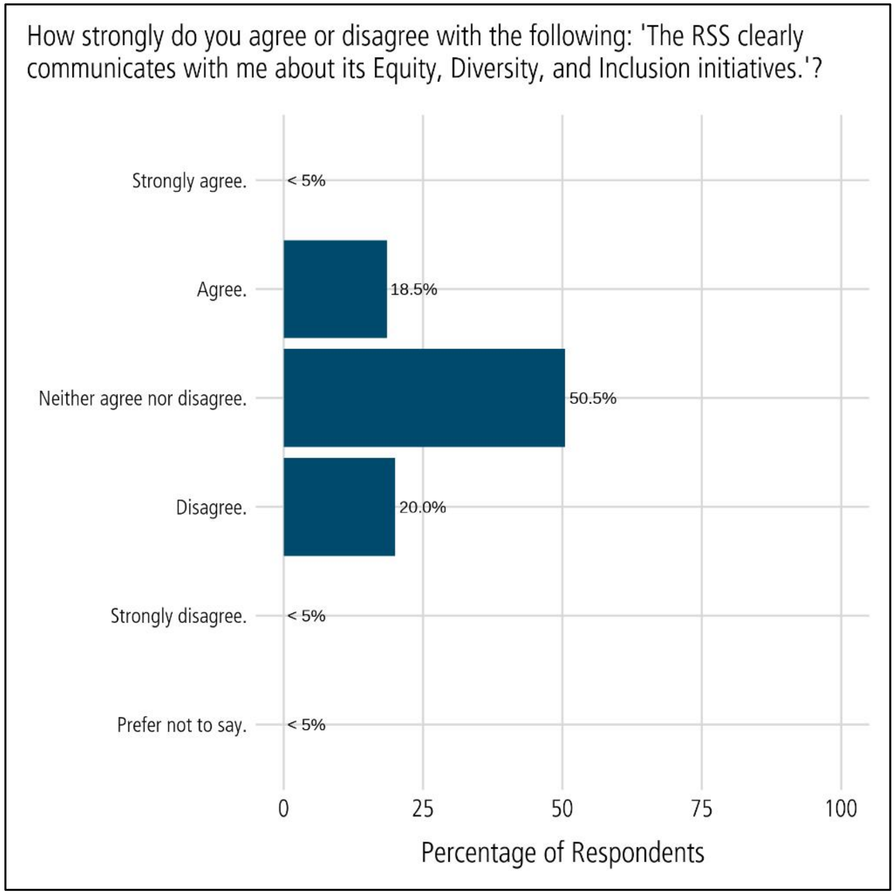
Here, there is clearly room for improvement in our communications. More members disagreed than agreed that we clearly communicate about EDI, and a majority neither agreed nor disagreed.[5]

In summary, taking into account our prior experience with collecting member data alongside these survey results, there are a few key points that merit emphasis:

1) Questions need to be asked at routine contact points with members, especially upon joining. Relying on members to spend their scarce time and attention completing their profiles at non-routine points is unlikely to produce good completion rates, particularly if there is no direct incentive.

2) The comparison of completion rates to non-mandatory questions like gender and a mandatory date-of-birth question reveals that non-mandatory questions can perform nearly as well as mandatory if asked when members join. This means even light-touch efforts can greatly improve completion.

Figure 11: Survey question on communication about EDI initiatives



---

[5] Members that had disagreed or strongly disagreed that EDI should be a priority in our strategic planning sometimes picked 'Neither agree nor disagree' when responding to this question—their elaborated answers emphasised their disagreement with EDI as a priority.

3) Members are generally willing to provide many of their personal characteristics on a member profile, although it is important to be sensitive to the nuances of the questions and to be as clear as possible about how we will use and secure the data.

4) Our members generally support the prioritisation of EDI initiatives in our strategic planning, but they do not feel that we are currently communicating effectively about EDI. While it could be that this is due to a lack of visibility on our current EDI initiatives, it could also have to do with the means by which we communicate with members. Indeed, given that most respondents were not aware that we have been interested in the information they provide on their member profiles, it could also be the case that this is indicative of ineffective communication with members more broadly.

4. USES FOR DIVERSITY DATA

This section discusses some key potential uses for diversity data in general. We run through several of the most practical uses for MOs, although we recognise that there are likely to be others.

MOs should determine uses for the member data that they collect. Wherever you collect personal data, data protection regulations require you to establish a clear justification for doing so. You can use this section to help guide your own thinking about justifications for collecting data. As will be clear from our examples, that exercise is intricately linked to the desired purpose of the data. The more explicit you can be, the better.

ESTABLISHING A BASELINE UNDERSTANDING OF MEMBERSHIP

The most fundamental use for member diversity data is to establish a baseline understanding of who your members are. Especially if this can be done via a CRM-based system that allows member profiles, an organisation can gain a lot of insight into the characteristics of their members and identify trends over time. This is the essence of diversity monitoring.

Having this basic information is crucial for more advanced efforts to understand the perspectives of your members. A survey might indicate that members overall are generally satisfied with their experiences with your organisation, but an age-based breakdown of responses might show that younger members are more dissatisfied than older members—a key insight that can help focus your future efforts.

Remember that you can target your communications at certain groups when you have a CRM to support your surveys, as well as monitor response rates by group. This can help you ensure that you get enough responses from target groups and help you identify patterns of non-response, which can help you take steps to mitigate bias in the results.

When thinking about which diversity questions to ask, it is important to determine what is the minimum information you will need to establish an acceptable baseline understanding of your membership. As explained in Section 2, we expect that, for most organisations, the minimal picture in terms of demographic characteristics will include at least:

- Sex and gender identity
- Ethnicity
- Age
- Country of residence
- Nationality.

More details about these and other questions about diversity characteristics is given in Section 5.

These factors are among the most commonly analysed in many forms of social research, and it is easy to imagine that the perspectives or other characteristics of your members will differ along these lines. Additionally, subgroups within these five characteristics are likely to be large enough to meaningfully analyse. Nationality and country of residence are particularly important for organisations that know they have a large international membership, as the ability to distinguish members along these characteristics can be critical for comparing your internal data to national comparators.

ENABLING BENCHMARKING

Useful data is often data that can be compared to external sources. This can be an invaluable tool for MOs that want to understand whether their members are reflective of wider society or the relevant sector that the MO serves.

Comparison of this kind generally requires that the data held internally by an MO about its membership align with the external data used for comparison. This means that the categories and questions used in both datasets should really capture the same information.

For example, suppose that Organisation A wants to assess whether it is attracting members from different ethnic groups at roughly proportional rates. One of their first tasks is to select the most relevant comparator data that will be used to define what proportional means. In other words, an appropriate benchmark needs to be chosen.

Many MOs that are based in the UK have memberships that are highly international. This can complicate efforts to choose a benchmark that can reasonably apply to the entire membership, because data for a specific sector may only be available at a national level, and each country is likely to have its own distinct practices.

Ultimately, the important point is to ensure that the comparison is being made on data that is as similar as possible between the datasets. Supposing that Organisation A wants to use UK Census data from England and Wales as the relevant comparator, they may want to zoom in on the portion of their membership that resides in England and Wales.

Organisation A would also want to ensure that their ethnicity question aligns with the question in the Census, which is discussed in Section 5. This involves looking at the wording of the question (also called the 'question stem') and its response options to ensure they are the same (or very similar).[6] If they are new to collecting ethnicity data, Organisation A could choose to copy the England and Wales Census questions exactly, including the same question stem, response options, and any guidance that accompanies them.

If Organisation A already collects ethnicity data but in a slightly different way to the Census, then they should see if there is a way to combine response options such that the datasets can be analysed side-by-side. This may be possible if, for example, Organisation A only includes 'white' as a response option instead of the white subgroups available in the Census question. In this case, comparison should be made after aggregating the responses for the detailed categories in the Census data to a singular 'white' category.

However, if Organisation A collects ethnicity data in a very different way to the Census, then a comparison may be impractical. This can happen if, for instance, they only ask members to self-identify their ethnicity with a free-text response box, and it is not always straightforward to map the responses onto Census categories.

In general, as long as a clear mapping of response categories is possible and consistent between two datasets that ask the same question, then comparison should be feasible, although it may require aggregating some response categories. When choosing the level of detail to ask of members, one important consideration is that it is often possible to combine response categories, but it is not possible to disaggregate a response category if more detailed information wasn't requested.

UNDERSTANDING YOUR MEMBERSHIP PIPELINE

A core business-relevant consideration for membership organisations is where their members come from and why they leave. Analysing the sources of new members and the push factors that make them cancel their memberships is important for maintaining robust revenue streams.

Collecting diversity data will allow you to determine whether different groups are more likely to leave your organisation than others. If you find that this is the case, you can conduct follow-up research that seeks to identify the reasons for the discrepancy and plug any gaps. Analysis does not have to be restricted to attrition rates, either—you can consider discrepancies at the joining end of the pipeline, too, or even renewals or upgrades in membership class.

For instance, you might use data on the number of university students graduating into your field as a comparator against your own data for new, career-young members. If you identify that you are under-subscribing from this group (or certain subgroups

---

[6] Comparison may also be possible when questions are very similar, but care must be taken when doing so, as seemingly small differences in how a question is worded can lead to potentially significant differences in responses. It is impractical to provide detailed guidance on such cases in a general way in this report, since every case is unique. For readers who want to prioritise their ability to benchmark, they should emulate the questions used in their desired comparator dataset as closely as possible, barring an overriding reason not to do so.

within it, like graduates of particular nationalities, for instance), then you can take steps to address this. One possible course of action would be more substantial partnerships with universities to promote membership. No matter what solution you determine to be appropriate, you will need to have the data to identify that there is a problem in the first place.

EVALUATING IMPACT OF EDI INITIATIVES

The diversity questions you choose to ask should be informed by the various EDI initiatives that your organisation runs or wants to run. Evidencing impact from these initiatives will be much easier if you have relevant data that can quantify their effects. For instance, an initiative focused on boosting the proportion of conference attendees from the LGBTQ+ community would naturally benefit from having data on sexual orientation for conference attendees.

Similarly, an initiative that sends tailored communications to members from different ethnic communities or nationalities when their subscriptions are close to lapsing would require having ethnicity or nationality information for members, and this would have to be linkable to their member profiles so that their subscription dates are known.

Clearly, any systematic evaluation (or even implementation) of these initiatives will require that your organisation collects a minimum set of information about members, particularly their diversity characteristics. But you should also think about how you collect that information; many analyses require comparison of data from before and after an initiative launches. You may need to compare the perspectives of members who joined before or after a new induction policy focused on inclusivity was adopted, for example. This would entail having the ability to reach out to members based on their time of joining the organisation, but it could also entail being able to group members by some diversity characteristic that is relevant to the induction policy.

The same information should be collected consistently throughout the periods that are relevant for your analyses. Otherwise, you may not be comparing like for like, and your ability to assess the impact of an initiative may be compromised. Consequently, part of your effort should be spent thinking carefully about the kind of questions that may be relevant to your organisation in the medium and long terms.

IDENTIFYING AND ENGAGING SPECIFIC GROUPS

Another objective of your diversity data might be to allow you to identify members of specific groups for the purpose of promoting engagement. Especially where you can link the diversity data with information about engagement with your services, events, or groups, you will be able to identify patterns of lower engagement for certain groups and adjust your programming or communications accordingly.

You will need to collect diversity data that is sufficient to meaningfully distinguish one target group from others. If you need to distinguish a nationality in one instance and a religion in another, separate questions on nationality or religion will be necessary.

Again, the ‘how’ is important here. If your goal is to simply get snapshots of the size or perspectives of these groups, then a survey is sufficient for that goal. But if the goal is to enable targeted communications or promote inclusivity on advisory panels and conference sessions, then you may need to link the diversity information to a member profile on a CRM.

In the latter case, you must take extra care to ensure that the additional risks associated with collecting potentially sensitive information and holding it on an internal database are weighed against the benefits of holding that data. Particular care should be taken where a group is at heightened risk of negative treatment if their information is leaked, such as those with minority sexualities, gender identities, and particular religious beliefs. Each of these factors should be considered on an intersectional basis as well, particularly as members from different nationalities may have radically different risks associated with the disclosure of certain information. See Section 6 for more information about keeping data secure.

Figure 12 (p. 32-33) presents a summary of the various uses of data discussed here.

Figure 12: Summary of potential uses for diversity data

Desired Outcome	Justification for diversity data	Which tools should you use, and how?
Building a baseline understanding of your members; Diversity monitoring; Benchmarking	Diversity monitoring is crucial for understanding whether certain groups are being treated fairly by the organisation. This is usually done by benchmarking internal data to external comparators to identify discrepancies. Without monitoring, you may miss key issues that affect specific parts of your membership, which raises concerns about both equity and revenue risks.	• CRMs offer the best option for monitoring and benchmarking. You can aim for responses from a large majority of members, giving you a firm picture of your members' characteristics over time. • Use your CRM to focus on core characteristics of interest to your organisation, like gender identity, age, ethnicity, nationality, and country of residence. Consider adding other questions if you have clear purposes for them. • Make collection of diversity data routine by mandating the return of forms at key contact points like joining and renewal—but remember to let members opt-out of any specific question and to decline to consent to any of their diversity data being used. • Ensure you choose an appropriate benchmark. This could be national-level data, but there may be industry-specific data that more closely resembles a population of relevance to your organisation. • Consider tracking more sensitive information that is not on your CRM via anonymous surveys that are distributed at regular intervals and universal contact points. • Complement your diversity monitoring efforts with surveys that track how sentiments vary across your diverse membership.
Understanding membership pipelines	A core business question for MOs is how to attract new members and keep existing ones. Collecting diversity data enables you to better understand the member 'journey' from prospective member to perhaps former member in greater detail. For instance, you may identify that you convert a lower percentage of prospective members from minority ethnic backgrounds, or that those members are more likely to leave the organisation each year.	• Analyse data from your CRM to understand trends in joining and attrition rates by specific demographic characteristics. For instance, current members are female as often as they are male, but 60% of joiners are male and 70% of leavers are female, possibly indicating an area for future inclusion work. • You can use CRMs to systematically understand where members come from before joining the organisation, or what interested them in joining. • You can use surveys to try to understand the reasons why people leave. You should distribute these surveys when people leave the organisation, possibly alongside incentives to stay. Recording data on the responses (and possible changes in behaviour) can be used in conjunction with diversity data to understand whether there are factors pushing certain members out of the organisation that can be addressed. • Consider conducting exit interviews with members if further detail is needed about why they don't want to stay.

Evaluating impact of EDI initiatives	Any organisation interested in promoting EDI initiatives should implement ways of tracking progress or measuring impact of an initiative directly. This virtually requires collecting diversity data about members over time.	• CRMs can be useful for tracking key behaviours that might be affected by EDI initiatives, like renewal or attrition rates. You may also use CRMs to log complaints data, and you could assess whether an EDI initiative has reduced complaints among specific groups, for example. • Well-designed surveys can be useful tools to understand changes in sentiment or opinion over time among key groups. Intentionally run your surveys before and after implementing an EDI initiative to try to pinpoint the change attributable to the new efforts. If you have a CRM, you can use it to contact the same members in multiple surveys to better estimate change. • Qualitative methods can be helpful when the impacts of an initiative are likely to be highly nuanced. You can use these tools to understand how you might change the initiative in the future or to understand effects that are not as easily quantifiable.
Member engagement	Member engagement can mean a lot of things, but organisations may be interested in offering services tailored to different communities within their memberships. This could include things like professional mentorship programmes aimed specifically at members from marginalised communities. Diversity data enables targeted communication that can build those networks and also directly pull in members from groups that are under-represented in your organisation's functions, like events, special interest groups, or advisory panels.	• Holding diversity data on a CRM means diversity information will be directly linkable to contact information (at least for staff with the relevant access permissions). This forms the basis of effective targeted communication, which itself will benefit the engagement of specific groups. • Related to diversity monitoring, you can see whether there is under-representation in specific parts of your organisation if you have diversity data held on a CRM that also has information on what groups members participate in. Use your CRM to identify gaps, like low representation of marginalised groups in special interest groups, and consider targeting communications to boost engagement. • Use surveys to gauge interest in potential engagement initiatives, but keep in mind that if the responses cannot be linked to specific members, then their usefulness will be limited. • Use surveys and qualitative methods to understand what members think you could do to boost their engagement. Also use these tools to understand member perspectives on strategic decisions and direction. Surveys will give a more representative picture, while qualitative methods will give a more detailed one.

GENERAL APPROACH TO DETERMINE WHICH QUESTIONS TO ASK

After you have determined what questions you want to answer as an organisation, what data you need to answer those questions, which tools you will use to collect the information, and a plan to use the data for clear purposes, you will eventually need to specify the precise wordings of the questions you pose to members.

This section presents a set of suggested wordings for various diversity-related questions likely to be asked by member organisations. Getting the questions right is important; members are more likely to engage when they see their identities reflected in the response options. They are also less likely to trust that organisations are interested in their perspectives when the questions are not sensitively phrased. Moreover, responses can often be sensitive to minor differences in the question wording or the response options.

We typically begin with reference to harmonised standards for these kinds of questions that have been developed by the Government Statistical Service (GSS) Harmonisation Team based in the Office for National Statistics (ONS). As a general rule, these standards are reasonable default options that otherwise unsure membership organisations should consider adopting. They have undergone extensive testing and form the backbone of questions in many public datasets that MOs will want to use for benchmarking.

Although we present suggestions for a variety of demographic categories, it is important to recognise that organisations can have differing objectives when collecting data, some of which are discussed in Section 4. The harmonised standards have largely been developed with public-sector institutions in mind, and consequently incorporate some considerations that are not strictly required for MOs. Here, we mention possible deviations from harmonised standards where we feel it is likely to be appropriate.

You should feel empowered to include characteristics other than those outlined in the limited space here. While we have tried to include a good selection, organisations are beginning to think about other kinds of information too, such as data about members' caring duties or marital status. Consider asking for more extensive information, and keep in mind that harmonised standards have been developed for more than what we have covered.[7]

Regardless of the specific characteristics you choose to include, diversity questions should strive to allow members to report their identities as faithfully as possible to how they view themselves. In other words, members should see themselves in the response options; they should not have to give themselves a label they are uncomfortable wearing unless you have a clear overriding reason to ask them to do so. Additionally, you should give members the option to decline to provide an answer to certain questions. Typically, this is best done via the inclusion of a 'Prefer not to say' option.

One approach to this is to enable open-ended, free-text responses to diversity questions, but this can sometimes frustrate analysis because analysts need more concise, manageable response categories. Therefore, if you choose to use free-text responses, you should consider doing it alongside questions with pre-defined options.

SEX AND GENDER IDENTITY

Context

Many of the organisations that we have spoken to have told us they are particularly unsure about how to gather data about members' sex and gender identity. This is a highly sensitive topic that can be challenging for organisations to navigate, and understandably most want to avoid participating in what has become a very polarised and difficult debate. Since the start of this research there have been two substantial changes to the landscape.

---

[7] https://analysisfunction.civilservice.gov.uk/government-statistical-service-and-statistician-group/gss-support/gss-harmonisation-support/harmonised-standards-and-guidance/

First, there was the Sullivan Review, which aimed to identify obstacles to accurate data collection on sex and gender identity in public bodies and the research system and to provide guidance on best practice. Second there was a Supreme Court Ruling on the meaning of "sex" in the Equality Act.

The Sullivan Review is helpful from the perspective of this report in two main ways. First, it clearly identifies some of the challenges with terminology in this debate – particularly highlighting the confusion around the use of the term "gender". Sullivan points to at least three different meanings for the term: sex, social structures and stereotypes associated with sex and gender identity. The review helps to clarify that the terms used when collecting data should be explained, and that the words sex and gender should not be used interchangeably.

Second, Sullivan provides good practice guidance for how to collect data on sex and gender identity: organisations requiring this information should ask a two stage question that first asks about a person's sex and then a follow-up question to capture gender identity. This makes sense from a methodological perspective, because it makes it possible to develop a better understanding of the influence of both factors and distinguish between outcomes.

We suggest that membership organisations follow the review's recommendations, asking a two stage question and avoid using wording that combines sex and gender identity in one question (see possible questions below).

There is a nuance here concerning how people with a Gender Recognition Certificate (GRC) should respond to a "sex" question. Prior to the Supreme Court ruling, it was understood (for example in guidance for the 2021 Census) that people with a GRC should answer these questions in line with their GRC. The ruling has cast some doubt on this – stating that, in the context of the 2010 Equalities Act – a GRC should not be seen as altering someone's sex.

Sullivan has published an addendum to the original report, providing additional guidance following the Supreme Court ruling. The addendum reinforces the advice captured in the original review, that 'data collection on sex… should refer to biological rather than certificated sex'. The addendum points to the importance of biological sex in ensuring organisations meet their Public Sector Equality Duty.

What does this mean for membership organisations? The Sullivan Review is a government-commissioned report designed to address methodological issues in the collection of sex and gender data, and is the most comprehensive research published on this to date. As such, our advice would be to follow the recommendations in the review.

We have stressed repeatedly in this report that organisations should be clear about why they want the data they are collecting and what they will do with it. Members should then be asked questions designed to find out that information. Data should be collected in a manner that is methodologically sound, and enables members to feel they are being included and recognised as themselves.

When considering what data to collect on sex and gender, as with all of the possible questions suggested in this report, membership organisations should consider their own needs and context, what the data will inform, and the balance of priorities, when deciding what to ask. For example, membership organisations do not generally have a Public Sector Equality Duty and you may decide that the appropriate target in your context for the question on sex is legal sex, which will enable members who have a GRC to feel recognised.

Possible questions

In line with the Sullivan Review, you should first ask a question that directly asks about "sex". That question should be: "What is your sex?", with response options: Female, Male, Prefer not to say; and you should provide clarity on the target of the question.

Membership organisations, as set out above, should decide if the target is sex at birth or legal sex, and you should provide guidance on how you want your members to answer. Either way, you should indicate that a question about gender identity will follow so that members know they will have an opportunity to express their identity.

We would recommend asking a second question to establish whether your members are trans, non-binary or gender diverse. Though you may not be planning to use this information, if you just ask a sex question then you do not give your members an opportunity to tell you their identity. The question should be explicit – as set out in the Sullivan Review – and ask: '.eg, "Are you transgender, non-binary or gender diverse?' with response options: Yes, trans woman; Yes, trans man; Yes, non-binary; Yes, Other, please specify if you wish; No; Don't know; Prefer not to say.

The Office for National Statistics is developing harmonised standards – if you would like to compare your data with national statistics, you should follow the ONS standards when they are released.

ETHNICITY

For ethnicity data, a good starting point is to first ask members to provide their ethnicity using categories via a tick-box or 'search-as-you-type' system that aligns with the 2021 Census for England and Wales. There are varying levels of detail that you could use, but you should consider using either a version with nineteen categories or the fully detailed version, which has 287, depending on what is feasible in your system.[8]

MOs should note that the ONS is working on potentially updating their list of detailed response options, as they have received feedback that more detail is desirable by many users. This means that if you are interested in benchmarking to Census data, you should be ready to update your response options in the future when these updates are introduced.

Although you should consider asking for the detailed response options, we anticipate that for some of them, the number of members falling into certain categories will be very small and therefore not strictly analysable on their own. Despite this, collecting more detailed categories is important for at least the following reasons:

- It can be hard to know for sure in advance which categories will be too small to analyse, and you should be open to surprise findings.
- Insights from the higher-level ethnicity categories may be limited.
- Members should see themselves in the response options as much as possible. Offering more categories to this effect can improve your response rates and build trust with members.
- For analysis, you can aggregate categories that have small numbers to a less detailed category, but you cannot go the other way if you did not collect more detailed information.

Some members may not find it straightforward to categorise themselves according to even the detailed response options. This can be because ethnicity is sometimes perceived as subjective, fluid, or context-dependent, or because membership organisations can have a large number of international members who may not fit neatly into ONS categories. To address this, MOs might consider a second ethnicity question to accompany the first, namely an open-response question asking members to state their ethnicity as they see it. We suggest providing some guidance to members alongside the pair of ethnicity questions explaining why you are asking both questions.

The first question has to do with collecting data that is easily comparable to what is found in comparator datasets. You should clearly communicate that collecting this data is therefore important for benchmarking purposes.

The second question is about collecting data that is as faithful as possible to the identity of the person responding. Collecting this information can be helpful when designing services tailored to the needs of specific communities. For instance, for most

---

[8] For information about the categories used for the Censuses, a good starting point is available here.

member organisations that target any of their programming, they may want to rely on self-described identities when trying to determine who to target.

When you analyse and report ethnicity data externally, you will likely want to aggregate the detailed options to 7 categories (white British, black, Asian, mixed, white other, other, prefer not to say) rather than the 5 that are reported for the Census data.[9] Most of the time, we expect that comparison of minority groups to the white British group will be more appropriate than comparison to the full white group, and thus the white category should be separated accordingly.[10] It is also good practice to give members the option to explicitly decline to respond via a 'prefer not to say' option.

When direct comparison to the five aggregated Census categories is required for benchmarking, the white British and white other categories can be combined, and the 'prefer not to say' respondents can be dropped from analysis if item non-respondents are not included in the Census figures. When benchmarking is not the immediate goal of an analysis, all seven categories should be used.

For some organisations, ethnicity data may be more useful when collected alongside religion and national identity, and organisations should consider collecting all three. This is because members of some groups or communities may not be identifiable without additional information. When asking all three, the question ordering should be national identity, then ethnicity, then religion.

However, given the small size of some population in member datasets, it may not always be appropriate to collect such detailed information, as use-cases may be more limited, and the benefits may be outweighed when considered against the additional risks of processing this data and the additional burden it places on members. We leave it to the discretion of individual organisations to decide in their own cases what is most appropriate regarding religion and national identity.

Regardless of whether religion or national identity data are collected, it is good practice to collect data on individuals' country of residence, as this has considerable value apart from identifying small groups, including when benchmarking other data against national data sources. See the subsection on national identity for more detail.

DISABILITY

As with other types of data, if collecting disability information, MOs should be clear on why they are doing so. If this is primarily to inform how you deliver your services – rather than with a view to external benchmarking – your purpose may be most effectively served by the question: 'Do you consider yourself to have a disability or long-term health condition or illness?' with 'Yes,' 'No,' and 'Prefer not to say' as response options. You might also consider questions that asks directly about neurodiversity or other kinds of conditions if you have a clear purpose for this information.

If included on a CRM, a practical follow-up question could be a free-text response option that asks members to describe any reasonable accommodations your organisation can make to improve their membership experience. Such a question might be less appropriate on an anonymous survey as you will not be able to reasonably action or respond. However, wherever you ask this, you will need to ensure expectations are managed, as you may be unable to fulfil every type of request for all individuals.

These questions are aimed at understanding subjective experience, regardless of any exact duration or effect on someone's life. We suspect that the subjective experience of

---

[9] Of course, external reporting using even more detailed categories might be appropriate if there is sufficient data to avoid running into privacy concerns for the smaller groups. For internal purposes, sharing the more detailed data may be less sensitive, but you should still be careful to avoid sharing identifiable information with unauthorised individuals.

[10] Interested readers can view the RSS response to the Race Disparity Unit's Consultation on Standards for Ethnicity Data from August 2022, available here.

disability and illness is more relevant to most MOs than disability that meets any specific criteria.

However, MOs should consider using the harmonised standards for disability, especially if they are interested in benchmarking. In their most recent review,[11] the GSS highlighted various pros and cons to the currently recommended approach, but ultimately suggest continued use of the standards for the time being. The current standards have been extensively tested and will likely outperform untested approaches, at least in so far as an organisation is trying to capture the same sort of information as the standards.

The current harmonised standard for disability data involves several questions, but there are two 'core' questions:

- 'Do you have any physical or mental health conditions or illnesses lasting or expected to last for 12 months or more?' This comes with response options of 'Yes,' 'No,' 'Don't know,' and an option to refuse to answer.[12]

- If someone responds 'Yes' to the previous question, then they are asked: 'Does your condition or illness / do any of your conditions or illnesses reduce your ability to carry out day-to-day activities?' This has response options 'Yes, a lot,' 'Yes, a little,' and 'Not at all.' [13]

Whatever your requirements around benchmarking, you should consider whether there are more important things to capture through questions around disability. For instance, the Royal Society[14] have recommended that national statistics bodies should start prioritising questions around functional challenges rather than identity alone. Some MOs may find that a similar approach would be appropriate in their case.

Furthermore, the GSS review found that people who are not in the 'core' disabled population, meaning people who are neurodiverse or who have progressive or fluctuating conditions for example, are not well captured. If you determine that capturing that level of detail is relevant for your own purposes, then you should consider adapting your question list accordingly.

SEXUAL ORIENTATION

For sexual orientation, you could follow the UK Census 2021 question wording for sexual orientation if your organisation has determined that it wants to collect data for this characteristic. The ONS has conducted extensive testing on this question in comparison with the wording found in the GSS harmonised standard.[15] Although the two suggested approaches are very similar and will likely produce similar results, the Census is most likely to be the data source used for benchmarking if not a sector-specific dataset.

The question is: 'Which of the following best describes your sexual orientation?' The response options are:

- Straight or Heterosexual
- Gay or Lesbian
- Bisexual
- Other sexual orientation (with a free-text response option available to those who select this option).

---

[11] https://analysisfunction.civilservice.gov.uk/policy-store/review-of-disability-data-harmonised-standards/#methods-of-research

[12] https://analysisfunction.civilservice.gov.uk/policy-store/long-lasting-health-conditions-and-illness/

[13] https://analysisfunction.civilservice.gov.uk/policy-store/activity-restriction/

[14] https://royalsociety.org/-/media/policy/projects/disability-technology/disability-technology-report.pdf

[15] https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/sexualorientationquestiondevelopmentforcensus2021

- Prefer not to say

There are other sexual orientations that some members may identify with, such as asexual or pansexual. These were not included in the Census question because consultations with representatives of some of these identities did not reveal a user need for the prevalence of these categories to be estimated. These orientations are grouped into the ‘Other’ category in the Census data.

As far as benchmarking is concerned, including additional sexualities in the list of response options should not dramatically change the comparability of the data. These groups can always be combined for analytical purposes, although it may make the question very slightly more burdensome to complete if respondents have to read more options. Indeed, we expect that, for most organisations, each group responding with any of these minority sexual orientations will be small as a percentage of all respondents. Therefore, it will probably be necessary to group all sexual minority respondents into an ‘LGB+’ category when conducting analysis of survey results.[16]

Therefore, listing additional sexual orientations beyond what is provided by the Census will mostly be a matter of making members feel recognised in their identities and enabling targeted communication with them rather than enabling an understanding of their unique perspectives. Still, you may determine that additional categories are appropriate for your purposes.

SOCIOECONOMIC STATUS

The GSS harmonised standards suggest eight total questions to be asked to measure socio-economic status (SES). The questions are intended to capture socio-economic background as defined by the Social Mobility Commission:

> ‘”Socio-economic background” is the term to refer to the particular set of social and economic circumstances that an individual has come from. It permits objective discussion of the influence of these circumstances on individuals’ educational and career trajectories; and it can be objectively measured by capturing information on parental occupation and level of education.’

The Social Mobility Commission offer detailed guidance on how to approach collecting data on socioeconomic status.[17] Their guidance is aimed at employers collecting employee data, although most of its recommendations should also apply to membership organisations looking to collect this information from members. It includes guidance on which questions to ask and how to use and interpret each one.

MOs looking to measure SES may wish to follow this guidance, but we reiterate that organisations should have a clear purpose for this data before beginning to collect it. This could include an intention for your organisation to facilitate mentoring groups within your sector, for instance, or to offer additional resources to members from lower SES backgrounds. You should keep in mind that you may not need all of the questions, and adding all of them to surveys in particular may add considerable length. Nevertheless, these extensive questions are a sound and commonly used option for organisations looking to collect SES data.

---

[16] When considering whether a group is too small, you should think about whether you can meaningfully do anything with data for the group’s expected size. You should also consider whether small errors in responses (say, 1% of people mistakenly select the wrong option) will lead to dramatic levels of inaccurate data for a small group. If so, you might have to take steps to verify that members have accurately recorded themselves, likely with follow-up studies. Consider whether that will be feasible in your case. Consideration should also be given to whether public reporting of small groups will risk identifying individuals. And, as always, the risks of sensitive personal information being wrongly disclosed always needs to be factored into whether collection of data is appropriate.

[17] https://socialmobility.independent-commission.uk/resources/socio-economic-diversity-and-inclusion-employers-toolkit/

RELIGION

The GSS harmonised standard for religion differs slightly for each of the nations within the UK. As England is the largest of these and we expect most interested MOs will have a larger share of members from England than elsewhere in the UK, they will likely want to follow the harmonised standard for England.[18] [19]

Another reason for following the English standard is that it captures a 'weak' religious affiliation rather than a 'strong' one. The difference is that a weak affiliation can be understood more as a community affiliation, whereas a strong affiliation may relate more closely with religious belief.

We expect that, for most MOs, a weak affiliation is of more practical interest, although individual organisations may differ in that assessment. Organisations looking to capture a strong religious affiliation might instead rely on the Scottish or Northern Irish wording, although this may reduce their ability to effectively benchmark: 'What religion, religious denomination or body do you belong to?'

Following the English standard, the wording is: 'What is your religion?' The response options are:

- No religion
- Christian (including Church of England, Catholic, Protestant and all other Christian denominations)
- Buddhist
- Hindu
- Jewish
- Muslim
- Sikh
- Any other religion, please describe.
    - This option should be accompanied by a free-text response box.
- Prefer not to say

NATIONAL IDENTITY, NATIONALITY, AND COUNTRY OF RESIDENCE

National identity refers to one or more self-identified countries or nations that an individual feels most connected to. It is distinct from nationality, which refers to the countries in which an individual has citizenship, and from country of residence, which has to do with where an individual resides or is domiciled.

MOs will find that data on their members' country of residence is of considerable practical value. Knowing this information will not only allow the organisation to plan events more effectively around their membership, but also to help them benchmark against data that are only available at national levels. In particular, we expect that benchmarking exercises will usually require UK-based MOs to identify their members that reside in the UK before engaging in any comparisons. Therefore, country of residence will be useful information regardless of whether other data is collected.

If desired, data on national identity should be collected using a list of the most common identities your organisation expects from its members. For UK-based organisations, it will be appropriate to follow the GSS harmonised standard on national identity,[20] which uses a question stem of 'How would you describe your national identity? Please choose all that apply.' The response options are:

- British
- English
- Scottish
- Welsh

---

[18] https://analysisfunction.civilservice.gov.uk/policy-store/religion/

[19] The question used for the 2021 Census in England and Wales differs slightly from the harmonized standard, but this variation is not expected to meaningfully impact the comparability with data collected using the harmonized standard.

[20] https://analysisfunction.civilservice.gov.uk/policy-store/national-identity-harmonised-standard/

- Northern Irish
- Irish
- Other, please describe.
- Prefer not to say

The 'Other' option should be accompanied by a free-text response where members can write their own answers.

Organisations should consider their own knowledge of their members and feel free to add other, pre-defined response options for nations that they believe will have high levels of identification. This will reduce the burden on analysts to combine free-text responses into analysable categories.

Regulations on the processing of personal data stem from two key sources of legislation: the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018. For the purposes of this report, we refer simply to 'GDPR' as a blanket term for the entire data protection regime in the UK.

Compliance with GDPR was flagged as one of the areas of greatest concern among membership organisations that attended our roundtable events for this report. There is pervasive fear among MOs about overreaching in their data collection efforts, both from a legal and an ethical perspective.

We have put together this section to offer some support on how to improve member data in the context of GDPR. Although we are not in a position to produce a comprehensive guide to complying with GDPR, there are some good practices that you could adopt when improving your member data, and these have some relevance for GDPR compliance as well.

However, readers should note that the RSS is not equipped to offer legal or other formal advice on this matter. The discussion in this section is based on our understanding of what constitutes good practice around the use of data and how we handle our own membership data. Readers in need of legal guidance should seek advice from legal professionals.

As a practical first step whenever you have questions about something related to GDPR, the Information Commissioner's Office (ICO) publishes helpful guides on its website[21] that you should consult.

JUSTIFYING DATA COLLECTION

It is critical that you establish a clear justification for collecting your members' personal data. As a first step, this involves applying the principle of being clear about your objectives, needs, and abilities. If you cannot identify clear reasons for collecting the data, then you should avoid collecting the data in the first place.

Section 4 of this report runs through possible use cases for diversity data with accompanying guidance about which tools to use for each case and how to implement them. Interested readers should consult that section for ideas for how they might use data for concrete purposes.

When thinking about justifying the collection of personal data, you have a duty to comply with the data minimisation principle of GDPR. This principle dictates you should only collect data that is adequate, relevant and limited to what is necessary for your purposes.[22]

Before adding a question about, for example, ethnicity or disability on a membership form or survey, you should ask: 'What will we use this information for, and is that use necessary and fair?' You should be able to explicitly justify how collecting that data will serve your organisation's EDI objectives (or another legitimate purpose). For example, you might determine that knowing members' ethnic backgrounds is necessary to identify if certain communities are underrepresented in your field, which informs outreach programs to relevant institutions that feed into your membership, like universities—a clearly defined purpose. On the other hand, if you cannot pinpoint a use for a particular question, you should not ask it.

In addition to necessity, consider relevance and proportionality: is the depth of information appropriate for the stated purpose? For example, asking for a detailed medical history would likely be disproportionate if your aim is just to know how many members have a disability. Here, a simple 'yes or no' question would suffice. Of course, you may identify other reasons for asking for more detailed disability information. The key is to always align questions with a legitimate purpose and ensure the scope of data is not excessive.

---

[21] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/

[22] Further detail about the data minimization principle is available here.

Consent

When you process any personal data, you are required to identify at least one lawful basis for doing so.[23] Although there are multiple options, their applicability will depend on your specific circumstances. We focus here on consent, which we believe is the most likely to be available to most MOs. It also best adheres to the principle of prioritising trust.

Organisations using consent to process data must meet certain standards for it to be valid. That is, the consent must be:

- Freely given: Members must have a genuine choice to provide diversity information, meaning there cannot be pressure or consequences if they choose not to provide their personal data. It should be clear that providing it is optional and not a condition of membership or services.
- Specific and informed: Explain what data will be used for (e.g. for aggregate diversity statistics to improve inclusion programs) so that consent is tied to a well-defined purpose.
- Unambiguous and affirmative: Use an opt-in mechanism – for example, an unticked box or a question where the member actively chooses to provide the information.
- Prominent and separate: The consent request should be separate from other terms or agreements. It should not be hidden in a privacy policy or bundled with unrelated consents.
- Documented: You should keep records of what each individual consented to, and when/how they consented, in case you need to demonstrate this later.[24]
- Easily withdrawn: Members must be able to withdraw consent at any time. They should be informed upfront how to revoke their consent (for example, via an email to a certain address or an account settings page).

Relying on consent will also be helpful for enabling you to collect more sensitive data , known as 'special category data,' which includes information on someone's racial or ethnic origin, religious or philosophical beliefs, sexual orientation, health (ie disability), or trade union membership, among other categories. Processing special category data requires that you satisfy one of the conditions listed in Article 9 of GDPR on top of what is typically required. Properly-obtained consent can serve as both a lawful basis and an Article 9 condition.

DOCUMENTING DATA PROCESSING ACTIVITIES

There are good practices to follow to ensure that you are keeping appropriate records of your data-related actions. Following these practices will help you formally understand the breadth of information that your organisation has on members and help organise thinking around the data you use.

At a minimum[25], you should keep an up-to-date record of:

- what kinds of personal data you collect;
- your lawful bases for processing that data;
- your organisation's purposes for the data;
- where the data is stored;
- links to records of consent, if applicable;
- who is responsible for the data;
- the types of external parties that will receive the data (if any);

---

[23] For more detail, please refer to https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/.

[24] The most practical way of recording consent is usually digital. Ideally, you can simply include a field that captures someone's consent in the raw data. Here, the 'link' to this information can just be a reference to where this record of consent is kept (ie in the raw data files), and you can include a note (in the raw data or the ROPA) explaining how the data in the consent field is collected (ie via tick-box or a direct question included on the original form or survey).

[25] Most of these are derived from ICO guidance available here.

- details of any data transfers to other countries;
- information about the groups to which the data pertain (ie, whether the data is about members, employees, or another group)
- retention policies; and
- a description of security measures in place.

All of this information should be compiled into one place in the form of a Record of Processing Activities (ROPA)[26]. Keeping things together makes it easier to map your data-related activities more clearly in the future, which will aid decision-making and strategic thinking.

KEEPING DATA SECURE

Data Protection Impact Assessments (DPIAs)

Performing a Data Protection Impact Assessment (DPIA) is a requirement before starting any processing that is likely to result in a high risk to individuals.[27] We assume that membership organisations will likely be engaging in activity that warrants a DPIA when they collect special category data. However, apart from any legal requirements, it is good practice to conduct a DPIA in the course of your data-related activities, since it will help you think clearly about how you will use data and keep it secure.

A DPIA involves systematically describing the project (what data you'll collect, how, and why) and then assessing the potential risks to individuals and how you will mitigate them. For a diversity data initiative, the DPIA process should consider questions like:

- Could individuals suffer harm or distress if this data is misused or leaked? For example, consider that revealing sexual orientation or gender identity could in some cases pose serious personal risks to an individual's safety. The DPIA should flag this and specify how your organisation will go about mitigating this risk.
- Are we collecting any data that is not necessary or overly intrusive?
- What security measures are in place?
- How will we ensure transparency of our processes and decisions to members?
- What is the plan if a breach occurs?

By identifying such risks early, you can take steps to mitigate them. For instance, you might decide to encrypt the data, or use a coded system where names are stored separately from individuals' information (also called pseudonymisation) to lessen the impact of any potential leak. You may also simply assess that the risks outweigh the benefits and decide not to collect as much data. We encourage organisations to think carefully about the trade-offs to avoid both overly zealous and overly risk-averse approaches to member data—there are potential costs to both.

Access Restrictions and Confidentiality

Access to raw diversity data should be limited on a need-to-know basis. In practice, this means only certain staff or volunteers who are charged with analysing the data should be able to view individual responses. Thus, you could implementing role-based access controls on databases or spreadsheets containing the data. For instance, if the data is collected via an online survey tool, ensure that the results portal or survey account is password-protected and available only to the relevant team members.

If data is stored in an internal drive or system, use folder permissions such that only the assigned personnel can open it. In most CRMs, it should be possible to set access restrictions to certain views within the member database, such that everyone can see basic contact details for members, but not everyone has access to the full suite of information available on members' profiles.

---

[26] The ICO has published a template ROPA with an example. It is available for download here.

[27] More detailed information about DPIAs, including a template from the ICO, is available here.

Secure Data Handling and Storage

Beyond controlling who can access the data, you must secure the data against external threats and accidental loss. Technical measures are crucial when handling diversity information; consider implementing the following measures when storing data and performing analysis:

- Encryption: If the data is stored electronically, especially in portable form, it should be encrypted. Encryption ensures that if a device is lost or if an unauthorised person somehow obtains the files, they cannot read the data without the decryption key.
- Pseudonymisation: Where possible, consider separating directly identifying information from the sensitive data. For example, you might simply hold member IDs in the raw data file, rather than including members' names as well. This can help secure against breaches of information to external parties, although internal access restrictions will still be necessary to avoid unauthorised personnel from being able to link individuals between datasets.

Regularly review your security measures. Simple steps like keeping software up to date, using strong passwords, and two-factor authentication can prevent many breaches. You should consult IT professionals about more specific ways of securing your data.

Data Sharing and Reporting

Given the sensitivities around personal data, MOs should share data externally only when it is necessary to do so. For smaller organisations that lack in-house analytical expertise, external sharing may be unavoidable for the data to be useful. However, you can potentially share data only in aggregated form, such that data is reported at the group level rather than for individuals, depending on the level of detail required by the analysis.

If you are collaborating with an external partner and group-level data will not suffice, you should have a data sharing agreement or contract in place. This agreement can limit what they receive to the minimum necessary and help ensure that they keep the data secure.

If you plan to publish diversity statistics (for example, in an annual report or on your website to demonstrate transparency about your organisation's diversity), you should use aggregated and anonymised data. This will usually mean reporting results only at the group-level. You should also ensure that the aggregation is done in a way that individuals cannot be re-identified.

A common rule is to suppress or avoid reporting very small group sizes. For example, if fewer than 10 members identified as a certain ethnicity, do not publish a statistic specifically about that subgroup, and, if appropriate, try to combine with a larger subgroup. The risk with small numbers is that readers could guess who they are or identify themselves in the data, and thus that the data is not truly anonymous.

If you anticipate that you will need to share members' data externally for any reason, you should explain this in your privacy statements, including the reason for the sharing and, if known, some information about the partner organisation.

Data Retention

You should have a clear policy regarding data retention periods, which includes specific periods after which data will be securely deleted from your systems. Consider the purpose of your diversity data collection and determine when that purpose is fulfilled, or when the data become outdated.

If you conduct a one-off diversity survey or census of your members to get a snapshot, you might only need detailed personal responses until you have analysed and reported the results. After that, keeping the identifiable responses may not be necessary. One approach is to anonymise or delete the raw data after analysis is complete and validated. You could retain the aggregate statistics for historical comparison, as these are not personal data once fully anonymised. If you plan to do the survey at regular intervals, you might keep the individual data until the next couple of survey cycles are done, so that you can compare or measure changes for those who participated across time. However, you should still specify that there will be a cutoff for the data retention even in this case, such as deleting the data after two or three cycles have passed.

If diversity attributes are stored as part of each member's profile in your database, you should have a policy for regular review and deletion. For example, if a member leaves the organisation, you might decide to remove their diversity data after a standard retention period that meets your needs. This could be one year if year-end diversity monitoring is something your organisation does, or longer depending on your specific purpose. Again, remember that you keep aggregated data indefinitely so long as it does not contain personally identifiable information.

Additional Considerations

After you complete a cycle of data collection and analysis, review what went well and what didn't. Maybe you found that some questions made participants uncomfortable or that your security measures could be enhanced. Use that feedback to improve future practices.

Additionally, ensure those involved in the project have at least basic training in data protection. They should understand that diversity data is sensitive. Volunteers or committee members should be briefed on confidentiality and the importance of following the procedures set out.

Communicating with Members

One of the core principles of this report is that you should prioritise building and maintaining trust with your members. You can do this in the data protection space by clearly showing that you have robust practices that guide how you process members' data. Generally, this should be done via an official privacy statement that accompanies any form of data collection you distribute to members (or other individuals). You can link to part of your website when the distribution happens digitally in surveys, or on joining and renewal forms that gather data for your CRM.

Privacy statements need to include: what you collect, the lawful basis you are using for processing, information about your retention policy, and an explanation of your members' data protection rights. You should rely on guidance from the ICO when determining how to present this information.[28] Additionally, well-executed privacy statements will include at least the following:

- A clear statement of what the data will be used for and how it will contribute to your organisation's objectives. In so far as you know the exact outputs from the data collection, you should include details of them here.
- Information about your justifications for collecting the data. If you are relying on consent, you should state that explicitly and reiterate that there will not be negative repercussions if they do not consent.
- Any organisation with which the data will be shared should be referenced in the privacy statement, if known. If you do not intend to share the data externally, then you should mention that fact as well.
- Information about your security measures.
- You should include information about how a member can reach your project team for more information, and how they can withdraw consent (ie via an email address).

If relevant, you can reference the fact that you have completed a DPIA in your communications. This will help build trust that you have followed all necessary procedures and that you are thinking carefully about how to protect members' data.

---

[28] See further guidance available here.

There are tools that analysts can sometimes use to help alleviate the burden of incomplete information about an organisation's members. With the rapid proliferation of AI technology throughout society, we have heard interest during our roundtables for some guidance about how these tools could be leveraged to help organisations that otherwise struggle to get satisfactory engagement with their surveys and CRMs.[29]

This section offers some brief guidance about the use of these methods by MOs. The goal is not to provide a detailed explanation of these methods or how to apply them in practice, in part because such a discussion is beyond the scope of the present report, and a thorough discussion can quickly become overly technical. Rather, the goal is to offer a brief overview of some of the ways in which applied statistics might (or might not) benefit data-related projects in MOs.

We begin by noting that, although AI tools contain the promise of greater accessibility in data analysis, we cannot make a general recommendation that untrained staff use these tools in isolation to try and produce organisational insights. The state of these technologies at the time of writing this report is such that human technical expertise is still required to ensure proper application.

Consequently, we caveat this section by reiterating that interested organisations should seek the support of statistical experts. Additionally, it is important to reemphasise that you should have clear purposes for the data generated by these tools. They are not substitutes for careful thinking about your objectives or the tools you will use to collect data by default.

IMPUTATION METHODS

Missing data is a common issue that many organisations seek to overcome. In the context of diversity monitoring forms collected via a CRM, missing data can be at the level of a member, in which case someone has simply not completed a form, or at the level of a specific item on the form, in which case their data will be available for some questions but not all.

Imputation methods are among the most standard for addressing item-level missing data. These generally aim to fill relatively small holes in datasets that are otherwise mostly complete.

There are various methods of imputing data, each of which makes its own assumptions about the nature of missingness in the data. Imputation methods do things like build statistical models to predict someone's missing information based on individuals for whom data is available.

These methods are not straightforward to apply without some level of statistical expertise guiding the work. The validity of the results depends on the validity of the assumptions of the imputation method, and not all assumptions are created equal. Indeed, there are situations in which it is not appropriate to use imputation. These can include when missing data is actually the norm—as is the case for many member organisations. Put another way, statistical methods are not a magic bullet for solving issues in the data collection pipeline. They can be useful for filling gaps, but they cannot conjure a comprehensive picture of your membership from scratch.

It is more appropriate for organisations that struggle with persistently low response or disclosure rates to invest resources into improving data collection practices such that non-response is tackled directly. Accomplishing that will also generally lead to data that is more reliable than those derived from imputation.

Imputation may be valuable in limited circumstances where some survey respondents have selectively chosen not to provide certain information, but the majority of respondents have filled out all questions. Unfortunately, many of the limitations of imputation still apply under these circumstances, so again we emphasize the need to seek guidance from a data professional.

---

[29] More information about data collection tools is available in Section 2.

Additionally, imputed personal data still counts as personal data for the purposes of GDPR, even if it is incorrect. You therefore need to take care to ensure you are handling the data carefully, being sure to communicate with members and implement appropriate security protocols.

AI TOOLS

AI tools have a lot of potential to expedite and advance the analytical capacity of organisations, but as with imputation methods, we recommend consulting statistical experts before using them to analyse your data.

Nevertheless, AI tools do offer the potential for expanded capacity in some areas. Consider, for instance, the use of free-text response options to various diversity questions. Normally, these are tricky for analysts to deal with because they are not immediately analysable in a systematic way without first translating the responses into more manageable, well-defined categories. Analysts may have to spend considerable time recoding or classifying free-text responses for that purpose, sometimes to the point of impracticality.

This is one area where AI and other applied statistical tools can be potentially useful in so far as they can expedite the process of grouping responses into more manageable categories or themes. If you choose to go this route, you will want a human analyst to review the outputs from that exercise, as these methods are not foolproof. However, when properly utilised, they can save analysts hours of time and frustration while reducing errors.

Another area where AI tools and other applied statistical methods could be helpful is in making inferences about diversity from more limited information about members, like their names. Some tools perform reasonably well at predicting ethnicity and gender from names, and application of these tools can be done using readily available statistical packages in standard statistical software.[30]

The advantage to inferring characteristics from names is clear: your organisation is likely to have names for all of its members, and you can quickly gain an estimate of the full membership's characteristics.

There are concerns with following this approach. Firstly, there are ethical considerations that should not be overlooked. Inferring characteristics in this way at an individual level could be seen as offensive, but even more seriously, if there is a data breach and inaccurate information is revealed about sensitive personal characteristics, this could pose security risks in some cases. We therefore advise against using these methods to produce individual-level records.

You might consider merely aggregating the predictions to the membership level and then deleting all individual-level data. This approach could be helpful for providing a model-based estimate of your membership's gender or ethnicity distribution, but also for estimating these characteristics for attendees of certain events, advisory groups, or other activities for which you have a list of named attendees.

However, we generally advise against relying on these methods except as either a supplementary tool to be used alongside normal data collection methods or as a stopgap in the face of serious short-term data deficiencies. You should prioritise enhancing your ordinary data collection tools so that they give you a robust picture of your members.

---

[30] Research into how free R packages perform at prediction find that they are far better than random chance at predicting ethnicity and gender.

From past to present...

The image of the wheatsheaf first appeared in our original seal. Being the end product of the harvesting and bundling of wheat, it was a pictorial way of expressing the gathering and analysis of data: the foundations of statistical work.

It also implied that statistical practice comprises more than the collection of data: it consists of active interpretation and application as well (threshed for others, if the rural analogy is sustained). Rigorous data gathering is still at the heart of modern statistics, but as statisticians we also interpret, explain and present the data we collect.