

STAGE 1: KEY INFORMANTS

Introduction

- Thanks for your time. This interview should take between 45 and 60 minutes.
- Here is a brief summary of our study
 - This study seeks to understand the current skills requirements, training gaps and needs across the UK's sensitive data infrastructure in order to inform training and to develop a community of data professionals. In this first stage of the research, we will be asking key informants to suggest what questions we need to ask the wider data professionals community to gain a thorough understanding of the skills and training landscape.
 - We have come to you as you are someone with expertise and experience working with sensitive data in a Trusted Research Environment or Secure Data Environment. Your insights will help us to inform the design of a survey that will go out to a wider group of people working with sensitive data, as well as help us develop more in-depth questions for further interviews and focus groups.
- Thank you for returning the consent email - do you have any questions?
- A reminder - everything will be anonymised including details about you and others you talk about.
- Something we think is important in this study - We might use certain terms, but you might not know your work by these terms - so we might need to work through that to ensure we understand each other.

*****PRESS RECORD*****

About the participant role and workplace

- How would you describe your role, and your job title?
 - [Do these terms line up with our terminology? Negotiate.]
- Please tell me about your role and responsibilities at [your workplace] as it relates to handling sensitive data

- Do you work as part of a team or as a standalone role?

Perspectives on skills for sensitive data

DEFINITIONS

- How would you define sensitive data?
- What does 'data curation' of sensitive data mean to you?

YOUR WORK

- What type of sensitive data do you mainly work with? (Health data – Social science data)
- What type of data do you work with? (semi structured, structured (tabular), imaging, data entry)
- **When** do you need to curate data in your role?
- What methods do you apply to curate data? (Semi-automatic or manual tasks / solution / process / quality control)
- Does your organisation have a set of standard methodologies that are implemented when dealing with data?
 - Are these accredited, e.g. ISO standards
- What software are you using? (Python, R, Stata, SPSS, etc)
- Are your methods driven by a standard (like an ISO standard) or experience from your previous work?
- What are the challenges you routinely face when curating sensitive data?

YOUR WIDER TEAM

- Who in your organisation is a data steward? (everyone in the team, a specific role/person/are you the expert in a group of non-data professionals?)
- What is the division of labour within your team regarding data curation?

Thank you for providing that really useful context. Now onto a discussion of skills.

THINKING ABOUT SKILLS

- What skills are required to do data steward work? (are they different within a TRE/SDE vs elsewhere)
 - Domain knowledge?
 - Technical (digital) knowledge?
 - Soft skills, such as communication, teamworking?
 - Knowledge of processes and standards
 - Data-driven mindset? (Ability to be objective while reading the data / having the ability of supporting your decisions with data).
- Out of these skills, which do you think are the most important?
- Which are most frequently used?

(These two points are not the same: there might be a lesser-used skill that is absolutely vital).

- Thinking of these skills
 - Which of them does your team (or those in 'data professional' roles more widely) possess?
 - Which skills are missing?
 - Can you explain why this might be?
 - Do different groups of workers have different skills?
 - Different education
 - Different roles
 - Amount of time spent on data curation (proportion of FTE)
- Thinking about training
 - How do data professionals acquire and keep up-to-date on the necessary skills?
 - What training have you/your team found most beneficial for curation of sensitive data?
 - What format does this take (in-person, online)?
 - Do you draw on networks to share knowledge and improve skills?
 - Is there value in learning from other data modalities? Could there be a level of standardisation across different data modalities, or are they too different?

- o What additional training around sensitive data curation would benefit you/your team?
- o What should the format of the training be?
 - Paired programming
 - Workshops?
 - Online, etc.
- o What are the **barriers** to putting this training in place?
- Do you know any guidance and guidelines available to rely on and do you feel it keeps up with changes?

FUTURE CHALLENGES

- What upcoming advances/changes will require new or enhanced skills?
 - o How do you stay aware of upcoming changes?
- What are the different type of training that would be beneficial for meeting the needs of your work in the future? e.g., training on the growth of AI use in big data
- Are you aware of any upcoming constraints on training, such as reduced budgets, or other issues?
 - o What would the impacts be of any constraints?

BUILDING A COMMUNITY OF DATA PROFESSIONALS

- Would it benefit you to be called a data steward and join a community of data stewards?
 - o How could this help?
 - o What might this look like? (A slack channel, a conference, meetups, podcasts, etc?)
 - o Could this affect your career in adverse ways? Why?

CAREER PROGRESSION

- How do data professionals achieve recognition/career progression for their skills development?

ANYTHING ELSE?

- Are there any issues that we have not talked about regarding the skills needed for data curation of sensitive data?
- Is there anything that we should be asking the wider community?
- What would you like to know as [role] that would help you with your job?

Wrap up

- Thank you for taking part in this study. The recording we made will now be professionally transcribed, and de-identified/anonymised. We will analyse the data from our key informant interviews, which will then inform the next stages of the study, namely a survey to the wider data professional community, more interviews, and focus groups.
- Following that, we will write a report which will inform training to support people working with sensitive data. We will email you with updates on the report and the training, informing you how you can read it and find out more about the training.