Skills for the curation of sensitive data

Start of Block: Welcome

Q1 Welcome to the 'Data curation skills for sensitive data' survey 2025! Thank you in advance for taking the time to respond to this survey. This survey is part of the ’Data Curation Skills for Sensitive Data’ project— a collaboration between Health Data Research UK, the Usher Institute at the University of Edinburgh, University of Dundee, the Software Sustainability Institute at the University of Southampton, and Secure Data Access Professionals at the University of Essex. This project has received funding through the UKRI Digital Research Infrastructure Programme. The aim of this survey is to understand the skills and knowledge needed to support high-quality data curation in Trusted Research Environments (TREs) and other secure settings **in the United Kingdom**. Your input will help us map current skills and activities, identify training needs and inform the development of new, flexible training materials. This is an anonymous survey. Please make sure you do not include any identifying information in your answers. The survey should take around 10 minutes to complete; you can save your progress and return to the survey at any time before submitting your responses. As a thank you for completing the survey, you have the opportunity to enter a prize draw for one of three £50 Love2Shop vouchers. A link is provided at the end of the survey. The survey closes at 23:59 BST on Sunday, 28th September 2025.

End of Block: Welcome

Start of Block: PIS and Consent

Q2 **Combined Participant Information Sheet and Consent Form for Anonymous Online Surveys for Adult Participants** **Study Title:** Data curation skills for sensitive data **Researchers:** Ben Thomas - University of Southampton (Researcher) Simon Hettrick - University of Southampton (Project co-lead) Hollydawn Murray - Health Data Research UK (Project Lead) Sarah Cadman - Health Data Research UK (Project co-lead) Amy Tilbrook - University of Edinburgh (Project co-lead) Brittany Blankinship - University of Edinburgh (Project co-lead) Michelle Evens - University of Edinburgh (Project co-lead) Leigh Tate - University of Essex Kathy Harrison - University of Edinburgh Magalie Guignard Duff - University of Dundee **University email:** b.thomas@soton.ac.uk **Ethics/ERGO no:** 100680 **Version and date:** Version 3, 19/5/25 **What is the research about?** My name is Dr Ben Thomas and I am a research fellow at the University of Southampton. I am inviting you, on behalf of the whole research team, to participate in a study regarding skills requirements and training needs for the curation of sensitive data within Trusted Research Environments (TREs) and Secure Data Environments (SDEs). The aim of the study is to identify and address training needs through a pilot training package. This study was approved by the Faculty Research Ethics Committee (FREC) at the University of Southampton (Ethics/ERGO Number: 100680). **What will happen to me if I take part?** This study involves completing an anonymous questionnaire which should take approximately 10 minutes of

your time. You do not have to complete the survey in one go, as it will save your progress. If you are happy to complete this survey, you will need to tick (check) the box below to show your consent. As this survey is anonymous, the research team will not be able to know whether you have participated, or what answers you provided. **Why have I been asked to participate?** You have been asked to take part because you are someone who works in data curation of sensitive data in a TRE or SDE **in the United Kingdom**. We are aiming to recruit as many survey participants for this study as possible. **What information will be collected?** The questions in this survey ask for information in relation to your experiences of conducting data curation of sensitive data, and the skills that are required for this. It will also ask you questions about any training that you have had to do this work, what training you think would be helpful in doing this work in the future, and any barriers or enablers to taking part in training. There are also some demographic questions to help us understand the context of your answers and to better plan any interventions to improve training in the future. Some of the survey questions contain text boxes where you will be asked to type in your own answers. Please note that in order for this survey to be anonymous, you should not include in your answers any information from which you, or other people, could be identified.

What are the possible benefits of taking part? If you decide to take part in this study, our participation will contribute to knowledge in this area of research. In addition, you will have the opportunity to enter a prize draw for one of 3 £50 Love to Shop vouchers. To enter, you will be provided a link to a separate form so that your contact details are not linked to your survey responses. **Are there any risks involved?** It is expected that taking part in this study will not cause you any psychological discomfort and/or distress, however, should you feel uncomfortable you can leave the survey at any time. You may also consider having a support person with you while you fill this survey. If at any point you feel distressed you can also contact the mental health support services at your institution or check the available resources and support provided by Togetherall. **What will happen to the information collected?** All information collected for this study will be stored securely on a password protected computer and backed up on a secure server. In addition, all quantitative data will be pooled and only compiled into data summaries or summary reports. Anonymous qualitative textbox data may be quoted in reporting. The information collected will be analysed and written up in a project report, which will be used to inform the development of a training package. The University of Southampton conducts research to the highest standards of ethics and research integrity. In accordance with our Research Data Management Policy, data will be held for 10 years after the study has finished when it will be securely destroyed. **What happens if there is a problem?** If you are unhappy about any aspect of this study and would like to make a formal complaint, you can contact the Head of Research Integrity and Governance, University of Southampton, on the following contact details: Email: rgoinfo@soton.ac.uk, phone: + 44 2380 595058. Please quote the Ethics/ERGO number above. Please note that by making a complaint you might be no longer anonymous. More information on your rights as a study participant is available via this link. **Thank you for reading this information sheet and considering taking part in this research.**

Q3 Please tick (check) this box to indicate that you have read and understood information on this form, are aged 18 or over and agree to take part in this survey.

○ I agree (1)

End of Block: PIS and Consent

Start of Block: Definitions

Q4 **What do we mean by sensitive data and data curation in this survey?** In this survey, sensitive data refers to confidential information that, if disclosed or accessed without authorisation, could lead to harm or adverse consequences for individuals. This includes personal data such as health information and other special category data as in GDPR , but also individual's financial details, or commercially sensitive data. You can find the full GDPR definition here. We also include any data stored or processed within a Trusted Research Environment (TRE). A TRE is any digital system—or group of systems—designed to securely manage sensitive research data. Other terms for these environments include Secure Data Environment (SDE), data safe haven, and similar. Examples of sensitive data typically managed in these environments include: genomic or genetic data, medical imaging and electronic health records, commercial or financial data, administrative data (e.g. tax, benefits, or social care), education data (e.g. school or university records), and more. **This survey is aimed at people involved in curating or managing sensitive data within these types of environments.** Data curation refers to the process of managing and preparing data so that it is accurate, well-documented, and suitable for analysis — often described as making data “research-ready” or “analysis-ready”. This may involve cleaning, transforming, integrating, documenting, and safeguarding data throughout its lifecycle. Please note, this is our working definition for the purposes of this survey. We recognise that data curation is a complex and evolving practice, and definitions may vary across contexts. **If you consider any part of your work to involve sensitive data curation, we encourage you to take part.**

End of Block: Definitions

Start of Block: Info about participant

Q6 What is your job title?

_____

Q5 What are your responsibilities? Select all that apply.

☐ Strategic or senior management (25)

☐ Managing a team (26)

- [ ] Managing projects (27)
- [ ] Making decisions on strategy for a TRE/SDE (28)
- [ ] Making decisions on how to handle data (e.g., free text or image data) (29)
- [ ] Liaising with organisations involved with the data source (30)
- [ ] Liaising with researchers and/or other clients (31)
- [ ] Requirements gathering (32)
- [ ] Information governance (33)
- [ ] Ensuring adherence to organisational standards (e.g., ISO27001) (34)
- [ ] Gaining information governance approvals (35)
- [ ] Arranging data sharing agreements (36)
- [ ] Disclosure checking (input and output checking) (37)
- [ ] Dealing with data protection incidents and breaches (38)
- [ ] Data acquisition (39)
- [ ] Transferring data (40)
- [ ] Data cleaning (41)
- [ ] Linking data (42)

- [ ] De-identifying, pseudonymisation, and anonymising data (43)
- [ ] Data minimisation (44)
- [ ] Quality assurance checks (45)
- [ ] Gathering or producing documentation of data curation process (46)
- [ ] Creating extracts of data (for researchers within a TRE/SDE) (47)
- [ ] Gathering or producing metadata (48)
- [ ] Research and clinical activity (49)

Q7 How many years' experience do you have working in and around sensitive data curation?

- ○ (1)
- ○ 1-5 (2)
- ○ 6-10 (3)
- ○ 11-15 (4)
- ○ >15. (5)

Q8 What do you consider to be your career stage?

- ○ Phase 1 - Junior (Trainee) (1)
- ○ Phase 2 - Early (2)

- ○ Phase 3 - Mid/Recognised (3)
- ○ Phase 4 - Established (4)

Q9 What sector do you work in?

- ☐ Academic institution (1)
- ☐ Government body (4)
- ☐ Non-profit sector/charity (5)
- ☐ Private sector/industry (6)
- ☐ Other (please specify) (7)

_____

Q10 What types of data do you work with/support? Select all that apply.

- ☐ Health data (1)
- ☐ Education data (2)
- ☐ Social security data (3)
- ☐ Census data (4)
- ☐ Geospatial data (5)
- ☐ Biological data (6)

- [ ] Social attitudes data (7)
- [ ] Business/economic data (8)
- [ ] Commercial data (9)
- [ ] Other (please state) (10)

_____

Q11 What types or formats of data do you work with? Select all that apply.

- [ ] Textual data (1)
- [ ] Numeric data (2)
- [ ] Device data (3)
- [ ] Image data (4)
- [ ] Categorical data (e.g., genomic) (5)
- [ ] Geospatial (6)
- [ ] Other (please state) (7)

_____

Q12 In your workplace, which of the following aspects do you consider part of the curation of sensitive data, as distinct from general data curation? Select all that apply.

- [ ] Legal and/or information governance complexity (1)

- [ ] Application of de-identification or pseudonymisation processes (4)
- [ ] Attention to the quality of the curated data (5)
- [ ] Data minimisation practices (6)
- [ ] Use of secured, restricted environments regulated by Information Security Management System (ISMS) standards (7)
- [ ] Other (please state) (8)

_____

End of Block: Info about participant

Start of Block: Skills

Q13 In your experience, what skills are required for doing data curation work?

- [ ] Awareness of and ability to interpret data within its external and contextual influences (1)
- [ ] Domain knowledge (2)
- [ ] Knowledge of specific datasets (3)
- [ ] Knowledge of standard data models, ontologies, and vocabularies (e.g., OMOP, DPV, ODRL) (4)
- [ ] Understanding of changes in technology that can affect methodologies (5)
- [ ] Understanding of how terminology is used within a certain context (6)
- [ ] Understanding of researchers' methods (7)

- [ ] Ability to do disclosure checks (also known as output checking) (8)
- [ ] Knowledge of legal obligations (e.g., GDPR, disclosure control) (9)
- [ ] Knowledge of organisational processes and standards (e.g., ISO27001) (10)
- [ ] Understanding of anonymisation processes (11)
- [ ] Ability to adapt to diverse data sources, formats, and types — including working with linked data (12)
- [ ] Ability to apply a methodological approach for data cleaning and preparation (e.g. identifying and correcting errors, handling missing or inconsistent values, standardising formats, and transforming data to ensure it is accurate, usable, and analysis-ready) (13)
- [ ] Ability to apply AI tools in data curation, with an understanding of their value and limitations in a TRE/SDE context (14)
- [ ] Ability to apply statistical analyses using available software tools and packages (15)
- [ ] Ability to structure data curation work so that it is reproducible, transparent, and reusable (including clear documentation, version control, and workflows that reliably reference the same underlying data) (16)
- [ ] Understanding of how to combine data from different sources and formats/types for use by researchers (17)
- [ ] Understanding of how to write code for data curation (applying domain knowledge to programming) (18)
- [ ] Understanding of programming concepts (knowing languages, able to learn new languages) (19)

- [ ] Understanding of TRE/SDE infrastructure literacy (including how secure environments are technically configured, how permissions and access controls are managed, and how infrastructure design impacts data access and workflows) (20)
- [ ] Ability to create a searchable data catalogue (21)
- [ ] Ability to create metadata (e.g. a data dictionary) (22)
- [ ] Ability to translate requirements into a workflow to achieve desired outputs (23)
- [ ] Ability to write documentation describing the process of curation (24)
- [ ] Confidence to question data when it does not look right (25)
- [ ] Understanding of how to ask questions of data that help in addressing research aims (26)
- [ ] Understanding of how to use data to achieve a goal (27)
- [ ] Understanding of what is feasible within constraints - understanding data and communicating limitations of data with clients (28)
- [ ] Ability to build trust with clients/researchers (29)
- [ ] Ability to support researchers (30)
- [ ] Ability to understand what researchers need to get out of the data (31)
- [ ] Ability to work collaboratively with clients/researchers (32)
- [ ] Customer service skills (33)

- [ ] Ability to do effective knowledge exchange (34)
- [ ] Ability to present to different different audiences (35)
- [ ] Ability to ask for help (36)
- [ ] Ability to learn on the job (37)
- [ ] Ability to problem-solve (38)
- [ ] Ability to project manage (39)
- [ ] Time/Task management (for own work or that of a team) (40)
- [ ] Ability to manage people (41)
- [ ] Ability to work in a team (42)
- [ ] Other (43) _____

Carry Forward Selected Choices from "Q13"

Q14 Of the skills you have selected, select the **top five** most important to your role.

- [ ] Awareness of and ability to interpret data within its external and contextual influences (1)
- [ ] Domain knowledge (2)
- [ ] Knowledge of specific datasets (3)

- [ ] Knowledge of standard data models, ontologies, and vocabularies (e.g., OMOP, DPV, ODRL) (4)
- [ ] Understanding of changes in technology that can affect methodologies (5)
- [ ] Understanding of how terminology is used within a certain context (6)
- [ ] Understanding of researchers' methods (7)
- [ ] Ability to do disclosure checks (also known as output checking) (8)
- [ ] Knowledge of legal obligations (e.g., GDPR, disclosure control) (9)
- [ ] Knowledge of organisational processes and standards (e.g., ISO27001) (10)
- [ ] Understanding of anonymisation processes (11)
- [ ] Ability to adapt to diverse data sources, formats, and types — including working with linked data (12)
- [ ] Ability to apply a methodological approach for data cleaning and preparation (e.g. identifying and correcting errors, handling missing or inconsistent values, standardising formats, and transforming data to ensure it is accurate, usable, and analysis-ready) (13)
- [ ] Ability to apply AI tools in data curation, with an understanding of their value and limitations in a TRE/SDE context (14)
- [ ] Ability to apply statistical analyses using available software tools and packages (15)
- [ ] Ability to structure data curation work so that it is reproducible, transparent, and reusable (including clear documentation, version control, and workflows that reliably reference the same underlying data) (16)

- [ ] Understanding of how to combine data from different sources and formats/types for use by researchers (17)
- [ ] Understanding of how to write code for data curation (applying domain knowledge to programming) (18)
- [ ] Understanding of programming concepts (knowing languages, able to learn new languages) (19)
- [ ] Understanding of TRE/SDE infrastructure literacy (including how secure environments are technically configured, how permissions and access controls are managed, and how infrastructure design impacts data access and workflows) (20)
- [ ] Ability to create a searchable data catalogue (21)
- [ ] Ability to create metadata (e.g. a data dictionary) (22)
- [ ] Ability to translate requirements into a workflow to achieve desired outputs (23)
- [ ] Ability to write documentation describing the process of curation (24)
- [ ] Confidence to question data when it does not look right (25)
- [ ] Understanding of how to ask questions of data that help in addressing research aims (26)
- [ ] Understanding of how to use data to achieve a goal (27)
- [ ] Understanding of what is feasible within constraints - understanding data and communicating limitations of data with clients (28)
- [ ] Ability to build trust with clients/researchers (29)

- [ ] Ability to support researchers (30)
- [ ] Ability to understand what researchers need to get out of the data (31)
- [ ] Ability to work collaboratively with clients/researchers (32)
- [ ] Customer service skills (33)
- [ ] Ability to do effective knowledge exchange (34)
- [ ] Ability to present to different different audiences (35)
- [ ] Ability to ask for help (36)
- [ ] Ability to learn on the job (37)
- [ ] Ability to problem-solve (38)
- [ ] Ability to project manage (39)
- [ ] Time/Task management (for own work or that of a team) (40)
- [ ] Ability to manage people (41)
- [ ] Ability to work in a team (42)
- [ ] Other (43) _____

Carry Forward Selected Choices from "Q13"

Q15 Are any of these skills unique to, or particularly important in, sensitive data curation?

- [ ] Awareness of and ability to interpret data within its external and contextual influences (1)
- [ ] Domain knowledge (2)
- [ ] Knowledge of specific datasets (3)
- [ ] Knowledge of standard data models, ontologies, and vocabularies (e.g., OMOP, DPV, ODRL) (4)
- [ ] Understanding of changes in technology that can affect methodologies (5)
- [ ] Understanding of how terminology is used within a certain context (6)
- [ ] Understanding of researchers' methods (7)
- [ ] Ability to do disclosure checks (also known as output checking) (8)
- [ ] Knowledge of legal obligations (e.g., GDPR, disclosure control) (9)
- [ ] Knowledge of organisational processes and standards (e.g., ISO27001) (10)
- [ ] Understanding of anonymisation processes (11)
- [ ] Ability to adapt to diverse data sources, formats, and types — including working with linked data (12)
- [ ] Ability to apply a methodological approach for data cleaning and preparation (e.g. identifying and correcting errors, handling missing or inconsistent values, standardising formats, and transforming data to ensure it is accurate, usable, and analysis-ready) (13)
- [ ] Ability to apply AI tools in data curation, with an understanding of their value and limitations in a TRE/SDE context (14)

- [ ] Ability to apply statistical analyses using available software tools and packages (15)
- [ ] Ability to structure data curation work so that it is reproducible, transparent, and reusable (including clear documentation, version control, and workflows that reliably reference the same underlying data) (16)
- [ ] Understanding of how to combine data from different sources and formats/types for use by researchers (17)
- [ ] Understanding of how to write code for data curation (applying domain knowledge to programming) (18)
- [ ] Understanding of programming concepts (knowing languages, able to learn new languages) (19)
- [ ] Understanding of TRE/SDE infrastructure literacy (including how secure environments are technically configured, how permissions and access controls are managed, and how infrastructure design impacts data access and workflows) (20)
- [ ] Ability to create a searchable data catalogue (21)
- [ ] Ability to create metadata (e.g. a data dictionary) (22)
- [ ] Ability to translate requirements into a workflow to achieve desired outputs (23)
- [ ] Ability to write documentation describing the process of curation (24)
- [ ] Confidence to question data when it does not look right (25)
- [ ] Understanding of how to ask questions of data that help in addressing research aims (26)
- [ ] Understanding of how to use data to achieve a goal (27)

- [ ] Understanding of what is feasible within constraints - understanding data and communicating limitations of data with clients (28)
- [ ] Ability to build trust with clients/researchers (29)
- [ ] Ability to support researchers (30)
- [ ] Ability to understand what researchers need to get out of the data (31)
- [ ] Ability to work collaboratively with clients/researchers (32)
- [ ] Customer service skills (33)
- [ ] Ability to do effective knowledge exchange (34)
- [ ] Ability to present to different different audiences (35)
- [ ] Ability to ask for help (36)
- [ ] Ability to learn on the job (37)
- [ ] Ability to problem-solve (38)
- [ ] Ability to project manage (39)
- [ ] Time/Task management (for own work or that of a team) (40)
- [ ] Ability to manage people (41)
- [ ] Ability to work in a team (42)
- [ ] Other (43) _____

Carry Forward Selected Choices from "Q13"

Q16 Thinking about yourself, **select up to five skills** you feel least well-equipped with for your current role.

- [ ] Awareness of and ability to interpret data within its external and contextual influences (1)
- [ ] Domain knowledge (2)
- [ ] Knowledge of specific datasets (3)
- [ ] Knowledge of standard data models, ontologies, and vocabularies (e.g., OMOP, DPV, ODRL) (4)
- [ ] Understanding of changes in technology that can affect methodologies (5)
- [ ] Understanding of how terminology is used within a certain context (6)
- [ ] Understanding of researchers' methods (7)
- [ ] Ability to do disclosure checks (also known as output checking) (8)
- [ ] Knowledge of legal obligations (e.g., GDPR, disclosure control) (9)
- [ ] Knowledge of organisational processes and standards (e.g., ISO27001) (10)
- [ ] Understanding of anonymisation processes (11)
- [ ] Ability to adapt to diverse data sources, formats, and types — including working with linked data (12)

- [ ] Ability to apply a methodological approach for data cleaning and preparation (e.g. identifying and correcting errors, handling missing or inconsistent values, standardising formats, and transforming data to ensure it is accurate, usable, and analysis-ready) (13)
- [ ] Ability to apply AI tools in data curation, with an understanding of their value and limitations in a TRE/SDE context (14)
- [ ] Ability to apply statistical analyses using available software tools and packages (15)
- [ ] Ability to structure data curation work so that it is reproducible, transparent, and reusable (including clear documentation, version control, and workflows that reliably reference the same underlying data) (16)
- [ ] Understanding of how to combine data from different sources and formats/types for use by researchers (17)
- [ ] Understanding of how to write code for data curation (applying domain knowledge to programming) (18)
- [ ] Understanding of programming concepts (knowing languages, able to learn new languages) (19)
- [ ] Understanding of TRE/SDE infrastructure literacy (including how secure environments are technically configured, how permissions and access controls are managed, and how infrastructure design impacts data access and workflows) (20)
- [ ] Ability to create a searchable data catalogue (21)
- [ ] Ability to create metadata (e.g. a data dictionary) (22)
- [ ] Ability to translate requirements into a workflow to achieve desired outputs (23)
- [ ] Ability to write documentation describing the process of curation (24)

- [ ] Confidence to question data when it does not look right (25)
- [ ] Understanding of how to ask questions of data that help in addressing research aims (26)
- [ ] Understanding of how to use data to achieve a goal (27)
- [ ] Understanding of what is feasible within constraints - understanding data and communicating limitations of data with clients (28)
- [ ] Ability to build trust with clients/researchers (29)
- [ ] Ability to support researchers (30)
- [ ] Ability to understand what researchers need to get out of the data (31)
- [ ] Ability to work collaboratively with clients/researchers (32)
- [ ] Customer service skills (33)
- [ ] Ability to do effective knowledge exchange (34)
- [ ] Ability to present to different different audiences (35)
- [ ] Ability to ask for help (36)
- [ ] Ability to learn on the job (37)
- [ ] Ability to problem-solve (38)
- [ ] Ability to project manage (39)

- [ ] Time/Task management (for own work or that of a team) (40)
- [ ] Ability to manage people (41)
- [ ] Ability to work in a team (42)
- [ ] Other (43) _____

Carry Forward Selected Choices from "Q16"

Q17 Which of these skills do you expect to improve by yourself, and which need additional training?

Improve by myself	Need additional training
_____ Awareness of and ability to interpret data within its external and contextual influences (xx1)	_____ Awareness of and ability to interpret data within its external and contextual influences (xx1)
_____ Domain knowledge (xx2)	_____ Domain knowledge (xx2)
_____ Knowledge of specific datasets (xx3)	_____ Knowledge of specific datasets (xx3)
_____ Knowledge of standard data models, ontologies, and vocabularies (e.g., OMOP, DPV, ODRL) (xx4)	_____ Knowledge of standard data models, ontologies, and vocabularies (e.g., OMOP, DPV, ODRL) (xx4)
_____ Understanding of changes in technology that can affect methodologies (xx5)	_____ Understanding of changes in technology that can affect methodologies (xx5)
_____ Understanding of how terminology is used within a certain context (xx6)	_____ Understanding of how terminology is used within a certain context (xx6)
_____ Understanding of researchers' methods (xx7)	_____ Understanding of researchers' methods (xx7)
_____ Ability to do disclosure checks (also known as output checking) (xx8)	_____ Ability to do disclosure checks (also known as output checking) (xx8)
_____ Knowledge of legal obligations (e.g., GDPR, disclosure control) (xx9)	_____ Knowledge of legal obligations (e.g., GDPR, disclosure control) (xx9)

_____ Knowledge of organisational processes and standards (e.g., ISO27001) (xx10)

_____ Understanding of anonymisation processes (xx11)

_____ Ability to adapt to diverse data sources, formats, and types — including working with linked data (xx12)

_____ Ability to apply a methodological approach for data cleaning and preparation (e.g. identifying and correcting errors, handling missing or inconsistent values, standardising formats, and transforming data to ensure it is accurate, usable, and analysis-ready) (xx13)

_____ Ability to apply AI tools in data curation, with an understanding of their value and limitations in a TRE/SDE context (xx14)

_____ Ability to apply statistical analyses using available software tools and packages (xx15)

_____ Ability to structure data curation work so that it is reproducible, transparent, and reusable (including clear documentation, version control, and workflows that reliably reference the same underlying data) (xx16)

_____ Understanding of how to combine data from different sources and formats/types for use by researchers (xx17)

_____ Understanding of how to write code for data curation (applying domain knowledge to programming) (xx18)

_____ Understanding of programming concepts (knowing languages, able to learn new languages) (xx19)

_____ Knowledge of organisational processes and standards (e.g., ISO27001) (xx10)

_____ Understanding of anonymisation processes (xx11)

_____ Ability to adapt to diverse data sources, formats, and types — including working with linked data (xx12)

_____ Ability to apply a methodological approach for data cleaning and preparation (e.g. identifying and correcting errors, handling missing or inconsistent values, standardising formats, and transforming data to ensure it is accurate, usable, and analysis-ready) (xx13)

_____ Ability to apply AI tools in data curation, with an understanding of their value and limitations in a TRE/SDE context (xx14)

_____ Ability to apply statistical analyses using available software tools and packages (xx15)

_____ Ability to structure data curation work so that it is reproducible, transparent, and reusable (including clear documentation, version control, and workflows that reliably reference the same underlying data) (xx16)

_____ Understanding of how to combine data from different sources and formats/types for use by researchers (xx17)

_____ Understanding of how to write code for data curation (applying domain knowledge to programming) (xx18)

_____ Understanding of programming concepts (knowing languages, able to learn new languages) (xx19)

_____ Understanding of TRE/SDE infrastructure literacy (including how secure environments are technically configured, how permissions and access controls are managed, and how infrastructure design impacts data access and workflows) (xx20)

_____ Ability to create a searchable data catalogue (xx21)

_____ Ability to create metadata (e.g. a data dictionary) (xx22)

_____ Ability to translate requirements into a workflow to achieve desired outputs (xx23)

_____ Ability to write documentation describing the process of curation (xx24)

_____ Confidence to question data when it does not look right (xx25)

_____ Understanding of how to ask questions of data that help in addressing research aims (xx26)

_____ Understanding of how to use data to achieve a goal (xx27)

_____ Understanding of what is feasible within constraints - understanding data and communicating limitations of data with clients (xx28)

_____ Ability to build trust with clients/researchers (xx29)

_____ Ability to support researchers (xx30)

_____ Ability to understand what researchers need to get out of the data (xx31)

_____ Ability to work collaboratively with clients/researchers (xx32)

_____ Customer service skills (xx33)

_____ Ability to do effective knowledge exchange (xx34)

_____ Understanding of TRE/SDE infrastructure literacy (including how secure environments are technically configured, how permissions and access controls are managed, and how infrastructure design impacts data access and workflows) (xx20)

_____ Ability to create a searchable data catalogue (xx21)

_____ Ability to create metadata (e.g. a data dictionary) (xx22)

_____ Ability to translate requirements into a workflow to achieve desired outputs (xx23)

_____ Ability to write documentation describing the process of curation (xx24)

_____ Confidence to question data when it does not look right (xx25)

_____ Understanding of how to ask questions of data that help in addressing research aims (xx26)

_____ Understanding of how to use data to achieve a goal (xx27)

_____ Understanding of what is feasible within constraints - understanding data and communicating limitations of data with clients (xx28)

_____ Ability to build trust with clients/researchers (xx29)

_____ Ability to support researchers (xx30)

_____ Ability to understand what researchers need to get out of the data (xx31)

_____ Ability to work collaboratively with clients/researchers (xx32)

_____ Customer service skills (xx33)

_____ Ability to do effective knowledge exchange (xx34)

_____ Ability to present to different different audiences (xx35)

_____ Ability to ask for help (xx36)

_____ Ability to learn on the job (xx37)

_____ Ability to problem-solve (xx38)

_____ Ability to project manage (xx39)

_____ Time/Task management (for own work or that of a team) (xx40)

_____ Ability to manage people (xx41)

_____ Ability to work in a team (xx42)

_____ Other (xx43)

_____ Ability to present to different different audiences (xx35)

_____ Ability to ask for help (xx36)

_____ Ability to learn on the job (xx37)

_____ Ability to problem-solve (xx38)

_____ Ability to project manage (xx39)

_____ Time/Task management (for own work or that of a team) (xx40)

_____ Ability to manage people (xx41)

_____ Ability to work in a team (xx42)

_____ Other (xx43)

Q18 Are there any skills you do not currently need but expect to require in the next 1–2 years? This might be in relation to AI tools in curation workflows, or for cross TRE collaboration.

_____

_____

_____

_____

_____

End of Block: Skills

Start of Block: Training

Q19 What qualifications and experience have you gained prior to taking up work in your current organisation? Select all that apply.

- [ ] High/secondary school qualifications at 16 (1)
- [ ] High/secondary school qualifications at 18 (4)

- [ ] Undergraduate university degree (5)
- [ ] Postgraduate university degree (6)
- [ ] Doctoral degree (7)
- [ ] Postdoctoral fellowship (8)
- [ ] Internship (9)
- [ ] Professional qualification (please state) (10)

_____

- [ ] Other (please state) (11)

_____

Carry Forward Selected Choices from "Q19"

Q20 Which of these qualifications provided you with some of the skills you have needed in your data curation work? Select all that apply.

- [ ] High/secondary school qualifications at 16 (1)
- [ ] High/secondary school qualifications at 18 (2)
- [ ] Undergraduate university degree (3)
- [ ] Postgraduate university degree (4)
- [ ] Doctoral degree (5)
- [ ] Postdoctoral fellowship (6)

- [ ] Internship (7)
- [ ] Professional qualification (please state) (8)

_____

- [ ] Other (please state) (9)

_____

Q21 In your role, is training in data curation available?

- ○ Yes, extensive training is available (1)
- ○ Yes, some training is available (2)
- ○ Only informal or on-the-job training is available (3)
- ○ No, training in data curation is not available (4)

Skip To: End of Block If Q21 = No, training in data curation is not available

Q22 What formats is this training available in? Select all that apply

- [ ] Self-led Online (1)
- [ ] In-house (TRE/SDE) training (2)
- [ ] University-led training (3)
- [ ] Funder-led training (4)
- [ ] External courses (5)

- [ ] Communities of practice (6)
- [ ] Through attending conferences (7)
- [ ] Informally through networks of data professionals (8)
- [ ] Other format (please state) (9)

_____

Q23 How is training funded?

- [ ] Organisation (1)
- [ ] Grant and/or bursary money (2)
- [ ] Self-funded (3)
- [ ] Other (4) _____
- [ ] No funding (5)

End of Block: Training

Start of Block: Barriers

Q24 What barriers exist to you taking part in skills training? Select all that apply.

- [ ] No time to train during working hours (1)
- [ ] Money to pay for training (4)
- [ ] No organisational support (5)

- [ ] Unsure what skills I need (6)
- [ ] Lack of appropriate courses for training needs (7)
- [ ] Structural barriers to do with my circumstance and/or aspects of my identity (9)
- [ ] Other (please state) (10)

_____

Display this question:

If Q24 = Structural barriers to do with my circumstance and/or aspects of my identity

Q25 What structural barriers do you face? Select all that apply.

- [ ] Language barriers (8)
- [ ] Limited access to location or remote work (9)
- [ ] Caregiving or other personal responsibilities/commitments (10)
- [ ] Disability-related barriers such as accessibility (11)
- [ ] Discrimination due to aspects of my identity (12)
- [ ] Other (please state) (13)
- [ ] Prefer not to say (14)

Display this question:

If Q25 = Discrimination due to aspects of my identity

Q26 Which aspects of your identity are discriminated against?

- [ ] Gender identity (1)
- [ ] Sexual identity (2)
- [ ] Race/ethnicity (3)
- [ ] Age (4)
- [ ] Class identity (5)
- [ ] Other (please state) (6)

_____

- [ ] Prefer not to say (7)

Carry Forward Selected Choices from "Q24"

Q27 Please rank the barriers you have chosen, with 1st being the biggest barrier.

_____ No time to train during working hours (1)
_____ Money to pay for training (2)
_____ No organisational support (3)
_____ Unsure what skills I need (4)
_____ Lack of appropriate courses for training needs (5)
_____ Structural barriers to do with my circumstance and/or aspects of my identity (6)
_____ Other (please state) (7)

Page Break

Q28 Would better training in data curation support help your career prospects?

- ○ Yes (1)
- ○ No (2)

- ○ Not sure (3)

Q29 Would you be interested in a structured skills framework or professional pathway for sensitive data curation roles?

- ○ Yes (1)
- ○ No (2)
- ○ Not sure (3)

End of Block: Barriers

Start of Block: Wrap up

Q30 Thank you for taking the time to complete the survey! As a thank you, we invite you to enter a prize draw for one of three £50 Love2Shop vouchers. Click here to enter.

End of Block: Wrap up