

# Focus Group Planning: Objectives and Questions

## Objective

To build on survey findings and inform development of new training by identifying (1) which skills gaps to prioritise and (2) how training should be designed and delivered.

## Structure

Two workshops for each audience (ECRs and established researchers):

Workshop 1: Prioritising skills gaps

Workshop 2: Training landscape and design

Each workshop lasts approximately 90 minutes.

## Workshop 1: Prioritising skills gaps

### Objective

To explore why key skills gaps exist, validate which should be prioritised for training, and understand views on the low priority given to AI in data curation.

### Key survey findings to reference

- Most respondents have received only informal or on-the-job training in data curation.
- The largest barrier to training is the *lack of appropriate courses*.
- Top general skills identified: Knowledge of legal obligations; Methodological approach to data cleaning; Ability to problem solve
- Skills seen as unique to sensitive data curation: knowledge of legal obligations, anonymisation processes, and disclosure control/output checking.
- Use of AI in data curation was ranked lowest in importance.

*<multivariate analyses quick insights added 27 Oct>*

- TRE/SDE literacy: only skill listed as a *gap across all career stages*.

- Knowledge of legal obligations: only skill ranked *top 5 in importance* across all stages.
- Domain knowledge (mid-career): *only overlap* between a top 5 skill and a skill gap.
- Knowledge of specific datasets: top 5 *gap* for *ECR* and *mid-career*.
- Knowledge of standard data models, ontologies, vocabularies: top 5 *gap* for *mid-career* and *established*.
  
- Knowledge of legal obligations: top 5 skill across all sectors (except private).
  - Only listed as a gap in the academic sector.
- Confidence to question data: top 5 skill across all sectors (except private).
- Domain knowledge: common top 5 skill across sectors.
  - Also a gap for the non-profit sector.
- Private sector: unique focus on
  - Anonymisation: top 5 skill and gap.
  - TRE/SDE literacy: top 5 skill and gap.
  
- Knowledge of legal obligations: top 5 skill across all data types.
- Common shared gaps:
  - Knowledge of specific datasets
  - Domain knowledge
  - Standard data models / ontologies / vocabularies
- Census data: Knowledge of specific datasets: top 5 skill and gap.
- Commercial data: TRE/SDE literacy: top 5 skill and gap.

## **Data to collect**

- Which skills researchers see as highest priority for training investment.
- Why legal, anonymisation, and disclosure skills remain challenging.
- Why AI in curation ranks low—lack of use, lack of relevance, or lack of awareness.

- Differences in priorities or skill gaps between ECRs and established team members.
- How informal learning currently operates and where it fails.

## **Discussion prompts**

### **1. Informal learning**

- How do you usually learn about curating sensitive data?
- What works well about learning on the job, and what doesn't?
- When has informal learning been insufficient or risky?

### **2. Persistent gaps**

- The survey showed legal knowledge and anonymisation as key skills but also major gaps. Why do you think these are still difficult areas?
- How do you currently access advice on disclosure control or legal compliance?
- Are these skills viewed as specialist or everyone's responsibility?

### **3. AI in curation**

- The survey showed AI ranked lowest in importance—why might that be?
- Is AI simply not part of your workflow yet, or is it seen as unnecessary or too complex?
- Do you expect AI to become relevant to your work in the next few years?

### **4. Prioritisation**

- If training could focus on only one skill, which should come first?
- What benefits would improving this skill bring for your work, team, or institution?

- How might priorities differ between ECRs and senior team members?
- How might priorities differ across contexts?

## 5. Looking ahead

- What kinds of support or resources would help close these skills gaps most effectively?

## Workshop 2: Training landscape and design

### Objective

To understand current access to training, barriers to participation, and preferences for training content, format, and delivery.

### Key survey findings to reference

- The largest barrier is a *lack of appropriate courses* to meet training needs.
- Most researchers rely on informal or on-the-job training.
- Researchers value problem solving, legal knowledge, and the ability to learn on the job.

### Data to collect

- What training currently exists and how effective it is.
- Specific reasons existing courses are viewed as inappropriate.
- Main barriers to accessing or completing training.
- Preferences for delivery format, duration, and level.
- Motivations and incentives for engaging with training.
- Differences in training needs between ECRs and established researchers.

### Discussion prompts

## **1. Current training landscape**

- What formal or structured training on sensitive data and curation is available to you?
- How relevant or practical have you found it?
- How does it compare with what you've learned informally?

## **2. Barriers to access**

- What prevents you or others from taking part in training—time, cost, relevance, or awareness?
- When researchers say “lack of appropriate courses,” what makes a course inappropriate—content, level, or accessibility?
- Are there institutional or cultural factors that discourage participation?

## **3. Designing better training**

- If we were to create new training on legal obligations, anonymisation, or disclosure control, what format would work best?
- Would you prefer short online modules, workshops, mentoring, or integrated sessions within your day job/projects?
- Would practical case studies or scenarios make training more relevant?

## **4. Motivation and incentives**

- What would make you more likely to engage in training—certification, recognition, compliance requirements?
- How could training be incorporated into existing development or supervision processes?
- How could we make sure training remains relevant over time?

## **5. Career-stage differences**

- (ECRs) What support from supervisors or departments would help you access training more easily?
- (Established) What would make senior staff or PIs more likely to participate or encourage their teams to do so?
- Should advanced or AI-related topics be introduced now or only when use cases emerge?

## **Expected outputs**

1. A prioritised list of skills to address in new training (e.g., legal obligations, anonymisation, disclosure control).
2. Insight into why informal learning dominates and where it is insufficient.
3. Understanding of barriers to accessing training and what “appropriate” training would look like.
4. Preferences for training formats, duration, and delivery.
5. Motivators for participation and ways to embed training institutionally.
6. Context for the low prioritisation of AI in data curation and expectations for future relevance.