The Safe Data Access Professionals (SDAP) Competency Framework

Prepared by:

Carlotta Greci (Health Foundation), Richard Welpton (Cancer Research UK), Christine Woods (UK Data Archive, University of Essex)

Introduction

More researchers and analysts are demanding access to more detailed data about individuals and organisations than ever before. The benefits for analysis and research are obvious: more robust and innovative analyses can be undertaken using data of such a detailed nature; analysts and researchers can explore themes and delve into strands never before uncovered. This work better supports policy-makers who use the results to plan and create policy.

Such data are now routinely accessed. However, the level of detail is such that these data are considered 'personal, sensitive, confidential' data (organisations differ in how they formally apply the definition: for example, the Office for National Statistics applies the label 'Controlled'; Public Health England consider such data to be 'Personally Identifiable'). As such, access to these data are typically made available in a 'Safe Setting' (also known as a Safe Haven, Research Data Centre, or Secure Data Environment) to ensure the confidentiality of the data is preserved.

Researchers and analysts (users) can access the Safe Setting to view and analyse the data. The results of their analyses are returned to them, subject to a review of Statistical Disclosure Control by service staff, to make sure the results do not reveal the identity, and/or, confidential information about a data subject.

A number of Safe Settings now exist, and include the Office for National Statistics (ONS) Virtual Microdata Laboratory (VML) or Secure Research Service (SRS), the UK Data Service Secure Lab and the HM Revenue and Customs (HMRC) Datalab. Charities including The Health Foundation and Cancer Research UK have established their own Safe Settings in order to acquire confidential data for use by their analysts[1].

The number of staff employed by these services has grown, particularly since the VML was launched in 2004. Staff deal with a variety of issues, including (but not limited to):

- Training users about how to use the Safe Setting
- Undertaking SDC of outputs

[1] Throughout this document, 'analyst' and 'researcher' may be used interchangeably. Generally, some organisations will provide a Safe Setting for their 'analysts'; 'researchers' may be accessing a Safe Setting not attached to their university, but which is provided as a 'service'. Either way, the term refers to somebody accessing confidential/sensitive data for statistical purposes.

- Day-to-day management of users

The Working Group for Safe Data Access Professionals (SDAP) network was established in 2011 to bring about staff working in these services, as a way to share experience and learn from each other. In late 2016, the group pooled knowledge of all the activities staff are engaged in, and considered how these could be described in a 'Competency Framework'. This work recognised that while staff gain substantial experience and expertise when working in these services, there is often little in the way of formal professional development. As described at the 2005 IASSIST conference session 'Discovering a Profession: the Accidental Data Librarian',[2] staff working in Safe Settings often find themselves in their positions through curiosity, accident, selection beyond their control, or a combination of these reasons. As such, they may not necessarily be in a position to focus on their skills development. This is because a lack of a framework for development.

The consequences of this lack of opportunity may result in:

- High turnover of staff (as professional development is sought elsewhere)
- Loss of morale
- Variation of services offered to users
- Stagnation of services which do not keep pace with developments

This document sets out the Competency Framework as developed by the Working Group for Secure Data Access Professionals, and it is intended to guide staff working in Safe Settings explore areas of professional development that will benefit them and add value to the services they work for, and the users they serve.

About the Framework

The Competency Framework sets out the competences for staff working in Safe Settings, and how this applies at different stages of their career. Competencies are the skills, knowledge, behaviours and activities associated with effective performance.

The Framework has been designed primarily for staff development: as a tool for setting objectives, identifying strengths and areas for development, assessing achievements and performance, and preparing for promotion. Additionally, it can be used during the recruitment process e.g. in writing job descriptions and assessing interview performance. See 'Using the Framework' below for further guidance.

[2] http://iassistdata.org/conferences/archive/2005?page=5.

Note that the services offered by Safe Settings and the roles of Working Group members differ considerably and this Framework has been designed accordingly.[3] The Framework covers the range of work undertaken by staff in Safe Settings (from data acquisition to data management and through to carrying out statistical disclosure control) however the competencies that are relevant to each individual will depend on their role. For example, individuals may be involved in data acquisition in one Safe Setting, but not in another, and so the competencies associated with this area will only be relevant for some.

Individuals should therefore select the elements of the Framework that apply to their role. Individuals are not expected or required to demonstrate all of the competencies within a given level.

The Framework is split into three stages/levels:

1. Beginner level – those beginning the role
2. Mid-level – those in the role for between 6 months and 2 years
3. Advanced level - those in the role for more than 2 years

Using the Framework

Staff development[4]

When beginning a new role or at the start of the performance year, we recommend that you use the Framework to identify the competencies that apply to your role and agree these with your line manager. For example, start by looking at the different competency areas (i.e. understanding analysis and analysts/researchers, data acquisition, data transformation and statistical disclosure control) and identify which of these are relevant to your role. Once you have established these, the Framework can be used in the ways shown below.

Remember that the Framework covers the wide range of activities carried out by a variety of Safe Settings. Individuals are not expected to demonstrate all of the competencies within a given level, so you should select only the competencies that apply to your role.

- Setting objectives
  Use the competencies to identify the knowledge, skills and behaviours that you wish to demonstrate throughout the year and incorporate these into your objectives for the year.

[3] This is in line with the approach adopted by other professional competency frameworks. See, for example, 'The Government Statistician Group (GSG) Competency Framework' (https://gss.civilservice.gov.uk/learning-and-development/statistician-competency-framework-2/).
[4] Parts of this section on staff development are adapted from the 'GSG Competency Framework: User Guide' (https://gss.civilservice.gov.uk/wp-content/uploads/2016/03/User-Guide.pdf).

- ### Identifying your strengths and areas for development
  Use the competencies to identify your strengths and areas for development. Think about how you can develop competencies in those areas requiring development. Competencies may be developed on-the-job (e.g. reading of relevant documents and literature, discussions with colleagues, experience of carrying out new work), as well as through the SDAP network, or more formally via training courses. Agree the learning activities with your line manager and incorporate them into your Personal Development Plan (PDP) or equivalent.

  After you have completed a learning activity, reflect on how it has contributed to your personal development and performance and record this for your next review.

- ### Assessing achievements and performance
  When preparing for your reviews, use the Framework to help identify your achievements and performance and to provide evidence of your progress.

- ### Preparing for promotion/progressing to the next stage
  When preparing for promotion or progressing to the next stage (e.g. from mid-level to advanced level) use the competencies from the next stage to set your objectives, identify any areas for development and subsequent learning activities, and to demonstrate that you are developing in line with the appropriate level.

Recruitment

The Framework can be used when producing job descriptions for roles of staff working in Safe Settings. Select the competency areas that apply to the role and use the competencies to help describe the:

- Main responsibilities of the post (e.g. provides training to analysts/r on how to use the Safe Setting)
- Experience/knowledge and skills/abilities required (e.g. experience of delivering training courses, strong communication skills)

The Framework can also be used as a structure for assessing interviewees' performance and ability. Interviewers should use the Framework to assess the extent to which candidates demonstrate the required competencies for the role. Where candidates are being assessed for a beginner role and have limited experience of working in a Safe Setting, they will develop a lot of the competencies on the job, so an assessment should be made of their potential to do so based on their previous experience and existing skills/abilities.

Competency area	Beginner level (these include skills that staff beginning the role might have, or would develop)	Mid-level (as Beginner level, plus)	Advanced level (as Mid-level, plus)
Understanding analysis and analysts	Appreciates what analysis[3] is, its aim and why some analysis requires data of different levels of detail (from previous analysis experience).	and experience of advising analysts accordingly, ensuring that analysts use data of the right detail for their project (e.g. through checking project proposals and answering data queries). Assists analysts with the application process and communicates application proposals to different stakeholders (e.g. data providers, lay members, other analysts).	Is able to identify the right data for projects, and suggest methodological implications from using data for analysis (such as advantages, pitfalls etc., data quality aspects and limitations etc.).
	Develops an understanding of how analysis project cycles work (e.g. funding cycles, team composition, how analysis project teams work, hours of work) and the impact that this may have on analysts' behaviour when using data. Works flexibly to support analysts when required.	Good understanding of how analysts work. Builds good rapport with analysts.	Is aware of funding cycles, funding council analysis priorities (and data required to answer these). Is able to advise on funding opportunities. Develops effective relationships with key stakeholders (e.g. sponsors, heads of department, analysis programmes etc.).

Understanding data legislation and licencing	Develops knowledge and understanding of: ○ the Data Access Spectrum; ○ data protection principles and other legislation, such as GDPR; and ○ the difference between personal and non-personal data and 'sensitive' versus 'non-sensitive' data sources. Trains analysts on the above (e.g. safe data access training) with support from colleagues. ○ Understand licencing requirements (inc. how they vary across Data Access Spectrum)	Is able to train analysts independently and respond to questions confidently during training sessions. Is able to use experience of job role to present real-life scenarios. Is able to contribute to development of new course materials.	Keeps up-to-date with and communicates legislative changes to colleagues, including implications for the Safe Setting. Supports management in implementing data protection when necessary (e.g. data security incidents). Leads on the development of new training resources.
	Develops knowledge of the most appropriate means of data distribution, including attending and contributing at relevant meetings (e.g. meetings with data suppliers).	and experience of advising data suppliers for less complex datasets (based on characteristics of data).	Is able to advise key stakeholders on data infrastructure, provide advice to other organisations looking to provide access to their data, and on new and complex data sources.
Understanding data	Develops familiarity with data of different types (e.g. business, social science, health, administrative, survey, population). Develops knowledge of the data held in the Safe Setting and provides support to analysts of these various data types (e.g. answers data queries, carries out SDC of analysts' outputs).	Good knowledge of the different data held in the Safe Setting, with experience of supporting analysts of these various data types (e.g. answers data queries, carries out SDC of analysts' outputs). Is able to make presentations to stakeholders about data.	Strong knowledge of the different data held in the Safe Setting, with extensive experience of supporting analysts of these various data types (e.g. answers data queries, carries out SDC of analysts' outputs). Shares knowledge across the team. Is able to derive new sources of data from existing; implement quality improvements to existing data; add value to existing sources (e.g. add commonly derived variables).

Understanding data setting	Develops an understanding of different environments for different data 'sensitivities.	Confident providing advice about the use of different data settings. Supervises safe use of data, ensuring relevant procedures are followed.	Contributes to development of the Safe Setting (e.g. best practice procedures, analyst experience, security etc.).
	Develops basic knowledge of how IT works and provides support for IT related queries.	Operation knowledge of how IT works and experience of providing support for IT related queries. Ability to communicate to IT about analyst needs, and communicate results back to analysts. Is able to advise analysts on basic IT set up. ○ Understanding of encryption techniques	Is able to provide advice to IT manager about future IT needs (e.g. software required, hardware issues). Reports statistics on IT (number of analysts etc.) to Management.
	Develops knowledge of how analysis software provided in the Safe Setting operates (e.g. Stata, R) and answers queries related to this.	Working knowledge of software programmes in RDC (e.g. Stata, R) and experience of providing support for software related queries. Is able to learn new software as required.	Shares software knowledge across team. Is able to undertake analysis using software.

Data acquisition	Develops knowledge of data negotiation process e.g. through discussions with colleagues and attendance at relevant meetings. Develops knowledge of data sharing principles.	Supports colleagues in negotiating for access to less complex data sources: collects the relevant information (e.g. reasons why access is required, key legal gateways); presents a persuasive business case; liaises confidently with analysts and data suppliers (F2F, telephone and email), ensuring key messages are understood and good relationships maintained. Supporting drafting of proposals to data suppliers, setting out the key messages.	Supports beginner-level and mid-level staff on negotiating for access. Helps managers with negotiation to advanced data sources. Prepares reports on data use for data suppliers. Experience of drafting contracts. Able to advise on repository options for own or third party data.
Metadata and data documentation	Is able to locate data sources and advises analysts accordingly. Develops knowledge of data cataloguing standards.	Compiles a set of records (metadata) to describe data sources.	Applies for and gains access to data for analysts. Advises data suppliers on data access issues. Applies skills to new and innovative sources of data (e.g. Big Data)
	Develops knowledge of metadata – why they're important, different types, practical uses – and produces basic metadata.	Good understanding of metadata. Reviews existing metadata, identifies gaps and creates new metadata where required.	Supports beginner-level staff in producing metadata.

	Reviews data documentation and guidance and identifies gaps. Creates basic documentation and guidance in consultation with data suppliers and analysts.	Creates more complex documentation and guidance in consultation with data suppliers and analysts. Produces templates for metadata, documentation etc. Presents webinars on data.	Supports (and leads) beginner-level staff in compiling new documentation. Provides training on data sources.
Quality assurance	Understands data quality measures (from previous analysis experience). Develops understanding of how quality assurance is carried out in the Safe Setting.	Investigates data quality. Is able to add value to data by improving quality.	Influences data collection process to improve data quality.
Data management	Develops knowledge and understanding of best practice/how the Safe Setting stores and processes data[5]. Stores and processes simple data according to the Safe Setting's procedures.	Stores and processes more complex data according to the Safe Setting's procedures. Provides training and advice to others on storing (including encryption) and processing data. Understanding of data sharing practices	Develops procedures for storing and processing data. Assists managers in overseeing the storing and processing of data. Assists managers in overseeing the determination of differences between 'Safe Setting' datasets and licensed/downloadable datasets. Assists managers in implementing changes in regulations concerning data handling requirements.

[5] For example, see Corti et al (2014) for the standards and procedures for storing and processing secure data at the UK Data Archive.

Data transformation	Develops basic statistical programming skills (e.g. R, Stata, SPSS) – on-the-job or via relevant course.	Further develops statistical programming skills, to level required for the role (dependent on services offered by the Safe Setting).	Further develops statistical programming skills e.g. identify and implement automation and quality assurance of code.
	Derives simple new variables to existing data sources (for benefit of analysts).	Derives more complex variables to existing data (for benefit of analysts).	Advises team and works with data suppliers to introduce data quality measures.
	Develops knowledge and understanding of data linkage carried out by the Safe Setting (e.g. different sample frames, units of analysis and how data can be linked together), including familiarisation with programming code. Advises analysts on the linking process.	Carries out data linkage for analysts and generates new programming code where appropriate, with support where required e.g. generation of pseudo-anonymised identifiers. Undertakes analysis of linked data (e.g. tabulations, distribution).	Carries out more complex data linkage requests for analysts, including generation of new programming code. Trains and provides support to beginner-level and mid-level staff carrying out linking. Is able to present webinars etc. and produce documentation, about linking. Is able to advise on linkage feasibility and quality.
Support function	Answers simple queries about data (by reading data documentation, analysing the data and/or contacting data suppliers). Answers more complex queries with support from colleagues.	Answers a range of data queries, including more complex ones (by reviewing the documentation, analysing the data and/or contacting data suppliers).	Shares knowledge with colleagues. Supports data suppliers (liaises with them about data queries and recommends actions, e.g. data or documentation improvements).
Understanding analysis findings and outputs	Confident interpreting analysis findings and develops knowledge and understanding of a range of analysis results (through SDC of outputs).	Uses this knowledge to carry out SDC effectively, promote data sources (e.g. at webinars, training) and advocate and negotiate for access to more data sources (see ‘Data acquisition’ for more details).	Participates and contributes knowledge of subject in papers, conferences etc.

Statistical disclosure control (SDC)	Develops knowledge and understanding of the Safe Setting's approach to SDC. Analyses and applies SDC techniques to analysis outputs and data sources, with support where required.	Analyses and applies SDC techniques to a wide range of analysis outputs and data sources. Seeks further support on difficult or complex results.	Provides support to beginner-level and mid-level staff. Is able to advise and make decisions about complex outputs. Contributes to wider knowledge in this area (e.g. present at SDC conference).
	Confident speaking to analysts/researchers about their outputs and providing advice on basic SDC issues (e.g. re-coding or suppressing values).	Confident speaking to analysts/researchers about their outputs and providing advice on more complex SDC issues (e.g. methods of transformation).	Provides bespoke training to individual analysts/researchers with certain queries/needs.
Effective analyst management and customer service	Reads relevant literature e.g. Desai, T. and Ritchie, F. (2009) and discusses approach and best practice with colleagues. Delivers Safe Setting training (e.g. safe use of data training) according to 'research management' approach, with support from colleagues. This includes presenting analysts as collaborators and ensuring that they have a clear view of their responsibilities and the service's. Works with analysts to ensure that SDC is a collaborative process. This includes speaking to analysts to understand their output (where necessary), ensuring they provide the required information, and ensuring that safe outputs are released (with support where required).	Discusses 'research management' approach with colleagues and shares best practice. Demonstrates good customer service. Independently prioritises requests from analysts. Resolves minor queries; triages serious queries to more senior staff.	Mentors new staff. Resolves serious queries. Identifies and implements new ways of working to better manage analysts. Devises and communicates policy changes to analysts.

	Develops knowledge of the RDC's approach to customer service (e.g. service level agreement). Deals with a range of customer queries, with support where required, and provides strong customer service.	Deals with a range of customer queries (including more challenging cases, with support where required) and provides strong customer service. Supports analysts by identifying potential needs in advance. Manages analysts' expectations. Is able to apply the right incentives to best-manage analysts.	Communicates to analysts on a number of issues. Organises workshops when required (to deal with specific needs). Identifies right incentives.
Advocating for data		Participates in workshops etc. advocating data. Contributes to consultations about proposed changes to data and national scenario on data sharing.	Influences stakeholders. Drafts consultations about proposed changes to data (led by managers). Responds to consultations about proposed changes to data and data sharing opportunities and challenges.
Records and auditing and data security	Develops knowledge and understanding of the Safe Settings's record keeping processes and relevant certification (e.g. ISO 27001 or Information Governance Toolkit).	Populates and maintains relevant records (e.g. data applied for, received access etc.). Is able to respond to a potential data security threat. Contributes to internal audits.	Prepares for audits. Carries out internal audits. Assists managers with data security events/investigations. Helps managers to implement recommendations from audits/investigations.

Training and education	Attends different training delivered by SDAP in the RDC (e.g. safe use of data training, data webinars etc.). Delivers the above training, with support from colleagues.	Delivers training independently. Contributes to development of training.	Supports beginner-level staff in delivering training. Leads on design of training.
Data re-use and retention	Develops knowledge of data retention.	Implements data retention e.g. applies for renewal. Supports drafting of data retention plans for analysts' or own data.	Leads on data retention for less complex data sources/standard contracts. Support management in data deletion, archiving and re-sharing.

[3] This may be analysis, research or service evaluation.

Glossary

Analyst (or researcher)

Person who has been approved to access and undertake statistical analysis of the data. Could be an academic researcher visiting the Safe Setting, or a member of staff in an organisation who has been granted access to the data.

Confidential Information

Data about Data Subjects that have a significant probability of re-identification. May include data that have been 'de-identified' (direct IDs removed) but still possible to infer the identity of a Data Subject (or the data are detailed enough that they pertain to one Data Subject). Apart from de-identification, the data have not been perturbed in any way (e.g. variables banded, outliers removed etc,). May also include data with direct identifiers.

Data Cataloguing standards

Pertaining to international standards about creating metadata and cataloguing (e.g. Data Documentation Initiative, DDI).

Data Retention

This refers to a policy about long-term preservation and storage of a data source. This policy will differ depending on the Data Supplier, legal and ethical frameworks etc.

Data Re-Sharing

The process by which data used by Analysts/Researchers are prepared for further use by others. This might be to enable results to be replicated and scientifically validated; or might be for the intention of furthering analysis by sharing with others. Generally, data containing Confidential Information are not shared to ensure confidentiality is preserved.

Data Subject

Individual observation in dataset, about which variables in dataset relate to. Could be an individual person or enterprise (company).

Data Supplier

In this context, the Data Supplier is the organisation with responsibility for authorising access to the data. It may be the organisation that collects data, and may also be an organisation that supplies data to an intermediary organisation for further dissemination of access.

Information Governance (IG) Toolkit

A similar information governance accreditation to ISO27001, the IG Toolkit is considered a 'sub set' of the former. Devised by the Health and Social Care Information Centre (NHS Digital) for organisations that process patient records, it is often required by organisations that undertake analysis of patient data.

Internal Audits

The process of undertaking checks that systems and processes are working as they should be (and accord with pertinent documentation). Normally required for ISO27001 and IG Toolkit accreditation.

ISO 27001

An internationally recognised information governance accreditation. While the scope of accreditation will vary with organisation (i.e. the areas of the organisations assessed and accredited), information security systems are generally within the scope. Requires regular auditing to ensure the standards of the accreditation are maintained.

RDC

See Safe Setting.

Safe Setting

Sometimes referred to as a Secure Data Environment, Safe Haven, Research Data Centre. A secure computing environment for accessing confidential data. Full description provided by Schiller and Welpton (2014). Essentially data are analysed in Safe Setting. Only statistical results are released, subject to a check (see Statistical Disclosure Control).

Statistical Disclosure Control

In the context of a Safe Setting (where access to Confidential information which hasn't been anonymised is permitted), Statistical Disclosure Control (SDC), is the process of checking statistical results that have been generated from the information, to ensure that Data Subjects can't be re-identified, and/or that no confidential information about them is released.

Statistical Programming

The process of using specialist statistical software to shape, recode, clean etc., data, and execute statistical commands to produce statistical results.

Acknowledgements

Much of the content in this document was prepared by the following individuals: Carlotta Greci (The Health Foundation), Richard Welpton (Cancer Research UK), James Scott and Christine Woods (UK Data Archive, University of Essex).

In addition, we are grateful for feedback and advice from the particular individuals: Scott Summers (UK Data Archive, University of Essex), Arne Wolters (The Health Foundation); and other members of the Working Group for Secure Data Access Professionals.

Date: July 2018

About the Working Group for Secure Data Access Professionals

The Working Group for Secure Data Access Professionals was established in 2011, and acts as a group to share expertise, experience and best practice for managing access to confidential/sensitive data in secure facilities. This includes, but is not limited to, issues around:

- Building and developing secure settings to access data
- Understanding the data landscape (including emerging sources of data, new legislation, new technologies and new methodologies for protecting data)
- Developing staff skills
- Statistical Disclosure Control
- Uses of confidential/sensitive data

The Working Group meets quarterly, and is attended by staff of all levels and experience from a number and range of organisations, including government departments, universities, and charities. If you would like to know more about the group, please visit securedatagroup.org