Data Personas

Report for Scottish Government

October 2021

Report For Scottish Government

[UV2618]

Submitted by: User Vision and Effini

May/June 2021

Contacts:

User Vision	**Effini**
Chris Rourke	**Dr Jo Watts**
CEO	Founder and CEO
0131 225 0851	07753610456
chris@uservision.co.uk	jo@effini.com

Executive Summary

This project supports the development of a Data Transformation Framework to improve and enable data reuse in the Scottish Public Sector. Individuals and organisation will undertake data maturity assessments to measure their current maturity level and identify pathways to increase knowledge and skills. To support the pathway development a set of core personas are required to bring the framework to life for users.

This project has reviewed sets of existing personas in use both internally and externally to the public sector to identify a set of suitable potential personas that cover the end-to-end activities around data within the public sector. A survey and interactive workshops have then been carried out to validate these potential personas.

The output of this project is a set of final personas that cover the breadth of users of the Data Transformation Framework. 11 final personas have been identified and are documented below. 5 priority personas have been identified to provide initial focus. These are highlighted with **s below.

Final persona	Persona summary
Data novice**	A public sector worker who is just beginning to understand the potential that data can bring to their role and their organisation. They are often users of systems that capture and create data.
Data consumer	A non-technical data user that consumes visual information through reports or metrics to extract insights and support decision-making, research or planning activities.
Data analyst**	A technical data user that manipulates data to create output analysis and visuals to extract insights, support decision-making, diagnose issues and answer questions. The types of data analysed can be general or specific, such as GIS data.
Data scientist	A data professional who uses modern data science techniques, such as AI and machine learning, to develop analytical solutions to problems.
Data engineer	A data professional who uses software engineering techniques to create data pipelines that support the automation and reproducibility of data solutions
Leader**	A strategic business leader who wishes to use data to realise value for their organisation. Will be responsible for overall organisational strategic direction with access to budgets to deliver change.

Data transformation lead**	A team (or department) leader who champions the effective use of data within their area of the organisation. They may be responsible for programmes of work, or leading teams to deliver change, but without the scope or budget align activities across the entire organisation.
Data product/service owner	A solution innovator that is developing products or services that are data rich. Although they may not be highly data literate, they understand the importance of good data management in developing a scaleable solution.
Technology specialist**	A technology professional who is responsible for the creation and maintenance of tools, systems and applications used to capture and store data. They also work to ensure the integrity, confidentiality, and security of all data assets within these systems and applications.
Data governance specialist	A worker with responsibilities for the policies and processes that ensure data can be managed as an asset within an organisation. This includes ensuring compliance across legal and regulatory frameworks.
Data management specialist**	A specialist responsible for one of many activities required to manage data as an asset. This includes the standards, structure, integration, quality and definition of data, including processes to manage these.

Further work is still required to validate these in detail and bring them to life with detailed needs, behaviours, and goals.

Contents

1 Background to the project

The Data Standards Team (DST) in Scottish Government are currently in the process of developing a Data Transformation Framework (DTF). This is included in the vision of Scotland's refreshed digital strategy Realising Scotland's full potential in a Digital World', which aims to put digital at the heart of everything. The DTF has a strategic role to encourage organisations to become data-driven. It will provide:

▶ **Strategic engagement** – connecting leaders and peers to drive data transformation, building data skills in organisations to realise opportunities for data improvement, reuse and innovation
▶ **A learning environment** - support and resources for upskilling and training, to create a collaborative community around sharing best practise and common approaches
▶ **A knowledge base** – a repository of strategic drivers and guidance on policies and initiatives, co-designed with the data community to ensure it is populated with current and relevant resources, connections, and communities of practice.

Data transformation is the journey towards improvement and increased capability in using data. The Data Transformation Framework is being developed to support organisations in understanding not only where they are currently, in terms of data maturity and use, but how they can improve this, realising more of the value potential through data use.

There will be multiple ways into the DTF, one of which will be through data maturity assessments. Other routes may include a diagnostic tool to allow a quick assessment or a fast-track option. Utilising data maturity assessments (and other diagnostic routes) within the public sector will help to create a shared view on the level of individual and organisational data maturity, identifying practical pathways for upskilling and retraining, encouraging data improvements and reuse. The creation and use of data personas will enable the framework to be tested and validated for all the different user types at different current levels of data maturity. It will help to foster a common understanding within the public sector of what 'good data' looks like and how organisations can improve their collection, usage and sharing of data.

There are a number of existing sets of data personas in use both within and outwith the Scottish Government. The aim of this project is to review and validate these existing personas within the scope of the public sector users that will be engaging with the DTF

and supporting data maturity assessments with a view to understanding the completeness of any set for this purpose and identifying any gaps.

This project involved specifically:

1. Carrying out research in consultation with Scottish Government to review existing data personas/ user models in use for the public sector that may be relevant.
2. Carry out a survey and workshops and develop data personas.
3. Evaluate the personas using peer to peer or other methodology.
4. Produce a final report detailing the methodology, activities, outcomes and next steps.

The outputs included:

- A research paper reviewing the existing relevant data personas/user models in use for the public sector.
- Conducting at least two workshops with target users to carry out discussion and activities to build a better understanding of potential users and then evaluate current personas for the Data Transformation Framework.
- Producing a report reviewing the workshop findings and sense checking the personas (this document).
- Presenting outputs to the Scottish Government with a draft and then final report detailing methodology, activities and outcomes.

2 Background research findings

Personas and project scope

Personas are fictional composite characters that are used to represent the different user types that might use a product or service. They help to bring to life the different user's needs, behaviours and goals. As archetypical models of groups of individuals, different personas will aim to have distinct needs, behaviours and goals which aid future service design.

Once complete, personas are often documented as:

▶ Who they are, including what they are likely to say;
▶ What motivates them – their goals;
▶ What they want – their needs;
▶ Their behaviours and preferences including what they like and don't like.

This project looked at data personas, or personas with a focus on how the character interacts with data in their existing roles. This covers the end-to-end lifecycle of data within public sector organisations from data creation to storage, governance, management, analytics, communication, and insights. This broad scope meant that there was an expectation that a number of personas may be required to ensure completeness.

Existing data persona sets

The project commenced with a review of existing data personas that are in use both within and outwith the public sector. The sets of personas that were reviewed and documented in detail were:

▶ **Digital Scotland**: These were developed in 2018 by the Digital Transformation Service to understand their service users and their analytical needs
▶ **Archetypes**: These were developed as part of a UK Geospatial Commission project to understand the users of spatial data
▶ **ONS website** (https://style.ons.gov.uk/category/writing-for-the-web/personas/): a set of personas identifying the characteristics of their website users
▶ **Datacamp** (https://www.datacamp.com/community/blog/persona-driven-learning): the four most common personas they encounter, though they have a full list of 11 (see below).
▶ **Data Skills for Work** (https://dataskillsforwork.com/what-is-data-literacy): this is a framework developed by EKOS and SDS to help to develop data skills pathways
▶ **Draft DTF personas**: initial work done within this broader project to identify the motivations and needs of the Data Transformation Framework

The following personas were also visited, but their characteristics were not documented in a detailed manner since there was considerable alignment with those already reviewed:

- **Monte Carlo Data** (https://www.montecarlodata.com/which-of-the-six-major-data-personas-are-you/): a blog covering the major types of data user in organisations
- **Aryng data literacy** (https://aryng.com/blog/data-literacy-personas/): a blog covering the skills and data literacy levels within an organisation
- **Qlik data literacy** (https://www.qlik.com/us/-/media/files/training/global-us/qlik-education-data-literacy-program-strategy-and-framework.pdf): a framework to identify different data literacy levels and therefore training needs within an organisation
- **Sustainable Development Goals** personas (https://sdgdata.gov.uk/user-personas/): a set of personas designed to understand the website users.
- **Data Quality Hub**: a set of personas developed by the Data Quality Hub within the Office of National Statistics around the types of individuals they engage with.

Documenting the existing personas

The high-level areas that were investigated for each existing data persona set were:

- Their needs when working with data
- Their skills
- The tools, systems and devices they use
- The data types and formats they access
- Their current knowledge around data and data management

The high-level areas were then broken down into more detailed topics with a set of standard category responses. The standard categories were developed from existing frameworks and lists of standard characteristics for data users. The use of standard categories allowed different sets of personas to be documented in a similar manner and similarities across each to be identified. This allowed a superset of draft high-level personas to be identified.

Draft persona identification

The draft high level personas are documented below with a draft name and summary that describes their motivations and interactions with data.

Through existing experience working in the world of data there were additional personas that were expected to be identified but weren't included within the persona sets reviewed for the background research. We have called these inferred personas. Some of these have been documented by Datacamp in their Data Science Industry infographic:

https://www.datacamp.com/community/tutorials/data-science-industry-infographic. The reason these have not been identified within the reviewed persona sets is that the data persona sets used have had a focus on data usage rather than data management, governance, system integration or security.

High-level persona	Persona summary	Seen or inferred?	Need	Skills	Tools	Data	Data management
Citizen	A public consumer of open unbiased data to answer personal questions, find topical stories or hold government to account	Seen	Metric consumer Visualisation consumer	Data interpretation Societal data use Communication & storytelling Geospatial mapping		Datasets Metrics	
Data user non-technical	A consumer and interpreter of visual information or metrics to support research or benchmarking on a monthly basis	Seen	Visualisation consumer Metric consumer Metadata	Communication & storytelling Reporting Visualisation Data interpretation Data terminology	Spreadsheets Dashboards Point and click analysis	Metrics Surveys Datasets Documents	Data security Data quality Metadata

Role	Description					
Data user technical	A daily consumer and creator of visual information with limited analysis skills to support local decision-making, research and planning	Seen	Visualisation creator Metric creator Metadata	Data interpretation Tools and visuals Visualisation Reporting Data discovery Communication & storytelling	Spreadsheets Visualisation packages Dashboards	Datasets Metrics Documents
Analyst	A data professional who focuses on the use of data to support and monitor progress, diagnose issues and answer problems	Seen	Visualisation creator Analyst Metric creator Raw datasets	Tools and visuals Data interpretation Infographics Reporting Visualisation Programming Value from data Problem solving and delivery Communication & storytelling	Spreadsheets Dashboards Statistical packages Programming languages	Datasets Systems	Data warehousing & business intelligence

Modeller/statistician	A data professional with strong mathematical and programming skills to work directly with raw data to develop analytical solutions	Seen	Visualisation creator Modeller Metric creator Analyst Raw datasets	Visualisation Reporting Geospatial mapping Programming Predictive analytics Statistics Data management Communication & storytelling Problem solving & delivery Data protection & legislation	Programming languages Statistical packages Point and click analysis Visualisation packages Dashboards Spreadsheets	Datasets Surveys Systems Text	Data warehousing & business intelligence Data security Data quality Data modelling and design Document and content management
Leader/strategist	A business leader who wishes to use data to extract value for their organisation. May be responsible	Seen	Decision-maker Metric consumer Visualisation consumer Owner	Value from data Data strategy Visualisation Data protection & legislation Problem solving &	Dashboards Spreadsheets	Metrics Documents Visualisations & infographics Dashboards	Data governance

Role	Description						
	for data resources, own personal data or deliver data products.		delivery Managing data resources Data strategy & leadership Data interpretation Data governance				
Product owner/Project manager	A solution innovator that is looking to develop products or manage projects that are data-rich.	Seen	Visualisation creator Product creator	Value from data Data interpretation Reporting Visualisation	Spreadsheets	Visualisations & infographics	Metadata
IT/software	A technology professional who builds systems and products with clean and consistent data	Seen	Policies Metadata Raw datasets	Programming Data engineering Data systems & architecture	Programming languages	Systems Datasets	Data modelling & design Data storage and operations Data integration & interoperabilit

							y Data architecture
Data governance	Responsible for the policies and processes that ensure data can be managed as an asset within the organisation	Seen	Owner Policies	Data management Ethics Data protection & legislation Data governance	Spreadsheets	Documents Surveys Systems	Data modelling & design Data storage & operations Data security Data integration & interoperability Document & content management Reference & master data Data warehousing & business intelligence Metadata Data quality Data architecture Data governance

Data management	A specialist who understands the activities required to manage data, including its quality and definition in structured manner	Seen	Metadata Policies	Data governance Data management Data protection & legislation	Spreadsheets	Datasets Metrics Surveys	Reference and master data Metadata Data quality Data governance Data security
Data architect	A specialist responsible for designing and maintaining the structure and integration of data within an organisations systems and tools	Inferred	Data creator Product creator	Data systems & architecture Data management	Graphical tools Programming languages	Systems Datasets	Data modelling & design Data architecture

Role	Description		Archetype	Knowledge areas	Tools	Objects	Data management
Data engineer	A data professional with strong software development skills who focus on the automation of data solutions	Inferred	Data creator Product creator	Data engineering Data management Software development	Programming languages	Systems Datasets	Data integration & interoperability Data security Data modelling & design Data warehousing & business intelligence Data quality Metadata Reference & master data Data storage & operations
Market research	A specialist who extracts insights from data utilising research techniques. Often undertakes qualitative research.	Inferred	Visualisation creator Visualisation consumer	Data terminology Data interpretation Tools and visuals Infographics Problem solving & delivery Benchmarking	Spreadsheets Dashboards Point and click analysis Visualisation packages	Datasets Metrics Surveys	Data warehousing & business intelligence

Data security	A specialist responsible for ensuring the integrity and confidentiality of data assets within an organisation	Inferred	Policies Metadata	Data systems & architecture Data management	Programming languages	Systems	Data security Reference and master data Data integration & interoperability Document & content management
Data rookie	A worker who has previously not considered their role requiring data skills. Similar to a non-technical user, but with no current training	Inferred	Visualisation consumer Metric consumer Metadata	Data terminology Data interpretation Visualisation Communication & storytelling	Spreadsheets Dashboards	Metrics Surveys Datasets Documents	Data security

3 Workshop preparation and approach

Participant selection

The background research and draft high-level personas were used as a guide to select workshop participants. Relevant participants were drawn from various Government departments and Local Authorities. 18 individuals were invited to take part with the ambition of 16 individuals being successfully recruited to take part in one of two workshops. The participant selection was designed to cover as broad a range of the high-level personas identified in the background research as possible. It is acknowledged it was not possible within such a small group to cover them all. To aid discussions the participants mostly had some experience of working with data.

The breakdown of participants is as below:

Participant	Public sector area	Role title	Possible high-level personas
1	Council	Digital Transformation Manager	• Leader/strategist
2	Council	Digital Transformation Lead	• Product owner/project manager Data governance
3	Council	Data Intelligence Manager	• Leader/strategist
4	Council	GIS Officer	• Data user technical
5	Council	Principal Officer	• Product owner/Project manager Data management
6	Council	Lifelong Learning Strategy Officer	• Data user non-technical
7	Council	Strategic programme delivery	• IT/software

8	Council	IT Team Leader	• Product owner/Project manager
9	Council	IT Co-ordinator	• Data user non-technical Product owner/Project manager IT/software
10	Scottish Government	Scientific Adviser	• Data user non-technical Data rookie
11	Scottish Government	GIS Operations Manager	• Analyst Data engineer
12	Scottish Government	Head of Management Information Procurement	• Leader/strategist Data architect Market research
13	Scottish Government	Technical Architect	• IT/software Data management
14	Scottish Government	Statistician	• Modeller/statistician Data governance
15	Improvement Service	Data and Engagement Manager	• Data user technical • Product owner/Project manager
16	Scottish Government	Analytical Services Digital Manager	• Data user technical

Survey development

To maximise the effectiveness of the workshops for nuanced discussion, factual information was captured in advance from the participants through the use of a survey. The questions were designed to understand the participant's current engagement with data and validate their possible persona.

The following questions and possible answer sets were utilised in the survey:

▶ Can you provide us with a brief description of your role? [*Free text*]

▶ Please rate your experience of working with data, with 1 being the lowest [*Scale 1-5*]

▶ Have you completed any formal training around using data? [*Yes/No*]

- If you answered Yes, please provide a brief overview [*Free text*]

▶ How often are you involved with activities in the area of: [*never, daily, weekly, monthly, annually, ad-hoc*]

- Data creation eg surveys, questionnaires, research
- Report creation using graphs, tables or other data views
- Data governance
- Data quality and/or cleansing
- Data consumption eg reading other peoples reports
- Data analysis
- Data integrations
- System setup/configuration
- Data extracts
- Data strategy
- Collation of data

▶ Please specify any other data, or information, related activities you are involved with, not already mentioned above, and the frequency you do them [*Free text*]

▶ From the following list of data types, select all you make use of in your current role [*geospatial, demographic, open data, public data, private data*]

▶ What types of areas is the data you work with? Select all that apply. [*Local government, Your organisation, The internet, National agencies, Academia, National government, Subscription providers, Private companies*]

▶ Which of the following tools have you used in your current role? Select all that apply [*Spreadsheets, Visual Basic, DAX, Tableau, SQL, Power BI, Python, R, Statistical packages, Other*]

▶ From the following list, which areas do you feel additional training would most benefit you? [*Point and click analysis tools, Visualisation tool, Statistical packages, Data governance, Data analysis techniques, Programming languages, Data quality and cleansing techniques, Data ethics, Working with spreadsheets, Report automation processes, Other*]

▶ In the scope of your current role, how do you discover the data sources you need? [*Web based research, Colleagues, Experience, External sources, Other*]

▶ In the scope of your current role, what methods do you use to access data? [*Request to another person or department, via an API, File extracts or downloads, Other*]

- If there is anything else around your use of data/information in your current role you would like to share with us, please add it here [*Free text*]

Workshop discussion guide

The Scottish Government project group, Effini and User Vision collaborated to develop a discussion guide for the two workshop sessions to ensure good coverage of all of the main areas of interest for the project.

The discussion guide was designed to build on the factual information captured through the survey with more detail and personal experiences. Alongside an introduction to the Data Transformation Project itself, the workshop attempted to widen the scope beyond the participants own role to that of their team and stakeholders. It also delved into the barriers and challenges around their use of data and their views on improvements that could be made.

The following workshop discussion guide was agreed:

- **Background/Introduction**

  - Familiarisation with DTF project – introductory talk by Shona Nicol and Sally Kerr
  - Introduction to workshop format/familiarity exercises with Miro platform

- **Role of data – For you**

  - **Task**: "There are lots of different activities that you may do with the data. We'd like to find out more about those with this exercise and discussion. Please describe in your own words what you do with data then we will have a group discussion about your responses."
  - **Purpose**: To identify the actions/activities that participants perform with data
  - **Format**:

    - Sticky notes created by each participant in Miro
    - Group discussion via Teams

- **Role of data – For your team**

  - **Task**: "Thinking more widely around the activities that your team does with data, can you add anything that has not already been covered? In particular, what does your team use data for? What do you do with that data and what wider purpose does it serve?"

- **Purpose**: To identify how the role of data within the participants' teams might differ from the personal perspective in the previous task.
- **Format**:
  - Sticky notes created by each participant in Miro
  - Group discussion via Teams

▶ **Barriers/issues/challenges**

- **Task**: "When you work with data are there any stages that are particularly difficult or time-consuming? Why? What could be done to resolve these issues for you?"
- **Purpose**: To identify factors that may be preventing or restricting participants ability to work with data
- **Format**:
  - Sticky notes created by each participant in Miro
  - Group discussion via Teams

▶ **Improvements**

- **Task**: "What are the top 5 things that you (or someone else) could do to improve how you and your team work with data? And what external factors would need to change to help you in this regard?"
- **Purpose**: To discover participants' perceptions of measures that could address any issues they experience in relation to data
- **Format**:
  - Sticky notes created by each participant in Miro
  - Group discussion via Teams

4 Survey results and analysis

Key findings
The survey was sent out to all participants in advance of the workshop. 15 out of the potential 16 participants responded fully to the survey. The key insights from the survey responses are documented below.

Current experience level

Experience level	Count
1	0
2	1
3	5
4	6
5	3

Most participants scored themselves reasonably highly (3-5/5) when it came to their experience around working with data. However only 60% of respondents claimed to have received any formal training around data. In most cases this formal training was academic, but some was also through internal courses.

Frequency of activities

Participants were requested to confirm how often they did certain activities. The high-level ad-hoc creation of data was expected, as was the need to undertake data quality and cleansing and analysis activities. However, the high number of respondents that also confirmed they undertook activities to configure and bespoke the systems they were using to their own needs was unexpected. The number of respondents who confirmed they undertook data strategy activities at least monthly was 9. This was particularly unexpected and has highlighted a need for an additional persona that includes strategic activities around data

Data types

What types of data do you make use of in your current role?
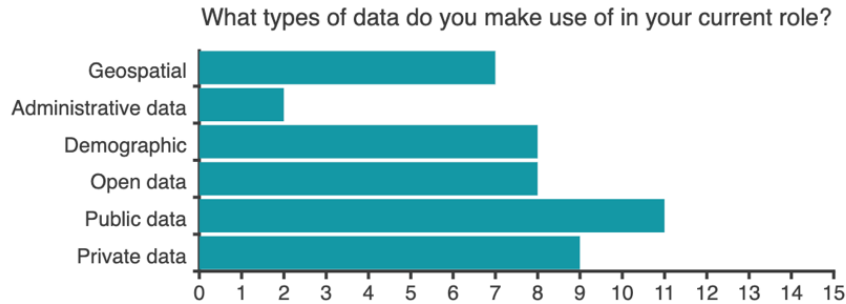
The types of data that participants reported working with, were as expected. There was a focus on geospatial and demographic data.

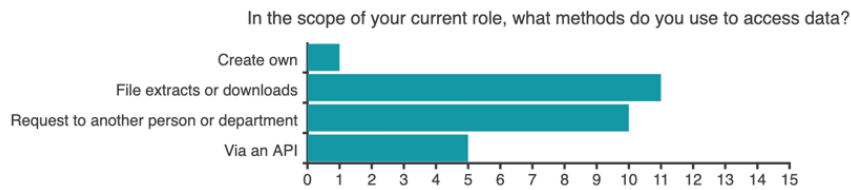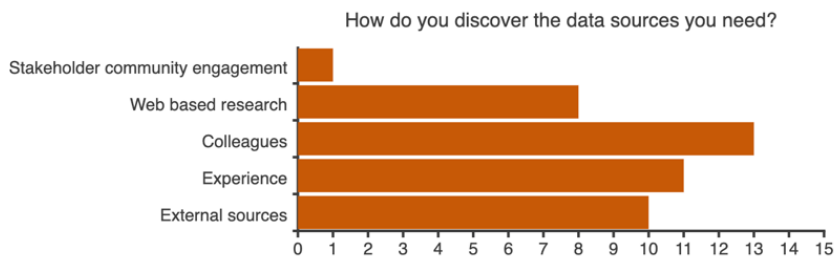## Data provenance

What types of areas does the data you use come from?
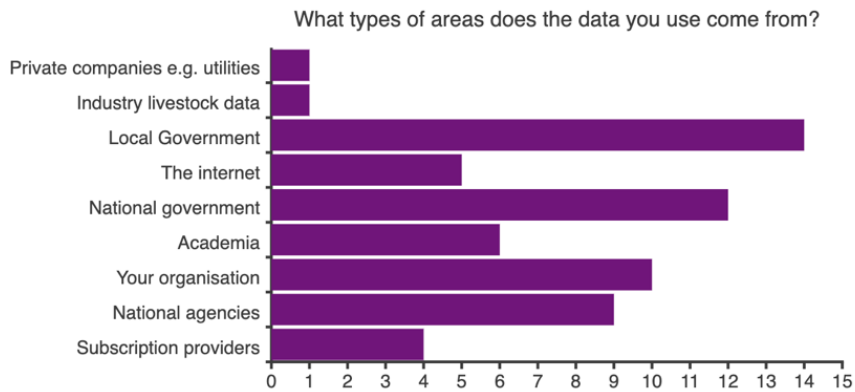
Government data, both local and national, is the most common source of data to work with.

Data discovery

Existing knowledge and networks are currently the most common ways to find the required data.

How do you discover the data sources you need?

In the scope of your current role, what methods do you use to access data?

Data access

Data is normally accessed through files or by requesting access existing networks.

Training needs

In which areas do you feel additional training would most benefit you?

The awareness of knowledge gaps around governance, data quality and ethics is interesting. It demonstrates an appetite for enhanced learning around key areas that will feature in the Data TransformationFramework.

Additional comments within the survey covered areas from legislation to metadata standards to the need to use and manipulate data in almost all roles.

Workshop input

Whilst participants summarised their roles within the survey, there was a clear need to delve into the difference between the individual's use of data and that of their wider team. This led to the inclusion of a discussion around the engagement with wider stakeholders within the workshop.

The decision to focus the workshop discussions around barriers and suggested improvements was also influenced by some of the additional comments included within the survey which began to touch on this.

5 Workshops analysis

Analysis approach

Two 2-hour workshop sessions were conducted remotely on Wednesday 12th May 2021. There were 8 participants in each session. The Miro online platform was utilised to capture participant comments and feedback, alongside detailed notes of the discussion by the workshop facilitators. Each participant was given their own colour to capture comments so that they could be traced back to the specific participant to enable persona validation.

Key discussion themes

There were several discussion themes that were highlighted by many of the participants as barriers to their own roles or potential areas for improvement. Although these are all currently well known within the public sector, they are worth documenting as potential areas of initial focus within the Data Transformation Framework. The areas of focus and challenge highlighted by participants also enabled personas to be validated and differentiated.

Metadata and standards

A comprehensive view of well documented available data was highlighted by many participants at all levels of current data maturity as a key driver behind high quality data use and reuse. The lack of good metadata is seen not only to create silos of knowledge and use, but also created a barrier to use, particularly around the appropriateness of use. It was also commented that legacy dataset design often also made some datasets particularly difficult to be used by those not familiar with them.

Data-driven decision-making

The use of data to drive decision-making was seen to be an area with significant scope for improvement. There were two aspects to this: the use of data within existing decision-making processes; and the purpose of specific data collection activities. It was generally agreed that a lot of data is captured, but not all of it is useful for the types of questions being asked. Data and information requests would benefit from additional context around the required decision and ensure consistent and complete data collection processes developed to support this. It was agreed that this would drive trust and transparency in the decision-making process.

Business case for data engagement

It was felt that the business case for capitalising on the value of data had not yet been clearly made. A large proportion of public sector employees are still not engaging with data. Factors that fed into this were the clarity of role profiles, prioritising time within existing roles for data-related activities and the need for upfront investment in upskilling, communication and additional roles.

Data literacy and upskilling

Much of the current training on offer was thought to focus on the tools required for manipulating and visualising data. It was felt that there was a significant need to develop general data literacy and communication across the organisation. The survey also highlighted a significant training need around the areas of data governance, data management, data quality and data ethics.

Tools, software and interoperability

To improve interoperability, reuse and sharing it was felt that the current plethora of tools, processes and systems was not helping. There was a clear wish to use a consistent system such as Power BI for reporting to enable organisations to provide consistent and traceable outputs.

At the same time there some frustration about some of the tooling restrictions in place. This was mainly around access to open-source software such as R or python, which can help to improve analytical capability and auditability for professional analytical teams.

Data sharing and communication

Improved data sharing was highlighted as an area that could break down silos between and within organisation. Visibility of the upstream and downstream data uses could help to bring focus onto the need for improved metadata amongst other things. It was commented that data protection concerns were often used as a barrier to data sharing, and therefore enhanced visibility of data quality issues.

There was a general request for more data to be made open, but the lack of clear guidance around data protection limited the opportunities here.

Data quality, governance and management

Data quality was discussed in detail, with requests for funding for roles for data quality monitoring and remediation to be recognised within any work commissioning process. The importance of ownership and therefore responsibility for data was highlighted repeatedly, but no clear path to achieve this.

Areas not covered

There were a couple of areas that we might have expected issues to be raised, however the lack of discussion on these topics was also insightful. It is acknowledged that the workshops did not cover the full range of expected personas, so other participants may have highlighted these if they had been present.

Data and cyber security

The need for processes around keeping data and systems secure were not mentioned. This is possibly considered as an activity for external teams; however, a growing number of organisations also place responsibility for data and cyber security on individuals.

IT and data engineering

We did not meet any participants who had responsibilities around system management, development or remediation. Data products often require data engineers to automate reporting or set up pipelines for data feeds. Systems often also require complex

configurations to ensure they are capturing the right data items and are managing the fully lifecycle of data from capture, storage, definition, access, archiving and deletion.

Evaluation of personas

The draft high level personas were validated by comparing the activities and needs participants identified through the workshop and survey with expectations. In many cases the possible high-level persona did not match reality once the breadth of the individual's role was understood in detail.

Persona validation

An initial observation was the surprising number of participants that reported they undertook strategic work daily within their role. Initially these individuals had been mapped to a Leader/Strategist persona, however it became clear that in general they did not have the required budget or full strategic oversight of an organisation to actively bring about the changes required. We identified the need for an additional persona who is more data-savvy than the Leader/Strategist, but with the scope only to deliver change within their area of the organisation. We have introduced the concept of a Data Transformation Lead to occupy this new persona.

Since the scope of these personas is limited to those that will be engaging with the Data Transformation Framework, it was felt that it was not relevant to include Citizens. Although they will be end users of public sector data products, they should not be included within the framework. More important was to call out the vast majority of public sector workers who are either not yet engaged with data or at the start of their data journey. Therefore, we have introduced the Data Novice persona, to represent these users.

We had originally made a distinction between technical data users, whom we assumed worked with data significantly, but it was not the main aspect of their role, and those we had called analysts, whom the role was specifically focused on data. There was an assumption that the analysts had undertaken more formalised training than the data users, however the survey results proved differently, where in most cases formalised training had not been undertaken. We therefore do not see sufficient distinction between these two personas to separate them and propose to merge together. The market researcher should also fall into this same merged persona.

In addition, the term Modeller/statistician as it is currently used within Scottish Government and the wider public sector does not align to the type of activities this

persona would be expected to undertake, therefore we propose to rename this persona to be a Data Scientist. Their focus would be on developing data science solutions.

We had also originally made a distinction between data management specialists and data architects. On review the breadth of the data management specialist is sufficiently broad to include data architecture within its remit, therefore we propose to merge these personas together.

Final identified personas

The final proposed list of personas is documented below, alongside a summary of the type of their required needs and knowledge. This level of detail is not yet sufficient to bring the persona to life for users, which should follow on from this piece of work. The persona names may also require additional refinement to make sure they align to terminology used or desirable within the Public Sector.

Final persona	Persona summary	Participant number	Summary of needs/knowledge
Data novice	A public sector worker who is just beginning to understand the potential that data can bring to their role and their organisation. They are often users of systems that capture and create data.	6	• Understand the value of data • The importance of high-quality data • System use and data capture • Data literacy • Interpretation of graphical information • Ethics and responsible use • Data protection
Data consumer	A non-technical data user that consumes visual information through reports or metrics to extract insights and support decision-making, research or planning activities.	-	• Extracting insights from data • Making data-driven decisions • Ethics and social responsibility • Data protection • Searching for data sources • Interpreting metadata • The importance of high-quality data
Data analyst	A technical data user that manipulates data to create output analysis and visuals to extract insights, support decision-making, diagnose issues and answer questions. The types of data analysed can be general or specific, such as GIS data.	9, 11, 12, 14	• Data visualisation • Data quality management • Metadata management and custodianship • Tools & techniques for data manipulation • Data communication & storytelling

			• Data monitoring and validation
			• Survey design and analysis
Data scientist	A data professional who uses modern data science techniques, such as AI and machine learning, to develop analytical solutions to problems.	-	• Machine learning and AI techniques
			• Ethical data usage
			• Data interoperability and design
			• Tools & techniques for data manipulation
			• Data interfaces
			• Metadata management and custodianship
			• Data pipeline creation
			• Data product development
			• Data communication & storytelling
Data engineer	A data professional who uses software engineering techniques to create data pipelines that support the automation and reproducibility of data solutions	-	• Data pipeline creation
			• Software development processes
			• Data design & architecture
			• Metadata documentation and data standards
			• Data interfaces
			• Cloud environment use and security

			• Data monitoring and validation
Leader	A strategic business leader who wishes to use data to realise value for their organisation. Will be responsible for overall organisational strategic direction with access to budgets to deliver change.	-	• Understand the value of data • Data strategy • Making data-driven decisions • Ethics and social responsibility • Data literacy • Data and infrastructure risks • Data governance • Business case for data
Data transformation lead	A team (or department) leader who champions the effective use of data within their area of the organisation. They may be responsible for programmes of work, or leading teams to deliver change, but without the scope or budget align activities across the entire organisation.	2, 3, 8	• Business case for data • Data strategy • Data management • Data governance • Data design & architecture • Metadata documentation and custodianship
Data product/service owner	A solution innovator that is developing products or services that are data rich. Although they may not be highly data literate, they understand the importance of good data management in developing a scaleable solution.	1, 5, 15, 16	• Data management • Data literacy • Metadata documentation and custodianship

			• The importance of high-quality data • Ethics and social responsibility • Data protection • Data governance
Technology specialist	A technology professional who is responsible for the creation and maintenance of tools, systems and applications used to capture and store data. They also work to ensure the integrity, confidentiality, and security of all data assets within these systems and applications.	-	• Software development processes • Cloud environment use and security • System configuration and support • Data design & architecture • Metadata documentation and data standards • Data management
Data governance specialist	A worker with responsibilities for the policies and processes that ensure data can be managed as an asset within an organisation. This includes ensuring compliance across legal and regulatory frameworks.	10	• Data governance • Data protection • Data management • Understand the value of data
Data management specialist	A specialist responsible for one of many activities required to manage data as an asset. This includes the standards, structure, integration, quality and definition of data, including processes to manage these.	4, 7, 13	• Data management • Data design & architecture • Metadata documentation and data standards • Data quality management • Data governance

- Data visualisation

- System configuration and support

This is a total of 11 final personas. They all have a distinct set of needs and cover the breadth of activities contained within the Data Transformation Framework. However, the volume of personas was considered quite large and could perhaps become a barrier to adoption and use in the future. Therefore, it was decided to focus initial persona activities on a priority set.

The priority set identified to drive the most value was:

▶ Data novice
▶ Data analyst
▶ Data leader
▶ Data transformation lead
▶ Technology specialist
▶ Data management specialist

The priority personas have been identified from several different perspectives. They cover areas where there are largest volume individuals who with whom public sector employees would identify - data novice and data analyst. Areas where the most progress can be made to embed the Data Transformation Framework within roles - data transformation lead and data management specialist. Finally, areas where volumes are currently low, but where increased knowledge and adoption of the Data Transformation Framework can deliver significant changes in ways of working - data leader, technology specialist.

Summary and next steps

This is the first iteration in developing data personas for the DTF and as only 6 of these were validated within the workshops they should viewed as a work in progress, rather than the final set. Additional work will be required to validate all 11 personas and provide additional details to flesh out the needs, behaviours, and goals of the full set.

The next steps for the data personas project will include:

1. **Additional validation of the remaining final personas**: ensure that any gaps in validating the final personas are filled by specifically targeting individuals who are expected to map to these. In addition, it is important to identify any gaps, such as personas that are missing in their entirety within current public sector roles. The data scientist, data engineer and technology specialist are expected to be rare within existing job roles.
2. **Build out the personas**: bring all the personas to life by understanding in detail their needs, behaviours, and goals. This will involve additional validation with a

wider range of public sector employees. Given the large number of personas, it is recommended to focus initially on the priority personas.

3. **Map the personas to the data transformation framework**: This will involve mapping the required needs and knowledge of each persona to each of the data transformation framework categories, starting with the priority set of personas. It may also be important to understand the current knowledge strengths and weaknesses of each, in order to identify suitable learning interventions.

4. **Develop DTF maturity pathways**: The maturity pathways will involve a prioritised set of interventions for each persona, whereby knowledge and skills can be incrementally developed over time, encouraging data improvements and reuse.

5. **Develop case studies:** Case studies of maturity pathways can be developed for each priority persona to demonstrate the benefits. This could involve actual individuals. However, initially it may be easier to develop these using the personas and a fictionalised case study.

6. **Communicate and publish the personas**: Although it is important to communicate the personas throughout this project, to ensure that they continue to be validated and demonstrate their value. It is also important to finally publish the personas for wider visibility, demonstrating the direction of travel. It will then allow the maturity pathways to be incorporated into role profiles and performance evaluation frameworks.

Whilst finalising the personas it will be important to ensure the barriers to improvement identified within the workshops are also considered. In particular, as maturity pathways are developed, these should not just include upskilling in specific tools, but also focus on best practice data use, interpretation, communication and management alongside understanding the strategic drivers, policies and initiatives.

Documenting data, in particular metadata creation and standards is another area that should start to be embedded into existing processes, not just for open data creation, but also for internally shared datasets, critical reports and metrics used for decision-making and policy.

There is already quite a focus on data management processes within the public sector, however data quality improvement, such as a data remediation capability, data quality reporting and designing systems for high-quality data capture are also areas to consider when developing the maturity pathways.

It is expected that the choice of Technology Specialist as a priority persona will highlight the need to embed these individuals more closely to the business. This will ensure

systems are designed and developed for high-quality data capture, the data is designed for ease of sharing, reporting and interoperability, and the cyber and security risks are identified and more formally addressed.

All of these steps will support the aims of the data transformation framework to improve and increase public sector organisation's capability in using data.