JNCC Report
No: 590

Data Skills Framework: A generic approach to assessing and developing data related competencies and skills

Heeley, L. & Wilkinson, S.

December 2018

ISSN 0963-8091

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
http://jncc.defra.gov.uk

This report should be cited as:

Heeley, L. & Wilkinson, S. (2018) Data Skills Framework: A generic approach to assessing and developing data related competencies and skills, *JNCC Report No. 590*, JNCC, Peterborough, ISSN 0963-8091.

This report is compliant with the JNCC Evidence Quality Assurance Policy http://jncc.Defra.gov.uk/default.aspx?page=6675.

Acknowledgement:

JNCC would like to thank Elspeth Body and Rebecca Strickland, of the Environment Agency, for their review and revision of the original report document.

Summary

Data is a key asset to government and all organisations. Historically data-related skills have been held within specific areas of an organisation, but to make the most of our data we must develop these skills right across our organisations. At present there is no generic framework to support assessment of these skills at this scale.

The Data Skills Framework is designed to be generic and it covers the full range of data-related competencies from identifying data needs to visualisation of outputs. Seven core themes are described with three levels within each (foundation, intermediate and advanced). “Foundation” should be relatively easy to reach, this will encourage a wide number of staff to develop a broad understanding.

Within each of the competencies the precise knowledge of the individual will vary depending on their role and the domain they work within. Additionally, the familiarity with specific tools will be important. When implementing the framework additional context will be needed and an example is included to highlight this. The overarching framework helps to identify potential flexibility (e.g. potential to move staff between domains), it also allows the overall capability of the organisation to be assessed.

Contents

1 Introduction and overview

1.1 Background

Data is a fundamental aspect of all government business. We must have the right skills across the workforce to manage our data well and to realise the benefit of data to our organisations. To support the need for data skills across a wide range of disciplines we need a generic approach to assessing and building data competencies.

Some existing competency/skill frameworks cover selected data-related specialisms, sometimes as part of much broader IT-related skills frameworks (e.g. SFIA – Skills Framework for the Information Age). These tend to target a fairly narrow section of an organisation.

The Data Skills Framework is more generic and cross-cutting, it can be deployed more generally across an organisation. It is designed to be light and quick to apply, supporting both broad evaluations and strategic assessments of skills, to identify areas in need of development. It sits alongside profession-specific assessments, such as the Government Statistician Group Competency Framework, as it is designed to evaluate skills across an organisation.

The framework focuses on the basic competencies that individuals will need. However, there are many ways in which these competencies will be applied in different roles. Individuals will require specific subject knowledge and skills in addition to the competencies listed, depending on the exact nature of their role. This will often include specialist skills relating to tools/applications which are listed at the end of each section.

Data related competencies and skills are only one aspect of the overall skill set an individual requires to be effective in a role. Competency frameworks for other aspects can be found elsewhere and the Data Skills Framework is intended to compliment and be used in association with these.

1.2 Use of this framework

The framework can be used in four ways:

i. **Assessing individual competencies and skills** – first and foremost it is a tool to support individual self-assessment of competencies and skills. It provides a clear path for personal development within data themes. Individuals can work with their line managers to define target levels of competency within a role and the training/development to help them move to the next level.
ii. **Defining capability available within an area** – the framework can be used within a particular specialism to develop more tailored versions if needed. The framework is a common structure across these domains, this will support greater integration and flexibility when assessing specialist skills.
iii. **Assessing and tracking overall organisational competency** – the framework can be used to assess the overall competency of that organisation. It can also be used to target parts of the organisation for development and training. Assessment at this level can support enhanced flexibility and staff movement.
iv. **Supporting cross organisational exchange** – because of its generic nature, the framework is applicable across a very broad range of organisations and departments. Its can support broader flexibility and movement of staff.

In whichever manner you apply the Data Skills Framework, the data competencies and the skills in specific tools must be considered together.

1.3 The framework

Seven themes cover the range of skills for working successfully with data. Each theme includes three levels to encourage, assess and track professional development. Where relevant, tools applicable to each theme are listed at the bottom of the section. Experience with individual tools should be assessed in addition to the overall competency level.

The seven data themes are:

i. **Requirements and business analysis** – the ability to understand and prioritise user needs; and identify how data can be efficiently integrated into processes.
ii. **Data governance** – the responsibilities associated with collection, handling, ownership, publication and ultimately removal of data.
iii. **Data management** –knowledge of data concepts, including quality control, storage, and integration with other sources.
iv. **Access and security** – understanding of: the obligations and restrictions around granting and gaining access to data; the principles of open data; different licensing models; security; and the process of risk assessments
v. **Data manipulation** – manipulating, processing, cleansing and combining data for further analysis or use. Automation of complex manipulation on large data volumes
vi. **Analysis and modelling** – analysis of data and data modelling. Includes statistical knowledge, processing chains, machine learning and predictive analytics
vii. **Communication and visualisation** – interpreting, summarising and communicating data and various analytical outputs for different audiences.

There are three levels within each theme:

i. **Foundation** – individual has a good overall basic understanding of the key concepts. They can apply their skills and knowledge to their own work. They know when to seek help on the subject and where to find that help.
   Many people specialising in data will reach foundation level across all competencies to ensure a reasonable breadth of knowledge.
ii. **Intermediate** –individual has reasonable level of direct experience in the competency. They can apply their skills and knowledge to data outside their own area of work. The can resolve issues independently.
   Data specialists are typically intermediate in at least one of the themes.
iii. **Advanced** – the individual has a high level of experience in the theme. They may be a recognised technical expert or be setting strategic direction the field. Individuals at this level tend to be those who are providing a leadership and mentoring role within their organisation.

There are a huge range of tools for managing, manipulating, processing and visualising data, many of them are very domain specific. Broad categories of tool are given below each theme. When you apply the framework feel free to extend and refine the list to specifically identify software appropriate to your organisation and area of work.

The individual's experience level within each tool should be assessed as:

- **Basic** – individual has a basic knowledge of the tool. They are aware of what it does and can use the basic functionality.

- **Practitioner** – individual has made extensive use of the tool. They are familiar with the majority of the functionality it provides (typically 1 to 3 years of extensive use).
- **Expert** – individual is extremely familiar with the tool including the full range of its functionality. They are able to customise operations and they are capable of mentoring others.

2 Requirements and Business Analysis

By understanding the processes and data with which they work, a data specialist can effectively improve the efficiency and effectiveness of their work. To do this they need to understand the dependencies of their parts of the business and the importance of clarity in understanding data and other problems.

This covers process analysis (technical and business process), efficiencies and improvements.

2.1 Levels

2.1.1 Foundation

Understands the importance of continual improvement in processes. Understands the importance of user requirements and how priorities vary.

Individual understands processes relating to their own work. They understand the outcomes that these support and they're aware of how these process affect other parts of the organisation. The person can identify opportunities for improving processes and can evaluate the risks and effectiveness of these options.

Individual must meet all of the following:

- Identifies inefficient and ineffective processes relating to their own work
- Be aware of other parts of the organisation and key stakeholders that depend upon those processes/data
- Knows the people and roles with responsibility for those processes/data
- Asks questions of others to reach a broad understanding of a problem
- Presents options for improvements, including pros and cons
- Understands why it is important to clearly understand the needs of users
- Can prioritise user requirements relating to their own work

2.1.2 Intermediate

Identifies improvements beyond their own work. Is able to identify opportunities for significant improvements leading to financial or resource savings. Presents options clearly for customers and initiates improvements.

Individual is able to approach areas of business beyond their own, explore how existing processes operate and how data is currently used. They are able to review and challenge current systems and explore alternatives to provide higher quality or more efficient processes. They are also able to assess options including: identifying the relative impact on the business; and creating business cases to support any change.

Individual must meet most of the following:

- Understands business processes and opportunities for automation
- Applies a range of techniques to interact with users in order to understand and prioritise their requirements in line with organisational objectives
- Liaises effectively with both technical and non-technical, internal and external stakeholders to develop business opportunities
- Uses a variety of communication options to engage all stakeholders

- Has a good overall understanding of data processing and analysis options and their strengths and limitations
- Uses their knowledge to suggest alternative approaches to existing problems beyond their own work
- Aligns user requirements and business drivers and can draw up the costs and benefits of a proposal
- Collaborates with others to create viable specifications
- Maintains awareness of other similar projects to avoid duplication
- Maintains awareness of emerging technology and business drivers to ensure that delivery uses the most appropriate technology
- Is aware of open source technology and the potential for efficiencies and collaborative business opportunities
- Documents clear and testable user requirements
- Adapts to changing business requirements

2.1.3 Advanced

Excellent knowledge and insight into process improvement. Is able to drive potentially significant business opportunities, including convincing others of the benefits.

Individual is able to fundamentally review significant areas of an organisation and come up with innovative ideas for radical improvements in data/processes. They are capable of joining up disparate groups across the organisation, developing novel approaches to problems. They work at multiple levels of the organisation to develop these further.

Individual must meet most of the following:

- Rapidly builds a good understanding of new areas of business and their processes/data
- Generates radical ideas for how customers could make better use of data
- Establishes engagement for alternative approaches using a variety of techniques
- Establishes business cases to drive change
- Plans for implementation of improvements
- Identifies opportunities and efficiencies arising from new technology, methods or organisational change
- Promotes emerging methods and tools with relevant staff and stakeholders
- Is a respected authority on process improvement, can lead and support others

3 Data Governance

Data is a key asset for any organisation. It must be protected throughout the data lifecycle and data governance describes the ways this is done.

Data is protected in a number of different ways, including: establishing ownership; documentation of data; appropriate storage and sharing; and management of security risks. This section covers roles and responsibilities at all stages of the data lifecycle and measures to ensure that data is protected.

3.1 Levels

3.1.1 Foundation

Understands data is an asset and the basic actions that are required to maintain its value.

Individual understands that data is important and has value to their organisation. Recognises that there are risks involved in working with data and that data value can be eroded.

Takes appropriate responsibility in their own data work and follows the relevant processes for parts of the data lifecycle where they are involved.

Is aware of the responsibilities of others, seeks help and sign-off at an appropriate level.

Individual must meet all of the following:

- Describe their own responsibilities towards data
- Understands the purposes and benefits of data they work with
- Identifies where their work sits within the data lifecycle and be familiar with policies and processes relating to those parts
- Knows data owners/custodians for the data they work with
- Understands that there are risks attached to certain data content (e.g. personal data, national security, commercial confidentiality)
- Has completed the required data/information security training
- Understands that good metadata helps data to be found and used

3.1.2 Intermediate

Understands the data life cycle. Is able to advise on appropriate governance of datasets.

Individual has a thorough knowledge of the data life cycle and is able to take overall responsibility for data throughout this. This includes advising on handling/management procedures from collection all the way through to removal and archive.

Individual must meet most of the following:

- Has a thorough understanding of the data life cycle and risks that need to be managed through-out this to maintain the value of the data assets
- Manages versions of data sources using appropriate principles
- Evaluates the confidence they have in different data sources
- Good understanding of metadata standards and their benefits

- Applies relevant data management/ governance policies outside their own area of work
- Applies governance principles to resolve issues and make decisions (e.g. on ownership)
- Spots risks and contributes to policy improvements in the area of data management
- Shares data governance best practice with others
- Good understanding of data quality. Assesses quality and describes caveats/risks around use
- Identifies and assesses security and legal risks within data
- Ensures all data they are responsible for is only accessed by appropriate people
- Knows and can implement relevant metadata standards
- Identifies suitable data standards/reference data for re-use

3.1.3 Advanced

Excellent knowledge of principles and practice around data governance. Is able to set and refine organisational policies to ensure effective internal processes.

Individual is able to assess organisational policies and practice and ensure these are effective and appropriately balanced against risks. Is able to share knowledge and develop other staff to take responsibility for individual sources.

Individual must meet most of the following:

- Thorough understanding of the risks affecting the value of data and how these can be mitigated
- Good knowledge of relevant policies and best practice both within and outside the organisation
- Advises other areas of the organisation on how to improve their data governance
- Develops and maintains internal policies and processes which facilitate good organisational behaviour and compliance
- Develops staff to ensure that they have the necessary knowledge to maintain data and information for which they are responsible
- Advises on and develops plans to proactively improve data quality
- Meets at least foundation competency in Data Management, Access and Security and Requirements and Business Analysis of this Framework
- Advises on the use and creation of organisational data standards and/or reference data

4 Data Management

Data must be properly managed to ensure it remains usable. There are risks associated with data management and storage.

This section covers specialist knowledge of data and the skills necessary to ensure data is managed in a way that supports efficient and reliable onward use including linkage between disparate data sources.

4.1 Levels

4.1.1 Foundation

Understands different categories of data types and the main mechanisms that exist to support and manage them. Identifies appropriate support for addressing data management issues.

Individual has awareness of different categories of data, including less structured information sources. This includes more specialist sources such as geographic data, raster/imagery data, earth observation data and real-time monitoring feeds. Understanding of the advantages of good data management and the main ways this can be achieved.

Individual must meet all of the following:

- Understands a variety of data storage options within their organisation
- Understands that data has structure, can define elements of structures relevant to their own work (e.g. columns, rows, tables, databases, shape file components)
- Has a basic appreciation of different categories of structured data (e.g. sensor data, geographical information, regulatory data, imagery and earth observation data) and the ways in which each is generated and their differences
- Understands the value of raw data and importance of good dataset definition
- Is able to work with simple datasets (e.g. within a spreadsheet)
- Is aware of the impact of co-ordinate systems on data integrity and use
- Understands that the quality of data can be eroded
- If working with spatial data: Is aware of spatial geometry types (points, lines and polygons)

4.1.2 Intermediate

Able to design, interrogate and summarise more complex or inter-related datasets. Is able to address moderate data management issues independently.

Individual is confident working with more complex data structures which typically underpin various corporate systems and processes. Includes working within database systems, but also has familiarity with concepts to integrate remote sources (e.g. APIs, Web Services). As well as interrogating existing systems, the user is able to optimise the storage and retrieval of data by designing medium complexity solutions.

Individual must meet most of the following:

- Good understanding of database design theory and capable of establishing robust multi-entity databases

- Familiar with a range of query/coding constructs relevant to their organisation (e.g. SQL, SPARQL, VBA) and can competently create and edit statements from scratch
- Good understanding of the different join types including when they should be used for efficient and effective data management
- Familiar with the concept of accessing remote sources (e.g. APIs, Web Services) and how these can be used to access data remotely
- Identifies and addresses problems with existing datasets to improve data integrity
- Strong awareness of organisational data management policies and best practice, and how to apply them
- Able to form and interpret spatial queries
- Understands the concepts and maintenance of spatial databases
- Implements controls to maintain and/or improve data quality
- Identifies and uses organisational data standards where appropriate

4.1.3 Advanced

Excellent in-depth understanding of data storage approaches and able to apply this to improve performance of implementations. Is a known source of knowledge for resolving data management issues.

Individual has a deep technical understanding of the how data is stored and accessed and is able to apply this knowledge to tune and optimise the performance of systems/data processes. This may include refining structures, indexing, process design, and technology choices.

Individual must meet most of the following

- Tunes indexes to improve performance
- Identifies structural changes to improve data integrity or performance
- Locates slow running queries and diagnose performance issues
- Synchronizes data across multiple systems
- Identifies opportunities for improved data management processes to increase data accessibility, security or quality
- Contributes to establishing and communicating data management best practice
- Strong awareness of data and systems which have an impact upon organisation or field of expertise
- Manages and plans for integration issues between strongly differing data across multiple systems

4.2 Tools

- **Spreadsheets** (such as Excel and Open Office) – essential tools for quickly scanning and checking data as well as providing simple data manipulation and combination.
- **Geographic Information Systems** (such as QGIS and ESRI's Arc toolset) – tools for holding, managing and integrating spatial information.
- **Databases** (such as Access, SQL Server, Oracle, PostGres, PostGIS) – powerful tools for manipulating and querying larger volumes of data.
- **Database management tools** (such as Toad, SQL Developer) – provides clear access to database structure for exploring and querying databases.

5 Access and Security

Our data is a huge asset to our organisations and to our customers but there are risks associated with sharing data incorrectly. Data that cannot be freely shared must be appropriately protected.

This section covers awareness of risks and regulations relating to data security. This includes appropriate sharing of data, proactively and in response to requests.

5.1 Levels

5.1.1 Foundation

Understands the principles of data security. Is aware of key legislation and their personal responsibilities for security and sharing of data.

Individual is aware of the main policies around data security, sharing and licensing. This includes awareness of the Data Protection Act and the appropriate action to take in the instance of a suspected or actual breach.

Individual must meet all of the following:

- Knows and complies with the organisation's information security policies and practices
- Knows and complies with the government security classifications
- Applies appropriate controls to share data based on their own area of work
- Completes annual government security training
- Is aware of the Environmental Information Regulations (EIR), Freedom of Information (FOI) Act and any other security/sharing legislation relevant to their area of work (e.g. INSPIRE, RoPSI)
- Understand that there are risks attached to certain data content (e.g. national security, commercial confidentiality, intellectual property, third party permissions)
- Is aware of the concept of open data and knows what data from their own areas of work is available as open data
- Is aware that data can be shared with different types of licence which can restrict how it is used
- Is aware of the key people within the organisation who can advise on security and risk and the procedure to take in the instance of a potential or actual security breach
- Is aware of the Data Protection Act and their responsibilities when handling personal data

5.1.2 Intermediate

Has good understanding of data security, open data and is able to assess risks around data accessibility.

Individual understands and applies principles for data security to data beyond their area of work. Comprehensively assesses risks before sharing data. Embraces the publishing of open data.

Individual must meet most of the following:

- Consistently applies security policies and core security principles

- Assesses data security risks and applies appropriate controls around data accessibility and use
- Interprets licences conditions, and advises on the implications for data use and reuse
- Investigates issues around Intellectual Property and data licensing (including third party permissions)
- Understands any areas of risk specific to your organisation's data. Applies appropriate methods for mitigation
- Knowledge of the Data Protection Act and the implications for their organisation in holding, accessing and publishing personal data
- Understands the concept of, and can apply, the public interest test for FOI, EIR and other data sharing policies
- Understands the importance and value of making data open and supports others in publishing open data
- Understands the processes for publishing open data in their organisations. Applies principles for publishing open data that falls outside normal publishing routes
- Can assess open data against the five-star openness ratings and suggest options for increasing the openness of data

5.1.3 Advanced

Excellent knowledge of data security and sharing principles. Is able to set and refine organisational policies relating to data security and sharing.

Individual is able to assess the overall organisational policies and practice to ensure these are appropriate and effective. Is a known source of expertise in security and sharing. Individual is able to assess more significant security and sharing risks to the organisation and identify appropriate action following incidents/near misses.

Individual must meet most of the following:

- Excellent understanding of risk management and the risk hierarchy
- Produces clear and comprehensive support for less experienced staff carrying out work such as risk assessments
- Identifies and assesses organisational data security risks including the potential impact of new technologies
- Advises on appropriate licencing options for data sharing
- Manages and develops internal policies and processes for data sharing, licencing and/or security which facilitate organisational compliance and internal consistency
- Identifies appropriate options for mitigating legal risks (national security, data protection, third party rights, *etc*)
- Evaluates organisations options for data sharing (e.g. publication platforms)

6 Data Manipulation

Before data can be analysed or used, it often needs to be standardised, cleaned or otherwise formatted to bring it to a required standard.

This section covers reformatting, cleansing and combining of data from different sources for further analysis, storage or use. This includes simple processes to check and transform data, through to automating complex manipulation of large data volumes. This excludes analysis, in-depth domain knowledge and ongoing management of datasets

6.1 Levels

6.1.1 Foundation

Understands basic data processing and preparation. Able to perform simple data manipulations. Understands concepts of data structures.

Individual has a good overview of methods, functions and tools that exist for manipulating data and their strengths and limitations. They can perform basic data manipulations (e.g. within spreadsheets, simple databases, GIS files, CSV).

Individual must meet all of the following:

- Understands the broad capabilities of different data manipulation packages relevant to their area of work and which is suitable for different tasks
- Performs basic data manipulation (e.g. insert, delete)
- Identifies issues with datasets (e.g. duplicates bad data types, nulls, inconsistent formatting), understands the impact of issues, is aware of options for resolving them
- Able to follow defined methodology to cleanse/manipulate datasets
- Basic data cleansing and transformation techniques (e.g. spotting missing and outlying values, recoding items for consistency, parsing, concatenation and format changes)
- Can change the format of datasets between relevant file types (e.g. CSV to Excel, Excel to Access)
- Understands the use and principles of formulae and interpret basic formulae
- Understands that data has structure, can define elements of structures relevant to their own work (e.g. columns, rows, tables, databases, shape file components)
- If working with spatial: Is aware of spatial geometry types (points, lines and polygons)

6.1.2 Intermediate

Good knowledge of data manipulation options and principles. Able to combine data sources, create new functions and automate simple processes.

Individual is capable of manipulating data using a range of tools/packages, selecting the most appropriate option for the data. Capable of manipulating data in ways that can be relied upon to work in an automated or sequential process.

Individual must meet most of the following:

- Can develop and use and range of functions/formulae

- Able to manipulate data using methods such as: lookup tables, decodes, case statements, aggregation, (de)normalisation, spatial intersection, buffering, cutting, point-on-polygon, editing geometry
- Work with data in and from relational databases
- Able to automate simple processes
- Ensures errors are captured during automated manipulation
- Understands how to manipulate and process very large datasets (including newly emerging tools and techniques)
- Can create simple processing chains/workflows
- Identify issues and implement processes for standardising/correcting data
- Good understanding of processing methods for different data and able to create these using the most appropriate platform
- Can work with loops, functions, variables, and numeric and string manipulations in manipulation packages or scripting languages
- Able to appropriately integrate disparate datasets based upon common factors.
- Able to convert data between different spatial formats while maintaining data integrity
- Understands the principle of Create, Read, Update and Delete (CRUD) actions on a database

6.1.3 Advanced

Highly experienced in data manipulation. Creates automated processing chains and can manipulate large volumes of data. Is a known source of knowledge for data manipulation best practice.

Highly experienced data manipulator able to fully automate complex processing chains and create processing functions and packages. Individual is confident at linking, combining and manipulating multiple datasets including very large remote datasets.

Individual must be able to meet most of the following:

- Able to fully automate complex processes
- Anticipates failures in automated processes (system and data) and mitigates risks through good design and development
- Creates complex functions and packages for processing data
- Links, combines and manipulates multiple, complex datasets
- Good understanding and experience of working with very large volumes of data. Is aware of emerging technologies and techniques
- Establishes and manages large scale processing tasks for transformation of data including integrating different data categories (e.g. sensor data, geographical information, regulatory data, imagery and Earth Observation data)
- Develops and runs more complex processing (e.g. segmentation algorithms, heavy weight processing, extensive spatial integration)
- Good familiarity with coding in at least one language and familiar with major coding constructs
- Understands Object Orientated concepts and able to design and implement code using this approach
- Develops and implements organisational best practice/policies for data manipulation
- Develops and supports staff to ensure that they have the necessary skills for data manipulation

6.2 Tools

- **Spreadsheets** (such as Excel and Open Office) – essential tools for quickly scanning and checking data as well as providing simple data manipulation and combination.
- **Scripting platforms** (such as Python and R) – extremely powerful flexible platforms which can integrate packages of functionality built by others which can be used to script data transformations.
- **Geographic Information Systems** (GIS) (such as QGIS and the ESRI Arc suite) – these tools are designed to support flexible manipulation of geographic data such as polygons.
- **Databases** (such as Access, SQL Server, PostGres, PostGIS) – powerful tools for querying and manipulating larger volumes of structured data including, in many cases, spatial data.
- **Image analysis tools** (such as ENVI, ERDAS and eCognition) – include capability to visualise, manipulate and transform Earth Observation data.
- **Extract, Transform, Load (ETL) tools** (such as Informatica and FME) – powerful platforms designed to automate large scale data transformations including the ability to interact with a broad range of proprietary formats.

7 Analysis and Modelling

Analysis turns data into meaningful information. Statistical analysis and modelling can provide insights into the current and future state of our environment. Thoroughly understanding data, and the stories it can tell, ensures that the right conclusions can be drawn and the right decisions made.

This section covers analysis of data (including imagery and spatial data) and data modelling. It includes knowledge of analytical methods and, unlike the data manipulation section, will require domain knowledge of data. This section also covers statistical knowledge, environmental modelling, machine learning (automated identification of potential patterns in datasets) and predictive analytics (making predictions about future events).

This section does not include preparatory data processing which is covered by data manipulation skills

7.1 Levels

7.1.1 Foundation

Broad understanding of types of data analyses that are possible. Understands and applies basic statistical theory and techniques.

Individual has an understanding of the importance of good data analysis. They understand that selecting an appropriate method is important. They are able to work with data and perform simple manipulations and summaries. They can ask questions about data to better understand it.

Individual must meet all of the following:

- Awareness of the concept of forming and testing hypotheses to objectively assess ideas
- Broad awareness of a range of analytical methods and techniques suited to their area of work and how these are used in their organisation
- Understands different forms of data (e.g. quantitative/qualitative; discrete/continuous; count; dichotomous; time series)
- Understands summary statistics (e.g. mean, standard deviation, median, counts)
- Able to carry out basic data analysis using relevant software to form narratives from data
- Understands of the principles of uncertainty, probability, bias and variability
- Understands the importance of domain knowledge to support good analysis
- Recognises the importance of a good, representative sample
- Questions the strengths and limitations of data they work with
- Understands that extreme values within data influence analyses

7.1.2 Intermediate

Good understanding of the theory and application of a range of analytical techniques. Uses appropriate methodology when working with complex or unfamiliar data. Draws compelling narratives from data.

Individual has strong numerate skills and a good understanding of broad statistical and modelling concepts. They are familiar with the theory of a broad range of statistical

techniques and have experience of applying these in a particular domain. They can analyse large datasets and draw out meaningful messages.

Individual must meet most of the following:

- Accounts for uncertainty and bias in analyses
- Able to question the source, content and collation/processing of data to gain domain knowledge
- Understands that data drives analytics. Consistently selects the right tools and methodology for the data
- Able to combine sources of data, including datasets from outside their organisation
- Familiar with a broad range of data transformations and applies these appropriately
- Able to independently write functions and has a good understanding of variables
- Familiar with the concept of scripting to create repeatable processes and is able to generate simple scripts
- Correctly uses ideas of randomness, chance, and likelihood to make judgements about uncertain events
- Accounts for the dependencies and complexities of dealing with spatial data
- Broad understanding of appropriate quality assurance methods and uncertainty analysis
- Familiar with a range of techniques for identifying relationships between data and what can be inferred from these (linear regression and correlation). This includes regression techniques for a range of data types (e.g. logistic regression)
- Familiar with concepts and approaches of probability for comparing variables and relationships including the concepts of covariance
- Able to deeply question data and find causes of variation (e.g. a true trend or just a change in collection method)
- Identifies questions that can and cannot be answered due to data design

7.1.3 Advanced

Thorough understanding of a broad range of techniques for modelling complex and dynamic systems. Capable of establishing and running complex analyses. A recognised source of knowledge on analysis and modelling.

Individual is familiar with the theory and use of advanced modelling techniques suited to more complex and dynamic systems. Individual has the ability to assess the limitations of their techniques and provide insight into the outputs. Individual is capable of using programming languages to establish and run complex tasks.

Individual must meet most of the following:

- Excellent knowledge of data sources available (within and outside their organisation)
- Designs and creates compelling analyses to influence and inform policy/decision making by senior stakeholders
- Works across multiple sources of different data, combining structured and unstructured data
- Good understanding of methods for fast manipulation of data
- Very familiar with scripting, able to create new packages including interacting with external libraries
- Familiar with Object Orientated programming
- Understanding and experience of predictive analytics and machine learning

- Familiar with multivariate techniques such as Principle Components Analysis, Discriminate Analysis, Multidimensional scaling, *etc*
- Familiar with a range of techniques and theory for modelling dynamic systems (e.g. Lotka Volterra equations)
- Able to apply stochastic modelling of systems such as population modelling
- Able to interrogate data to great depth and communicate the narratives of the data with authority
- Develops and supports staff to ensure that they have the necessary skills for analysis and modelling

7.2 Tools

- **Spreadsheets** (such as Excel and Open Office) – can be used to perform quick exploratory or simple analyses.
- **Scripting platforms** (such as Python and R) – extremely powerful flexible platforms which can integrate packages of potentially advanced analytical functionality built by others.
- **Geographic Information Systems** (GIS) (such as QGIS and the ESRI Arc suite) – include functionality to run spatial analyses.
- **Databases** (such as Access, SQL Server, PostGres, PostGIS) – powerful tools for summarising and combining larger volumes of structured data including, in many cases, spatial data. Analyses tend to be fairly straightforward.
- **Image analysis tools** (such as ENVI, ERDAS and eCognition) – includes the capability to process Earth Observation data as well as combining with other sources.
- **Machine learning platforms** (e.g. H20, Apache, Amazon, Azure).

8 Communication and Visualisation

Data can be a highly technical subject and it can be difficult to draw out information for decision making. Clear and meaningful visualisations help to communicate stories from data and analyses.

This section covers summarising and communicating data and analyses for different audiences. It includes using appropriate means of communication and responding to user feedback.

8.1 Levels

8.1.1 Foundation

Understands how to summarise, communicate and present data outputs from their own area of work for a range of audiences. Is aware that different tools can be used to support visualisation.

Individual has a basic understanding of the range of options available for visualising data including charts, graphs, maps and networks.

They know how to interpret and explain visualisations. They can create visualisations to help people better understand data and analyses.

Individual must meet all of the following:

- Understands the importance of data visualisation in communicating to stakeholders
- Has basic understanding of a broad range of data visualisation methods
- Understands the benefits of using visualisation to communicate technical concepts to non-technical users
- Creates simple visual outputs such as: histograms, scatter plots and basic maps
- Understands the importance of good design to improve users' experience
- Asks appropriate questions about data to create representative visualisations
- Ensures their visualisations do not misrepresent data or analysis
- Puts visualisations into an appropriate context (e.g. including caveats)
- Clearly articulates the meaning of analytical outputs without jargon
- Takes account of accessibility issues relating to visualisations

8.1.2 Intermediate

Communicates with stakeholders about data and analyses using a range of tools and media for visualisation. Designs visualisations to suit customer needs.

Individual is able to design, create and present clear interpreted summaries of data and analyses. They communicate key concepts using textual and visual means to a range of technical and non-technical audiences.

Individual must meet most of the following:

- Creates visual displays of data through a range of tools and analytical packages
- Applies cartographic design principles to map outputs (printed or digital)
- Identifies simple improvements to data visualised in mapping services

- Good understanding of the importance of effective, targeted communication to different audiences to improve policies and decisions
- Creates clear blogs articles or reports on data-related subjects targeting a range of audiences
- Good understanding of the key design aspects of creating web content (e.g. audience, design)
- Good understanding of the technical design aspects of creating web content (e.g. uploading data, different user interfaces)
- Able to work with communication experts to improve visualisations
- Combines visualisations/text from multiple sources to tell a cohesive story
- Applies principles of good data visualisation outside their own area of work
- Designs content that is readily repeatable/updatable over time

8.1.3 Advanced

Excels at communicating about data and analyses through text and/or visualisations. Embraces a wide range of routes for communications. Is a recognised source of knowledge and leadership in good communication/visualisation of data.

Individual is able to work across large parts of an organisation and draw together content from widely differing sources. They are capable of communicating highly complex messages through clear and easily understood content. They find and lead on innovative solutions for communication and visualisation.

Individual must meet most of the following:

- Uses of a range of user experience methodologies to create better visualisations
- In-depth understanding of the importance of user experience
- Able to analyse web metrics and interpret results to improve services
- Thinks creatively about how to present output including working with external experts such as visual artists and digital marketers
- Builds and presents dynamic data driven content for the web
- Maintains and configures mapping services
- Optimises performance of mapping services and resolves issues
- Advises and supports others in clear communication of data and analyses
- Forms policies and guidance for visualisation
- Designs and creates compelling visualisations and text to influence and inform policy/decision making by senior stakeholders

8.2 Tools

- **Spreadsheets** (such as Excel and Open Office) – allow simple but quick visualisation through graphs.
- **Scripting platforms** (such as Python and R) – some such as R include good visualisation functionality associated with analytical functions. Also include capability to link well with web based delivery.
- **Geographic Information Systems** (GIS) (such as QGIS and the ESRI Arc suite) – include functionality to visualise spatial data in a range of ways.
- **Image analysis tools** (such as ENVI, ERDAS and eCognition) – include capability to display and visualise Earth Observation data in a range of ways.

Appendix 1: Example of applying framework

If you use this framework within a specialism, such as GIS, you can provide additional guidance. This is particularly important where specific tools are used.

This annex sets out such an example for the role profile of a GI Analyst working for a nature conservation organisation. The organisation is primarily using the ESRI toolset but currently looking at moving to the open source equivalent (QGIS).

The individual will work within a broader GI team but will be associated with a specific area of the business where they will act as a conduit for the requirements. They will deliver some requirements themselves and integrating with the broader GI team as needed.

	Entry	**Fully competent**	**Notes**
Requirements and business analysis	Foundation	Intermediate	Should include a broad understanding of nature conservation and knowledge of how data is currently being used
Data governance	Foundation	Intermediate	Familiarity with specific issues around spatial data versions and relations to other sources
Data management	Foundation	Intermediate	Having a basic knowledge of databases particularly for the storage and querying of spatial data
- MS access	Basic	Practitioner	Including building queries through the visual tools
- Post Gres	Basic	Practitioner	Being able to access spatial databases from standard tools and run basic queries within the database
Access and security	Foundation	Intermediate	Significant portion of the role is around supply of data and strong familiarity with third party ownership and access issues is important
Data manipulation	Intermediate	Advanced	Primary focus is on editing, manipulation and querying of spatial data
- Arc GIS (manipulation)	-	Expert	Strong familiarity with this tool is important
- QGIS (manipulation)	Practitioner	Expert	Strong familiarity with this tool is key
- Python (manipulation)	Basic	Practitioner	In particular building good knowledge of integration with libraries associated with GI tools and being able to automate processes
Analysis and modelling	Foundation		Basic processing within desktop tools but working

			towards being able to run processing within other tools
- R (spatial modelling)	Basic	Practitioner	Developing knowledge of the sorts of spatial processing possible within R
Visualisation	Foundation	Intermediate	Primary focus is the generation of maps
- QGIS (visualisation)	Basic	Practitioner	In particular being able to generate publication quality maps

Note that in addition to these data related competencies the individual would also be expected to have a range of more generic competencies such as communication, project management, leadership, *etc*.