

Unit 5 - Digital Modelling and Data Management

Overview

The candidate can understand and investigate some of the ways data is managed and how specific data, such as numerical and financial data, is modelled to make predictions. Learners will investigate the best ways to create models to examine budgets or similar systems. They will explore the means and methods for managing the increasing volume of data that is part of their everyday lives.

A work activity will typically be 'non-routine or unfamiliar' because the task or context is likely to require some preparation, clarification or research to separate the components and to identify what factors need to be considered. For example, time available, audience needs, accessibility of source, types of content, message and meaning, before an approach can be planned; and the techniques required will involve a number of steps and at times be non-routine or unfamiliar.

Example of context – create a system that tracks income and expenditure for a small business and outputs charts and graphs. Also saves in the right format to the right media.

Assessor's guide to interpreting the criteria

General Information

RQF general description for Level 2 qualifications

- Achievement at RQF level 2 (EQF Level 3) reflects the ability to select and use relevant knowledge, ideas, skills and procedures to complete well-defined tasks and address straightforward problems. It includes taking responsibility for completing tasks and procedures and exercising autonomy and judgement subject to overall direction or guidance.
- Use understanding of facts, procedures and ideas to complete well-defined tasks and address straightforward problems. Interpret relevant information and ideas. Be aware of the types of information that are relevant to the area of study or work.
- Complete well-defined, generally routine tasks and address straightforward problems. Select and use relevant skills and procedures. Identify, gather and use relevant information to inform actions. Identify how effective actions have been.
- Take responsibility for completing tasks and procedures subject to direction or guidance as needed.

Requirements

- Standards must be confirmed by a trained Gold Level Assessor or higher.
- Assessors must at a minimum record assessment judgements as entries in the on-line mark

book on the INGOTs.org certification site.

- The work in the unit is recommended in order for candidates to have covered enough depth and breadth in the topic to successfully carry out their controlled assessment and take the external exam.
- When the candidate has covered as much of this material as necessary to complete the controlled assessment element, they may be introduced to the topic
- This unit should take an average level 2 learner 25 hours of work to complete.

Assessment Method

This unit will be assessed synoptically via a controlled assessment and also through an external examination.

Expansion of the assessment criteria

1. Understand the range and type of data and data management

1.1 I can list some data types and their characteristics

Learners will list some of the data types they will use in this qualification as well as their main characteristics

Additional information and guidance

For this a working knowledge of digital file types will be required.

File types covered by this section could be presented as a table as

Below:

Type	Proprietary	Open	Other	Comment
Word processing	.docx	.odt	.rtf	Used primarily for documents
Numerical data	.xlsx	.ods	.csv	Used for numbers, but can have text
Image files	.bmp	.svg	.png	Files have different properties based on use and purpose

There are extensive file types and the above is only an example.

Other files to explore could include: .pdf, .txt, .pptx, .htm, .html, .jpg, .gif, .wav, .mpv, .mp4, .mov, .wmf, .mp3, .psd etc.

Learners will also be able to describe the advantages and disadvantages in saving information in proprietary versus open standard formats.

1.2 I can explain the range of data available and its purpose

Learners will show a clear understanding of the relationship between formats and purpose

Additional information and guidance

Many data formats have evolved in relation to what they do. The most obvious example here is the proprietary document processing format and the extension it has of .doc for an abbreviation for document. Many explain their purpose in their abbreviated file extension, for example .png (Portable Network Graphics), though it is not always clear what this means in practice.

This criterion extends the list of files gathered as part of 1.1 and allows learners to demonstrate a deeper understanding of their purpose. The above identified .png files suggest that they are easy to move around between networks, but their key purpose is arguably the ability to be used in layering as they can support transparency. Also portable are .jpg files as they are generally a small size, but this size is accomplished by using lossy compression which means the portability is a sacrifice of quality over size.

Learners can possibly extend the table in 1.1 to add these purposes to the files they identify and list.

1.3 I can explain the different data management requirements

Learners can demonstrate the range of data management needed

Additional information and guidance

Data is growing at an incredible rate. Just a few years ago, having a GB of space available to use would have been incredible, but now many people get 50GB of data free from online hosting companies. It is likely that most learners will have multiple data storage facilities for their different digital media. Smartphones generally come with a facility to store and manage pictures and other digital artefacts, but how do companies cope. TLM is a small company, but has to manage 0.5TB of data which is growing daily. This is a problem. If it is all managed online, what happens if that online company has an issue. It is rare, but does happen. If it is managed locally, what happens if something happens to the physical company. It is estimated that 60% of companies that lose their data will be closed within 6 months of the loss. If all of your life (short though it currently is) gets lost from an online site, how upset would you be?

These are some of the different data management issues that learners need to consider. Some questions to explore are:

- What are my own personal data management needs?
- What are the needs of an organisation like my school/college?
- What are the management needs of a company such as my bank?
- What are the management needs of my local doctor?
- What are the management needs of local or national government?

In each of these cases, what kind of data is required to be managed?, for what purpose? and for what time scale? etc.

1.4 I can describe the characteristics of “big data”

Learners can briefly describe the phenomenon known as big data

Additional information and guidance

As the previous criterion hopefully illustrated, there are enormous amounts of data being stored and managed for different purposes. With the advent of the Internet of Things, this will grow exponentially. Governments are already struggling to consider what to do with all of this data, though it undoubtedly has great potential for good or harm. The following site illustrates clearly the growth of big data in a graphical way.

<https://www.internetlivestats.com/> [1]

Many definitions of big data focus on their scale in comparison to previous data, such as analogue data. Wikipedia has a useful [5 V definition](#) [2] which seems useful:

Volume

The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.

Variety

The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

Velocity

In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

Variability

Inconsistency of the data set can hamper processes to handle and manage it.

Veracity

The data quality of captured data can vary greatly, affecting the accurate analysis.

There are other definitions based on different characteristics, but the essential thing is that this is very different to just a few years ago and is expanding exponentially which means all the management aspects are also increasing almost too quickly to work.

1.5 I can understand the need to manage data for different purposes

Learners will show they understand that that all data requires some form of management

Additional information and guidance

Learners will have demonstrated an appreciation for all the different forms of data and their characteristics. This knowledge will help clarify how this data can be managed. Most learners might well use something like a USB drive to move their own data around from home to school, or might use a cloud based service. This is fine for some sizes of data, but what about when the volume gets into “big data” territory. High Definition video is usually around 25Mbps/sec (video and sound), or 3.125 MB/sec. A 5 minute video would need almost 1GB of storage ((3.125 X 60) X 5).

Note, learners need to know the difference between bits and bytes. A byte is 8 bits, so a Megabit is 125,000 bits. Many measures are in Megabits.

This very small example shows the need for huge amounts of storage. In the UK, we have the most CCTVs in the advanced world. Each one of these cameras is recording video content 24 hour a day and 7 days a week. It is probably rolled over at some point, but this is massive amounts of data to store. Most homes are now introducing IoT devices which means even more data being sent and managed somewhere. Some of this can be stored on hard drives, but it is still cost effective to use tape drives. How cost effective is cloud based storage. Putting 1TB on your local network and managing it yourself is relatively cheap as a 1TB hard drive is currently about £50. However, adding 1TB to a cloud based system would cost a significantly larger amount. The cheapest dedicated server to host material on the cloud with 1TB of storage would be £50 a month.

Learners need to give some examples of different data requirements and management needs.

1.6 I can define the issues related to managing data

Learners will show an understanding of some of the main issues related to data management

Additional information and guidance

Some of this criterion will be incorporated into others above, particularly looking at the physical aspects of size and volume etc. However, there are other significant issues. For example, who does the data belong to? If you create an interesting games based app while at school using all of the school's facilities and software, does that app belong to you, or the school? Schools, for example, are expected to keep information relating to your GCSEs for a number of years. If your GCSE was in art and required you to generate multiple GB design files or video sequences, is it fair to ask your school to manage this material for you for several years? What are some of the other legal issues? Many students are not familiar enough with the law on copyright and quite often download material to their own drives in school to use for some purpose. Is this now the school's responsibility or the individual student? Who is ultimately responsible for the data?

The large data centres required for big data, such as the Google data centres, consume more power than large cities to keep cool. This means enormous amounts of energy which is likely produced by burning fossil fuels. A similar finding is now coming out for companies trying to mine Bitcoin as it is very CPU intensive so makes the computers run very hot. All of this waste energy is causing long term climate problems, but is the collection and analysis of data more important? Who decides?

Learners need to address and answer some of these questions to show that they fully understand the entire complexity of data management.

2. Plan and apply effective data modelling and management practices

2.1 I can make a plan for a data modelling solution

Learners will show that they can plan out a basic solution to a problem requiring modelling

Additional information and guidance

Data modelling varies a great deal, but in most instances, at least for this criterion, we mean the modelling of numerical data. Most organisations need to have some way of making short, medium and long range decisions based on what their current income is and what their income and expenses are expected to be in the future. A model using a spreadsheet is an excellent way to make plans and decisions for the future. Knowing the rough income versus expenses from the present, most companies can make accurate predictions about the future. If the company has had an increase of roughly 5% over the last 3 years, it seems likely, though not definite, that it will be 5% or so next year. They can then use this information to make a decision such as hiring a new person.

If learners are asked by a local company or charity to make a data model for them to track their finances, what sort of information will they need. What is the best information to gather to begin planning and how much and of what nature is the data they require to make a good plan.

Is it important to know timelines?

Should the plan be for the next 6 months, or can it be more ambitious and be something like the next 5 years.

2.2 I can gather the requirements from a client for their needs

Learners can demonstrate accuracy and competence in data gathering

Additional information and guidance

Listening to a client's needs and creating documents like questionnaires to get exactly what you need for a successful project are difficult skills but will make a huge difference to the project. It may be worth doing some cross curricular work with the English department on what makes clear questions and how to construct questions that will get the responses required. It may also be useful to do some pre-preparation on interview techniques to give learners confidence in asking questions and noting down responses.

Other devices such as digital audio recorders (or smartphones) can be deployed in this exercise to guarantee accuracy.

2.3 I can make a workflow diagram to explain the capture, processing and output of data

Learners need to demonstrate they understand the differences between these elements

Additional information and guidance

What is the best and most effective way to capture the data that you need? How will that data be processed and what impact will the processing have on the final output? If the system is a budgeting system, then the input is likely to be the income and expenditure of the company. How is it captured. Do people come in to a company and pay mostly by cash. If so, how will that process be captured. How is that captured data organised in order to make it more efficient to process. In most accounts systems they will use some type of debit and credit system. When someone pays for an item from a shop, how much of that payment comes back to the company? From that item, some money will go towards the wages of people working there, some will pay for materials to make the item, other money will go towards paying local and national taxes, some money will go to research and development of new items. All of this needs to be processed. Once this processing has occurred, the system should be able to answer "questions" for the company. "If we pay the employers more money, how will that affect our balance?", "How much money did we make in January because of the bad weather?" etc. The output from these questions will likely be in the form of a report on a word processed document, or a spreadsheet chart, or a presentation. How is the output managed?

2.4 I can explain the systems and processes for modelling and managing data

Learners need to explain how their solution will work

Additional information and guidance

The systems and processes here will vary, but in general will be some indication of the nature of the system, in most cases some form of spreadsheet, and an indication of the way data will transition through it. In the previous criterion learners will have explored the different inputs, processing and outputs, so this can be applied here as an illustration of the system. In terms of data management, this will be additional detail as to the technologies used to capture and manage data. The modelling side will be the way that data is managed in the spreadsheet, how it is tested and analysed and what outputs are produced. In terms of managing data, the learners should address where digital data is

stored, who has access to it and for what purpose, how long it is stored and in what way.

2.5 I can describe to a client how to manage the data from the system

The emphasis here is on the plans in place to make data manageable

Additional information and guidance

As investigated in other areas of this course, learners will hopefully have a good grasp of the characteristics of different data and have some idea of the mechanisms used to manage it. It also depends on what level of management their client needs. If they have highly sensitive or critical data that can not be leaked in any way, then this will determine what sorts of systems and services are required. If the data can in some sense be made public, then this will facilitate a completely different approach to management. How long does data need to be kept and in what order. Many people are now discovering that data that was made on something like a CD is degraded and can't be used as the life of that medium is actually quite short. Should everything be in the cloud? What are the implications of this.

2.6 I can establish systems and processes for data handling and storage

The learner needs to create a working system to be approved by a client

Additional information and guidance

Learner will create a working system to manage numerical data and have systems in place to deal with the data created for the storage requirements specified by the client.

3. Evaluate the effectiveness of tools and services for data modelling and management

3.1 I can evaluate the effectiveness of my solution

Learners will be able to demonstrate that they can evaluate the strengths and weaknesses of their developed solution

Additional information and guidance

All systems are designed to have a continuous development cycle as nothing is perfect and changes over time and changes in hardware will make system changes necessary. Learners should be able to understand aspects of their project that will inevitably change and be able to understand how and why they might change. This will be evidence by a reflective overview of their work. Statements such as, "My system was relatively fast, but the client's finances meant that the available computer processing power was a limitation, however, as power of CPUs increases and costs decrease, this will make the solution faster over time."

They might also reflect on their own skills and knowledge in terms of how they might improve perceived weaknesses. It needs to be stressed that a "weakness" is merely a stage in a journey. No person can do things perfectly all of the time and the learners should not think that signalling a weakness is a problem for them. There is more of a problem if you can't identify a weakness as you will continue making the same mistakes and never improve.

3.2 I can analyse the process used to develop my solution

Learners will show a level of understanding of analysis

Additional information and guidance

Analysing the process you adopted is to look for ways of being more effective or efficient. In other sections the learners will have tried qualitative and quantitative measures to have hard evidence of how well, or not, their solution is working. They could be looking at quantitative measures, such as

```
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)})(window,document,'script','//www.google-analytics.com/analytics.js','ga'); ga('create','UA-46896377-2','auto'); ga('send','pageview');
```

how many transactions the system can deal with, for example. Over time and with more experience learners will develop their own processes that work for them and there is probably not hard rule. The buzz word for developing solutions these days is **agile**. In the “good old days” people would have set a target and doggedly worked towards it, even if there were problems, assuming that it would all work out. With agile development, small teams work on small problems and solve these are part of the whole. This makes it much easier to react to issues and fix them rather than needing to do a large fix. This modular approach is far more effective and efficient. Learners just need to cite some examples from how they carried out their project and discuss why they did them and give some idea of how well they worked. There is not necessarily a right and wrong answer here as everyone works differently. Some people always leave things until the last minute and still get it done, but most people can’t make that work.

3.3 I can describe the strengths and weaknesses of files types for data management

Learners will describe in detail their understanding of file types and their attributes

Additional information and guidance

Learners will have already investigated some different file types for their project and be familiar with many in their day to day use of IT. What features are critical for certain tasks and why? In their own project, did they have to compromise in any way? For example, a small charity may have required a spreadsheet application to manage their accounts and also needed an office suite for general use. The proprietary systems are expensive and maybe not suitable, though they do have great features. You have recommended they use an open source alternative which works well, though it does not have all of the features they need. In this instance, the decision has been a pragmatic one: they have sacrificed some functionality for the sake of saving money over the long term. The money saved could be paid towards employing a developer to add extra features to the open source option or possible to give detailed help through training sessions. This compromise is due to an understanding of the features of different packages and their data types.

3.4 I can evaluate the best systems of data modelling and management

Learners will provide a rationale for their customer’s final product

Additional information and guidance

As with the previous criterion, the final product a customer gets, in most instances, will be a compromise between what they want and what they can afford. The system you specified and created may well not fit all of their needs, but no system will. How well did it meet certain needs and what further developments could be made to make it better over time. In an earlier criterion the idea was discussed that the development in technology will make systems better naturally, though this will involve cost and upgrade issues. Is this enough? Is this the designer’s concern? If you have specified a storage solution that was the best that they could currently afford, will this be enough in the future? Is it possible to future proof systems? One data storage solution online offers to give you “2TB for life” saying that is 99 years. Is that enough. Probably not.

Are there any good online systems for data modelling and management? What are their limitations. Issues of interoperability will always be important.

3.5 I can produce some support materials for data modelling and management

Learners will produce a short help guide for their product

Additional information and guidance

Most support desk personnel complain that people never read the user guide, which is mostly true, but it doesn’t mean the company should not produce one as it will be helpful. If the learner makes some screenshots and notes as they go along, they can quickly put together a short guide on the main features of their system. They may include elements such as key features of worked examples

of how to do certain functions. If they prefer, they can make these online or as a series of short videos.

3.6 I can prepare and deliver a presentation on data modelling and management

Learners will present their solution to an audience

Additional information and guidance

Any product needs to be sold to a potential or actual customer once completed and a “sales pitch” is a useful way to show off the system. Using a presentation format allows the learner to think about the best way to demonstrate it and practice presentation skills. This can be a presentation of aspects of their user guide, but may also include some of their reflections and thoughts on strengths and weaknesses if appropriate.

This presentation can be recorded if required as it will be useful for learners as part of their IT portfolio when applying for college places or career opportunities.

Source URL: <https://theingots.org/community/sil2u84x>

Links

[1] <https://www.internetlivestats.com/>

[2] https://en.wikipedia.org/wiki/Big_data