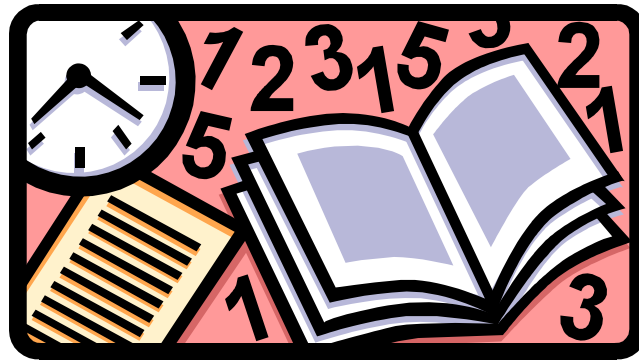
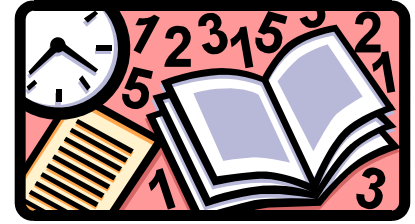


Data, Information & Knowledge





Forms of Information

In this topic, we will be looking at:

- What we mean by the word "data"
- The four basic types of data
- The difference between data and information
- Processing and encoding

What is Data?

Data is a stream of raw facts representing things or events that have happened.

In ICT, we usually say that data is made up from four basic **types**:

- Numbers
- Text
- Images
- Sound

Storing Data?

Inside the computer, however, all data is stored as numbers:

- Numbers...
- Text...
- Images...
- Sounds...

Storing Data?

Inside the computer, however, all data is stored as numbers:

- **Numbers** - are stored as numbers, obviously!
- **Text** characters are stored as a code that represents each - e.g. ASCII
- **Images** are stored as numbers representing the amounts of red, green and blue for each pixel
- **Sounds** are stored as numbers representing the loudness at given intervals

How are the numbers represented?

If you want to store and communicate numbers electronically, how can you do it?

- **Analogue** - you could use a range of voltages, e.g. 5V for the number 5
- **Digital** - where the number is represented in binary form, e.g. the number 5 would be stored/sent as 101 (using pulses of electricity)

Number Bases

- You will be familiar with the decimal system, and hundreds, tens and units, and a digit from 0-9 in each column
- **Binary** works in a similar way, except that you have units, twos, fours, eights, and each column only contains 0 or 1, e.g. 5 is 101 because it's one 4 plus one 1:

128	64	32	16	8	4	2	1
0	0	0	0	0	1	0	1

Bits and Bytes

- Each of the 0s and 1s is called a **bit** - binary digit
- Eight bits together form a **byte**
- The longest number a computer can handle in one go is called a **word** - e.g. a 32-bit computer has a 32-bit word length
- 1024 bytes is a **kilobyte**, 1024 kilobytes is a **megabyte**, 1024 megabytes is a **gigabyte**, etc

Why Use Binary?

- **Bits** are easy to represent electronically, with 0 being represented by no signal, and 1 being represented by a signal
- These signals could be **pulses** of electricity, flashes of light - even down to individual photons or electrons
- Analogue signals are affected by changing resistance, heat and interference

Adding Meaning

Here's an example of some data:

21 0769

But what does it mean?

Adding Meaning

Here's an example of some data:

21 0769

But what does it mean? It's just numbers - it's raw data!

If I were to add some formatting so that the numbers read *21/07/69* you can see that it becomes a date. The formatting has added meaning!

The meaning might only be clear to a European person, though - in the USA the date would be *07/21/69* and in Japan it would be *69/07/21*!

What is Information?

- Information is data that has been processed to make it *meaningful* and *useful*
- **Data + Meaning = Information**
- Another way to add meaning is to process the data. For example, individual exam marks are raw data, but if you were to process those to say that the average mark for the class was 53%, or that boys did better than girls, or that 76% of the students in your school got a grade A or B, then that is information!

Data or Information?

- The number of newspapers sold today
- The name of the best-selling newspaper
- The increase in house prices over the last year
- 352098527
- Your mark for this week's assignment
- The average mark for the assignment

Encoding Information

- Processing turns data into information
- Sometimes you might want to turn information into data - i.e. to store it - this is called **encoding**
- How do you code information to make it easy to re-process, without losing it's meaning?

Encoding Example

- Often surveys have questions like this:
 - A level ICT is brilliant!
 - Disagree strongly
 - Disagree
 - Neither agree nor disagree
 - Agree
 - Agree strongly
- How would you store the responses on a computer?
- What would the **user interface** look like?

Where is the data from?

Internal or External?

- **Internal** communication is communication with people inside the same organisation or company
- **External** communication is with people outside the company, such as suppliers or customers.

Direct or Indirect?

- **Direct** data is collected for the purpose of the processing being undertaken - e.g. time cards for pay
- **Indirect** data was originally collected for another purpose, but is now being processed to provide extra information - e.g. spending patterns from credit cards

Information Channels

Formal or Informal?

- **Formal** channels are the official (or reliable!) ones, such as memos, letters, the company noticeboard, etc.
- **Informal** channels are the unofficial ones, such as office gossip, informal meetings and rumours - these can often be unreliable.

The Value of Information

- It is often said that we are in the **information age**, and that information is a valuable commodity.
- Why is information valuable? Because:
 - It allows us to **plan** how to run our business more effectively - e.g. shops can stock what customers want, when they want it, and manufacturers can anticipate demand
 - Marketing materials can be **targeted** at people and customers that you know could be interested in your products and services
 - This can lead to increased **customer satisfaction** and therefore **profit**

Good Quality Information

- The characteristics of good quality information - it should be:
 - Accurate
 - Up-to-date
 - Relevant
 - Complete
 - On-time
 - Appropriately presented
 - Intelligible

Collecting Information

How is information about people collected?

1. Obviously you can ask people questions about their spending habits, etc. (but they might not like it!)
2. Or you can use a more **indirect** approach:
 - Supermarket loyalty cards
 - e.g. easily identify wine-drinking vegetarians!
 - Credit card transactions
 - amounts and locations
 - can help prevent fraud, too!

Coding Information

- Information stored in a computer is often **coded**
- Coding **categorises** information and can replace long, description strings with a few **letters** or **numbers** (or both!)
- You are probably familiar with examples such as F for female and M for male

Coding - Advantages

Information is often coded because:

- It is **quicker** to enter into the computer
- It require **less disc space** to store, and less memory to process
- It can make **processing** easier - or possible - as there will be fewer responses
- It improves the **consistency** of the data as spelling mistakes are less likely

Coding - Disadvantages

Coding also has some negative effects :

- Information is **coarsened** by forcing it all into categories - there might not be a category that matches what you want to record - e.g. hair colour
- The same can be true of rounding numbers - the intervals or numbers of categories is called the **granularity** - this needs to be chosen carefully to maintain the quality of the information

Knowledge

- Data and information deal with facts and figures
- Knowing what to do with them requires **knowledge**
- **Knowledge** = information + **rules**
- Rules tell us the likely effect of something
- For example: you are more likely to pass your A level IF you do your coursework and revise for your exam!

Benefits of ICT

Why use ICT to do your data processing?

- **Speed** of processing
- **Storage** capacity (and physical size)
- Flexible **searching** and **sorting**
- Real-time **response** - e.g. booking systems
- **Accuracy** of results
- Ease of **communication**
- Improved **image** - fonts, graphics, etc.

What are the drawbacks of using ICT?

- **Information overload** - having so much information you can't do anything with it all!
- The **systems may be inflexible** and not give you the information you require
- There may be **staff resistance** to the introduction of ICT
- Fewer staff may be required - **job losses?**
- Staff become **reliant** on ICT and can't work if the computer fails
- Expensive **training** may be required when systems are updated
- Complicated systems or slow hardware may cause **stress**
- The data may be coarsened by coding leading to **poor quality**

References

- 1. Ackoff RL. 1989. From data to wisdom. J Appl Syst Anal. 16, 3-9. [Google Scholar]
- 2. Rowley J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci. 33(2), 163-80. 10.1177/0165551506070706 [CrossRef] [Google Scholar]
- 3. Dammann O, Smart B. Making Population Health Knowledge, in Causation in Population Health Informatics and Data Science. 2019, Springer Nature: Cham, Switzerland. p. 63-77. [Google Scholar]
- 4. Mensah E, Goderre JL. Data sources and data tools, in Public health informatics and information systems, J.A. Magnuson and P.C. Fu, Editors. 2014, Springer: London. p. 107-131. [Google Scholar]
- 5. Audi R. The Cambridge dictionary of philosophy. 2nd ed. 1999, Cambridge; New York: Cambridge University Press. xxxv, 1001 p. [Google Scholar]
- 6. Brownson RC, Fielding JE, Maylahn CM. 2009. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health. 30, 175-201. 10.1146/annurev.publhealth.031308.100134 [PubMed] [CrossRef] [Google Scholar]
- 7. Dammann O, Smart B. Integrating Evidence, in Causation in Population Health Informatics and Data Science. 2019, Springer Nature Cham, Switzerland. p. 99-115. [Google Scholar]
- 8. Dammann O. 2018. Hill's Heuristics and Explanatory Coherentism in Epidemiology. Am J Epidemiol. 187(1), 1-6. 10.1093/aje/kwx216 [PubMed] [CrossRef] [Google Scholar]
- 9. Ichikawa JJ, Steup M. The Analysis of Knowledge The Stanford Encyclopedia of Philosophy 2013 03/07/2014]; Available from: <http://plato.stanford.edu/archives/fall2013/entries/knowledge-analysis/>.
- 10. Gettier J. 1963. Is justified true belief knowledge? Analysis. 23(6), 121-23. 10.1093/analys/23.6.121 [CrossRef] [Google Scholar]

References

- 11. Lucey KG. On knowing and the known: introductory readings in epistemology. 1996, Amherst, N.Y.: Prometheus Books. 437 p. [Google Scholar]
- 12. Fleck L. Genesis and development of a scientific fact. 1979, Chicago: University of Chicago Press. xxviii, 203 p. [Google Scholar]
- 13. Solomon M. Making medical knowledge. First edition. ed. 2015, Oxford: Oxford University Press. xiii, 261 pages. [Google Scholar]
- 14. Green LW, Ottoson JM, Garcia C, Hiatt RA. 2009. Diffusion theory and knowledge dissemination, utilization, and integration in public health. Annu Rev Public Health. 30, 151-74. 10.1146/annurev.publhealth.031308.100049 [PubMed] [CrossRef] [Google Scholar]
- 15. Broadbent A. Philosophy of epidemiology. New directions in the philosophy of science. 2013, Houndmills, UK: palgrave macmillan. [Google Scholar]
- 16. Feitelson DG. 2015. From repeatability to reproducibility and corroboration. Oper Syst Rev. 49(1), 3-11. 10.1145/2723872.2723875 [CrossRef] [Google Scholar]
- 17. Chaffey D, Wood SJ. Business Information Management: Improving Performance using Information Systems. 2005: Prentice Hall. [Google Scholar]
- 18. Choo CW. 1996. The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. Int J Inf Manage. 16(5), 329-40. 10.1016/0268-4012(96)00020-5 [CrossRef] [Google Scholar]
- 19. Heller RF. Evidence for population health. 2005, Oxford; New York: Oxford University Press. xii, 126 p. [Google Scholar]
- 20. Heller RF, Page J. 2002. A population perspective to evidence based medicine: "evidence for population health. J Epidemiol Community Health. 56(1), 45-47. 10.1136/jech.56.1.45 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
-