

BC

BEFORE COMPUTERS

On Information Technology from
Writing to the Age of
Digital Data



Handwritten characters in a cursive script, possibly representing an early form of data or code.



Stephen Robertson





<https://www.openbookpublishers.com>

© 2020 Stephen Robertson



This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text and to make commercial use of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work).

Some of the material in this book has been reproduced according to the fair use principle which allows use of copyrighted material for scholarly purposes.

Attribution should include the following information:

© Stephen Robertson, *B C, Before Computers: On Information Technology from Writing to the Age of Digital Data*. Cambridge, UK: Open Book Publishers, 2020, <https://doi.org/10.11647/OBP.0225>

In order to access detailed and updated information on the license, please visit <https://www.openbookpublishers.com/product/1232#copyright>

Further details about CC BY licenses are available at <https://creativecommons.org/licenses/by/4.0/>

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at <https://archive.org/web>

ISBN Paperback: 978-1-80064-029-0

ISBN Hardback: 978-1-80064-030-6

ISBN Digital (PDF): 978-1-80064-031-3

ISBN Digital ebook (epub): 978-1-80064-104-4

ISBN Digital ebook (mobi): 978-1-80064-105-1

ISBN Digital (XML): 978-1-80064-106-8

DOI: 10.11647/OBP:0225

Cover image: Katsushika Hokusai (1760-1849), *A merchant making up the account*. Wikimedia https://commons.wikimedia.org/wiki/File:A_merchant_making_up_the_account.jpg
Public Domain.

Cover design: Anna Gatti.

6. Organising information

Every act of communication involves organising information—choosing what to communicate, and how to express it, whether in speech or writing or some other method. All forms of writing, even writing in order to enhance your own memory (for example, a shopping list), require organisation—of ideas, connections, facts, words, numbers, feelings, desires, intentions, stories, opinions, or whatever. We have already seen how the earliest forms of writing were for such purposes as commerce and administration, and such writing is necessarily an act of organisation. Another purpose, which developed early, probably counts as the first scientific endeavour: the study of the heavens.

Astronomy

Observation of the stars, more particularly systematic observation and recording, began very early in human history. Much of what we know about it derives from written sources from the first millennium BCE, particularly Babylonian clay tablets, but these certainly include material from much older sources, now lost. One particular set of observations of the planet Venus probably dates to the seventeenth century BCE

Such observational data might reasonably be termed ‘information’ precisely because it is systematically collected and organised for recording. In fact, it may now provide us with information not envisaged by its authors. Despite various uncertainties about the accuracy of the copies we have and the exact interpretations of the record, these observations can now be used to validate aspects of historical chronology, because our present astronomical knowledge allows us to determine the exact positions of the planets in the second millennium BCE.

Babylonian astronomers constructed extensive catalogues of stars and

constellations. We have copies of two such catalogues, the originals probably dating from around 1200 and 1000 BCE respectively.

Astronomical matters are of course important for human affairs. Sun, moon and stars have been the most important resources for navigation across open seas ever since humans tried such navigation—only in very recent history replaced by satellite navigation. Astronomical navigation, as practised over the last two or three centuries, requires the preparation and distribution of nautical almanacs containing tables indicating the positions of sun, moon and 57 selected stars (as well as, famously, an accurate marine chronometer or clock).

The *Computus*

For an earlier example of the perceived importance of astronomical data, one of the questions that much exercised the early Christian church was when to celebrate Easter. This question brought into existence an entire subject of study called the *Computus*, concerned with the various astronomical events and cycles by which calendars are determined. Proper calculation of the date of Easter requires the taking into account of the length of the true solar year (approximately 365-and-a-quarter days—but the quarter is not exact), the true lunar month (again approximately 29-and-a-half days), and the week of seven days. The length of the solar year (then assumed to be 365-and-a-quarter days exactly) had been the basis for the introduction of the Julian calendar under Julius Caesar in the first century BCE. Various different versions of the Easter calculation were defined, but the one that came to dominate was formalized by the Venerable Bede in the eighth century, following a formula devised by Dionysius Exiguus in the sixth. Bede's great work on the *Computus*, *On the Reckoning of Time*, contains a number of tables based on astronomical predictions, and shows the date of Easter for many years in the future.

Much later, in the sixteenth century, the Gregorian calendar was introduced by Pope Gregory. The difference between the Julian and Gregorian calendars is to do with the difference between the assumed 365-and-a-quarter days and the true length of the solar year. But the specific reason for its introduction was to readjust the date of Easter in relation to the seasons, in particular to the spring equinox, to what it had been at the begin-

ning of the Christian era. Currently, the date of Easter as celebrated in most western churches differs from that used in most Orthodox churches. This is a consequence of the fact that the western churches generally converted to the Gregorian calendar, while the Orthodox churches stuck to the Julian calendar.

Tax collection

Another early example of information organisation was to do with taxation.

We know that there was a system of taxation in Egypt, early in the Old Kingdom, about 3000–2800 BCE. The easiest people to tax are the farmers, because typically both their means of production (fields and livestock) and what they produce are clearly visible to all. So the principle might be that 10% of the crop goes to the local governor or tax collector. Except that this is hard to police—you would have to have someone watching the farmer all the time. But you can measure his fields once or at long intervals, and count his livestock also. For the fields, you might assume that a field of a certain size will have a certain yield in a year, and tax the farmer on that basis. Maybe you need to distinguish between the very productive fields located in the Nile flood plain, and the somewhat less fertile fields on the hills. Then the farmer can be taxed not on what he actually produces, but on what the system assumes that he produces.

All of which requires the tax collector to keep records, in a standardised form. What area of fields, in each yield category, does this named farmer have? At once we see not only that the messy world has been manipulated into a tidy form, but also that this manipulation is not neutral. It is to the advantage of the farmer that his field on the edge of the slopes is classified as 'hill'—but to the tax collector, the advantage is reversed. Since the tax collector is the literate one who actually makes and keeps the records, his view is likely to prevail!

Census

One of the things a tax-collector needs to know is who the tax-payers are, and what they own. Governments have been conducting censuses as long as they have been systematically collecting taxes. There are of course other

purposes for conducting a census—knowing who to call for military service, all sorts of planning exercises that need statistical data, and so on. The word itself is Latin, and in Rome originally signified a list of those available for military service. But the concept is probably at least as old as tax collecting.

In England in the eleventh century, for example, William the Conqueror initiated a census of all his possessions, people included, called the Domesday (Doomsday) book. The book is primarily organised around land—the rural estates. In such feudal times, the people come with the land. But it includes the names (first names only) of under-tenants of the lord of the manor.

Modern censuses are normally tied to notions of ‘residence’ and ‘household’. A return is made for each household, and includes every person resident in that household. Both these notions are fuzzy at the edges. Nevertheless, the requirements of census-taking have played an important role in the development of ideas of information processing, as we shall see in Chapter 11.

History

In early human societies, history and mythology are irretrievably intertwined. One might argue that the same is true today, as in the saying attributed to Winston Churchill, that ‘history is written by the victors’. Nevertheless, we now associate the great classical Greek historians of the fifth century BCE, Herodotus and Thucydides, with the attempt to put history onto a more systematic footing, and to base it on carefully gathered evidence, in the process distinguishing history from mythology. Although I started this book by arguing that recorded history could not begin until we had developed writing, it is clear that this is not sufficient—we don’t immediately start the systematic recording of history because we have invented writing. These two Greeks had significant predecessors concerning whom less is known; but their role in developing historiography, the systematic study of history, is clear. Although they differed as to emphasis, between them they championed the meticulous gathering, analysis and evaluation of evidence, from witnesses and documents, about the events and circumstances they wanted to describe.

Libraries

We have already seen in Chapter 4 the importance of libraries in our story, as a method of communication. They also play a central role in methods of organisation of information.

Consider for example the great classical libraries that I mentioned: the Library at Alexandria, for example, or the House of Wisdom in Baghdad, or the library of one of the big medieval monasteries. In all these cases, scholars would arrive from remote places hoping to find enlightenment of some kind. The Alexandria library, for example, might have contained hundreds of thousands of items (the collection seems to have consisted mainly or entirely of papyrus scrolls; a single work might take up multiple scrolls). Either locating particular known items, or looking for multiple items on a subject, would have been a far from trivial task. The library was arranged by subject, each subject having a bin to contain the collection of scrolls. A tablet above the bin listed the contents of the bin, and each scroll had a tag attached to it, giving the author and subject. This kind of information was also the basis for what is supposed to be the first library catalogue, produced by a librarian called Callimachus for some of the material in the Alexandria library in the third century BCE. Just to indicate the scale of the finding problem, the catalogue ran to 120 scrolls.

The art of the library catalogue (thinking now of the present) is of interest to us for two reasons. The first is that it provides an organisation of the books or other materials in the library. It does this by collecting information or data *about* each book (sometimes referred to as metadata), specifying for example its author or authors, its title, when and where it was published, some codification of its subject matter, etc. It then provides access tools so that a book can be identified in a variety of ways, say by looking up the author. The location of a book on a shelf, quite likely as part of a subject arrangement, provides one (but only one) way of finding it. A catalogue typically provides multiple ways, suitable for different forms or types of enquiry. How it does this depends on other technologies available. Before the availability of computer-based library catalogues, various forms of index were needed—some were on cards, some printed on paper.

The second reason we may be interested in library catalogues is because of the organisation of the catalogue data itself. Consider for example the

data elements suggested above (author, title, publisher, date, subject). If an index based on any of these elements is required, each has to be treated in a consistent fashion across different items. For example, in order to make it easy (or even possible) to look up an author name in an index, the recording of the author name must follow a well-defined format and set of rules—and (ideally) be consistent if the same author has written multiple books. This might well require some manipulation of the messy real world.

Just for example, I have a book on the shelf next to me by the great physicist and Nobel prize-winner, Richard Feynman. Well, actually, the author's name appears as Richard P. Feynman. Elsewhere (not in this book) it is possible to discover that his middle name is Phillips. Another book, containing a collection of his writings, is titled *No Ordinary Genius: The Illustrated Richard Feynman*—which gives as author Richard Phillips Feynman (together with another person as editor). All of this would probably not matter very much, since author indexes are normally ordered by surname, and I would be quite likely to find entries for *Feynman, Richard*; *Feynman, Richard P.*; and *Feynman, Richard Phillips* quite close to each other. Besides, Feynman is a relatively uncommon name. And as I have only one Feynman in this book, I can get away with *Feynman, P.* in the Person index at the back. But names can cause much more serious problems than this—some further discussion below.

Forms

A very particular kind of organisation is required when you have to complete a form, whether on paper or online. Every time you fill in a form, you are slotting information that you have, about yourself and the world around you, into a kind of information-organisation devised by someone else. However messy the world around you, or the information that you have about it, the form makes you think about it in a particular way.

Let's take a name, for example. You have a name. More than that, I can say with some confidence (and any form you have to complete may well assume) that you have a surname that you inherited from your parents, or perhaps acquired later by marriage, and one or more given names—but if more than one, it's probably only the first that you actually use. So the slot in the form into which you are supposed to put your single used given name is quite likely labelled 'First name' (in my childhood, it was often labelled

'Christian name', though that obviously culturally biased terminology has largely disappeared). A form in the USA might ask for a 'Middle initial'.

But the entirety of the structure is culturally biased, of course. Chinese people coming to the West typically learn to reverse their two names—because by default in China, the surname comes first. Someone from the Indian subcontinent (I have friends like this) may have acquired only a single name as a child, and have had to invent a second for the purpose of filling in forms and (more generally) living in the west. I have several relatives who have two given names but actually use the second. I also have friends and relatives with double-barrelled surnames, not hyphenated but spaced, like the composer Ralph Vaughan Williams—that's not a problem when they complete a form themselves, but is definitely a problem for the library cataloguer. Other parts of the world have different practices—for example, in both Spain and Portugal, most people have double surnames. And of course if we go back in history as well as elsewhere in geography, the range of variations is huge. Often, the messy world has to be doctored in order to fit into a tidy form. And this is only the very first bit of the form!

Addresses

The next question on your form, after your name, is quite likely to be your address—though like your name, the form may require it to be split into multiple parts. This in itself is a slightly strange requirement in this day and age. If a friend writes down an address for me on a piece of paper, I will probably have no difficulty in parsing it—in distinguishing the house number, the street name, the town name and the postcode. It's a simple enough process, bound by rules, considerably simpler than long division—so I would expect a machine to be able to do it reasonably well. (If the form you are completing is not online already, it's likely to be fed into a machine shortly after completion.) So why isn't it left to the machine to do the parsing?

One reason might be that the form of an address is quite strongly history- and culture-dependent. More specifically, the national postal systems discussed in Chapter 2 have been very closely involved in the determination of standard address forms. Thus the standard varies considerably from country to country. Furthermore, although there have been attempts (since the

establishment of the Universal Postal Union in 1874) to define an international standard format for postal addresses, these now seem to have been abandoned. From the point of view of post alone, it probably doesn't matter very much—the Universal Postal Union is a federal structure, so as long as the postal service where the letter is posted can recognise the destination country, it can leave the rest of the address for interpretation by the local postal service in that country. However, it can cause problems for other uses of addresses (of which there are many).

One prime current example is the postcode. Although the first divisions of large cities into postal regions began in the nineteenth century (London 1857), and some more detailed attempts began in the 1930s, these mostly originate from the 1960s and '70s, a period that might just still be regarded as the heyday of the post, but perhaps its tail end. Many postcode systems provide a rather coarse level of granularity, a district containing many houses, but some are much more precise. In the UK system, for example, a postcode does not uniquely identify an address, but specifies a small group, up to a hundred but probably many fewer.

Postcodes (indeed, addresses generally) serve or contribute to a number of different functions other than postal deliveries. For example, postcodes are commonly used for satellite navigation (despite the fact that they were mostly devised before satellite navigation was invented). But for this purpose, one would like a very fine granularity. On the whole, the UK system works very well for this purpose, particularly in cities, but sometimes in the countryside it is not precise enough. But there is considerable variation between countries, even among those countries that have postcodes.

Returning to the parsing question raised above: what we do now find commonly in the UK is that the form-filler is invited to provide *only* the postcode, and then allow the computer to deduce almost fully the rest of the address, offering the user a small choice of house numbers and possibly street names. This deduction is based on a database, to which the computer has access, of postcodes and corresponding full addresses. To make sense of this statement, we need to talk a little about databases.

The concept of a database

Despite my comments above about the difficulties inherent in both, name and address data is often held up as a good example of a kind of information with a high degree of structure, a high degree of regularity, and a high degree of consistency. As a result, it is taken to be a good candidate for storage in a computer in what is commonly known as a database. If you keep your contacts on your computer or your phone or both, they will be held in a database. This means that even if some of the addresses take a slightly different form from others, or some data is missing from some, they are all held in a common structure. There are several reasons for doing it this way: essentially they revolve around how such data can be processed automatically, including for display to you. Thus, for example, you would expect to be able to see an alphabetical list of names. Once again, alphabetical sorting of names is not quite as straightforward as it might seem; nevertheless, you probably expect your computer, and your phone if it is even remotely clever, to be able to do that.

Databases, and computer programs that manipulate databases, are staples in the world of computing. Indeed, the maintenance and manipulation of databases is a vastly more important function of computers than calculation. Consider, for example, the computers in your bank, which look after your bank account. Clearly they have to do some calculation, when you add or withdraw funds or move them around—but by far their most important function is to maintain consistent records of all such transactions, as well as all the other information relating to this and every other account. Furthermore, you have probably never seen those computers make arithmetical mistakes—that's the easy part—but you are quite likely to have seen instances where, for one reason or another, transactions have gone AWOL.

If you frequently do online transfers, and have ever made a mistake, you may have discovered that a mistake in the destination account number can be much worse than a mistake in the amount. An account number is not really a number at all (nobody ever needs to do arithmetic with account numbers): it's a code identifying a particular set of database entries. As you may have read in many newspaper reports, if you transfer money into the wrong account, and the account holder is not willing to do anything about it, neither you nor your bank can recover the money. Until very recently, in

the UK at least, banks in these circumstances did not typically check names, only account numbers. This is probably because of all the issues discussed above with respect to names. If you do not know the exact form of the name of your payee, as held in the bank's database, then the chances are high that you would enter it in a slightly different form, so the banks prefer to rely on the code. Nevertheless, it is much easier for a human being to make a mistake with a long numerical code than with a name, so this logic can be counter-productive.

Varieties of database

Databases come in many different forms. In the present day, a database is generally assumed to be held on a computer. Many such systems follow the principle that data should be divided into its smallest coherent component parts, and that exact rules of inference should be specified, completely determining what can be learnt by recombining the data elements in new ways. This is a reductionist view, and has a strong analogy to the status of arithmetical calculation ever since the rules for this were codified. Some kinds of data are amenable to this approach, and it brings advantages in the ability to manipulate it in well-understood ways. However, not all data, let alone all information, can be treated in this way.

In the past, long before computers or the coinage of the word *database*, we have seen many collections of information that would now be called databases. Of those we have discussed in this chapter, all collections of completed forms, all library catalogues, all sets of census returns (and all tables derived from them), all tax collectors' records of the people and institutions that they tax, all banks' records of people, accounts, transactions, and so on and so on, can be seen as databases. All involve rules of organisation and of manipulation.

We return to the theme of calculation in Chapter 10, and the broader theme of information processing in Chapter 11.