

Changes in oceanographic data management technology since 1969

Roy Lowry

BODC Emeritus Fellow



50

In the Beginning

There were large computers

Programs/data were entered on punched cards

This one has human-readable headings but not all did

Programs and data were hand-written on coding forms

Converted to cards in central punch room

Card decks sent to computer centre

Operators ran the card job and filed the resulting line-printer output

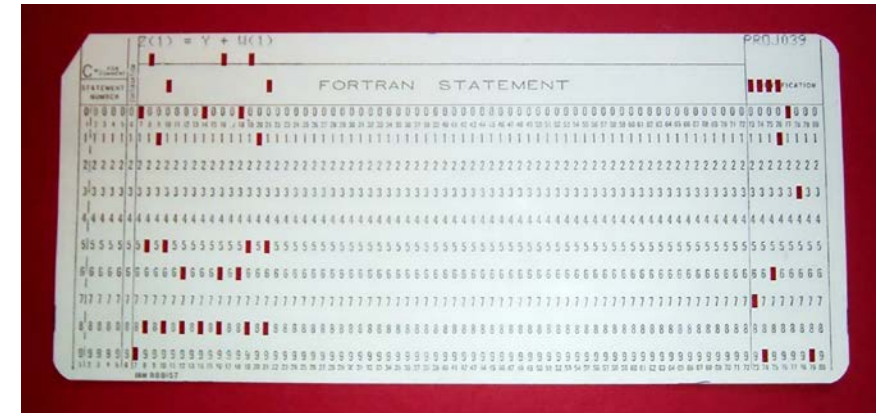
Output delivered from computer centre

Correction code forms written and sent to the punch room

Corrections merged into card deck and sent to computer centre

Output delivered from computer centre

When developing software three compilation runs per week was good going!!!



In the Beginning

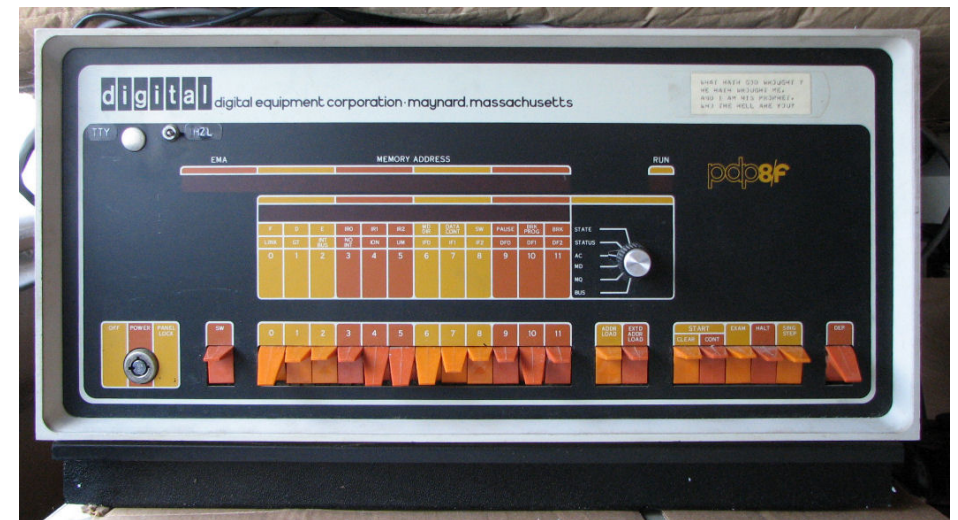
There were small computers like this DEC PDP-8

First computer costing less than \$20,000 when \$40,000 bought a house

Based on volatile magnetic core memory with no ROM or internal disk

Bootstrap entered byte-by-byte in machine code using front panel switches (12 as it was a 12-bit machine)

Bootstrap - if entered correctly - gave control to a peripheral such as disk or paper tape reader



In the Beginning

Communication was by post or landline telephone

Group secretary typed all communications

Paper copies circulated to the group once a week

Knowledge to support parameter labelling and data quality control was obtained by reading photocopies of papers or books

On-site library heavily biased towards physics and coastal oceanography

Checking the spelling of a taxon name required a postal inter-library loan request



Innovations During the 1970s

User-operated card punches cut card deck correction times from a day to minutes

Access to the mainframe from a room in the computer centre containing a card reader, line-printer and card punches cut time to get a program to compile from weeks to under an hour

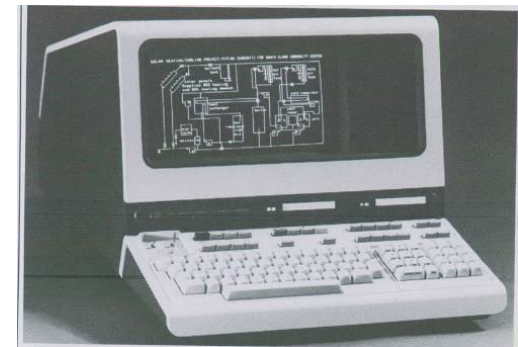
Time Sharing Systems (TSS)

Run small jobs (e.g. text editor or compiler) from a terminal

No more punching cards and fewer visits to the computer centre

At Royal School of Mines 50 staff and students shared 5 terminals

On arrival at Bidston in 1981 I had my very own HP2648!



The 1980s



The 1980s

Telex -based electronic mail available by the start of the 1980s

- Expensive so access was heavily restricted

- Mostly used for international communication

E-mail freely available by the end of the 1980s

- Text only – no attachments

- Addressing required a knowledge of Internet plumbing

Landline telephones augmented by mobile phones

- Expensive, dumb and built like a brick



1980s Data Ingestion Workflow

Log tape as an accession (paper form), register in tape library (form), copy and archive

Assess tape physical structure and contents from an octal dump (utility)

Produce a legible listing of the start and end of each file (Fortran)

Document the format (form) and prepare a Transfer (code data description file plus three Fortran subroutines)

Production Transfer run(s) generating screening resources

Three-tape (input, PXF output, ASCII output) jobs

Microfiche listings (Hobbs reprographics)

Paper plots (Calcomp drum plotter)

Metadata collation forms (line-printer)

Basis of the present-day BODC Series Schema



The 1990s



The 1990s

Communication

- E-mail developed straightforward addressing and attachments during the decade
- Small mobile phones capable of calls and texts
- Prototype smart phones under development

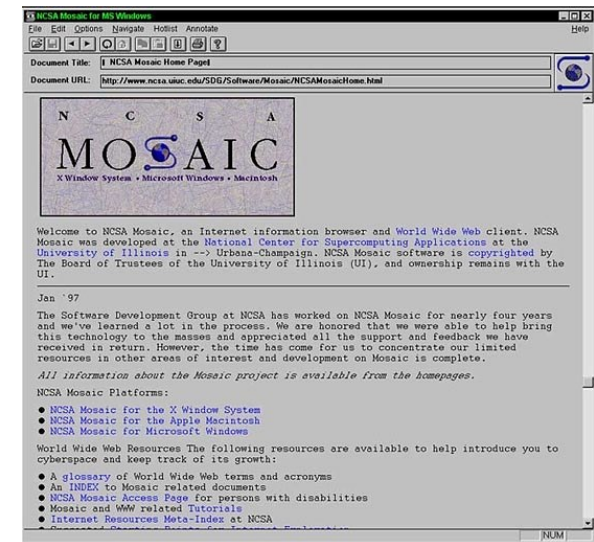
Access to knowledge

- On-line access to library catalogues and inter-library loan

Developing WWW

Content quality highly variable

Content hard to find as search engines were rudimentary



1990s Data Ingestion Workflow

Ascertain cruise data collection events

Load event information into Oracle to form a metadata skeleton

Process shipboard computer tapes (later CD-R) through bespoke system

Integrate data submissions from cruise participants

Physical standardisation (PC, Mac, Apple II, QIC, ExaByte, etc., etc.)

Conversion of proprietary formats to ASCII

Load data into Oracle

Link data into metadata skeleton

Publication of completed project datasets on CD-ROM

Basis of the present day BODC Samples Schema



The Noughties

BODC encounters computer science
UK e-Science and NERC DataGrid
Taking on computer scientist developers

Computer science delivered:
Externally developed standards
Single standard shared across projects and disciplines

Technologies
XML – revolutionary game-changer
Ontologies – basis for machine-managed knowledge
Semantic web – basis for internet-wide knowledge



The Noughties

Communication

E-mail fully mature

Smart phone revolution

Social media

Access to knowledge

WWW reaching maturity

Smart search engines

Authoritative content becoming available

ITIS and WoRMS for taxon names and taxonomies

CAS-based sites and ChEBI for organic chemistry nomenclature

The Google logo, featuring the word "Google" in its characteristic multi-colored font (blue, red, yellow, blue, green, red).

The Present Day



Cloud computing

Massive centralised off-site
computer resources

Users billed for resources
used

University of London Computer Centre (1970s)

Massive centralised off-site
computer resources

Users billed for resources
used

So what's changed ? 😊