Databases and database systems: introduction

- example database: HOLMES
- database vs file-based systems
- Relational database concepts

- *“The worldwide relational database management systems market saw a 12.6% growth spike in 2007 to $18.8 bln. Oracle took the top spot, capturing 44.3% of the market, IBM came in second with a 21% share, followed by Microsoft, with 18.5% of the market. iSybase and Teradata rounded out the top five, garnering market shares of 3.5% and 3.3%, respectively.”*

www.idc.com

Examples of database applications

- One patient, one record
- http://www.computerweekly.com/Articles/2007/09/03/226534/one-patient-one-record-unrealistic-says-bcs-group.htm

- iTunes (client and store)

- Googlemaps.com

database example: police manhunts

" *most of the investigation was done on paper. In the aftermath of it all, when information was fed into the computer, (his) name stood out like a sore thumb. Why? He'd been investigated by the police time and time again, but because all the information was on paper across several forces there was no reliable way of cross referencing the available data.* "

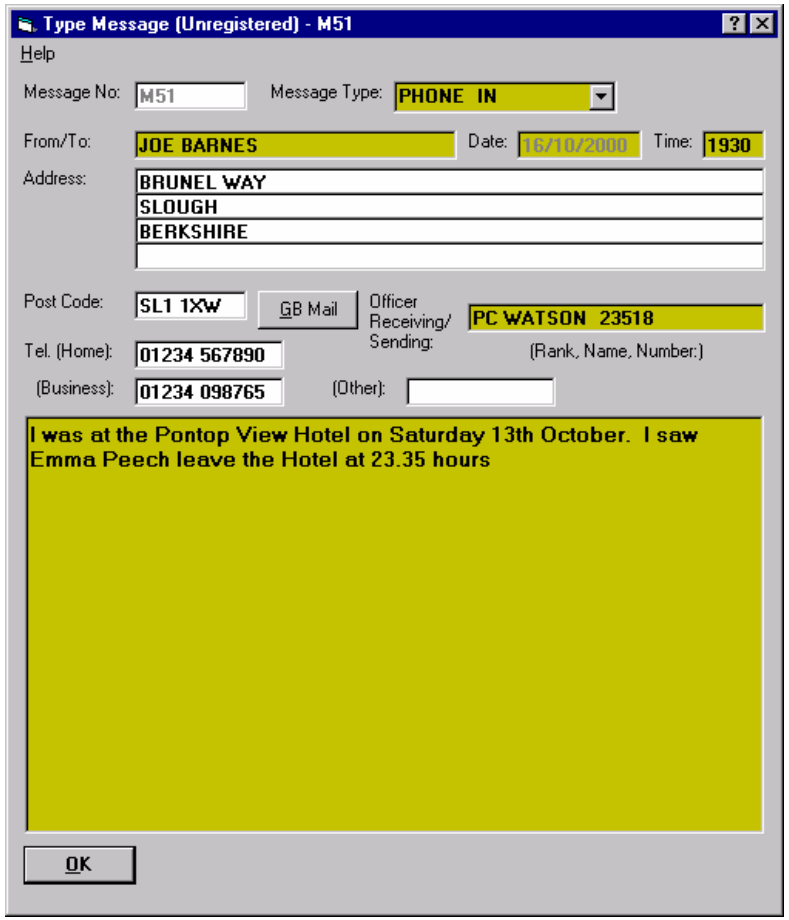
HOLMES 2: Home Office Large Major Enquiry System

- www.**holmes2**.com
- a case management system supplied by Unisys with Oracle database
- custom-built software and off-the-shelf components.
- the client-server architecture : workstations running Windows 2000 Professional or NT and Unix servers running Sun's Solaris
- each police force has its own HOLMES 2 setup, but they can be linked over a nationwide police extranet for cooperative work on a case
- an intelligent free-text search subsystem from Autonomy, using sophisticated algorithms to prioritize relevant parts of statements

- pieces of evidence are indexed and logged.
- closed-circuit television (CCTV) footage is also indexed and stored.
- Specific queries are then made of the system based on e.g. ages, descriptions etc
- The system can find common traits in the information, potentially raising the profile of those that police may want to focus on.
- investigators can make decisions about who to interview next, and that interview would eventually come back to the incident room for entry into HOLMES 2

Registering new information

- Each item of information into the incident room, eg a phone call from the public, is stored and can begin a new line of enquiry.

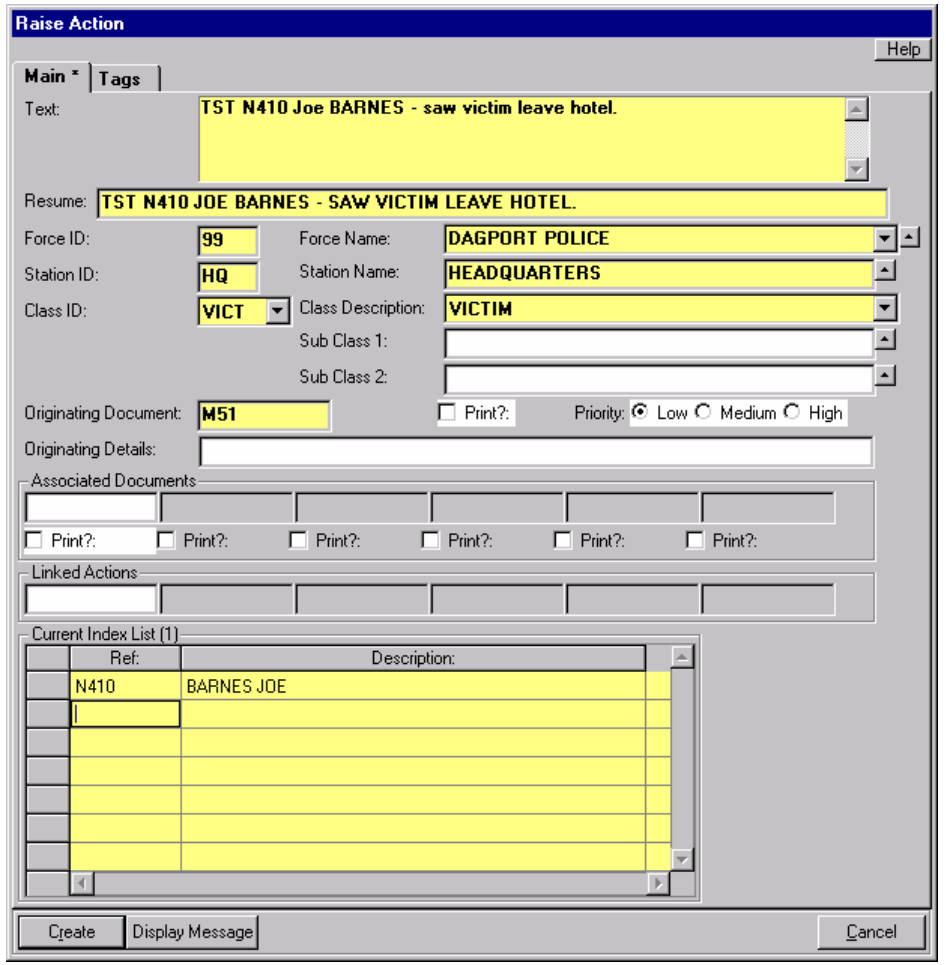


Task Allocation

- After a message has been recorded on the system, it is examined by an officer who creates an Action to investigate

- An appropriate enquiry officer is selected, and a copy of the Action is printed for the officer.

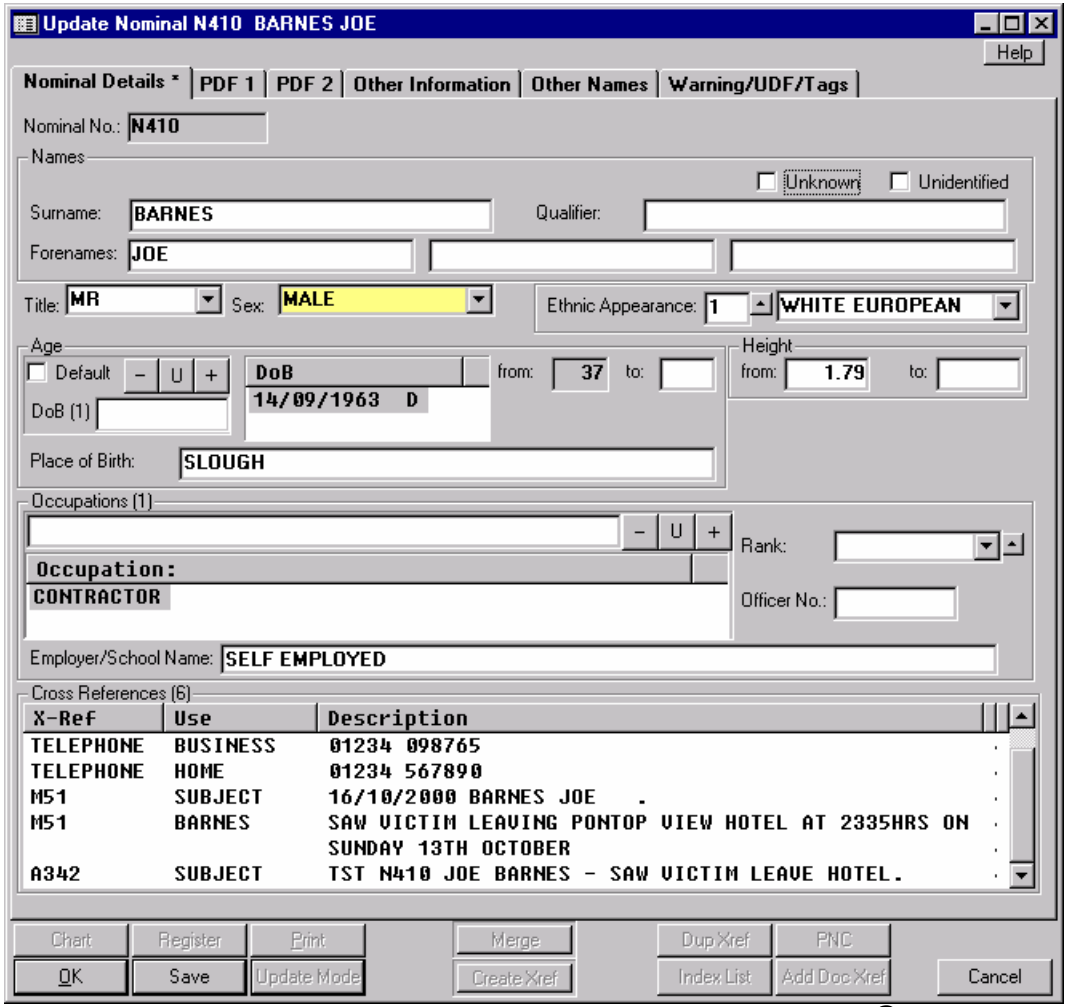
  The allocated officers investigate and obtain a full statement create a detailed description of the interviewee (Personal Descriptive Form (PDF)".



Nominal Records

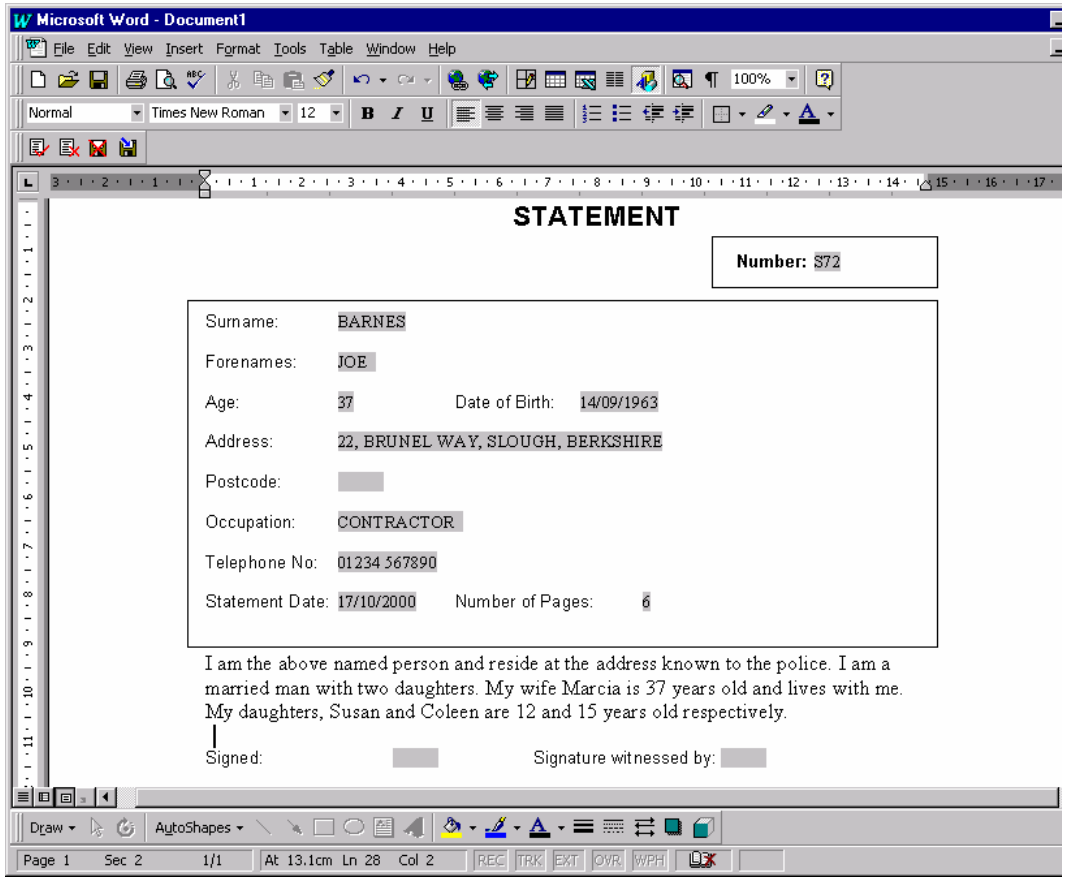
- Every person connected with the enquiry, either to the victim as a relative, friend, or colleague, etc., or to the enquiry as a witness, or mentioned in a Statement, is recorded in HOLMES 2 via a Nominal record.

- The Personal Description Form PDF collated by the enquiry officers is linked to the Nominal record



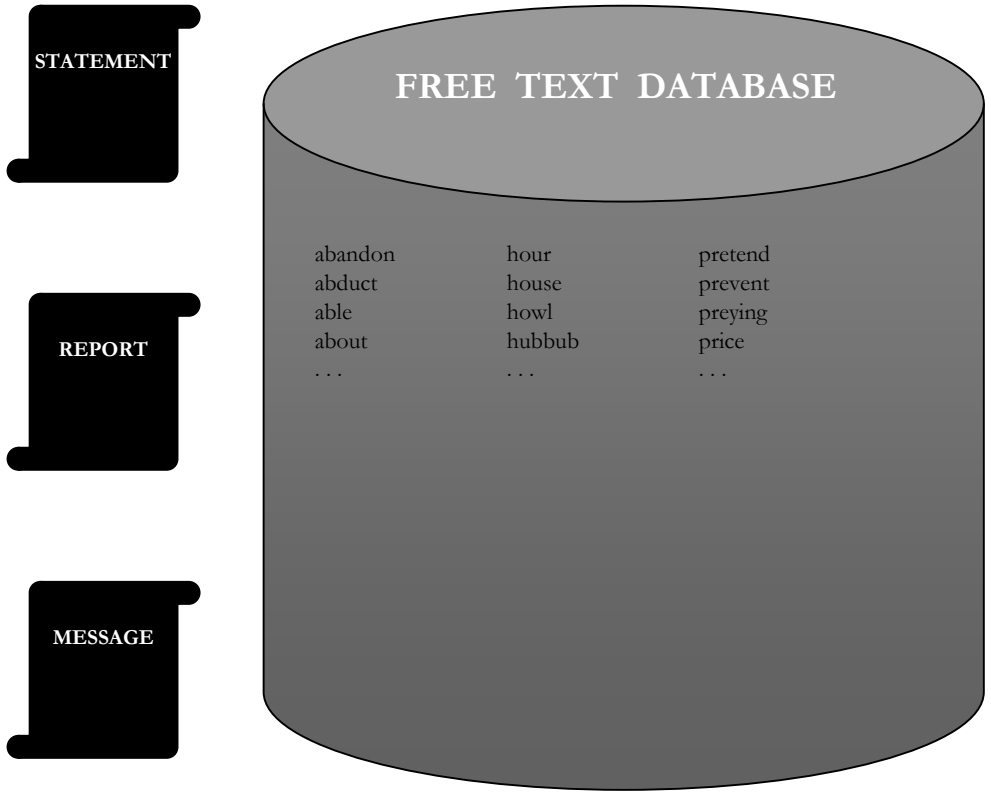
Statement is Typed

- Statements are transcribed into Word.Statements are linked to the Nominal record.
- HOLMES 2 uses Microsoft Word for all sored documents.



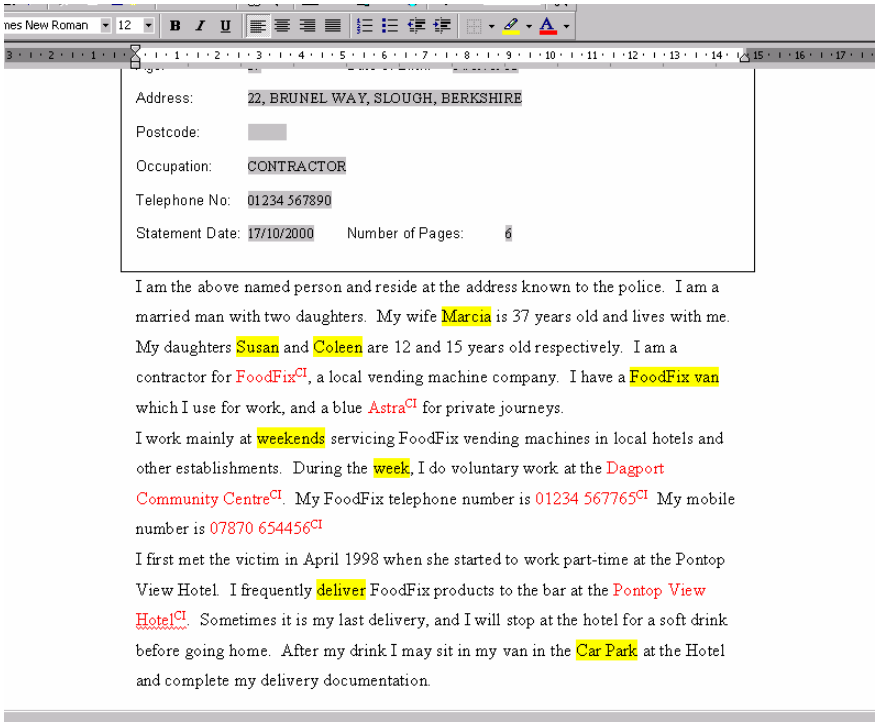
Investigation Information

- When the typing of a Document has been completed, it is immediately added to a "free text" database. This enables Detectives to start searching at the earliest opportunity for any document which may be connected to relevant information.



Document Mark-up

- The Statement now needs to be examined to identify the important items of information which need to be indexed separately.
- This is usually done by a responsible Investigating Officer who will also identify and prioritise any other Actions which are required as a result of the Statement.
- This process is known as Document Mark-up, and may be done on-line or off. The resulting Index Items and Actions are added to
- HOLMES 2 using the Graphical Indexing facility.



Address: 22, BRUNEL WAY, SLOUGH, BERKSHIRE

Postcode:

Occupation: CONTRACTOR

Telephone No: 01234 567890

Statement Date: 17/10/2000 Number of Pages: 6

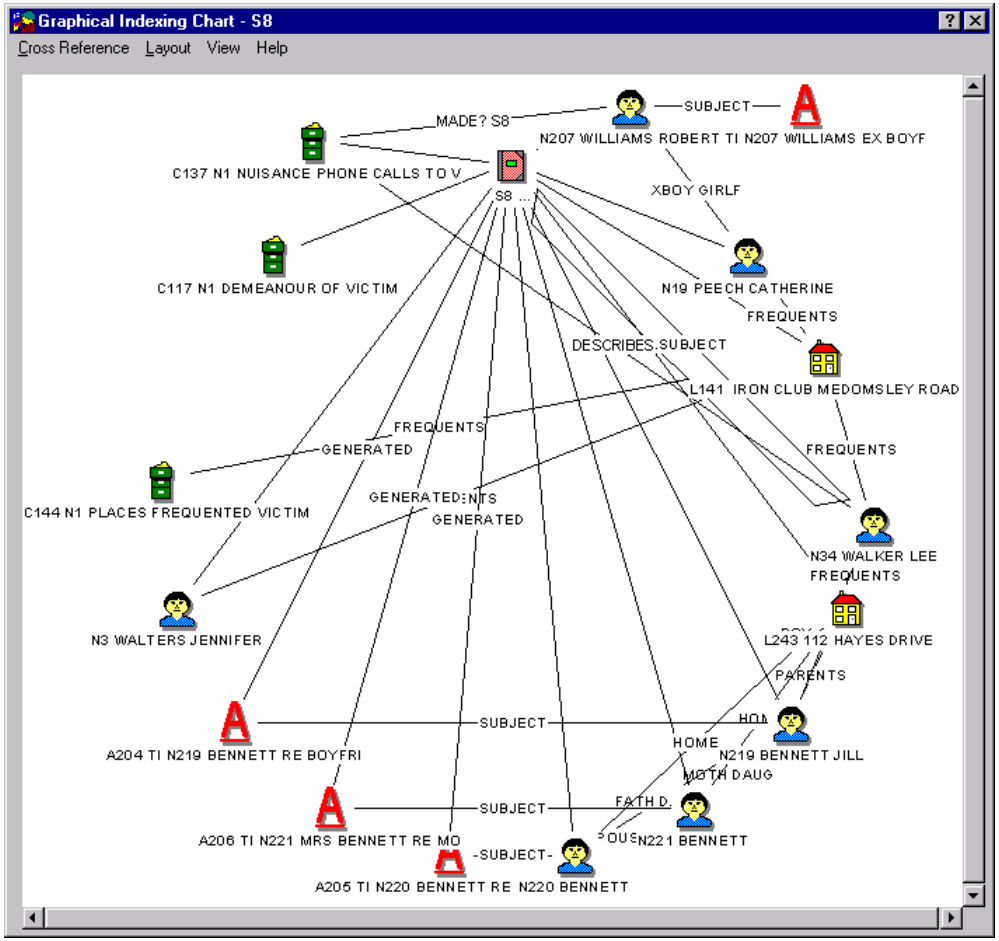
I am the above named person and reside at the address known to the police. I am a married man with two daughters. My wife Marcia is 37 years old and lives with me. My daughters Susan and Coleen are 12 and 15 years old respectively. I am a contractor for FoodFix[CI], a local vending machine company. I have a FoodFix van which I use for work, and a blue Astra[CI] for private journeys.

I work mainly at weekends servicing FoodFix vending machines in local hotels and other establishments. During the week, I do voluntary work at the Dagport Community Centre[CI]. My FoodFix telephone number is 01234 567765[CI] My mobile number is 07870 654456[CI]

I first met the victim in April 1998 when she started to work part-time at the Pontop View Hotel. I frequently deliver FoodFix products to the bar at the Pontop View Hotel[CI]. Sometimes it is my last delivery, and I will stop at the hotel for a soft drink before going home. After my drink I may sit in my van in the Car Park at the Hotel and complete my delivery documentation.

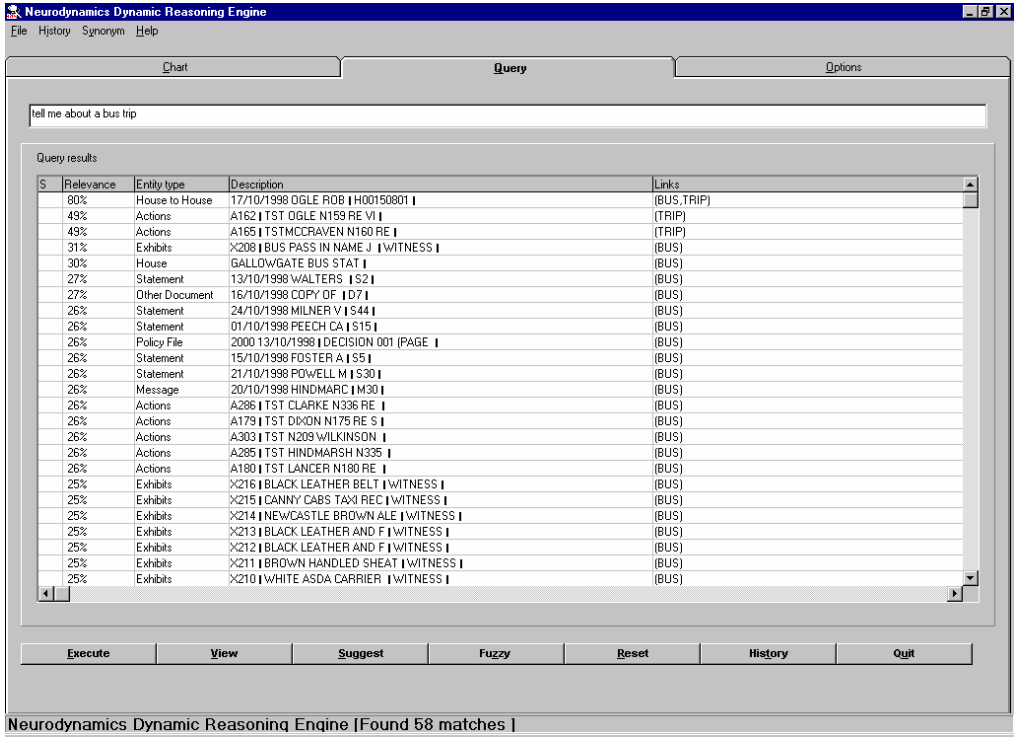
Graphical Indexing

- Items are added to the indexed database from a marked-up Document.
- As items are added (or confirmed), a graphical representation, called a Link Chart, of the information and relationships can be displayed.
- Index Items can be cross-referenced to each other, and to their originating Documents.



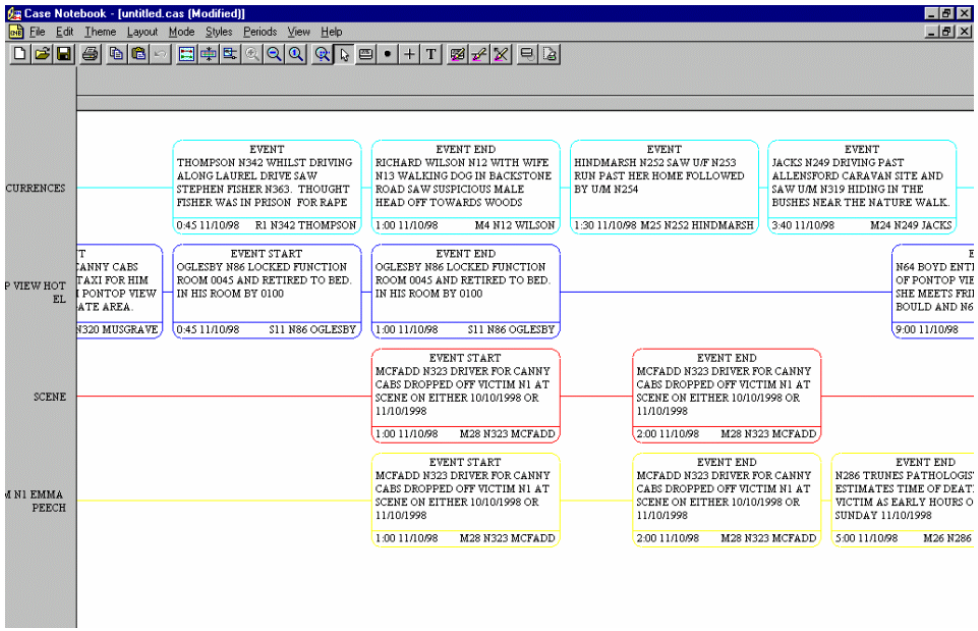
Free-Text Searching

- Now that a reasonable amount of information has been recorded in HOLMES 2, the enquiry team starts to research the data to identify new lines of enquiry.
- One method used is a free-text search of all document information to establish whether they have "missed" anything important.
- By searching on the key items, a list of information, ranked in order of relevance, is produced.



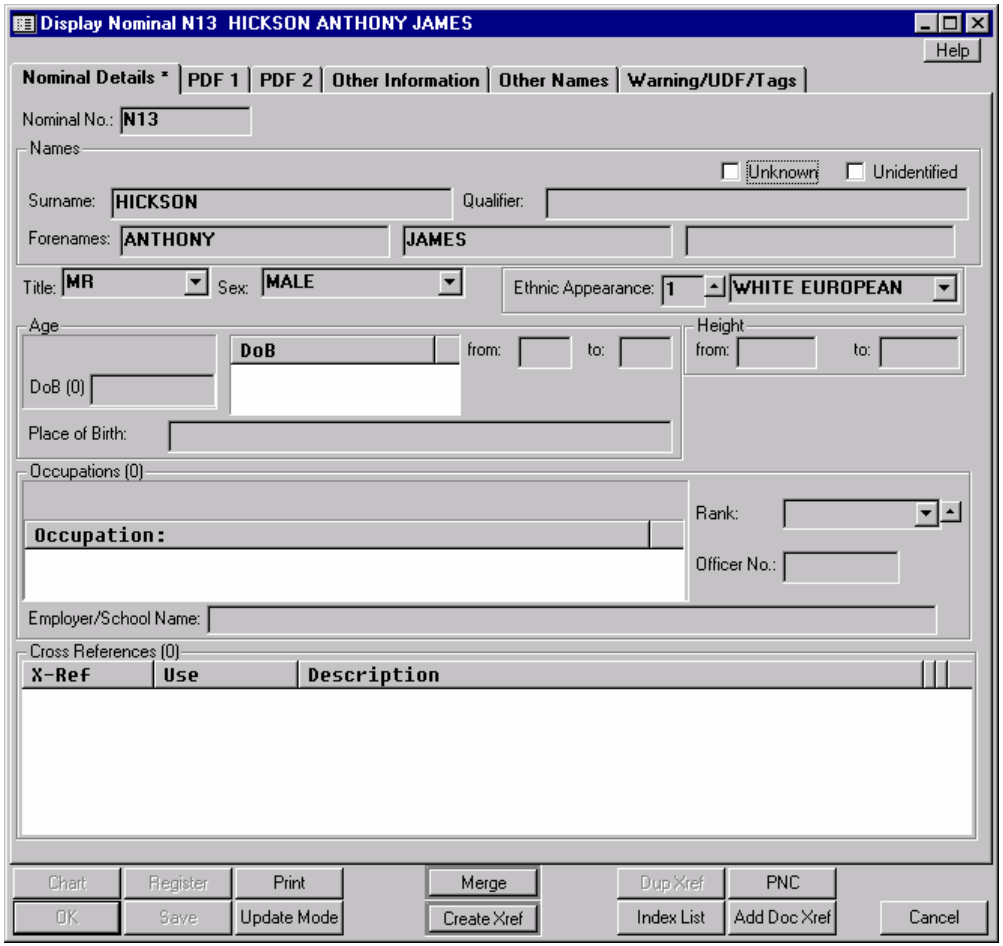
Graphical Sequence of Events

- At the same time, a detailed account of the crime scenario (eg, what the victim was doing at the time) can be built from friends' and witnesses' Statements.
- By comparing this Graphical Sequence of Events and the results of the free-text search, missing parts of the scenario can be inferred, eg, which people were in the vicinity of the victim at the time.



Index Searching

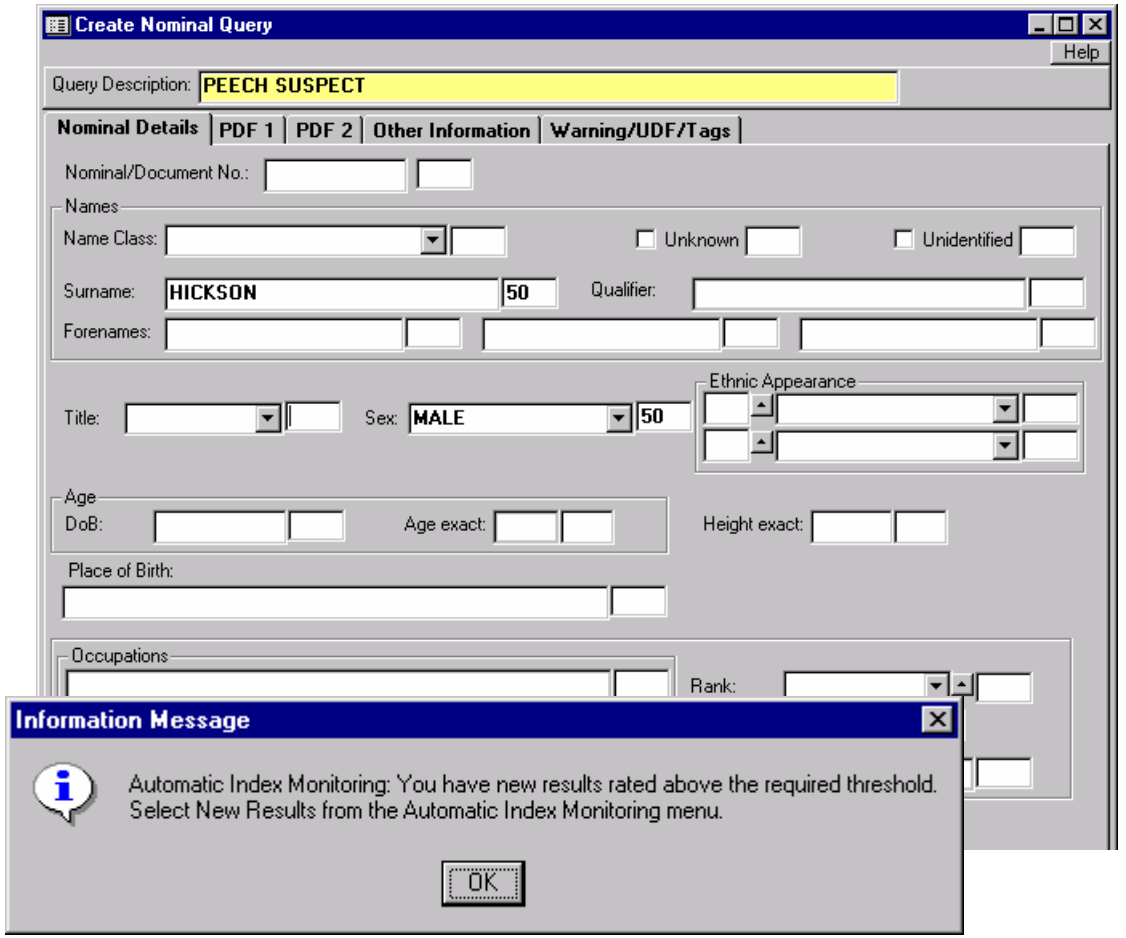
- The Analyst Charting tools are particularly useful in identifying new lines of enquiry, and in briefing the enquiry team.
- They are fully integrated into HOLMES 2, so a particular individual's index record can be retrieved directly from the Sequence of Events, or from a link chart.
- If there is missing information in a record, another Action is raised to gather that information about this Nominal.



Monitoring New Information

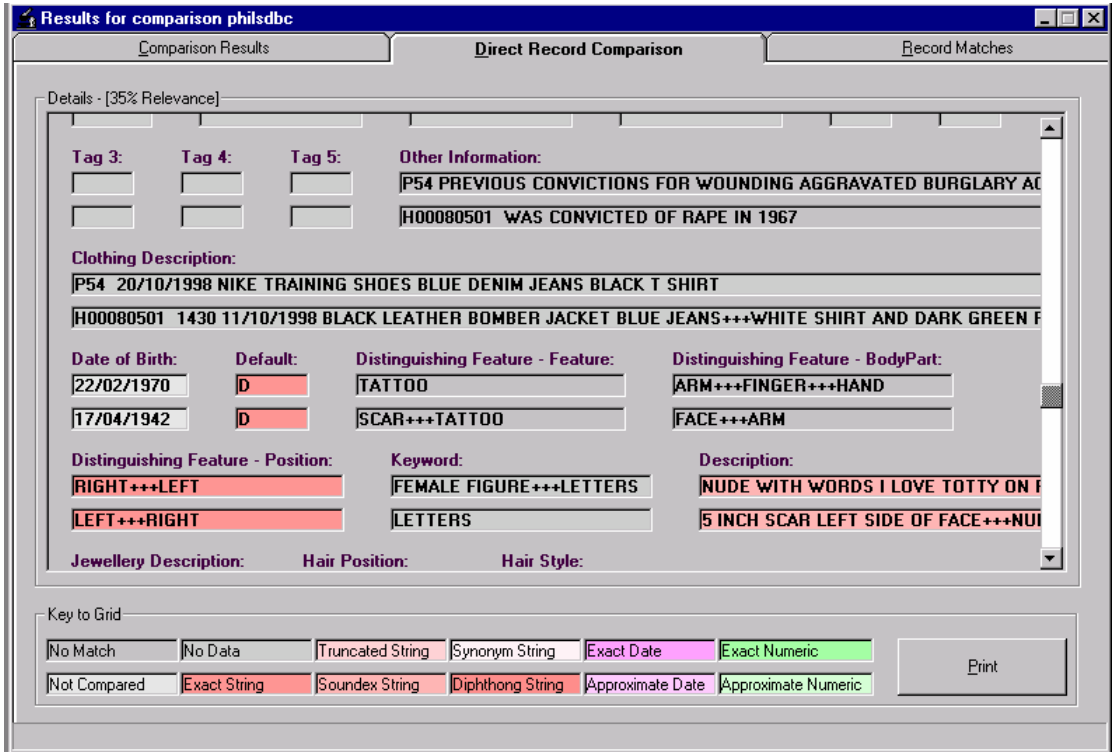
- Now that a person has been identified as an important Nominal, the Senior Investigating Officer wants to be notified of all new information regarding him.

- The Automatic Index Monitoring facility can be configured to recognise when new information is entered according to certain criteria. Every time a match is encountered, a message is produced.



Linking and Comparing Incidents

- two incidents can be linked to find similarities, and provide a further focus for the investigation.
- The HOLMES 2 Database Comparison facility automates the task of finding similarities by comparing individual records.



"*Very easily you would be able to find out all the information you needed to about that particular person and decide what you wanted to do about it,*"

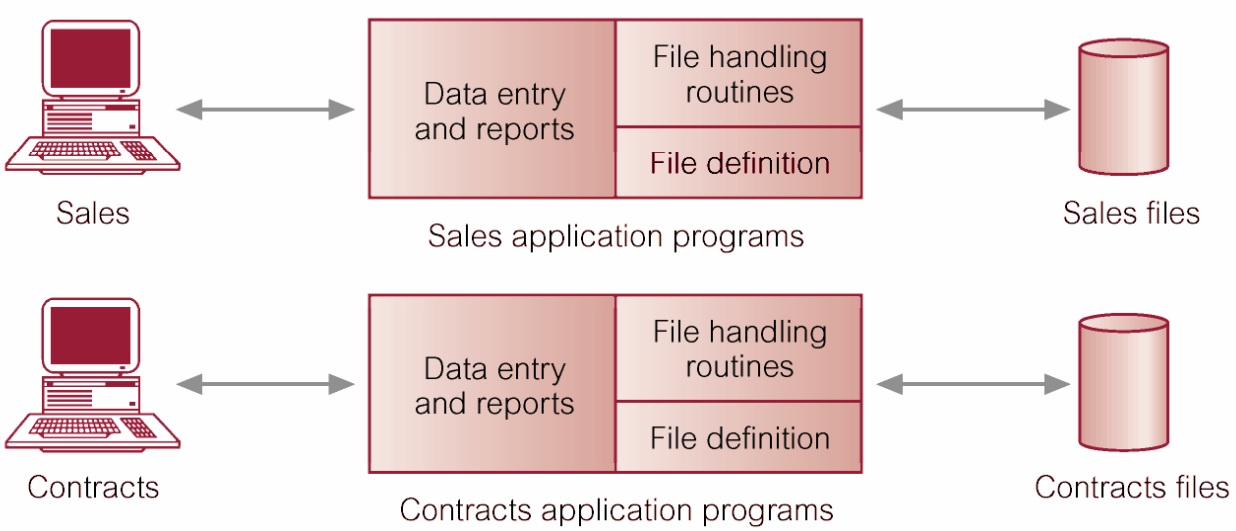
Helen Mylam, HOLMES 2 product manager..

File-Based Systems

- Collection of application programs that perform services for the end users (e.g. reports).

- Each program defines and manages its own data.

File-Based Processing



Sales Files

PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)
PrivateOwner (ownerNo, fName, lName, address, telNo)
Client (clientNo, fName, lName, address, telNo, prefType, maxRent)

Contracts Files

Lease (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)
PropertyForRent (propertyNo, street, city, postcode, rent)
Client (clientNo, fName, lName, address, telNo)

Limitations of File-Based Approach

- Separation and isolation of data
  - Each program maintains its own set of data.
  - Users of one program may be unaware of potentially useful data held by other programs.

- Duplication of data
  - Same data is held by different programs.
  - Wasted space and potentially different values and/or different formats for the same item.

Limitations of File-Based Approach

- Data dependence
  - File structure is defined in the program code.
- Incompatible file formats
  - Programs are written in different languages, and so cannot easily access each other's files.
- Fixed Queries/Proliferation of application programs
  - Programs are written to satisfy particular functions.
  - Any new requirement needs a new program.

Database Approach

- Arose because:
  - definition of data was embedded in application programs, rather than being stored separately and independently.
  - no control over access and manipulation of data beyond that imposed by application programs.

- Result:
  - the database and Database Management System (DBMS).

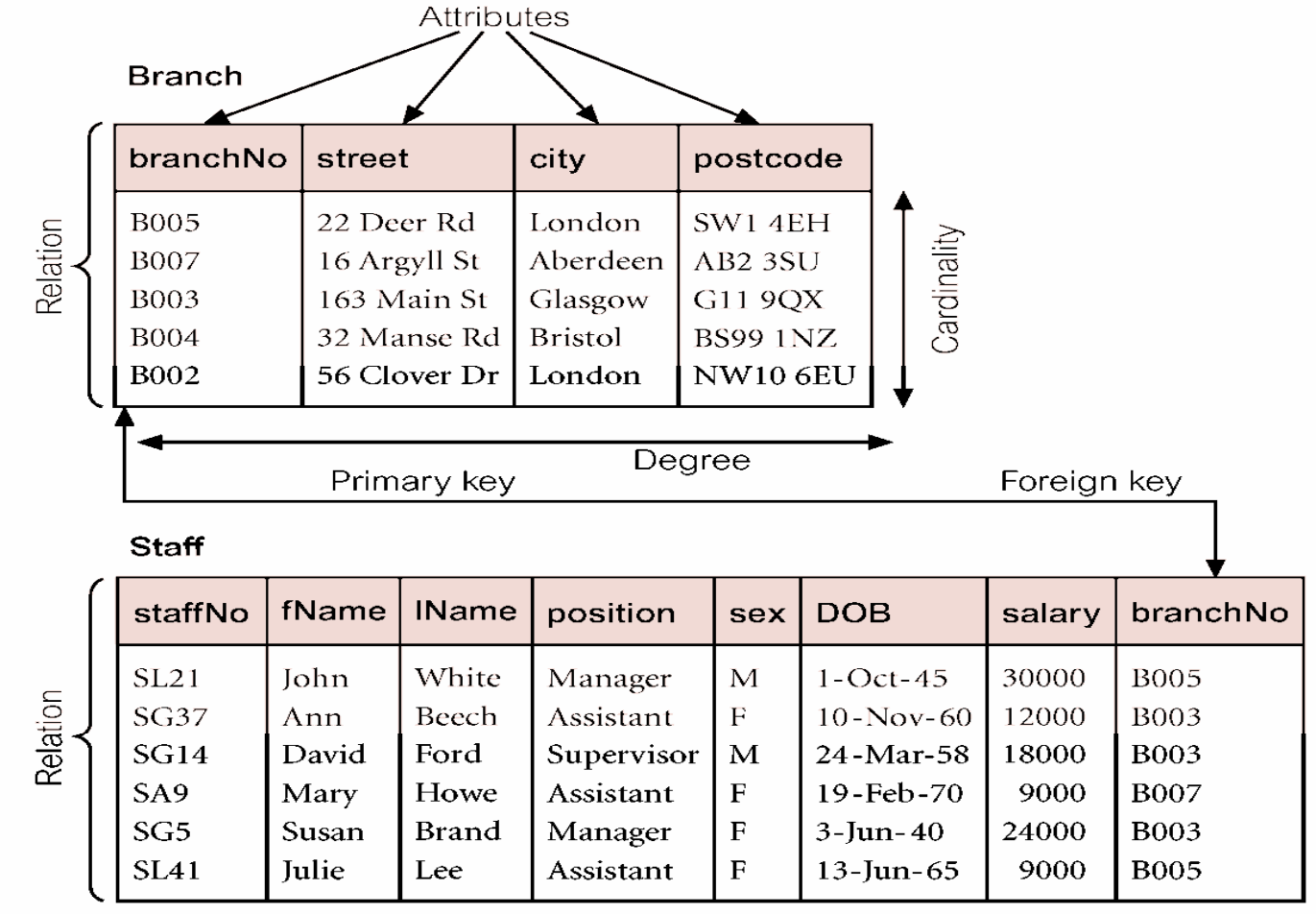
Database: definition

- a database is …
  - "a shared collection of logically related data (and a description of this data), designed to meet the information needs of an organization".
  - …in other words ….
  - a self describing collection of integrated records

Database: definitions

- data comprises entities, attributes, and relationships of an organization's information.
- the data conforms to a data model (the relational model in this course)
- the data catalogue maintains the description of the data of a particular database

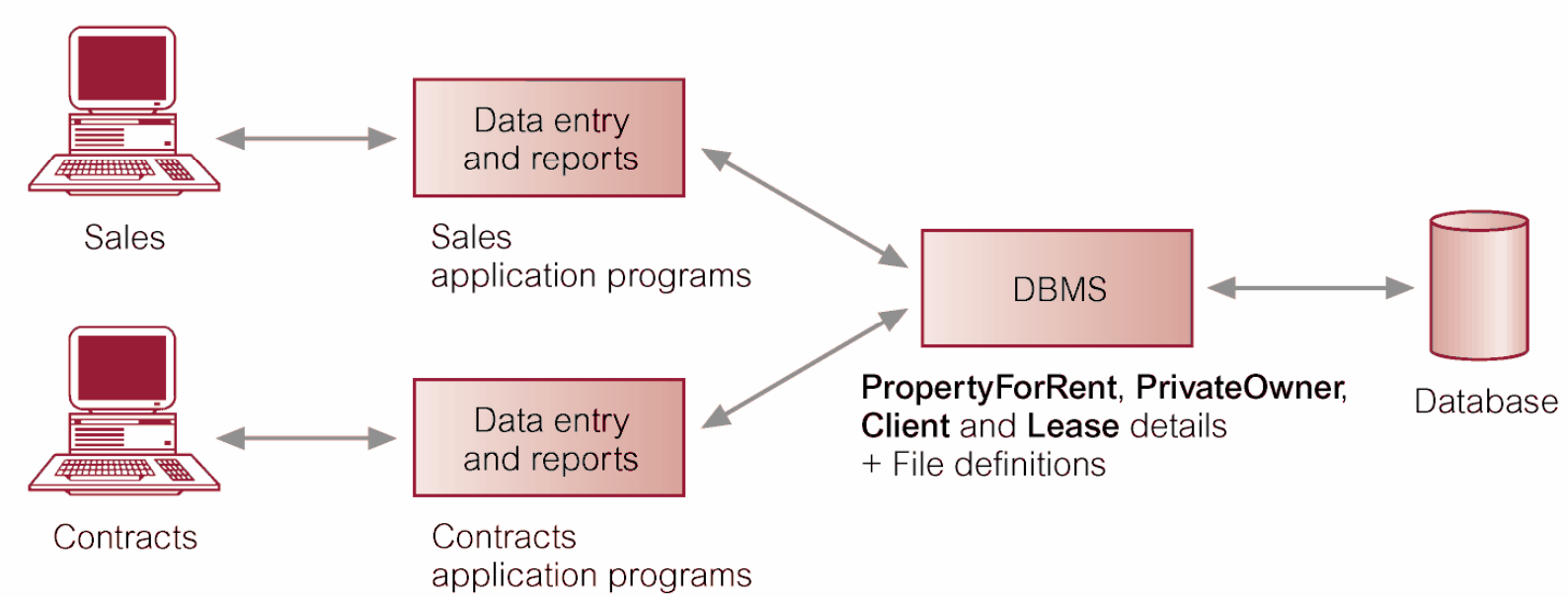
The relational data model



Database Management System (DBMS): a definition

: a software system that enables users to define, create, and maintain a database and that provides controlled access to this database.

Database Management System (DBMS)



PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)
PrivateOwner (ownerNo, fName, lName, address, telNo)
Client (clientNo, fName, lName, address, telNo, prefType, maxRent)
Lease (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)

Database Approach: data languages

- for defining databases
- For updating databases
- For retrieving data

Database Approach: controlling access to data

- A security system.
- An integrity system.
- A concurrency control system.
- A recovery control system.
- A user-accessible catalog.

Database Approach: Views

- provide each user with his or her own view of the database

- Benefits:
  - reduces complexity;
  - provides a level of security;
  - customises user interface
  - a consistent, unchanging presentation of the data, immune from changes to the underlying database.

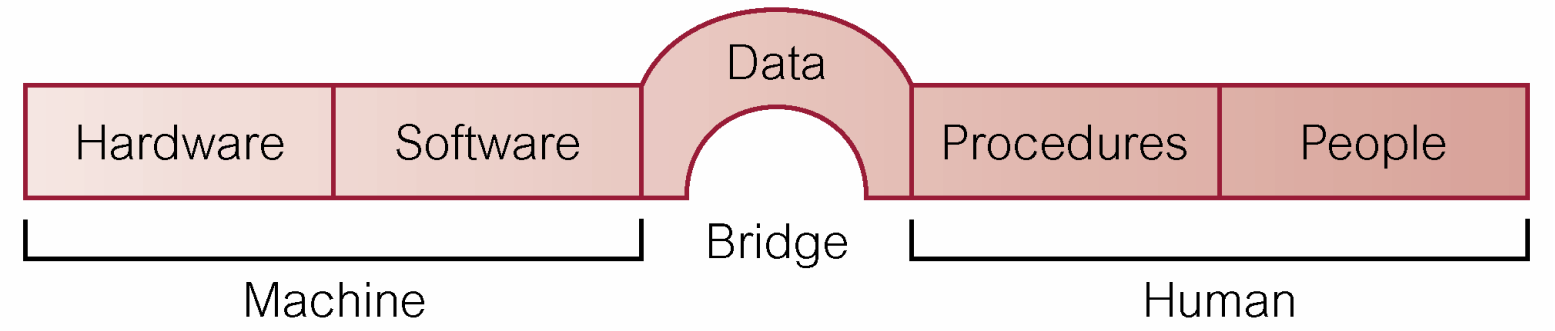
Database Approach: Views

- a virtual relation that does not necessarily actually exist in the database but is produced upon request, at time of request.

- contents of a view are defined as a query on one or more base relations.

- views are dynamic, meaning that changes made to base relations that affect view attributes are immediately reflected in the view.

Components of DBMS Environment



Components of DBMS Environment

- Hardware
  - Can range from a PC to a network of computers.
- Software
  - DBMS, operating system, network software (if necessary) and also the application programs.

Components of DBMS Environment

- Data
  - Used by the organization and a description of this data called the schema.
- Procedures
  - Instructions and rules that should be applied to the design and use of the database and DBMS.
- People

Roles in the Database Environment

- Data Administrator (DA)
- Database Administrator (DBA)
- Database Designers (Logical and Physical)
- Application Programmers
- End Users (naive and sophisticated)

History of Database Systems

- First-generation
  - Hierarchical and Network
- Second generation
  - Relational
- Third generation
  - Object Relational
  - Object-Oriented

Advantages of DBMSs

- Control of data redundancy
- Data consistency
- More information from the same amount of data
- Sharing of data

Advantages of DBMSs

- Improved data integrity
- Improved security
- Enforcement of standards
- Economy of scale
- Balanced conflicting requirements

Advantages of DBMSs

- Improved data accessibility and responsiveness
- Increased productivity
- Improved maintenance through data independence
- Increased concurrency
- Improved backup and recovery services

Disadvantages of DBMSs

- Complexity
- Size
- Cost of DBMS
- Additional hardware costs
- Cost of conversion
- Performance
- Higher impact of a failure