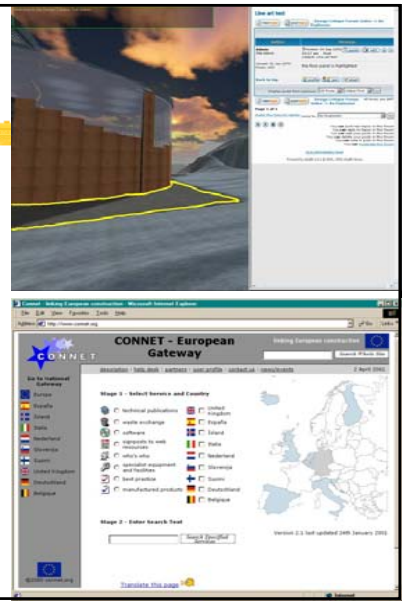
COMPSCI 732 FC – 2007 Data Mapping

Welcome!
Who Am I?
Lecture Outline
Introduction to Data Mapping

Who Am I?

- Robert Amor
  - Head of Department
  - Associate Professor
  - Computer science & Software engineering
  - 5 years in UK at Building Research Establishment
- Research interests
  - Construction IT (CAD, VR, Project workspaces)
  - Integration (Data mapping, distributed systems)
  - Interoperability (Internet portals, standards)



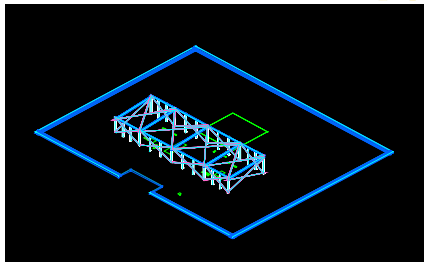
Outline of Lectures

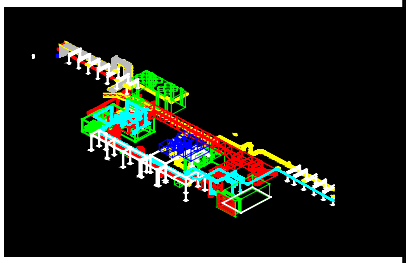
- Introduction to data mapping
- Types of mapping
- Approaches to mapping
- Mapping languages
- Specifying mappings (GUI)
- Automated generation of mappings
- Consistency management

Data Mapping

- Users and tools need information in their own specific formats
- Common data is represented differently in almost each tool
- Need to describe the transformation between representations
- Want verifiable and updateable mappings
- Want to transfer data in both directions

Views to Map





LAING

Data Mapping Issues

- Syntax
  - Data can be represented in many different encodings
  - E.g., XML, CSV, SQL, HTML, proprietary formats, etc
- Structure
  - Equivalent information can exist in vastly different structures
  - E.g., Point class in Java versus x, y, z variables
- Semantics
  - Meaning and scope of data representations are often incompatible
  - E.g., what does ‘door height’ encompass?

Where do we need mappings?

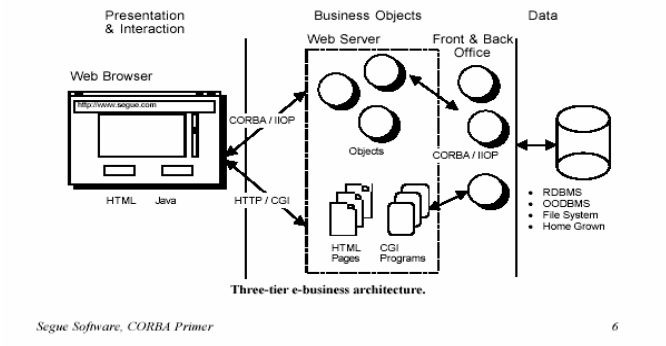
- Everywhere!
  - It is a constant task
  - Usually we don’t consider it independently
- Thinking about data mapping is another approach to understanding problems in software design
  - High-level specification (= analysis view)
  - Bidirectional data movement
    - No duplication of mapping specifications
  - Specification environment

Design the data structure for a contact database

- What information needs to be held on:
  - Contact’s details
  - Position they hold in the organisation
    - Sometimes they will have more than one job
  - Organisation they work for
    - Some data is common for the company no matter who works there

Multi-tier architectures

- Store and display different views of the same data



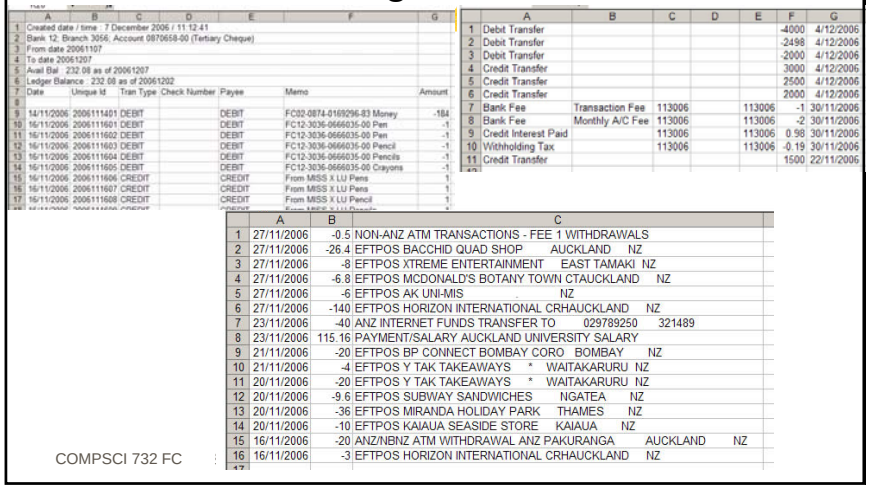
EDI (Electronic Data Interchange)

- All domains have their own messaging standards, and often several overlapping standards (e.g. medicine)

Semantic web

- Tim Berners-Lee's vision of a "machine understandable" sea of information
- Data describes itself
  - Points to a standard description of its schema
  - Tools that understand the description can use the data appropriately
  - When data is discovered it may have to be mapped to a suitable form
    - Conversion language
    - E.g., Operating range of equipment in deg F translated to deg C.

Online banking



	A	B	C
1	27/11/2006	-0.5	NON-ANZ ATM TRANSACTIONS - FEE 1 WITHDRAWALS
2	27/11/2006	-26.4	EFTPOS BACCHID QUAD SHOP AUCKLAND NZ
3	27/11/2006	-8	EFTPOS XTREME ENTERTAINMENT EAST TAMAKI NZ
4	27/11/2006	-6.8	EFTPOS MCDONALD'S BOTANY TOWN CTAUCKLAND NZ
5	27/11/2006	-6	EFTPOS AK UNI-MIS NZ
6	27/11/2006	-140	EFTPOS HORIZON INTERNATIONAL CRHAUCKLAND NZ
7	23/11/2006	-40	ANZ INTERNET FUNDS TRANSFER TO 029789250 321489
8	23/11/2006	115.16	PAYMENT/SALARY AUCKLAND UNIVERSITY SALARY
9	21/11/2006	-20	EFTPOS BP CONNECT BOMBAY CORO BOMBAY NZ
10	21/11/2006	-4	EFTPOS Y TAK TAKEAWAYS * WAITAKARURU NZ
11	20/11/2006	-20	EFTPOS Y TAK TAKEAWAYS * WAITAKARURU NZ
12	20/11/2006	-9.6	EFTPOS SUBWAY SANDWICHES NGATEA NZ
13	20/11/2006	-36	EFTPOS MIRANDA HOLIDAY PARK THAMES NZ
14	20/11/2006	-10	EFTPOS KAIAUA SEASIDE STORE KAIAUA NZ
15	16/11/2006	-20	ANZ/NBNZ ATM WITHDRAWAL ANZ PAKURANGA AUCKLAND NZ
16	16/11/2006	-3	EFTPOS HORIZON INTERNATIONAL CRHAUCKLAND NZ

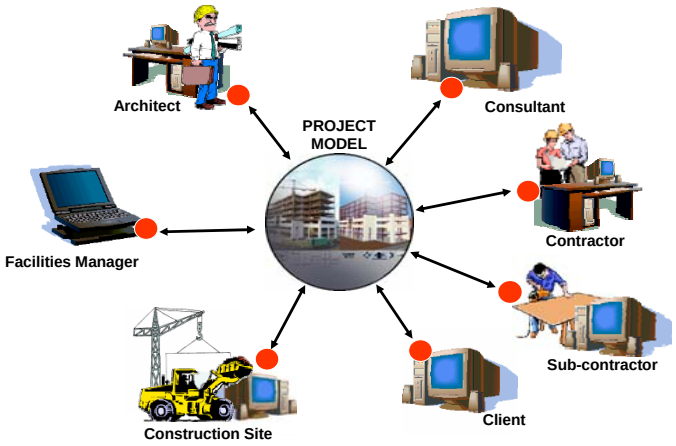
Data model standards

- Standard (ISO) data models exist, or are being developed, for many domains
  - E.g., in construction the IFCs describe major objects in a building. There are currently over 500 classes in the IFC standard.
- Tools in these domains need to map from their internal data representation to the standard, and vice-versa.
- Issues of verification and management of the developed mappings

Schema evolution

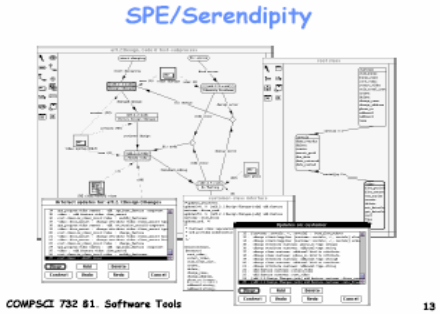
- Domain specific data models evolve over time
  - E.g., IFCs have a yearly update cycle
  - Tools need to handle the new data models
  - Tools need to map between previous versions of data models
  - Data files in old versions need to be mapped to the latest version
  - E.g., in construction domain there are over 4,500 companies developing software for sale

Integrated Environments



Software development tools

- Presentation of multiple-views of underlying form of the software



Learning Goals

- Appreciation of the importance of data mapping
- Understand the factors which impact on data mapping
- Able to specify mappings between disparate representations
- Knowledge of standards, languages, and frameworks that can be used for data mapping
- Knowledge of approaches to maintaining consistency between mapped data