



MISSION

Multi-Agent Integration of Shared Statistical Information Over the [inter]Net

Joanne Lamb,
CES, University of Edinburgh,
Scotland, UK

IASSIST/IFDO 2001 - A DATA ODYSSEY
AMSTERDAM, MAY 14 - 19, 2001

CES



Overview



- The Mission project
- Some concepts
- The Architecture
- Relevance to Data Archives
- Relation to other projects



The Project



- European Union R&D project - IST-1999-10655
- <http://www.epros.ed.ac.uk/mission>
- The Partners
 - » University of Edinburgh, Scotland, UK (co-ordinator)
 - » Office for National Statistics, UK
 - » Central Statistics Office, Ireland
 - » Tilastokeskus (Statistics Finland), Finland
 - » University of Athens, Greece
 - » University of Ulster, Northern Ireland, UK
 - » Desan Marktonderzoek BV, The Netherlands

IASSIST/IFDO 2001 - A DATA ODYSSEY
AMSTERDAM, MAY 14 - 19, 2001



Key Features



- Build a software suite allowing statistical data providers to publish data on the Web
- Grew from ADDSIA
- Table model and Presentation requirements
- Transporting aggregated data over the Web
- Combining aggregated data
- Publishing Methods
- Third party Metadata server

Objectives

- **Suppliers** can subscribe to an integrated network of datastores via an interface to their existing data
- **Suppliers** retain control over all aspects of access to their existing data
- **Users** can make requests in a declarative manner, with a minimum of understanding of statistics, or the domain area, and still retrieve meaningful results
- **Users** can tailor their environment, from simple requests to detailed in-depth analysis
- **Users** can build up individual profiles, accessing data and methods most relevant to their needs.



Objectives



- **Methods** of data manipulation and analysis can be retained, re-used and published
- **Libraries** of metadata and tables can be constructed and made available to other users
- A flexible architecture to allow third parties to act as **Independent Metadata Providers**
- Implemented on **independent, interoperable systems** running on different hardware platforms and accessing heterogeneous data storage systems.



Mission Concepts and Innovation

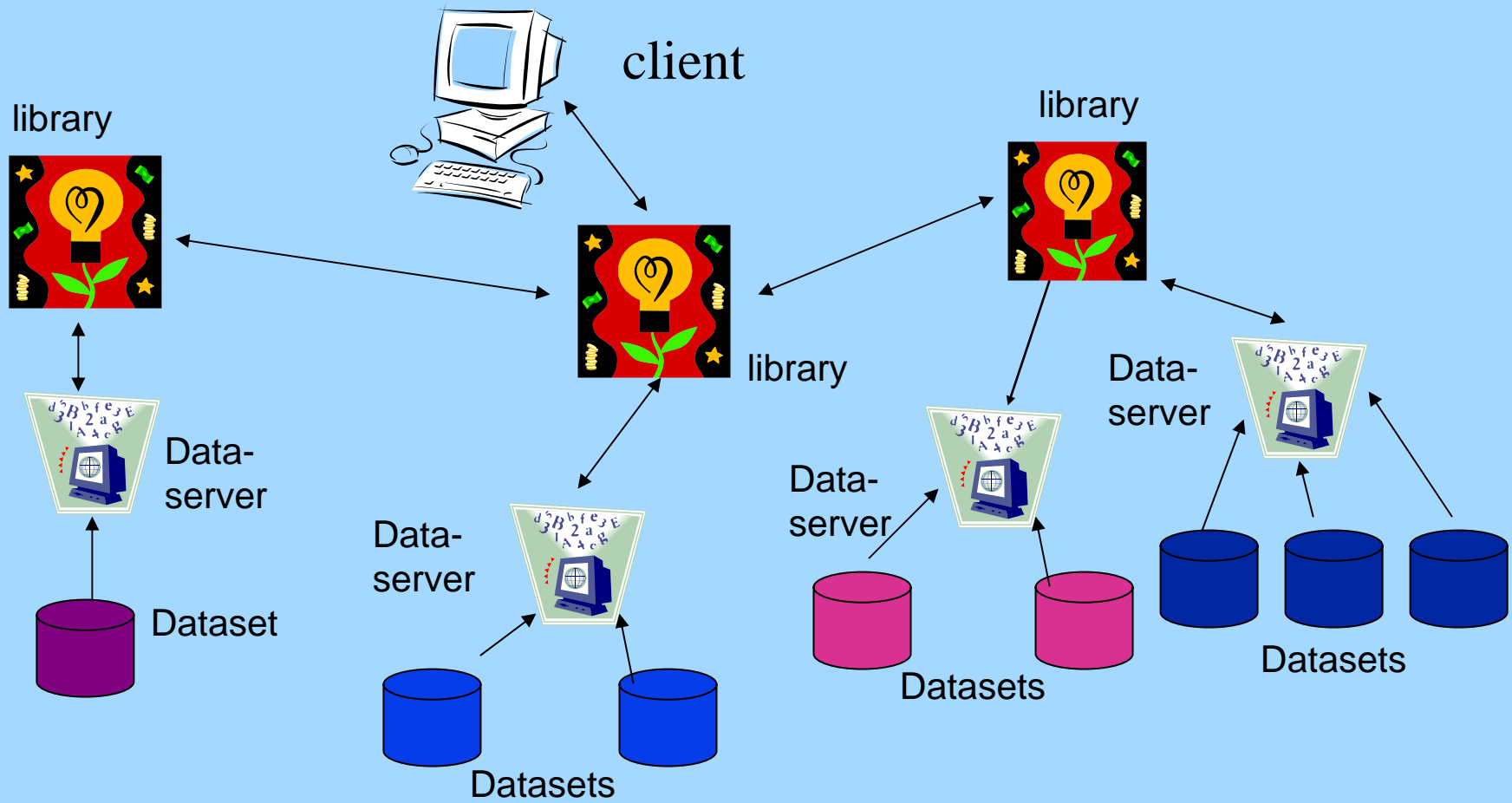


- Agent technology – ‘rough queries’ and optimisation
- Independent Metadata Libraries
- Visual specification of a query
- Experts sharing methodologies
- Three basic components

IASSIST/IFDO 2001 - A DATA ODYSSEY
AMSTERDAM, MAY 14 - 19, 2001

Key components

- Client
 - » Downloaded user interface that connects to a host Mission site (a Library)
- Library
 - » A repository for statistical metadata: access, methodological and contextual.
 - » A statistical processing engine
 - » Libraries communicate with each other via agents
- Data server
 - » The unit which gives access to the data
 - » Links to the data provider's data
 - » Giving control of access
 - » Only macro data sent to the Library
 - » Data servers connect to one or more Libraries





Implications



- The components can be deployed in different scenarios
- Public Users can browse metadata and published results
- Registered users can download software and do their own analysis
- The Client captures the metadata of the analysis process
- Metadata server is independent of the data provider



Data Archives and Data Libraries



- The inspiration for the Independent Metadata Provider scenario
- Typically a researcher gets data from an archive
- At the end of the project h/she publishes theoretical papers
- But the modified data is destroyed
 - » Legal reasons
 - » Economic reasons
- So the expertise in building up the methods applied to this data is lost
- But this information could be held for re-use in a metadata library



Transformations: different methods



Method	Comment
Computation	Uses an arithmetic expression to construct the result
Truncation	Removes one or more digits from the end of a code – usually for a classification
Rule-based	A series of instructions of the form IF xxx THEN result = y
Banding	A specification that a range of values should be collapsed to a single code
Transcoding table	A table of initial and final results
Complex	A combination of methods



Publishing 'methods'



- Allow users to publish developed methodology
- Other can use transformations identified by subject matter experts
- Benefit from shared expertise
- Can also publish tables derived via MISSION



Some other concepts



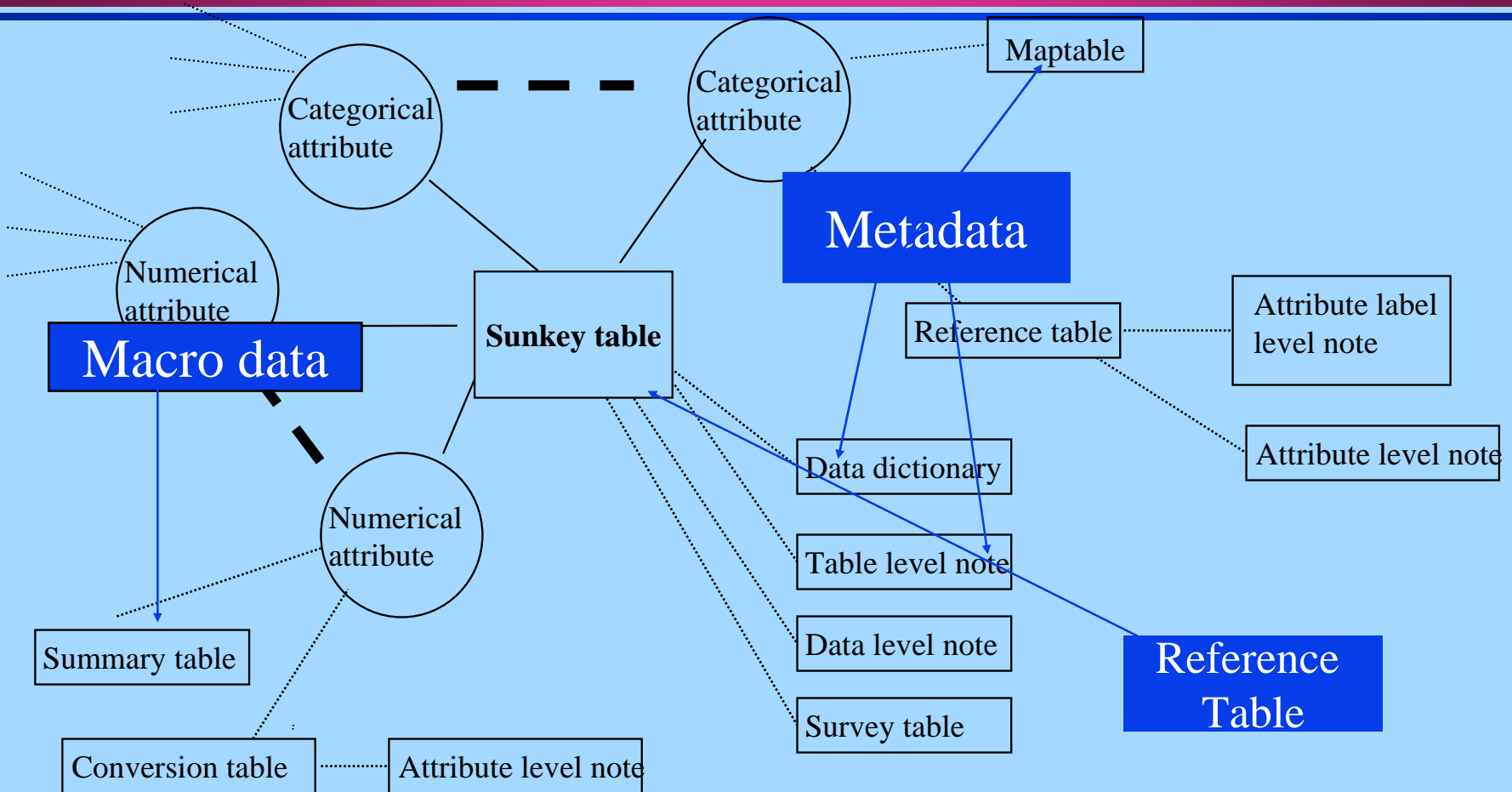
- The table driven user interface
 - » Specifying the query
 - » Modifying the results
- The data model
- Using agent technology

		Var 1			Var 2					
					Male			Female		
		Cat1	Cat2	Total	c1	c2	T	c1	c2	T
Course 1	School National									
Course 2	School National									
Course 3	School National									
Total	School National									

Merge Courses

Insert variable

Insert splitting variable: e.g. gender





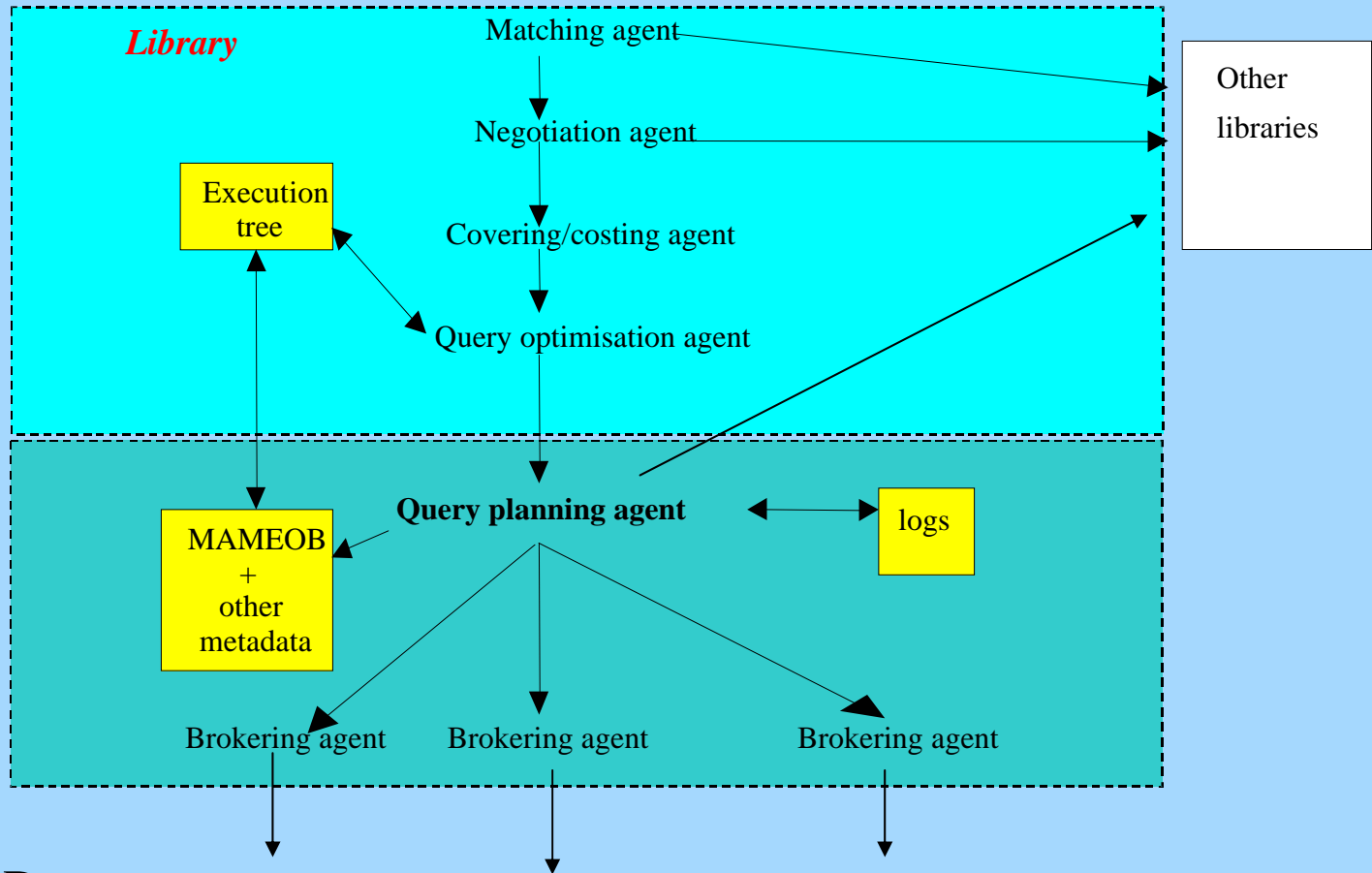
Mapping the data model



- We have demonstrated that this model maps to other common models,
- such as data warehouse structures, cubes and the CRISTAL model.
- Paper by Bryan Scotney
- <http://www.epros.ed.ac.uk/metanet/conferences/proceedings.html>

Query construction

Query execution



To Data Servers

IASSIST/IFDO 2001 - A DATA ODYSSEY
AMSTERDAM, MAY 14 - 19, 2001

- Test1: connectivity
 - » February 2001
- Test2: simple query on macro data
 - » May 2001
- Test 3= Prototype 1
 - » September 2001
 - » User can
 - browse metadata i.e., view dimensions of an aggregated dataset
 - produce 'simple' tables from one source
 - produce 'composite' tables from different homogeneous sources
- Prototype 2: March 2002
- Final version: December 2002

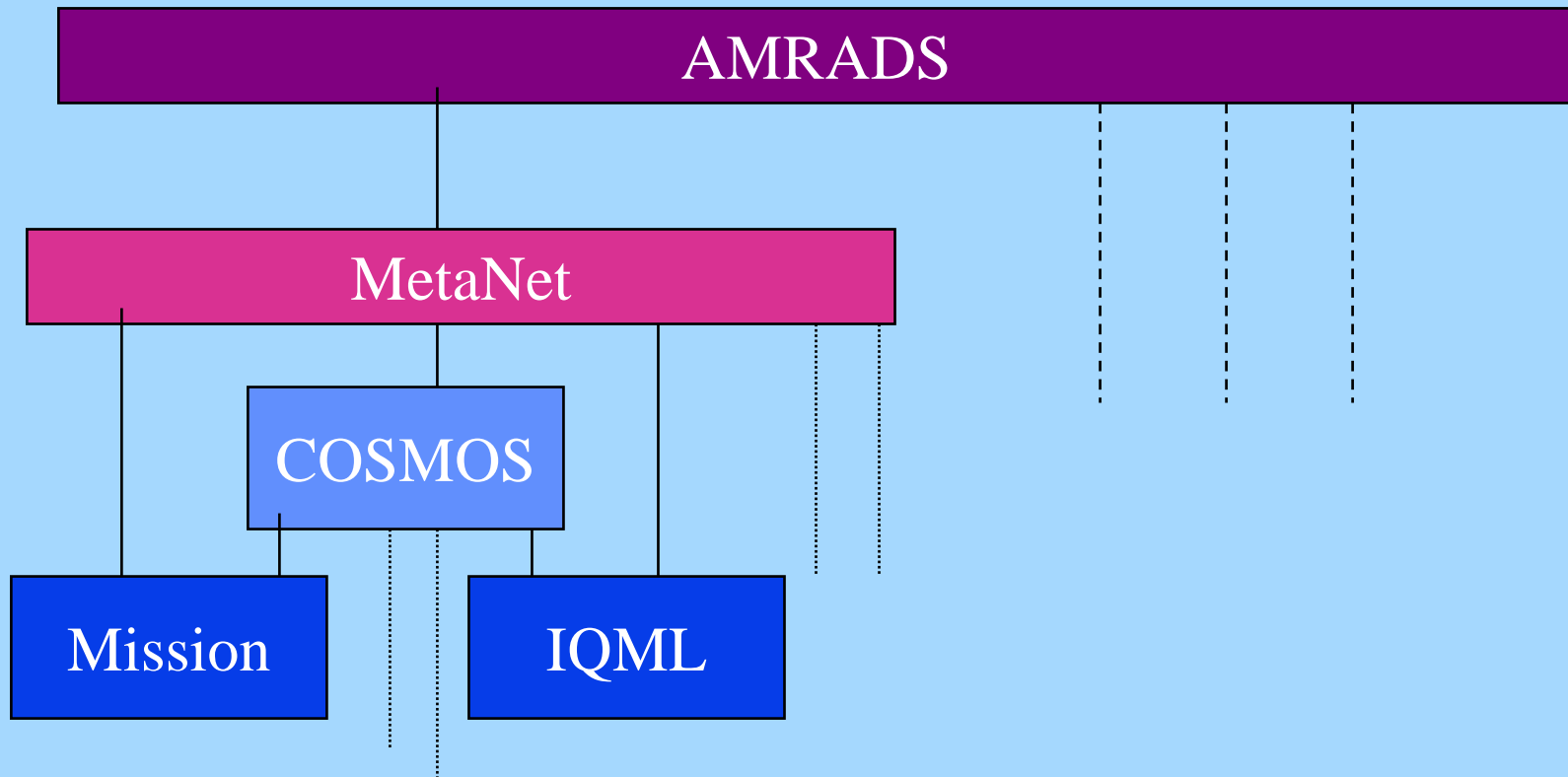
- Participating in 3 networking projects
 - » And one other R&D (IQML)
- MetaNet
 - » A network of excellence for harmonising and synthesising the development of statistical metadata
 - » Started November 2000
 - » First conference 2-4th April 2001
 - » Proceedings – end May
- AMRADS
 - » Accompanying Measure to R&D in Statistics
 - » UEDIN representing metadata
 - » ETK/NTTS2001 conference June 2001

- COSMOS

- » Cluster Of Systems of Metadata for Official Statistics
- » Starts Sept 2001
- » Projects are:
 - Mission
 - IQML
 - FASTER
 - Metaware
 - IPIS

- References:

- » <http://www.epros.ed.ac.uk>
- » <http://www.cordis.lu/ist/projects>



Conclusions

- MISSION is an open system, which aims to help users exchange methodologies as well as utilise data from different sources
- This give the opportunity for data archives and data libraries to host Metadata sites as well as access to data
- We deliberately called these site Libraries, since we feel that metadata knowledge is held at the brokering level, rather than by the data providers
- MISSION is also contributing to other activities aimed at getting a shared understanding of statistical metadata