European Spatial Data Research

July 2003

OEEPE Workshop on Next Generation Spatial Database – 2005

Editor: Keith Murray

OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management

Editors: M. Orhan Altan, Hayati Tastan

Official Publication Nº 45

The present publication is the exclusive property of EuroSDR

All rights of translation and reproduction are reserved on behalf of EuroSDR.
Published by EuroSDR

Printed by Gopher, Groningen, The Netherlands

Content

Steering Committee

Keith Murray, Arbind Tuladhar, Peter Woodsford, Ed Parsons:
Workshop Next Generation Spatial Database – 2005

Peter van Oosterom, Christiaan Lemmen, Rolf A. de By, Arbind M. Tuladhar:
Geo-ICT technology push vs. Cadastral market pull

Presentations:

Fred Finch – OS Ireland:
Geospatial Database from Proprietary to Open

Frank Fuchs, IGN, France:
New database developments in IGN

Jens Ole Jensen – Kort & Matrikelstyrelsen:
A Geospatial Datawarehouse to provide cadastral and other data

Xavier Lopez, Oracle:
Will your data service meet the customer needs in 2005?

David Maguire, ESRI:
Building applications with government data – needs for the future.

Gerhard Muggenhuber – Austrian Cadastre (BEV):
Cadastral needs and developments

Land Administration in Europe:
- Parameters
- Organizations
- Customers

Austria:
- BEV-Organization
- Spatial data structure
- Cadastre

Keith Murray:
Introduction

Ed Parsons, Ordnance Survey:
From the outside to the inside & back again

Peter van Oosterom (TU Delft) & Christiaan Lemmen (Kadaster) and Rolf de By (ITC), Arbind Tuladhar (ITC):
Geo-ICT technology push vs. Cadastral market pull

M. Orhan Altan, Hayati Tastan:
OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management

List of Publications

OEEPE Commission 4 – Workshop Report

Next Generation Spatial Database – 2005

Authors: Keith Murray, Arbind Tuladhar, Peter Woodsford and Ed Parsons.

Held at Ordnance Survey, Southampton on 22-24 May 2002

This Summary

This is a summary of the OEEPE workshop “Next Generation Spatial Database – 2005” hosted by Ordnance Survey in Southampton in May 2002. The summary provides an introduction to the accompanying presentations and highlights the issues raised by the presenters and participants in the following discussions. It is intended that together this summary, paper and presentations will help inform a wider audience of the current trends, directions and issues facing those whose role it is to ensure that large national spatial databases meet future strategic needs.

We believe that this is essential if spatial data is to fulfil a pivotal and leading role in developing the knowledge economy to support decision making at all levels. Further, this is particularly relevant in a world where much is rapidly changing - from users needs and applications, e-government (which is a major stimulus everywhere), technology, new ways of working and so on.

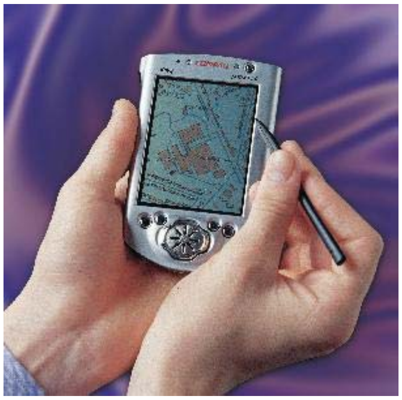
The Workshop

The workshop attracted over 30 participants from 12 different countries and built on the successful “NMA and the Internet” OEEPE workshop held in March 2000 at Ordnance Survey. This event followed a similar format with four sessions dedicated but different aspects of database development facing National Mapping Agencies and Cadastral organisations over the next few years. To stimulate thought each session included 2-3 presentations prior to a breakout session where teams addressed some of the key issues facing them. The meeting programme will be found at Annexe A.

Ordnance Survey Head Office, Southampton

Aim and Purpose

The demands of the emerging information and knowledge economies are already having a major impact on those agencies which supply data and information. New developments such as Location Based Services, web publishing, print on demand are now challenging the content and structure of national spatial databases.



Location Based Services

Print on demand mapping

People now need to *answer questions* such as “Where is my nearest?”, “How do I find/get to?”, “What will the local impact be?”.

Technology and competition are also having a significant impact. Many of the Mapping Agencies and Cadastral Agencies are advanced in addressing these issues and some are just starting. All too often those facing these challenges have to work in isolation with few peers at the national level. There is often no one to share some of these big strategic issues and discussions with.

This workshop was seen as an opportunity for this small community of experts to come together and share their experiences and future directions on key data management issues.

David Willey, Deputy Chief Executive of Ordnance Survey, welcomed the participants to the workshop and opened the meeting by illustrating some of the challenges we all face.

Session 1. Geospatial databases – the global drivers for change

Peter Woodsford (OEEPE Commission 5 Chair) chaired the first session which set the scene with an imaginative and stimulating keynote presentation from Tony Davison of IBM. Tony immediately moved us forward to the year 2025. In his animation of the future - we are all connected to a wireless “GRID” where keyboards and mobile phones are seen as relics from the past. In the near future communication and information for decision making is always at our fingertips. Geospatial data played a part in this world, but in a different way from today, it would be totally integrated into the fabric of the society he described. He proposed that the power of this technology will transform the way we conduct our lives, both domestic and in business.

Ed Parsons (Ordnance Survey) used his recent experience in industry and as a customer of mapping agencies, to highlight the issues facing the data providers as we all move towards the information society and the radical changes described by Tony Davison. Ed described OS MasterMap and recognised the need for ever greater attention to data management, data and system architectures.

Following the presentations the subject the teams addressed was to determine:

The top three issues facing NMA’s and Cadastral agencies & their information.

The main issues raised by the teams were:

- 1. Needs are changing (the only constant is change)
    - Living in a world of rapid change and no sign of this stopping
    - Information society will expect different products
    - NMA's future is not assured
    - We need to fit with the emerging "bigger picture" (new customers needs, National SDI's [spatial data infrastructures] and wider SDIs eg the EC INSPIRE initiative,)
    - Technological trends changing – need to take risks
    - Personal and cultural development is as critical
    - Need for greater cost recovery to support sustained investment
- 2. Customers
    - We all need to be more customer-driven
    - Need to understand business processes and integrate geography more centrally into those processes (internal organisational and external eg Location Based Services)
    - The right information in a timely and standard way
    - Includes 3D and time
    - Data Quality: Currency, Consistency, Completeness and Accuracy.
    - Concentrate on core data for today and tomorrow
    - Geography key to joining up disparate information, especially across government
- 3. Standards
    - Need to be "open"
    - Need to assimilate National, European and International, de facto and de jure
    - Currently hard to determine which "to back"

These three issues set the scene and recurred as the following two days of discussion while exploring different topics and areas of common interest.

Session 2. The National Mapping Agency – What data, why, when, how?

Keith Murray (OEEPE Commission 4 Chair) chaired the session dedicated to National Mapping Agency needs and developments. Fred Finch of Ordnance Survey Ireland reported on strategic developments over the past 18 months since the "NMA and Internet" workshop in March 2000. He specifically stressed the move towards an "open" architecture to support strategic flexibility in the future.

Jens Jensen of Kort & Matrikelstyrelsen (Denmark) described their database developments with TOP10DK and experiments and plans for electronic data services.

Frank Fuchs of IGN France provided a presentation of IGN's key needs and developments.

On this occasion teams tackled the following subject in the breakout session:

What architecture strategies (data & systems) are required to meet the needs of 2005?

The team findings were:

- We need two main forms of dedicated database, they have different roles:
    - 1. Data maintenance – ie the master database
    - 2. Publication – for delivering products from the master database,

- Other key issues:
    - Unique Identifiers (eg TOIDs) to provide information “hooks”, and support customers needs for data sharing
    - Data Modelling
    - Geography Mark-up Language,
    - Versioning, Long Transactions, Editing large objects,
    - Metadata,
    - The need for internal data integrity
    - Third party data linkages (and the fact that data conflation usually highlights discrepancies)
    - Open Standards for Publication are currently more advanced than for Data Maintenance

Session 3. The Cadastre – New services, new directions

Arbind Tuladhar (ITC) chaired and made the first presentation of this session on behalf of Peter van Oosterom (TU Delft), Christiaan Lemmen (ITC & Kadaster) and Rolf de by (ITC). He outlined the changing world in which the cadastre now operates in and the increasingly varied and demanding demands which the information and knowledge society make on it. The authors also include a paper – see Annexe A.

Gerhard Muggenhuber (BEV, Austria and Chair of FIG Commission 3 [Spatial Information Management]) showed that these changes are widespread across the world via summaries of the situations in several countries.

The question facing the teams after this session was:

What kinds of data models will ensure data integration for future applications?

The team findings were:

- Data models are critical for database consistency
- Uses of UML including Use cases and XML/GML seen as important base
- Object lifecycles need to be modelled
- Testbeds are useful to evaluate concepts
- Standards – international level is important, (ISO/TC211 and OpenGIS)
- NMA’s/Cadastral agencies should work more closer together – “as an expert group”

Session 4. Database Technologies – what can they do for us? When?

In the final session Ed Parsons (CTO, Ordnance Survey) sought an industry view from Oracle, ESRI and eXcelon. Xavier Lopez (Oracle) provided an overview of the directions and key developments at Oracle, while David Maguire (ESRI) showed how GIS functionality, database technology (such as Oracle) and data can combine to serve the needs of customers in the future. Adrian Marriott of eXcelon described the benefits of a true object-oriented database technology in meeting future needs.

The final breakout session subject was:

How will the database vendor’s solutions help me achieve my strategy?

The teams concluded that the key areas to concentrate on are:

- Data consistency - need to collect/edit once – use many times
- Topology needs developing more generically – role for OpenGIS here?
- Data management tools to be developed:
    - Conflict management
    - History/Time

    - 3D
    - Multi-resolution rendering (and generalisation?)
    - Object rules/behaviours
    - Change detection tools
    - Schema Evolution Tools
    - *While maintaining database performance*
- Vendors open up users ability, but can also limit (if they do not move forward in a desired direction)
- Backwards compatibility is required
- International Standards could help

Overall conclusions

As with the Internet Workshop the attendees found that attending this meeting and sharing issues such as those described above, common to each organisation, very helpful as we move forward with the same challenges in each country.

Three generic technical issues emerging were:

- Common Data Models need to be developed – maximise the effectiveness of our operations.
- Cadastral v NMA requirements are often different – but share common challenges
- "Openess" of databases is an aim – in this fast moving world

A recurring and key area of interest is knowledge sharing on data modelling across NMAs and cadastral agencies. Some organisations had done a lot of work on this and others less so. There are also several common activities emerging, such as the linking of topographic and cadastral databases in many countries across Europe and around the world. There is evidence of this where the NMA and cadastre is the same organisation and where they are separate.

Another common problem for many was the need to get from where they are today to where they know they need to get to for the future. Combining such a major step change in transition, while maintaining an uninterrupted supply of products to customers in the meantime is a major challenge. Data re-engineering to meet a future schema while "keeping the show on the road" requires major IT investment and commitment right across the organisation.

Next Steps

Taking all the above into account it is proposed that there would be benefit in some collaborative work to bring together experience with data models to determine what works and what does not. This could help shape "best practise".

This could be a follow on project if there was enough support from members. One approach might be that these subjects will be developed in the next step as joint research efforts within OEEPE and FIG [Commissions 3 and 7] and the results presented in joint sessions of both organisations in the future.

Finally we would like to thank all the participants for their participation in what proved to be both a specialist but very rewarding workshop.

We would also like to thank Ruth Williams (Ordnance Survey) who put in a lot of effort in organising and arranging the workshop.

Authors: Keith Murray, Arbind Tuladhar, Peter Woodsford and Ed Parsons.

Workshop Convenor: Keith Murray, Chair OEEPE Commission 4
August 2002

WORKSHOP PROGRAMME ANNEXE A

Wednesday 22nd May 2002

12.30 – 13.30	*Lunch*		*v1.02*
14.00	*Introductions*	*OEEPE*	
14.10	*Welcome*	*David Willey,* *Deputy CEO Ordnance Survey*	
Geospatial databases – the global drivers for change			
Chair	***Peter Woodsford***		
14.30	Implications for Geographic Information providers in the Information Age - Keynote	Tony Davison, IBM	
15.30	From the outside to the inside & back again	Ed Parsons, Ordnance Survey	
16.00 – 16.30	Tea/Coffee		
16.30	Breakout Session	Subject: Top three issues facing NMA's and Cadastral agencies & their information	
17.30	Reporting back		
18.00	Close		

Thursday 23rd May 2002

The National Mapping Agency – What data, why, when, how?			
Chair	***Keith Murray***		
09.00	Geospatial Database from Proprietary to Open	Fred Finch – OS Ireland	
09.30	A Geospatial Datawarehouse to provide cadastral and other data	Jens Ole Jensen – Kort & Matrikelstyrelsen	
10.00	New database developments in IGN	Frank Fuchs, IGN, France	
10.30 – 11.00	Tea/Coffee		
11.00	Breakout Session	Subject: What architecture strategies (data & systems) are required to meet the needs of 2005?	
12.00	Reporting back		
12.30	Lunch		

The Cadastre – New services, new directions			
Chair	***Arbind Tuladhar***		
14.00	Geo-ICT technology push vs. Cadastral market pull*	Peter van Oosterom (TU Delft), Christiaan Lemmen (Kadaster), Rolf de By (ITC); Speaker: Arbind Tuladhar (ITC)	
15.00	Cadastral needs and developments	Gerhard Muggenhuber - Austrian Cadastre (BEV)	
16.00 – 16.15	Tea/Coffee		
16.15	Breakout Session	Subject: What kinds of data models will ensure data integration for future applications?	
17.00	Reporting back		
17.30	Close		

* ”Geo-ICT technology push vs. Cadastral market pull” - paper included.

19.30	**Workshop Dinner**	**Southampton Football Club**

Friday 24th May 2002

Database Technologies – what can they do for us? When?			
Chair	***Ed Parsons***		
09.00	Will your data service meet the customer needs in 2005?	Xavier Lopez, Oracle	
09.30	Building applications with government data – needs for the future.	David Maguire, ESRI	
10.00	Realising the benefits of object technology	Adrian Marriott, eXcelon	
10.30-11.00	Tea/Coffee		
11.00	Breakout Session	Subject: How will the database vendors solutions help me achieve my strategy?	
12.00	Reporting back		
12.30 – 13.00	Summing up & close		
13.00 – 14.00	Lunch		

PRESENTED PAPER ANNEXE B

Peter van Oosterom (TU Delft), Christiaan Lemmen (Kadaster), Rolf de By (ITC) and Arbind Tuladhar (ITC):

Geo-ICT technology push vs. Cadastral market pull

PRESENTATIONS ANNEXE C

Fred Finch - OS Ireland:
Geospatial Database from Proprietary to Open

Frank Fuchs, IGN, France:
New database developments in IGN

Jens Ole Jensen – Kort & Matrikelstyrelsen:
A Geospatial Datawarehouse to provide cadastral and other data

Xavier Lopez, Oracle:
Will your data service meet the customer needs in 2005?

David Maguire, ESRI:
Building applications with government data – needs for the future.

Gerhard Muggenhuber – Austrian Cadastre (BEV):
Cadastral needs and developments

- Land Administration in Europe:
  - Parameters
  - Organizations
  - Customers
- Austria:
  - BEV-Organization
  - Spatial data structure
  - Cadastre

Keith Murray:
Introduction

Ed Parsons, Ordnance Survey:
From the outside to the inside & back again

Peter van Oosterom (TU Delft) & Christiaan Lemmen (Kadaster) and Rolf de By (ITC), Arbind Tuladhar (ITC):
Geo-ICT technology push vs. Cadastral market pull

Geo-ICT technology push vs. Cadastral market pull

Peter van Oosterom , TU Delft, The Netherlands
Chrit Lemmen, Kadaster/ITC, The Netherlands
Rolf deby, ITC, The Netherlands
Arbind M. Tuladhar, ITC, The Netherlands

OEEPE Workshop 2002, 22-24 May 2002, Southampton, UK

Agenda

- Introduction
- User requirements
- Geo-ICT Development
- Proposal and conclusions

Introduction: Technology push

- Geo-ICT developments:
  - Modeling standards
  - Database technology
  - Positioning systems
  - Internet development
  - Wireless communication
- Geometry accepted in mainstream ICT

Introduction: Cadastral Market pull

- User requirements of Cadastral systems change over time, due to:
  - Change in legislation
  - Governmental policy
  - New tasks for the organization
  - New technology
- Therefore, generic and flexible systems needed → Model Driven Architecture

Introduction: Geo-ICT Development

- Geo-ICT developments: Improve quality, cost effectiveness, performance and maintainability of Cadastral Systems
- GI part of integrated Information System architecture within organization
- Next step: information communities between organizations (GSDI)

Agenda

- Introduction
- User requirements
- Geo-ICT Development
- Proposal and conclusions

User requirements: overview

- UN/ECE land administration guidelines
- UN/FIG Bathurst Declaration
- Business requirements; based on observation from Netherlands Cadastre

UN/ECE: Land administration guidelines

- User=anyone interested in land matters
- Assess needs throughout lifetime
- Step-by-step approach (costs), aspects:
  - System components compatible→ standard
  - Data protection, key=unique parcel number
  - Avoid redundancy, update mechanism
  - Asses new technology (GPS) from economic angle
  - Secure data quality (topology, integration)
  - Orthophotomaps maybe (specific req's, cost, time)
  - 3D representation (strata titles) → in some places
  - Cost-effective: info sales, linking to other info
  - Public must understand 'presented' information

UN/FIG: Bathurst Declaration

- Importance of ICT is stressed
  – Constructing the infrastructure
  – Effective citizen access to information
- Maintain/upgrade infrastructure (system)
  – Flow of info in government (local - central)
  – Good land administration through incremental, inexpensive, user-driven systems
  – Multi-disciplinary, multi-national training courses
  – Territorial oceans (UN convention Law of the Sea) need support for marine resource administration

Business requirements: Netherlands Cadastre

- Information society, customers directive role
  - Registration process: clear, simple and speedy
  - Information access: complete, reliable, and rapid
- Chain-orientation, digitization → new and flexible products (standard based)
- Info 7*24 hours available (everywhere)
- E-government: Internet access, electronic conveyance, one-stop shopping
- Mobile computing: data collection, but also data distribution (applications)
- Land administration is part of (G)SDI

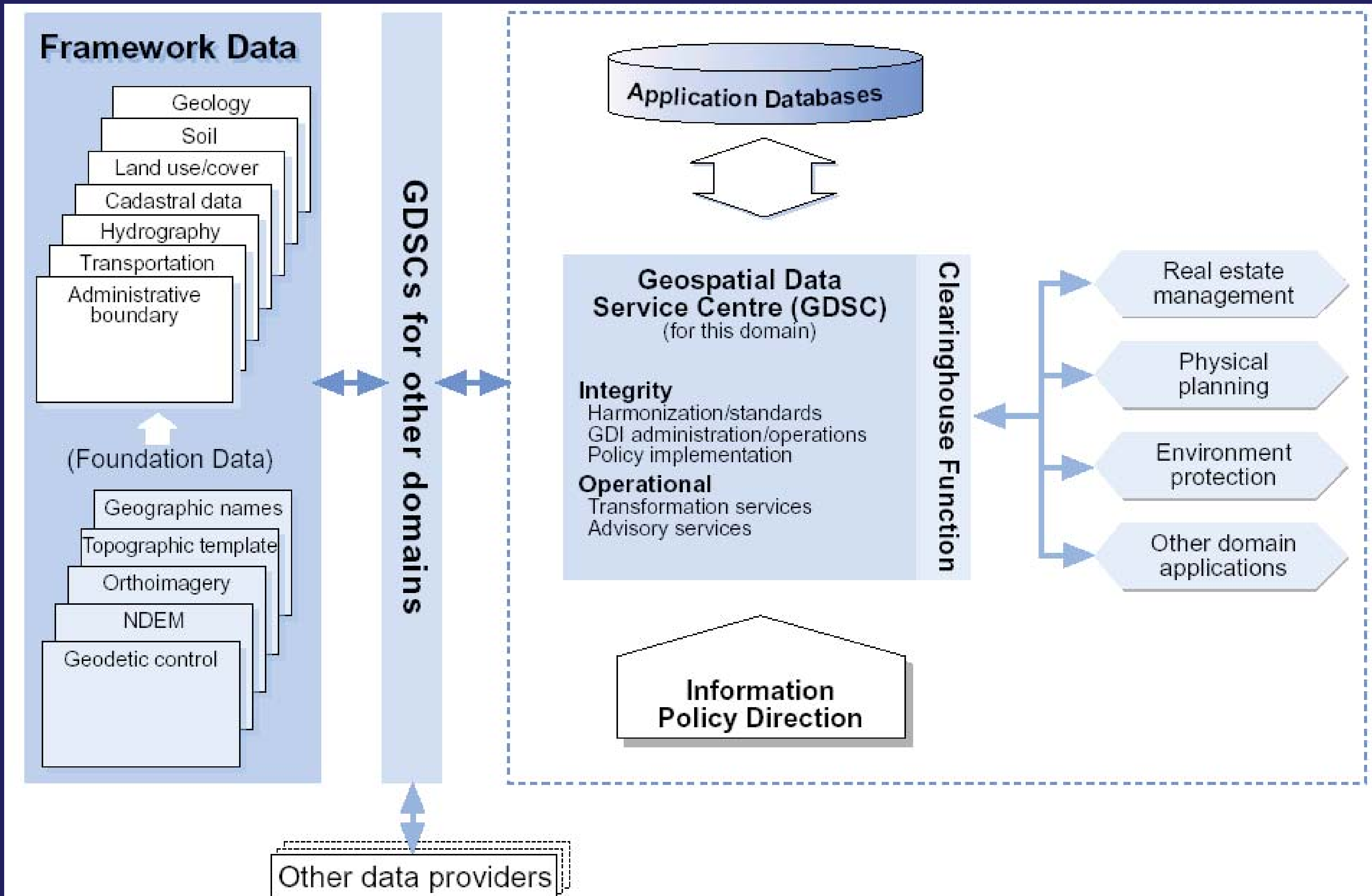
Agenda

- 1. Introduction
- 2. User requirements
- 3. Geo-ICT Development
- 4. Proposal and conclusions

Geo-ICT development: overview

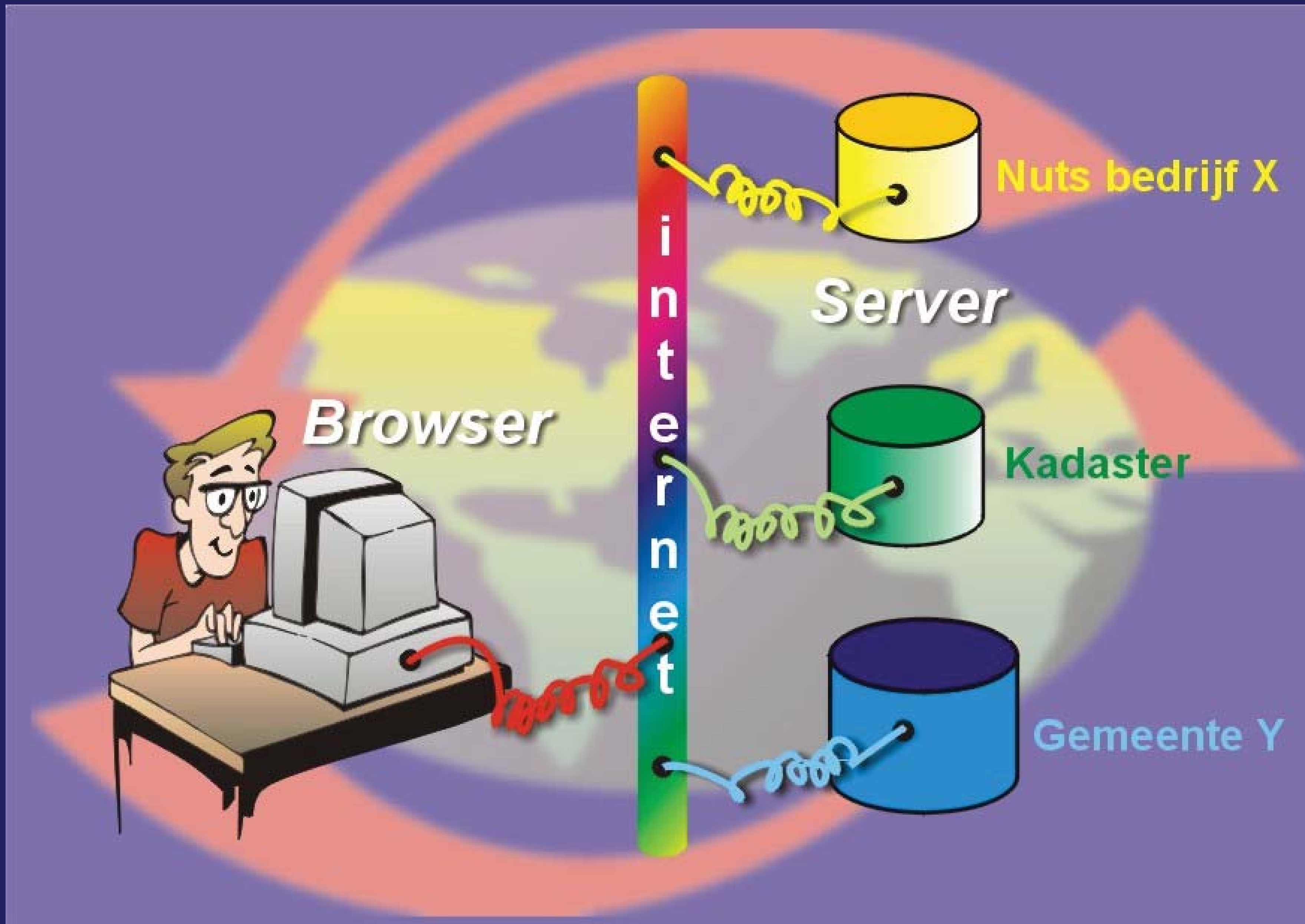
- GSDI & Cadastre
- OpenGIS standards
- Modeling standards (UML, XML)
- Database technology (geo)
- Location based services

GSDI & Cadastre (Groot & McLaughlin, 2000)



GSDI & Cadastre

- GSDI: (global) access to GI
- Based on 3 ingredients:
  - Geo-data sets (foundation, framework,…)
  - Geo-data services (e.g. Geo-DBMS)
  - Interoperability standards
- With organizational, legal, financial, technical aspects



GIS technology
Kadaster
TU Delft

OpenGIS standards: overview

- Consortium (industry, government, research): fast/effective standards
- Harmonization agreement ISO, TC211
- Two levels: abstract and implementation specifications
- Three topics in a little more detail:
  - Feature geometry
  - Meta data, catalog services
  - Internet GIS

OpenGIS standards: Feature Geometry

- Simple Feature Geometry: basic spatial types and functions (2D)
- Available for different types of computing platforms (Corba, OLE/COM, SQL) since 1998
- Certified products available
- Current work: implementation specs for 3D types, complex structures (topology), etc.

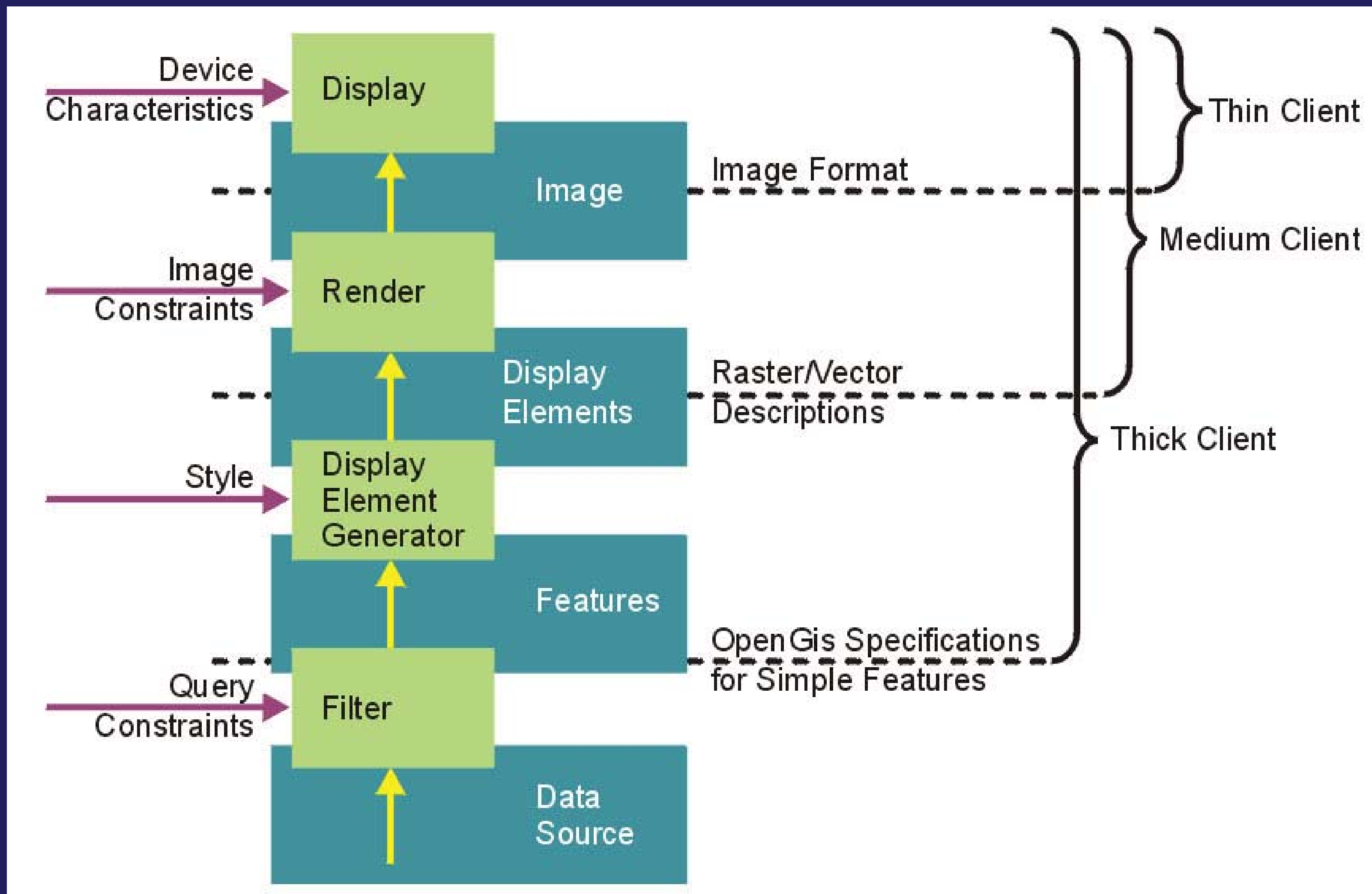
OpenGIS standards: meta data, catalog services

- Meta data: OpenGIS adopt work op ISO, tc211
- Attention on catalog services:
  - how to query a catalogue
  - How answers are send back
  - How to set up distributed catalogs
- Meta data and catalog services are the core of national clearinghouses for GI (NCGI’s)

OpenGIS standards: Internet GIS

- Standards for querying a server, e.g.
  - Web map server (GetCapabilities, GetMapm, GetFeature_info)
  - Web feature server: also for updating (distributed locking and transactions)
- Standards for returning the result, e.g.
  - Raster (JPEG,PNG), graphics (SVG) format
  - Data format: Geography Markup Language (GML)

OpenGIS Standards
Internet GIS: three levels



OpenGIS Standards
Internet GIS: GML example

```
<tdn:SpoorbaanDeel fid="TOP10.4200001">
 <tdn:aantalSporen>1</tdn:aantalSporen>
 <tdn:datum>6 juli 2001</tdn:datum>
 <gml:Polygon srsName="epsg:7408">
   <gml:coordinates>191008,447232 190990,447236
      … 191008,447232 </gml:coordinates>
 </gml:Polygon>

</tdn:SpoorbaanDeel>
```

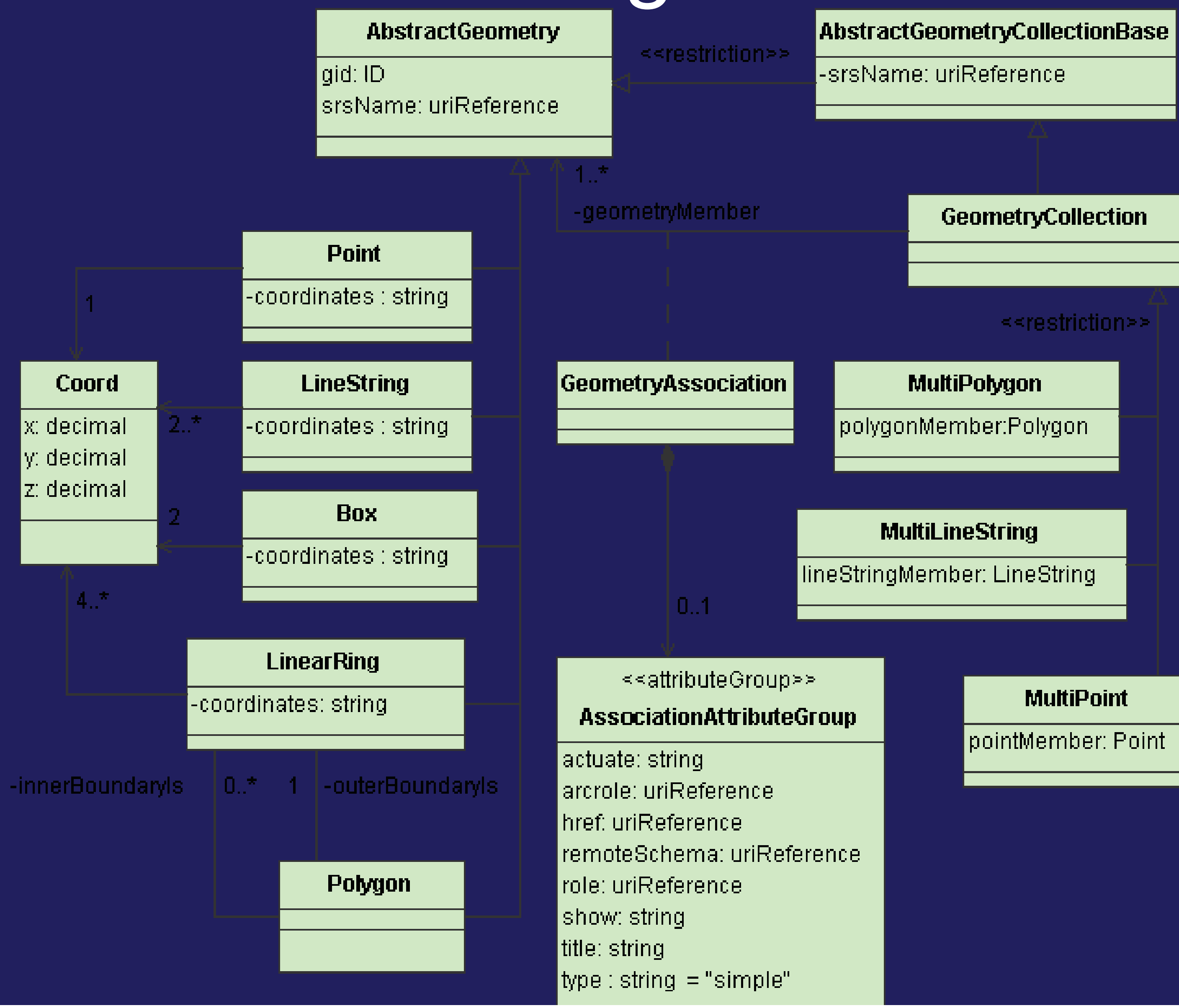
Modeling Standards :UML

- World-wide accepted standard: Unified Modeling Language (UML)
- After many decades with different methods
- UML has several types of diagrams for both the data (structural) and functional aspect of IS
- Class diagrams are one of the most well-known ones (roots in entity-relationship diagrams)
- Object Management Group (OMG) maintains the UML standard

Modeling standards: XML

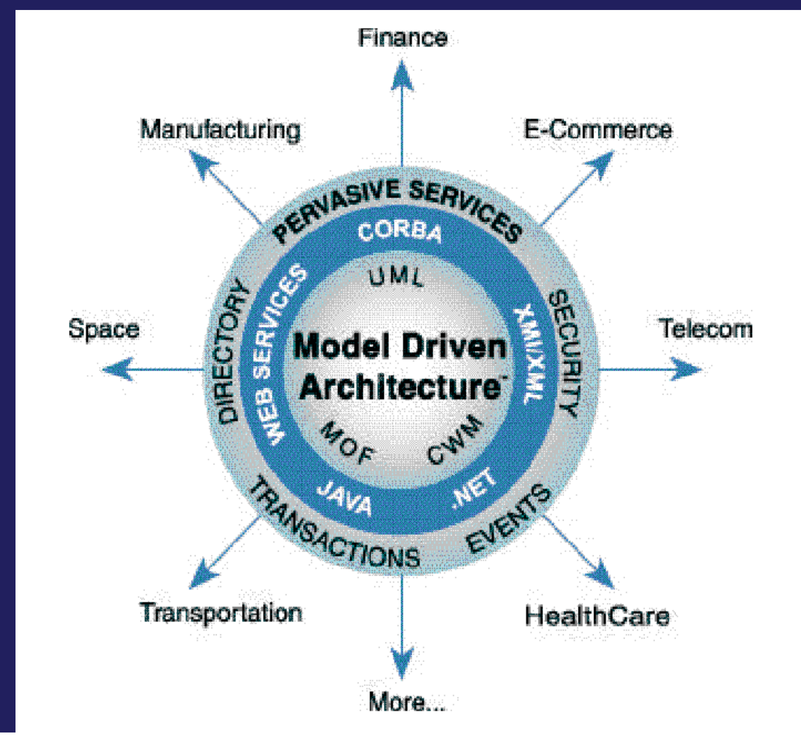
- Another world-wide accepted standard: eXtensible Markup Language (XML)
- Structured exchange of information
- Data models described by
  - Document Type Description (DTD) or
  - XML Schema
- XML documents must obey rules: well formed (syntax) and valid (schema)
- World-wide web consortium maintains XML
- UML can be encoded in XML

Modeling standards: UML class diagram of GML



Modeling standards
Model Driven Architecture (MDA)

- Cadastral systems must be generic and flexible (changing user requirements)
- This is the main principle behind MDA by OMG
- (UML) model is core
- Several domain models
- Only changes require additional work
- Standard implementations of components available
- Cadastral model needed!



TU Delft

Database Technology

- Cadastral systems need both traditional functionality (alpha-num) and spatial functionality at DBMS level
- Recently DBMSs are able to handle spatial data (types, functions, indexing) by Oracle, IBM Informix, Ingres
- Further requirements: topology structure management, raster geo-data types, 3D data types, spatio-temporal aspects, VLM (main memory DBMSs)
- GIS software uses Geo-DBMS (more often)

Location Based Services components from 3 disciplines

- 1. positioning (either GSM/GPRS/UTMS or GPS/GLONASS/Galileo)
- 2. wireless communication network based on GSM/GPRS/UTMS
- 3. GIS data and services form the GII

- Common aspect: geographic locations

Location Based Services
100m positioning accuracy (GSM)

GIS technology
Kadaster
TU Delft

Location Based Services Applications

- Data use (via wireless connections in the field) by
  - Cadastral employees (doing their jobs)
  - Others using cadastral data in the field (taxation, police, real estate)
- Data collection, of course in combination with advanced positioning (10 cm GPS based) and advanced ‘field office’
- Cadastral data services ‘chained’ with other services (coord transf, geo-coding, topo-map,etc.)

Agenda

- 1. Introduction
- 2. User requirements
- 3. Geo-ICT Development
- 4. Proposal and conclusions

Conclusions

- Well documented and accepted (but ever changing) user requirements
- Widely accepted Geo-ICT tools available for cadastral systems (and GSDI)
- Dramatic improvement in effectiveness of system development and functional possibilities
- Effective/flexible system development requires MDA and minimal Cadastral domain model

Proposal: What?

- Develop standard minimal Cadastral domain model, including:
  - Spatial part (geometry, topology)
  - Extensible frame for legal/admin part
  - Based on core object-right-subject model
- Object-orientation → express in UML
- Using world-wide accepted standards (XML)
- Accepted by large community: FIG, OGC, ISO
- Maximize co-operation, minimize double effort

ISO/TC211, resolution 203

- “ISO/TC211 appreciates the FIG proposal to develop a model of the basic contents and design of a cadastre using the ISO 19100 series of standards.
- It is acknowledged that this activity will serve to both test the 19100 standards and build on the exiting collaboration between ISO/TC211 and the FIG.
- ISO/TC211 encourages FIG to suggest how ISO/TC211 could assist in this activity.”

Proposal: How?

- Minimal cadastral domain model is not only a test vehicle, but standard itself
- FIG Commission 7, OEEPE commission 4
- ISO (through FIG standard network) and OGC
- Contact points
  - oosterom@geo.tudelft.nl
  - lemmen@itc.nl, lemmen@kadaster.nl
  - Deby@itc.nl
  - Tuladhar@itc.nl

GeoSpatial Database

Proprietary → Open

Fred Finch
Spatial Data Management
Ordnance Survey *Ireland*

Present System

- **Proprietary**
  - **Distributed - Small Scales and Large Scales**
  - **Raster (UNIX and Windows NT)**
  - **Vector (VMS and Windows NT)**
  - **Feature Codes with multi featuring**
  - **Attributes**
  - **One Reference System**
  - **Tile Based**
  - **Metadata**
  - **GeoDirectory Address Database (Oracle 8*i*)**
    - Automatic update of some fields using Oracle Spatial Cartridge
  - **Map Publishing – (Barco)**
  - **Digital Terrain Model**

New Database

- **Open – Oracle 9*i*** *July 2002*
  - Metadata (eCommerce)
  - Production Database (eCommerce)
  - Maintenance Database
- **GIS Vendor Independent**

Metadata

- **ISO standard**
  - Organisation details
  - Product description
  - Feature description
  - Currency
  - Accuracy
  - Etc.

Production Database -1

- Outside Firewall
  - Allows for Metadata viewing and downloading (XML)
  - Download of data to Key Account Holders
  - Download of small quantities of data to Public
    - Customised Maps, Raster and Vector data
  - Large quantities of data are downloaded in a batch process or are sent to Production Database 2
  - Commerce Server for authentication and payment
  - Viewing of thumb nails of Aerial Photographic Images
  - Security of data (Low resolution viewing with watermark)

Production Database -2

- Inside Firewall
- Allows for production of
  - large quantities of vector data outputs
  - large quantities of raster outputs
  - Aerial Photographic Images
  - Commerce Server for authentication and payment
    - Sales analysis
    - Automatic link to Account Systems

Maintenance Database -1

- Move from existing proprietary system
- Object/relational database
- New edit functions required
- Tendering process
- Re-engineering/ updating of flow-lines
- Re-training

Maintenance Database - 2

- Seamless database (not tiled)
- Feature level metadata
- Instantaneous translation from one projection to another
- Automatic update to production database

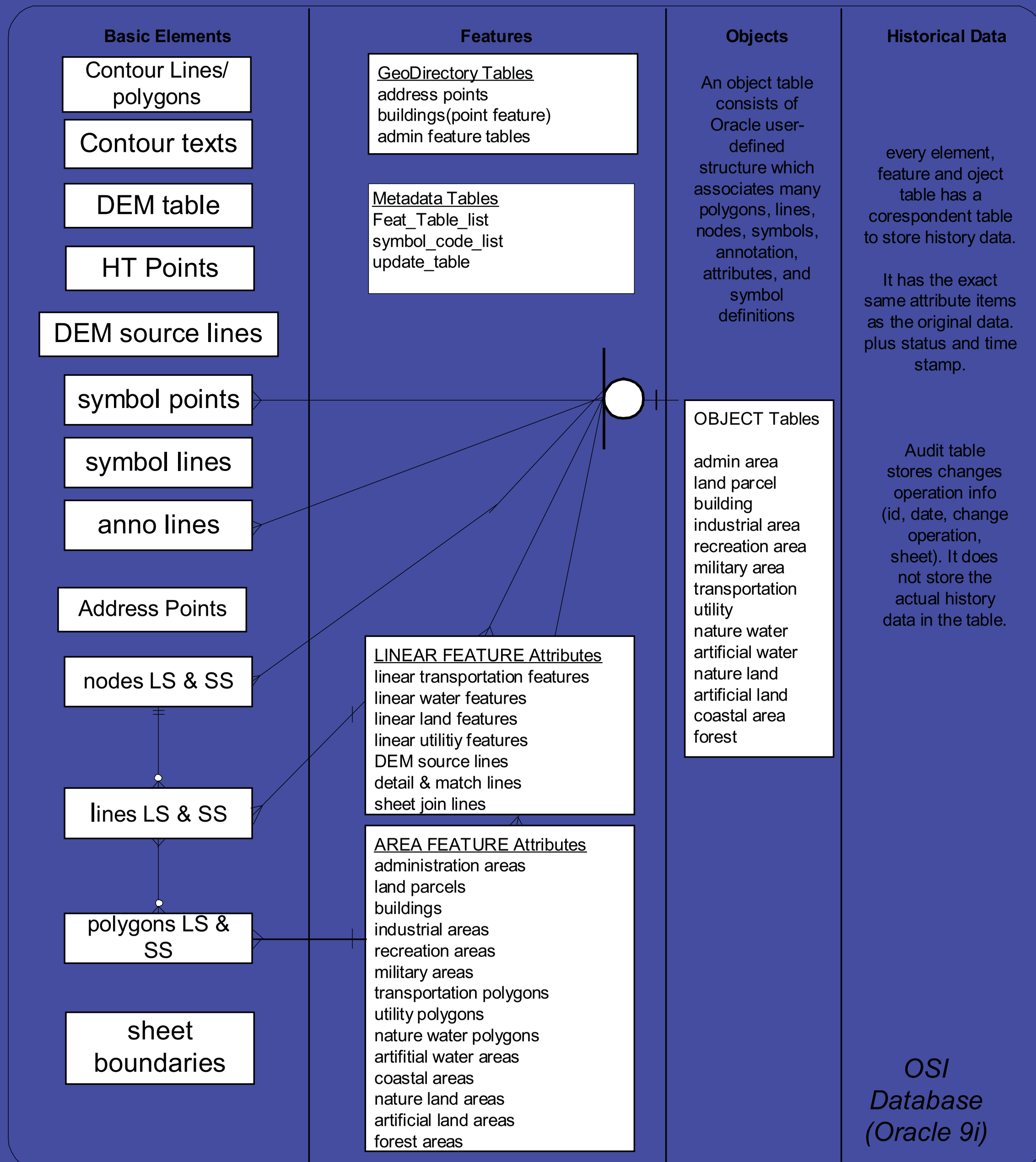
Data Conversion - 1

- Raster
  - Proprietary format (VIP) to GeoTIFF
  - Colour TIFF from Vector data
  - Ortho-photography
    - 1995 B/W (full coverage)
    - 2001 Colour (full coverage)
  - Versioning
  - Tile Based
- All new images will be produced automatically from the Production database to customer specification

Data Conversion - 2

- Vector
  - Translation from Proprietary Files to Oracle
- Translation to ITM coordinate Reference System
- Attribute Issues
  - Accuracy / scale
  - Quality
  - Including GeoDirectory intelligence
  - Etc.
- Topographical Identifiers
  - Unique 16 digit ID on each geographical feature/object
  - Unique identifiers within map sheet tiles wil be maintained until the maintenance database is fully operational

Open Database Schema



Ordnance Survey *Ireland*

Element Tables

- **Element tables** contain the spatial components of feature layers. A real world feature may consist of one or more elements, e.g. a river may consist of nodes, river sections or segments of river bank. The definition of an element is based on the individual feature created in the OSi proprietary system application.
- Each element has TOID and attributes, such as data source, quality, and valid period, scale range, etc.
- Elements break into various points (2D or 3D), nodes, lines, polygons in Oracle 9i spatial layers. There will be an element layer each for large scale and small scale data type.
- Address point data will be generated based on the easting/northing coordinates.
- Contour lines and polygons for both small scale (in 10 meter resolution) and large scale will be generated from DTM point data.
- DTM points are not stored as a spatial layer, but as an x,y,z attribute table.
- Large and small scale data will be stored in separate tables, i.e. there will be a polygon layer for large scale data, and a polygon layer for small scale data.
- Features in element tables will have feature codes to indicate the the spatial object category to which it belongs. If the feature is used by multiple categories, multiple feature codes will be associated.

Feature Tables

- **Feature tables** store the properties of spatial features. Each table contains information for one feature category or one sub-category in a hierarchy structure.

- Initially all tables will have the same attribute fields. Additional attribute items will be added at a later date when more intelligent data is available.

- Attribute records will be initially generated from the proprietary system application, thus one element feature coded in the spatial layer will result in the generation of one attribute record.

Object Tables

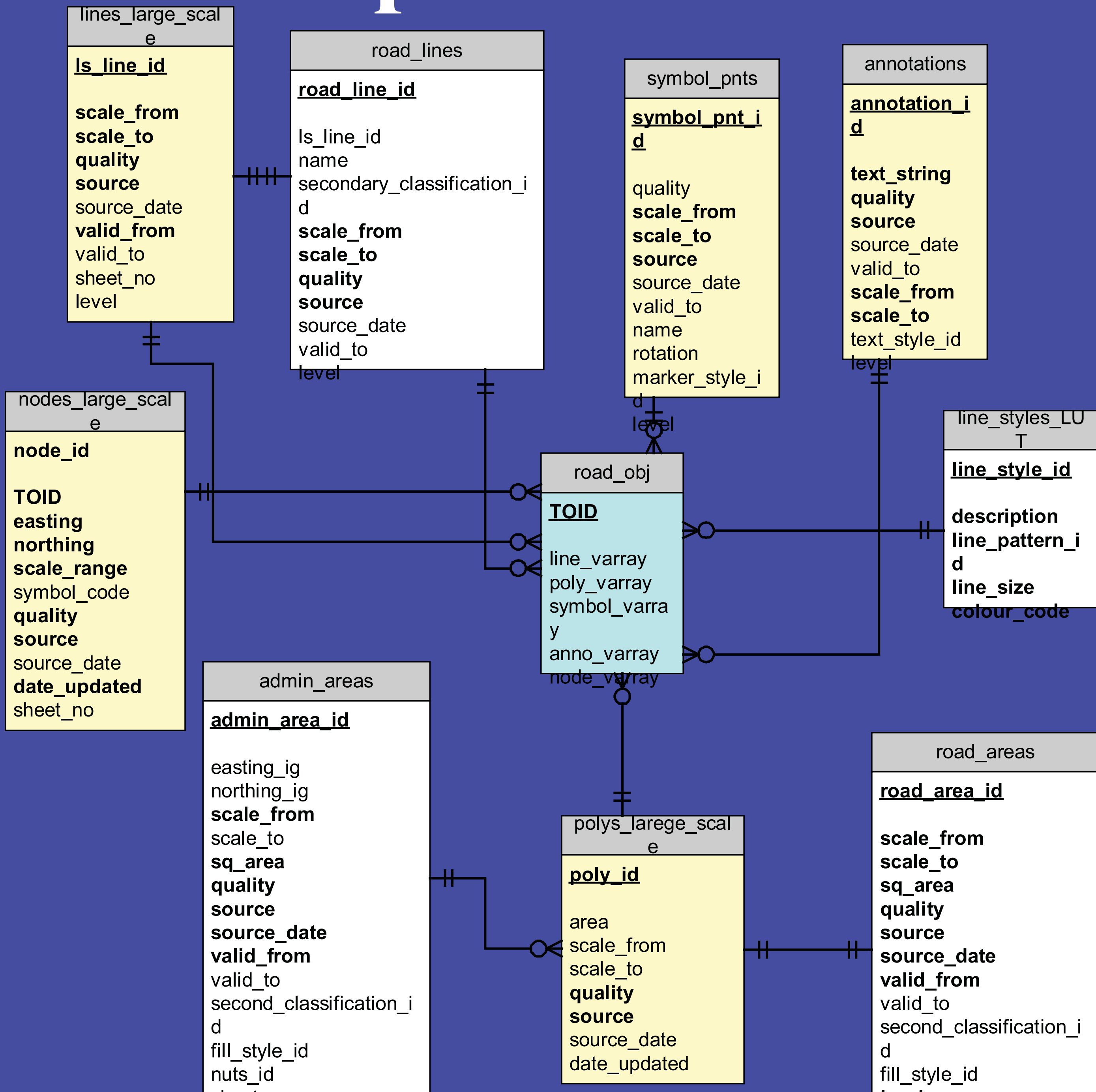
- **Object tables** will use the Oracle object structure through user-defined types.
- Records in the object table will associate various feature components from an element table (spatial components), attribute records, symbol data, annotations, symbols etc.

Audit and History Tables

- **Audit tables** store change information for all data in the database.
- The
  - record identifier,
  - feature type,
  - the type of operation, ie.
    - Insert
    - Delete
    - Update
  - time stamp,
  - sheet number if it has one,
  - operator's information.
- Audit table does not store history data.

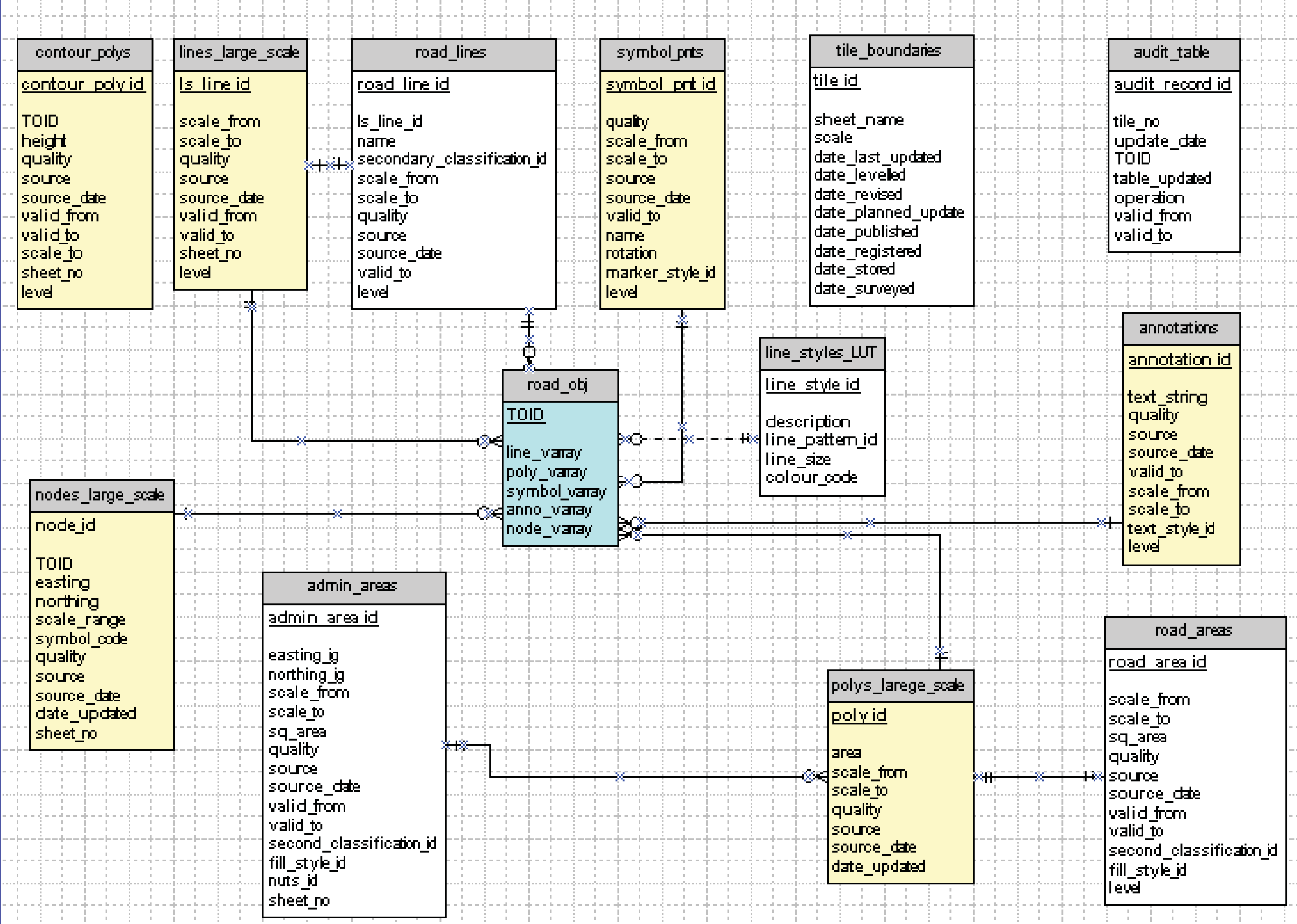
- **History tables** store records which have been updated or deleted.
- A history table will be created for every table containing an element and attribute component.
- The history table has the same structure as the current element and property tables, plus
  - the date of modification,
  - the valid period,
  - how the record was updated,
  - the user who modify the record
  - and more

Open Database Schema



Ordnance Survey *Ireland*

Open Database Schema (part of)



Platform

- Hardware/ Operating System
  - Microsoft Windows NT
- Software
  - Oracle 9*i*
  - ESRI ArcIMS
  - ESRI ArcGIS
  - ESRI Catalogue
  - Microsoft Commerce Server 2000

Data Output

- Raster
  - GeoTiff
  - Tiff (500,400,300,200 dpi)
- Vector
  - GML
  - DXF /DWG
  - NTF
  - UNLDST

Advantages of New Open Database

- Is it one or two?
- Single Database to OPEN standards with
  - a seamless database
  - all data referenced to GPS compatible reference system
  - one identifier (TOID) on each feature / object
  - more standard outputs (XML, GML etc.)
  - data more available to corporate and ordinary customers
  - more attributes at feature level
  - standardised Hardware and Software
  - no dependency on any one GIS vendor

GeoSpatial Database

Proprietary → Open

Spatial Data Management
Ordnance Survey *Ireland*

Fred Finch
fred.finch@osi.ie

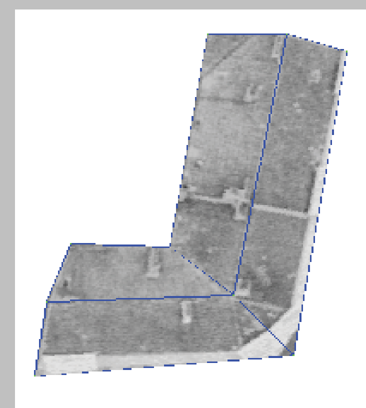
New database developments in IGN

Frank Fuchs,
Institut Géographique National
2-4 av. Pasteur,
94165 SAINT-MANDE Cedex
France

frank.fuchs@ign.fr,
Phone (33) 1.43.98.83.82
Fax (33) 1.43.98.81.71

Who's this guy ?

- IGN since 94 : training at ENSG
- 96-2002 : Research Dep. Digital Aerial Stereoscopic Imagery (Building Reconstruction)



- 2002-2005 : Project on DB Merging.

BDTopo : presentation

Context, needs and objectives

Pros / cons

Today

Tomorrow

Tasks

- Aims :
  - high precision reference geographic frame
- Themes
  - Road / rail / water network, buildings, vegetation, altitude, administrative boundaries
- Unstable : achievement 2006
- Will cover France
- Typical scale : >1/25 000
- Precision ≈ 1 m

Context, needs and objectives

Pros / cons

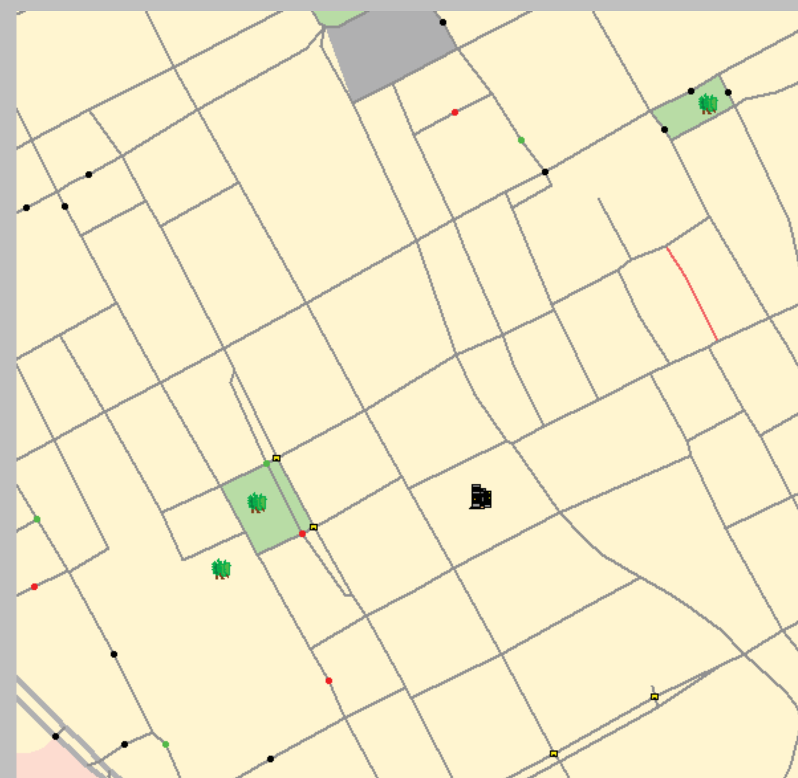
Today

Tomorrow

Tasks

Géoroute : presentation

- Aims :
  - itinerary computation,
  - on-board mapping / on-board navigation
- Themes
  - Road network, particular equipment (parks, hospitals, police), administrative boundaries
- Stable
- Covers >10k urban areas, with low scale links between these regions
- Typical scale : > 1/15 000
- Precision ≈ 10 m



Context, needs and objectives

Pros / cons

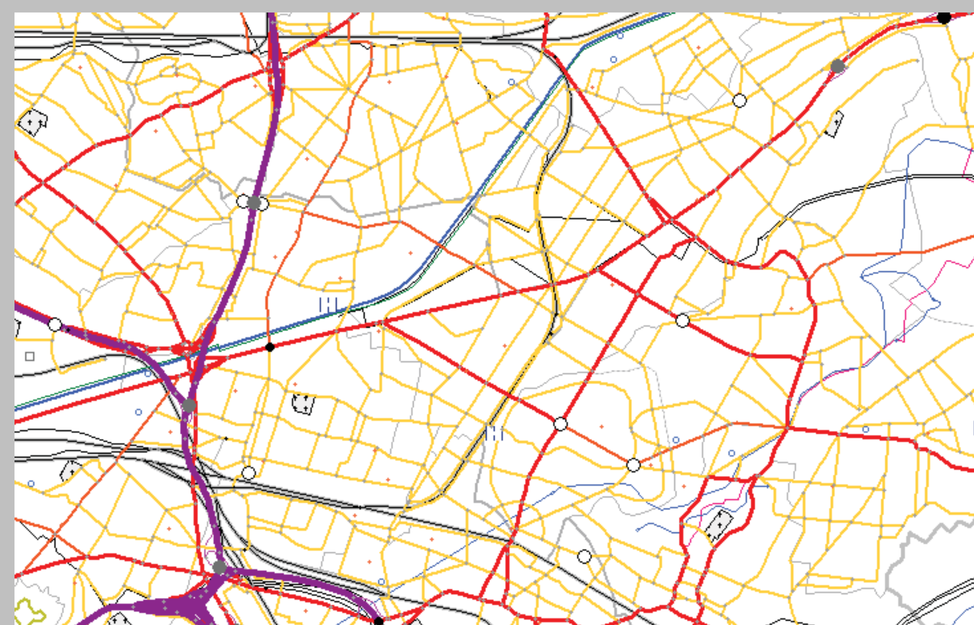
Today

Tomorrow

Tasks

BDCarto : presentation

- Aims :
  - Cartography at 1/100 000
  - Reference frame for customer's data
- Themes
  - Road network, hydrography, land use, administrative boundaries
- Stable
- Covers France
- Precision ≈ 30 m
- BDCarthage (EU)



Context, needs and objectives

Pros / cons

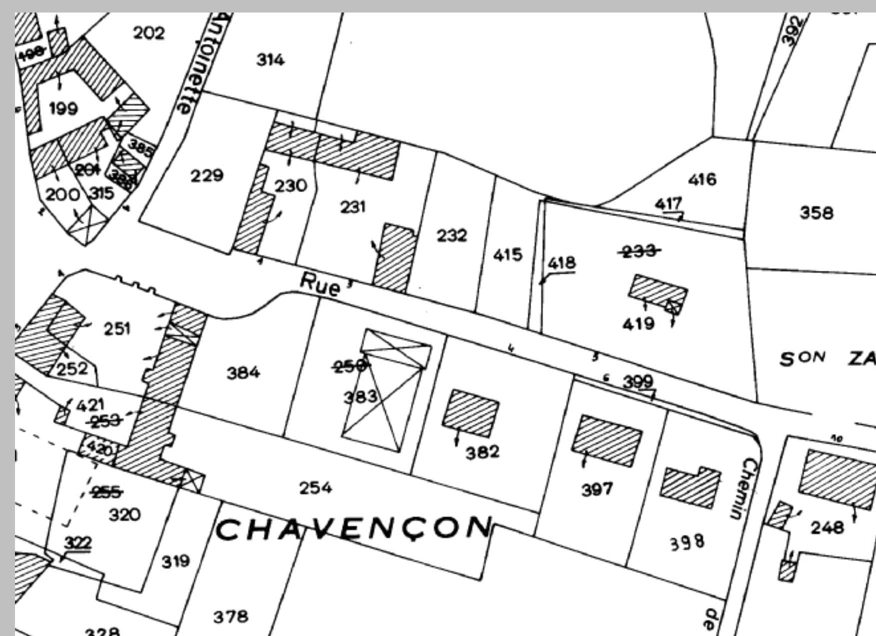
Today

Tomorrow

Tasks

BDParcellaire : presentation

- Aims :
  - provide link between geographic / topographic information, and several informations about the parcels
- Themes
  - parcels with identifiers (link->info), administrative boundaries
- Unstable : specifications (model) achieved. No data
- Partnership with cadastre
- Will Cover France
- Cadastre's scales : 1/500 ➟ 1/5000



Objectives

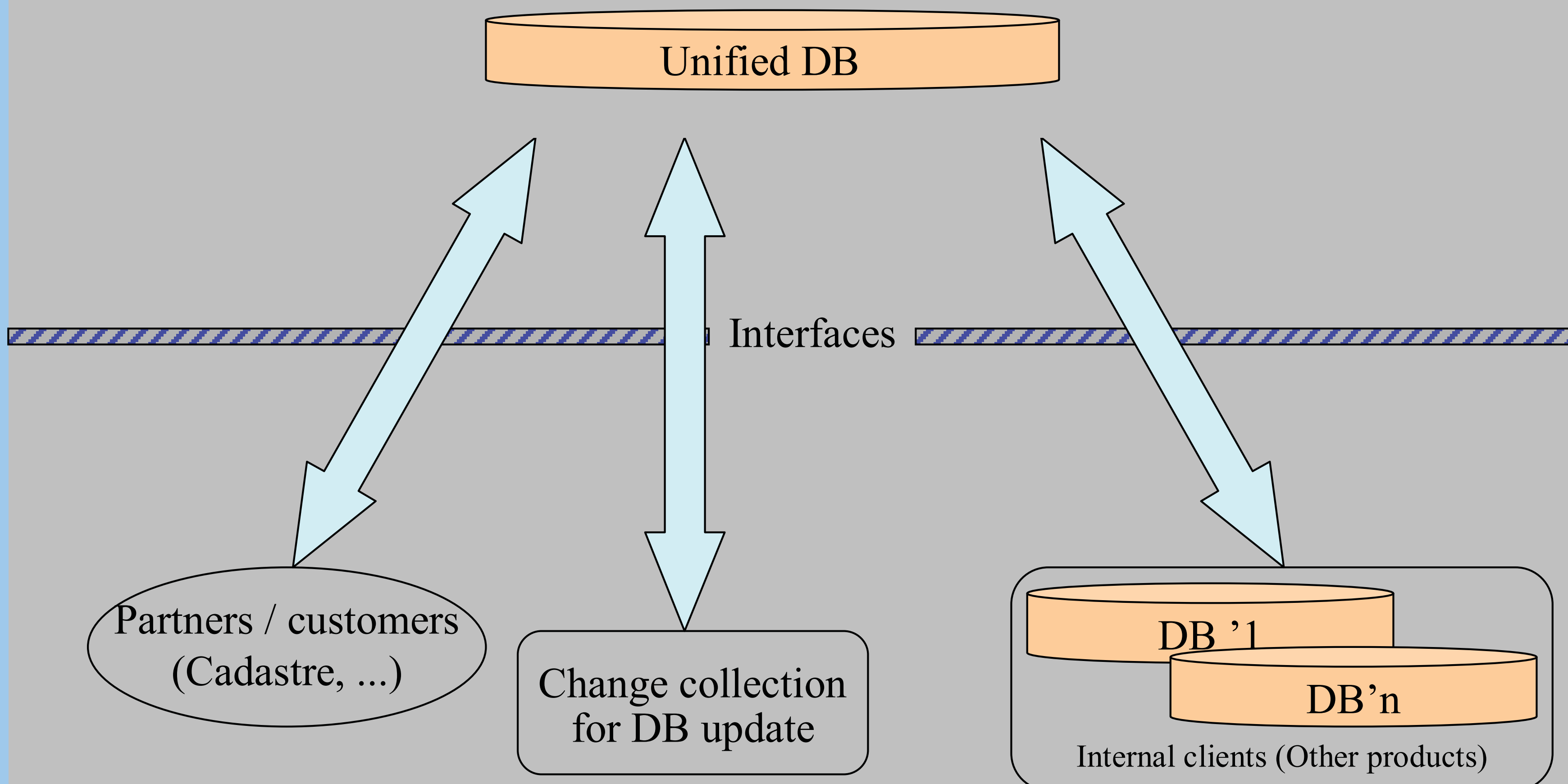
Context, needs and objectives

Pros / cons

Today

Tomorrow

Tasks



Context, needs and objectives

Pros / cons

Today

Tomorrow

Tasks

Pros

- Internal consistency (DB covering the country)

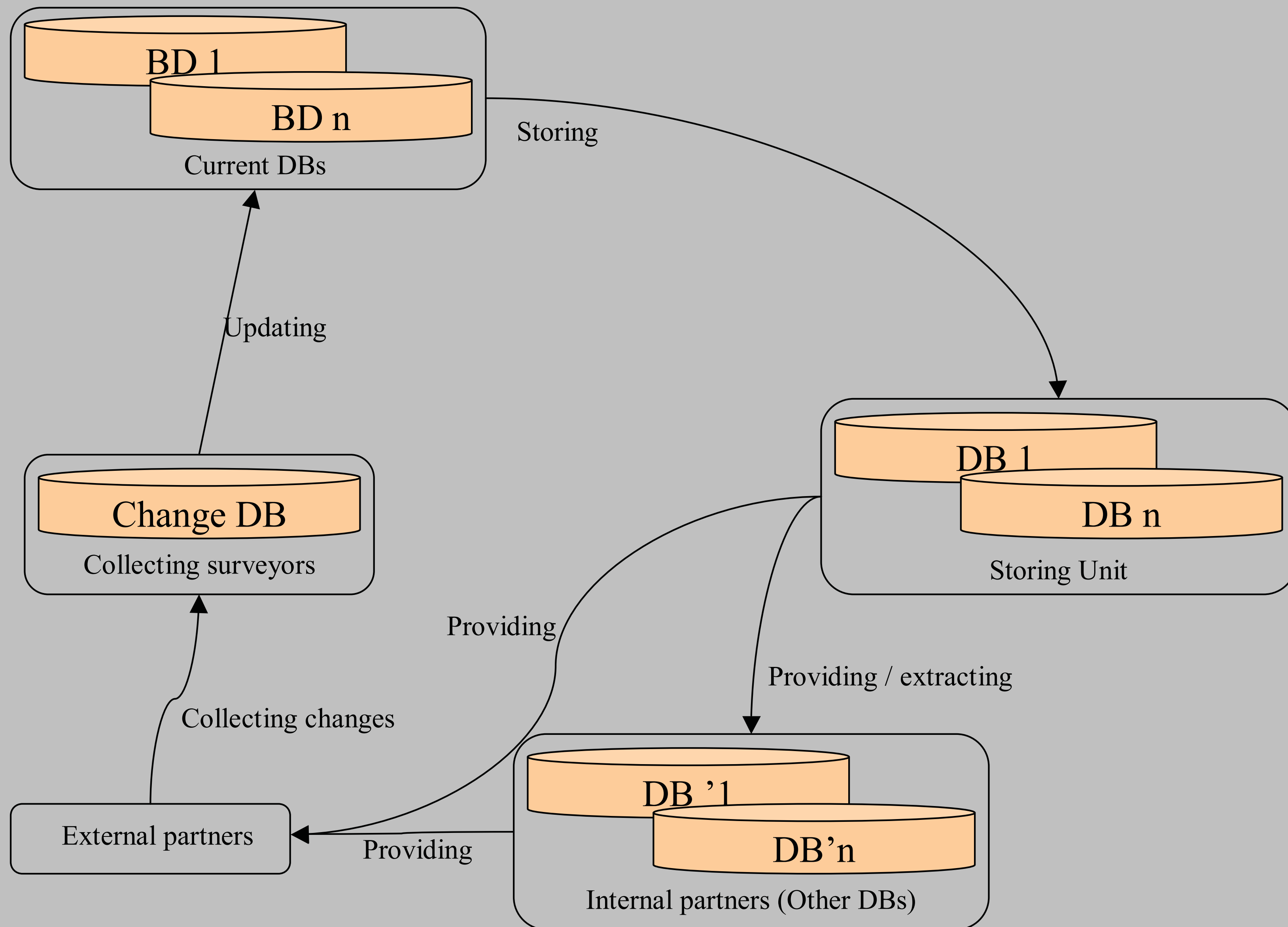
- Consistency between DBs
- Common update (cost)
- Deriving new products
- Homogeneous actuality

Cons : Risks

- Technology :
  - Matching tools
  - 2-scale representation of geographic data (including change derivation / generalisation)
  - Managing a DB on different places
- Input :
  - Variations of specifications of input DBs
  - Very different specifications (scales in particular)
  - Quality of data
  - Still non existing data (thus unknown)
- Human :
  - Move from existing different tools to one common tool

Current data flow graph

- Context, needs and objectives
- Pros / cons
- Today
- Tomorrow
- Tasks



Future data flow graph

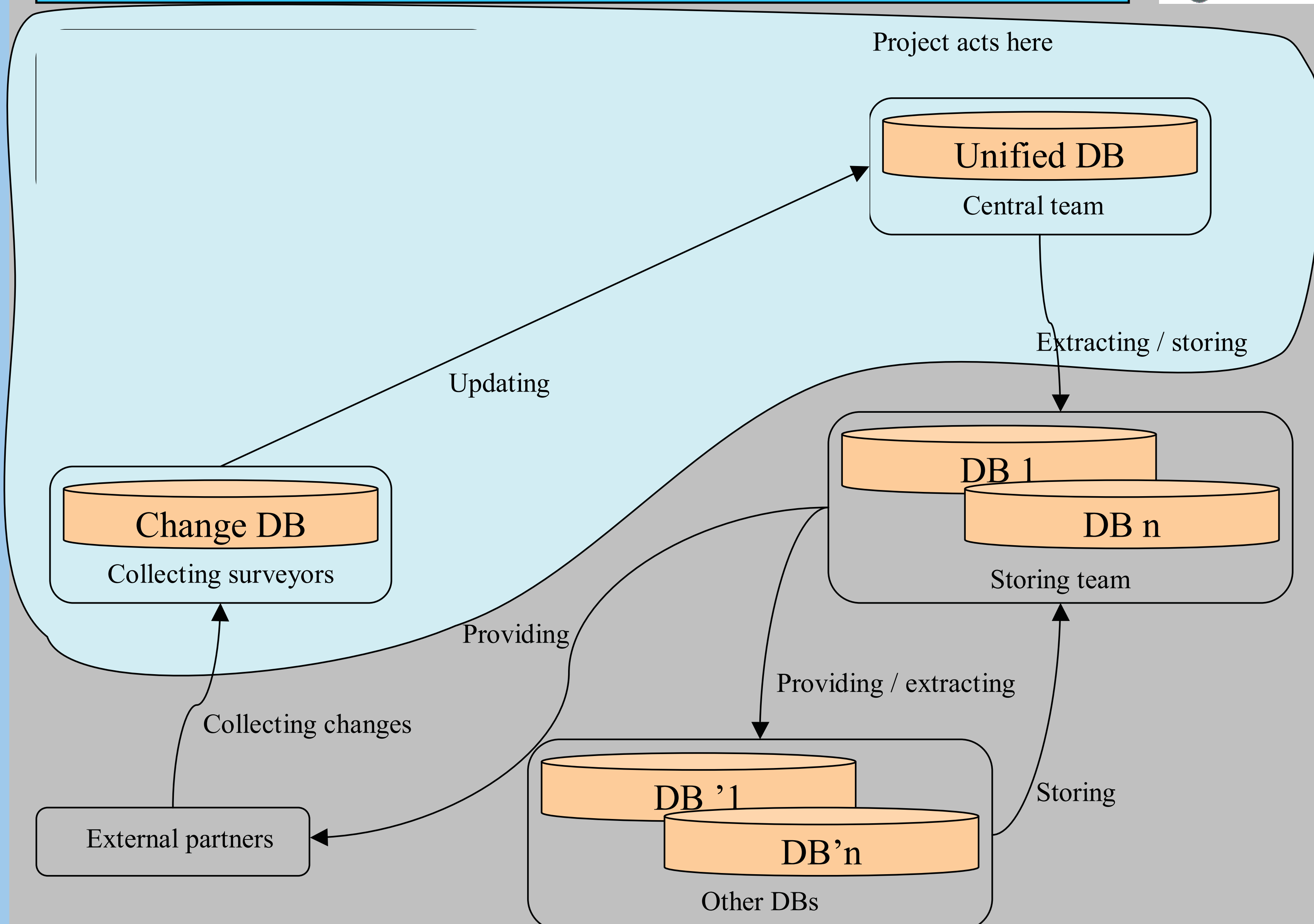
Context, needs and objectives

Pros / cons

Today

Tomorrow

Tasks



Context, needs and objectives

Pros / cons

Today

Tomorrow

Tasks

Task : merging

- Models

  Imagine the model of the unified DB.

  Be sure it fulfils the needs of the partners.

  Deal with 2 very different scales : EU Project : MurMur.

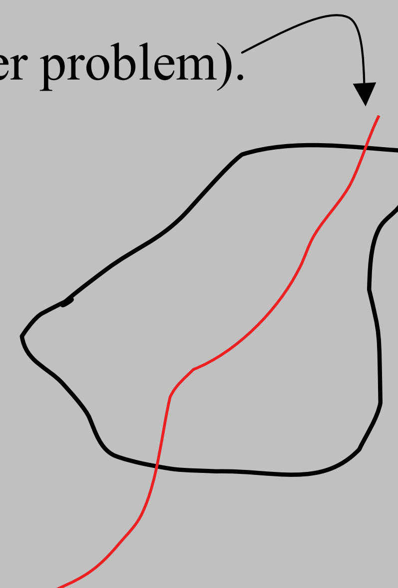
- Data

  [Devogele] : preparation, matching, merging

  Merge the pieces of data in one DB (the border problem).

- Notice :

  Merging is done only once.



Task : meet partners needs

- Context, needs and objectives
- Pros / cons
- Today
- Tomorrow
- Tasks

- Unplug current DBs and plug the Unified DB
  - Analyse the needs
  - Specify the new processes and tools (formats …)
  - Develop
  - Test

Context, needs and objectives

Pros / cons

Today

Tomorrow

Tasks

Task : continuous update

- Existing tools

  For BDCarto : Continuous update dep. (DB storing the changes)

  For Géoroute : strong experience of the continuous update around Paris

  Process under construction for common change collection

- Surveyor's point of view :

  Deal only with higher scale : propagation to lower scale (done by specialised update geographers)

Context, needs and objectives

Pros / cons

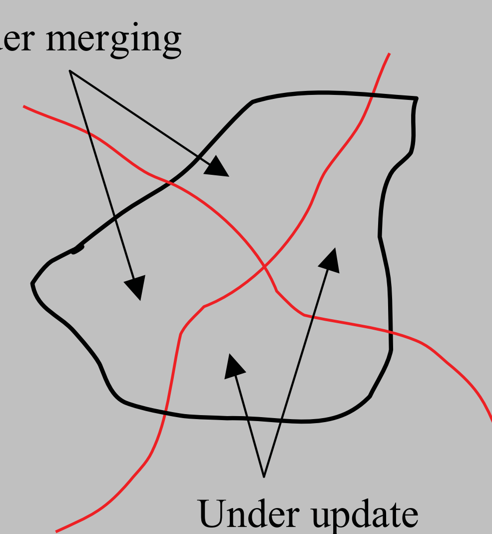
Today

Tomorrow

Tasks

Task : move from one process to another

- Deal with :
  - DB merging in progress (piecewise)
  - update merged regions.
  - Manage the new system while its load is growing
  - Helps the (mainly internal) partners deal with the new product
- Challenging co-operation between different departments of IGN



Task : analyse existing projects

- Searching POCs in other NMAs

Conclusion

- DB merging :
  - Better consistency
  - Data mixing -> new products
- Common update :
  - Lower cost
  - Homogeneous actuality

Questions

A Geospatial Data Warehouse to provide cadastral and other data

Jens O. Jensen,
Informatics Dept.

Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark

The Role of KMS

- KMS is the government authority in Denmark responsible for surveying, mapping, charting, cadastral registration and authorization of licensed surveyors

- Government research institute for mapping and geodata

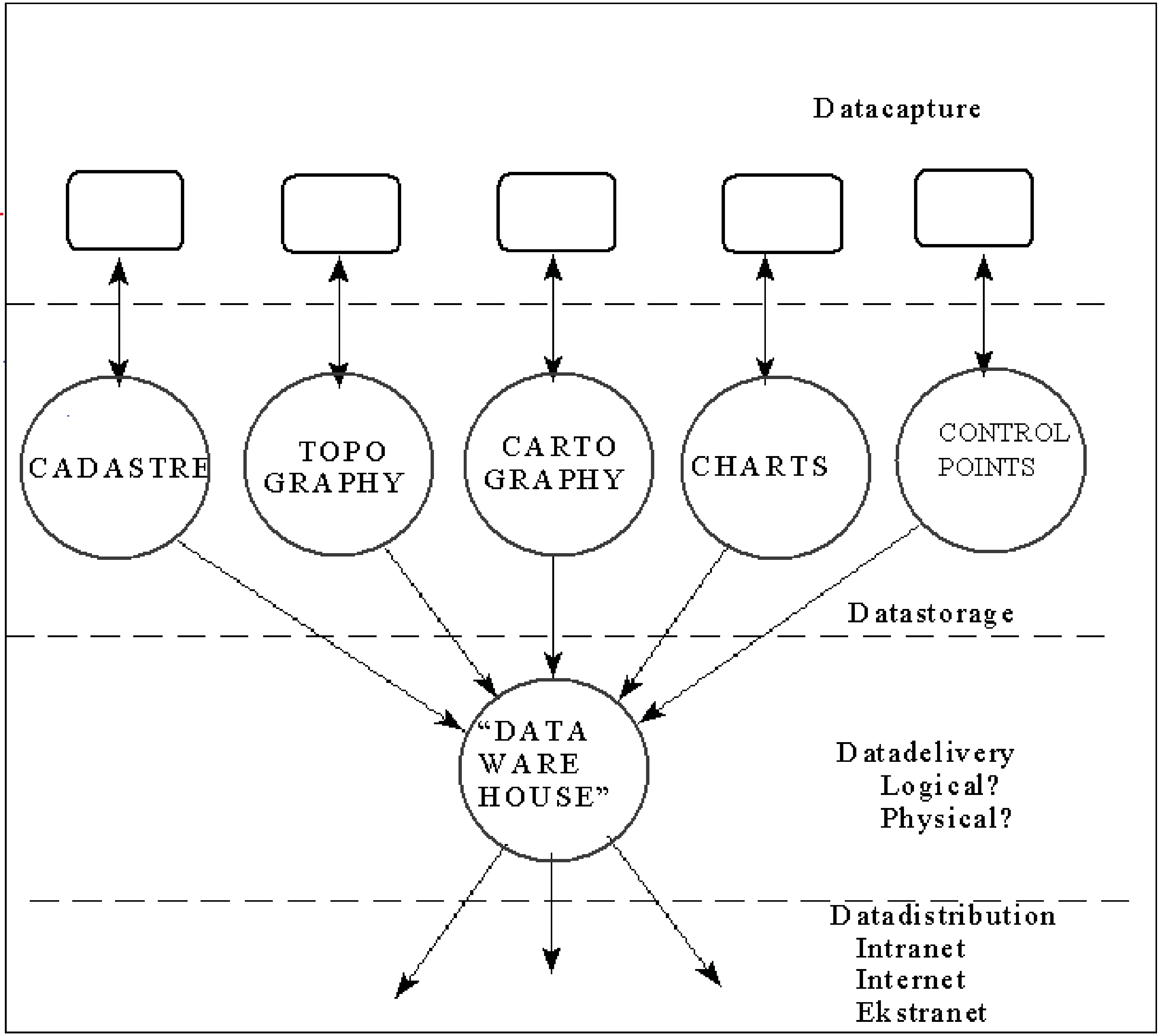
- To coordinate public surveying, mapping, charting and registration of spatial information.

Source: Vision 2009

Topics

- Introduction
- The LDS Data Warehouse project
- The data model
- The loading strategy
- Where are we now?
- Towards 2005

Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark



Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark

The LDS Data Warehouse project

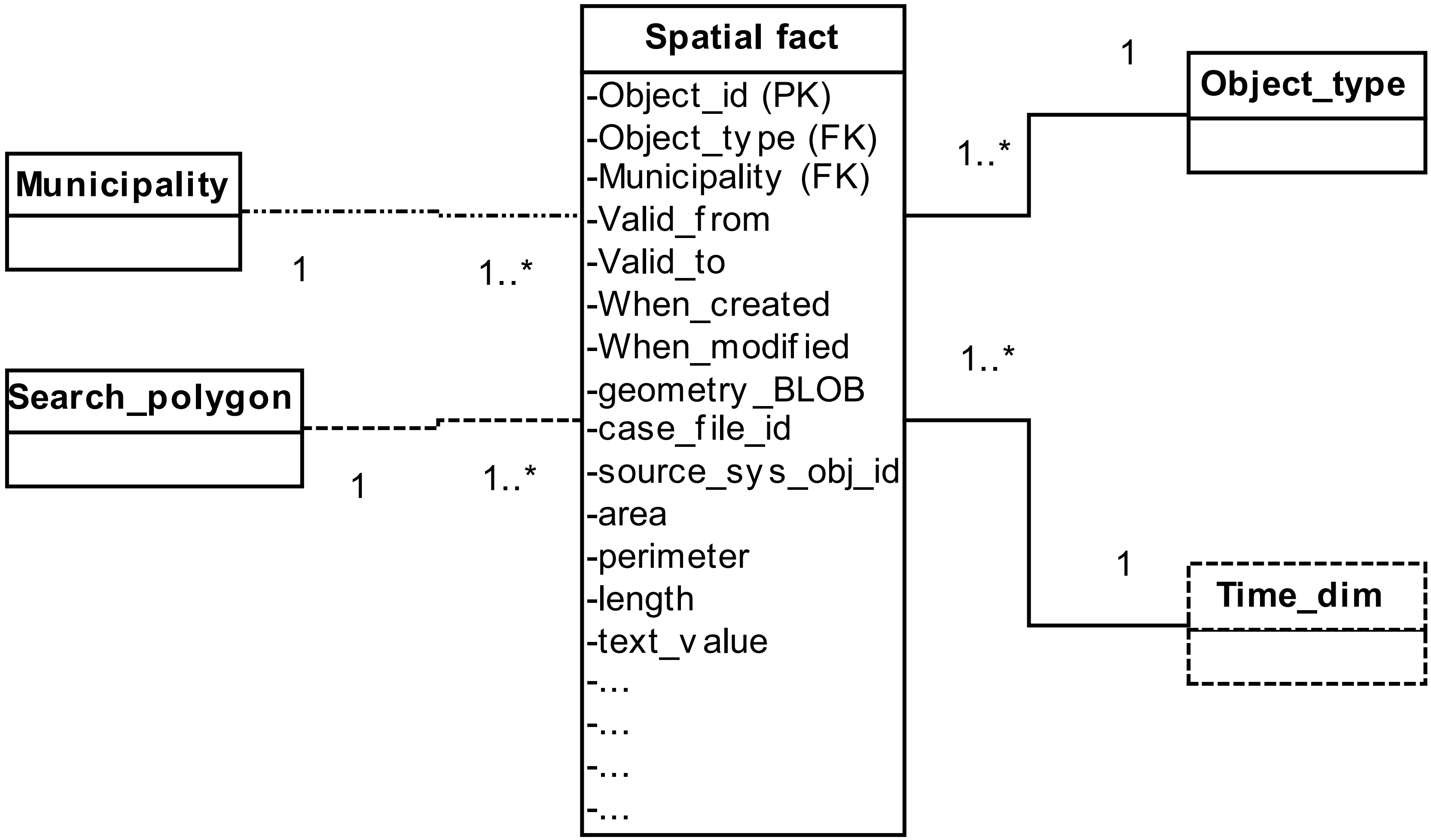
The server platform

- Sun Enterprise 3800 with 380 GB hard disc
- RAID5 allows 1 disc to be taken out
- Sun/SunOS 5.8 (64 bit) patch set 108528-12
- (32 bit) Oracle9i Spatial 9.0.1.3.0

Development tools

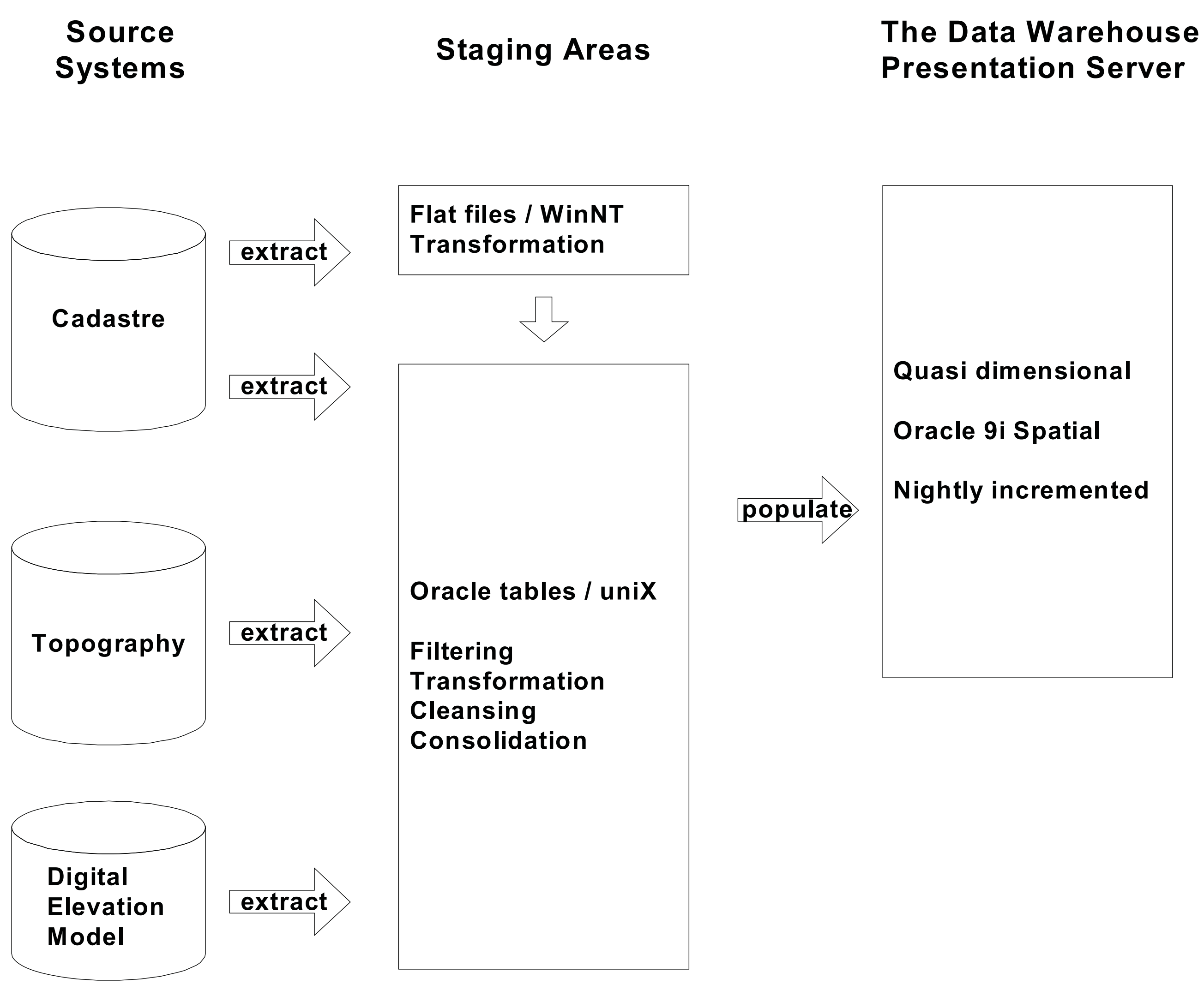
- PL/SQL packages
- TOAD from Quest Software
- Mixed in-house and external

The Data model 1



Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark

The loading strategy



The loading strategy 2

Performance: Load times

- 2½ million cadastral register rows in 3 hours
- 13½ mill. conv. from CAD-file format in 5 hours
- 40 million cadastral geometries in about 3 days. Includes generation of 2½ M. parcel shapes
- 270ooo elevation contour curve segm. in 30 min.
- 6½ million topographic geometries in 12 hours
- Gazetteer rows

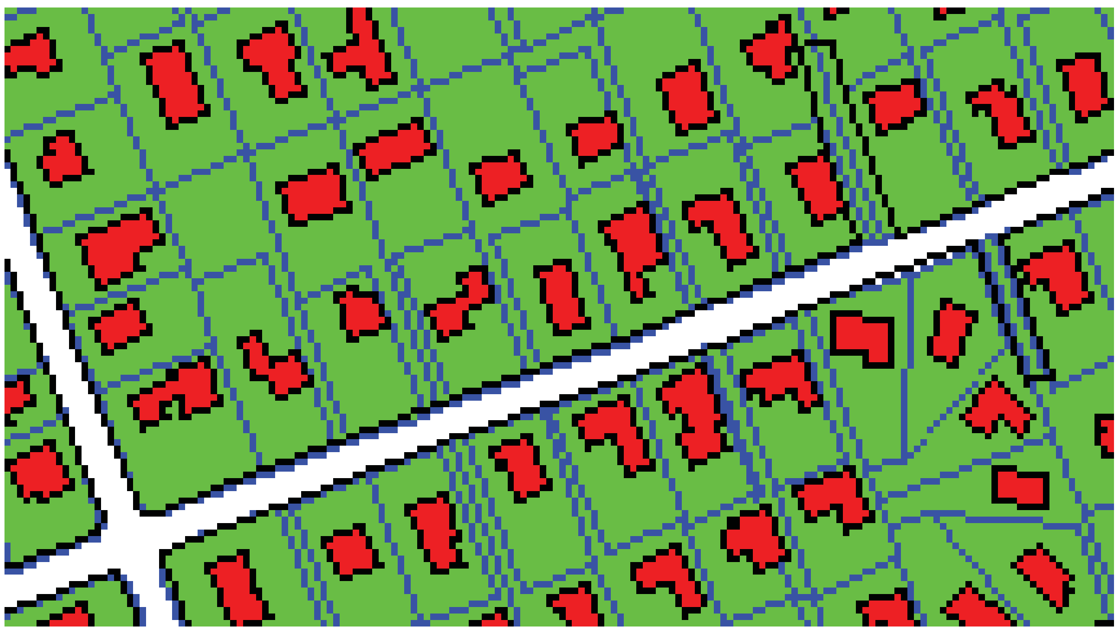
Spatial index generation (40M geom.) takes 4-5 hours

SDO_relate

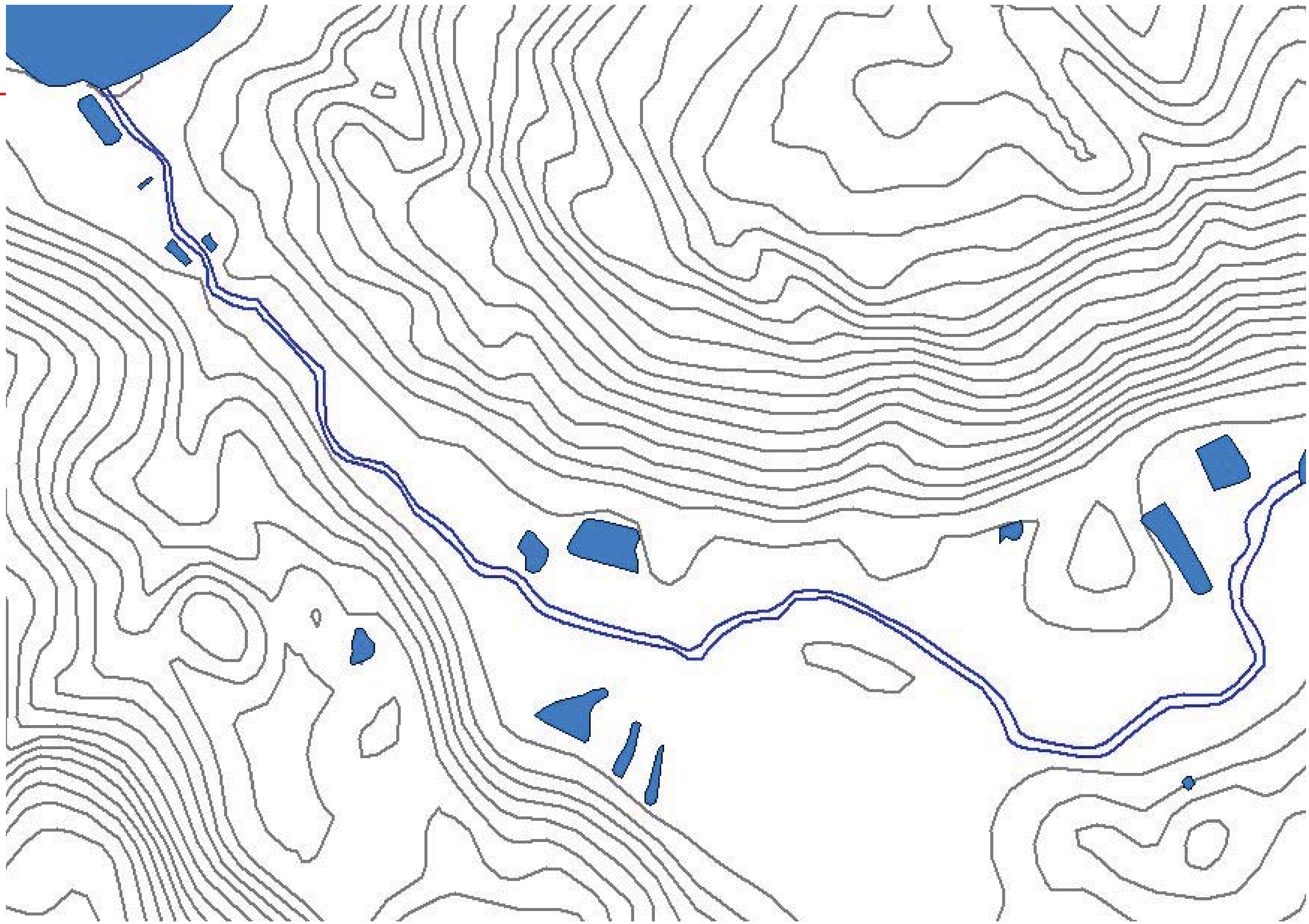
Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark

Where are we now? 1

Cadastral Parcels & Topo. Buildings



Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark



Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark

Towards 2005 1

<u>Phase 1</u> which is now

- makes vector data available

<u>Phase 2</u> in 2003 will add

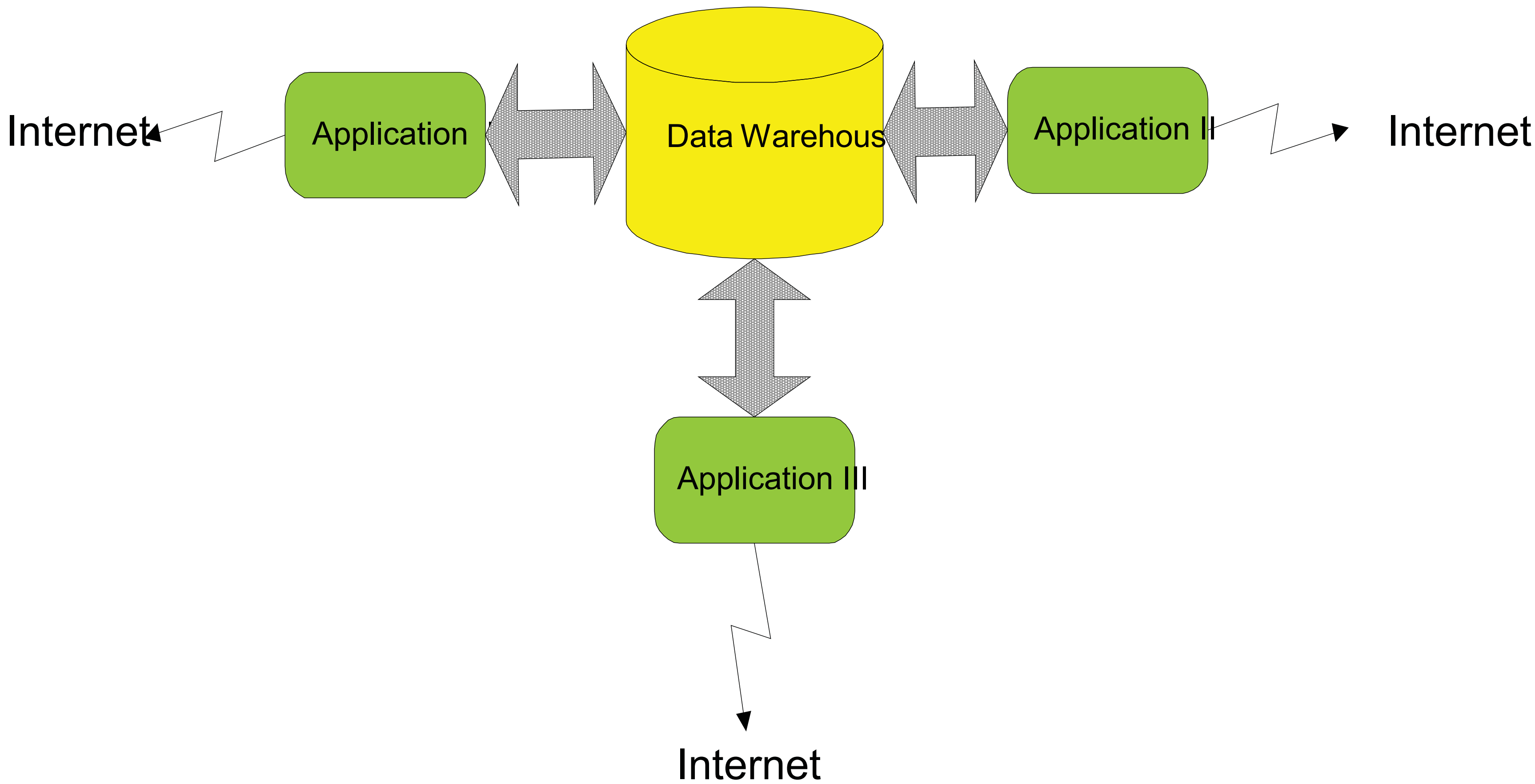
- new vector data sets
- Additional Data Marts are planned for selling
  1. Small-scale vector maps
  2. Raster data maps
- Extraction within a customer-defined polygon

<u>Phase 3</u> in 2004 may add

- cadastral measurement sheets,
- seismic data and other
- e-business to the public.

Kort & Matrikelstyrelsen
National Survey and Cadastre – Denmark

Towards 2005



A Platform for Managing Spatial Data

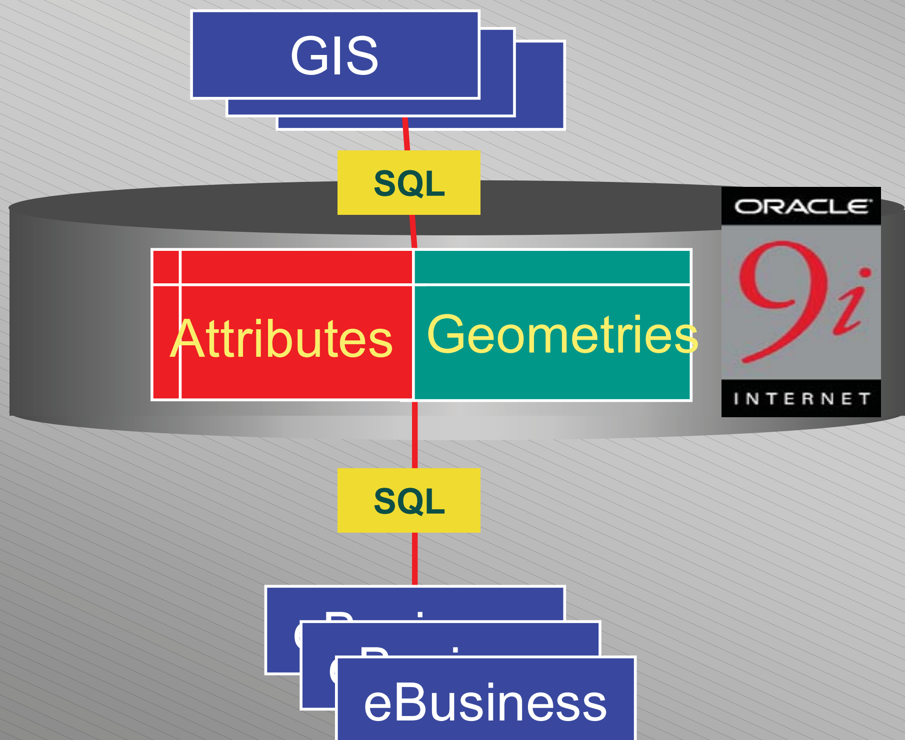
Xavier R. Lopez
Director, Location Services
Oracle Corp.

ORACLE

Overview

- Role of Spatial Databases
- Oracle9i Spatial
- Case Studies
- Partner Integration

Open Spatial Databases



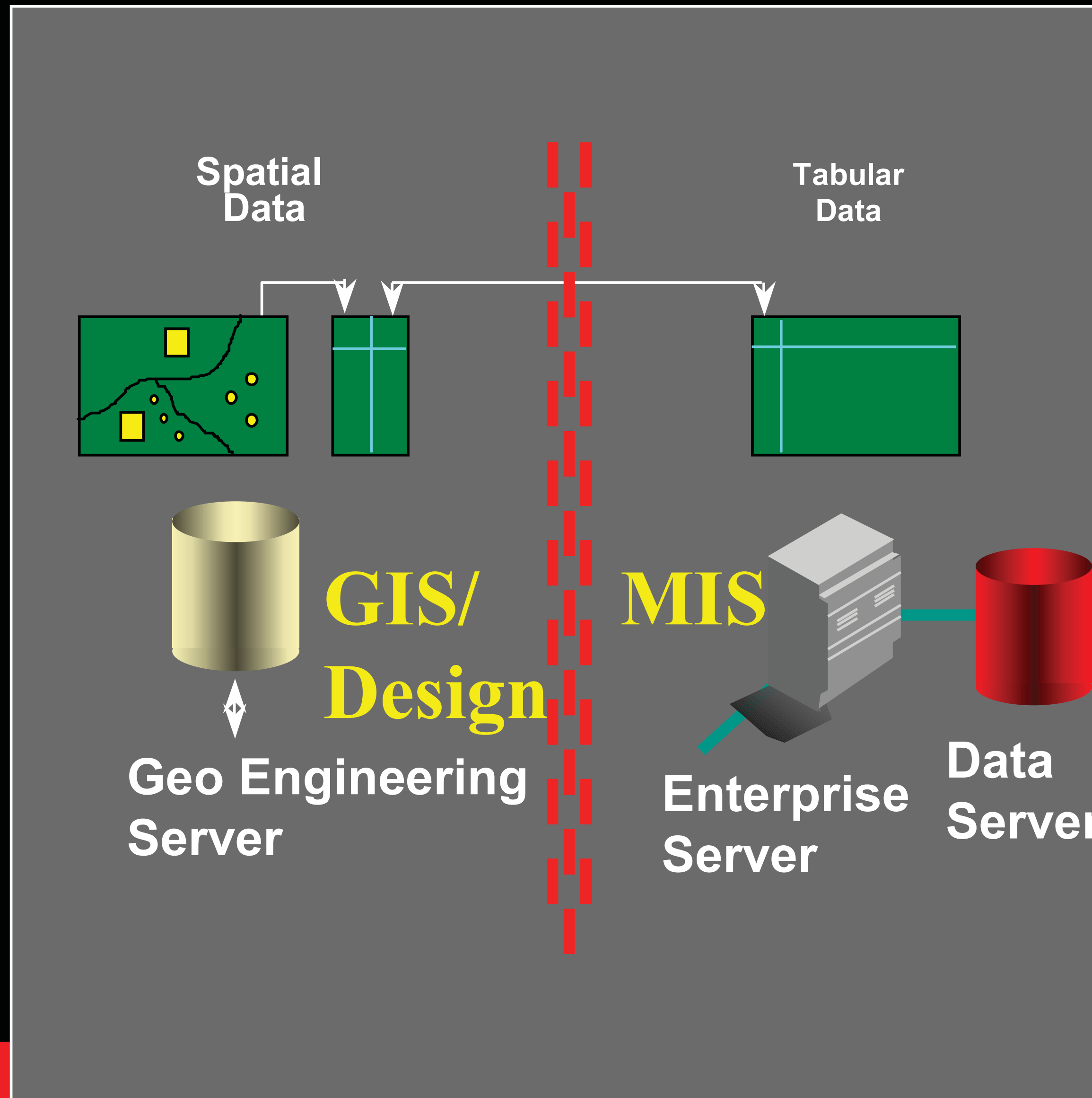
- Attributes and geometries fully integrated in the database
- Spatial is native DBMS type
- IT can access geometries via open standard interfaces
- Standard Query Language
- Supported by all GIS tools

Why use a Spatial Database?

- Integrated Data Management
- Seamless Datasets
- Supports Complex Workflows
- Easy to Manage & Program
- Fast - No Middleware
- Supports Terabytes of Data
- Supports 1000s of Users
- Easy to Program
- Multi-vendor Support
- Multi-platform Support
- Open/Standards-based
- Reliable
- Scalability
- Replication
- Partitioning
- Parallelism
- Bulk Load Utilities
- Version Management
- History Management
- Leverages Hardware

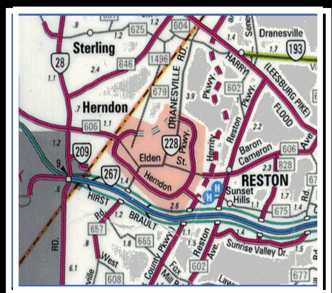
Bridging GIS "Stovepipes"

- Specialty Servers For Different Kinds Of Data
- Data Isolation
- High Systems Admin and Management Costs
- Scalability Problems
- High Training Costs
- Complex Support

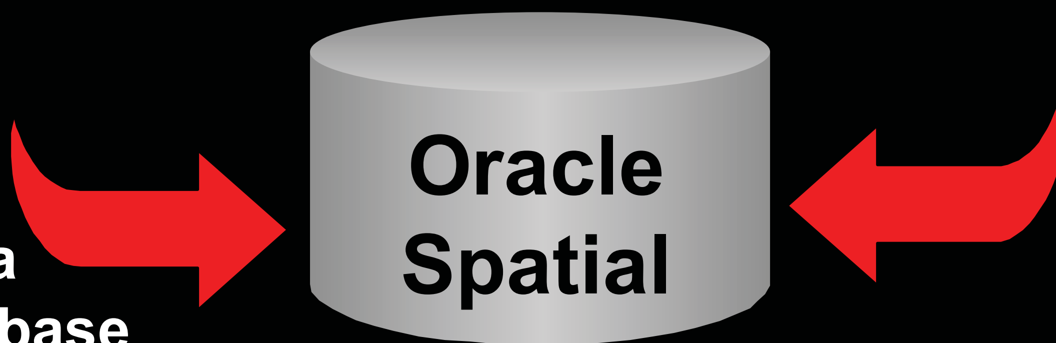


Oracle9*i* Spatial Capabilities

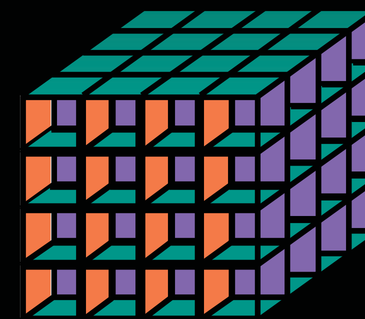
Spatial Data Types



All Spatial Data
Stored in the Database



Spatial Indexing

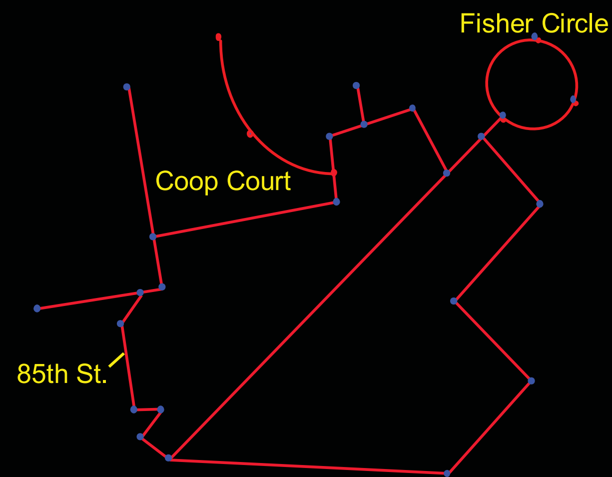


Fast Access to
Spatial Data

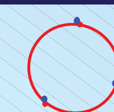
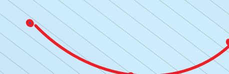
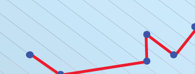
Spatial Access Through SQL

```
SELECT STREET_NAME FROM ROADS, COUNTIES
WHERE SDO_RELATE(road_geom, county_geom,
      'MASK=ANYINTERACT QUERYTYPE=WINDOW') ='TRUE'
AND COUNTYNAME='PASSAIC';
```

Spatial Data in Oracle Tables

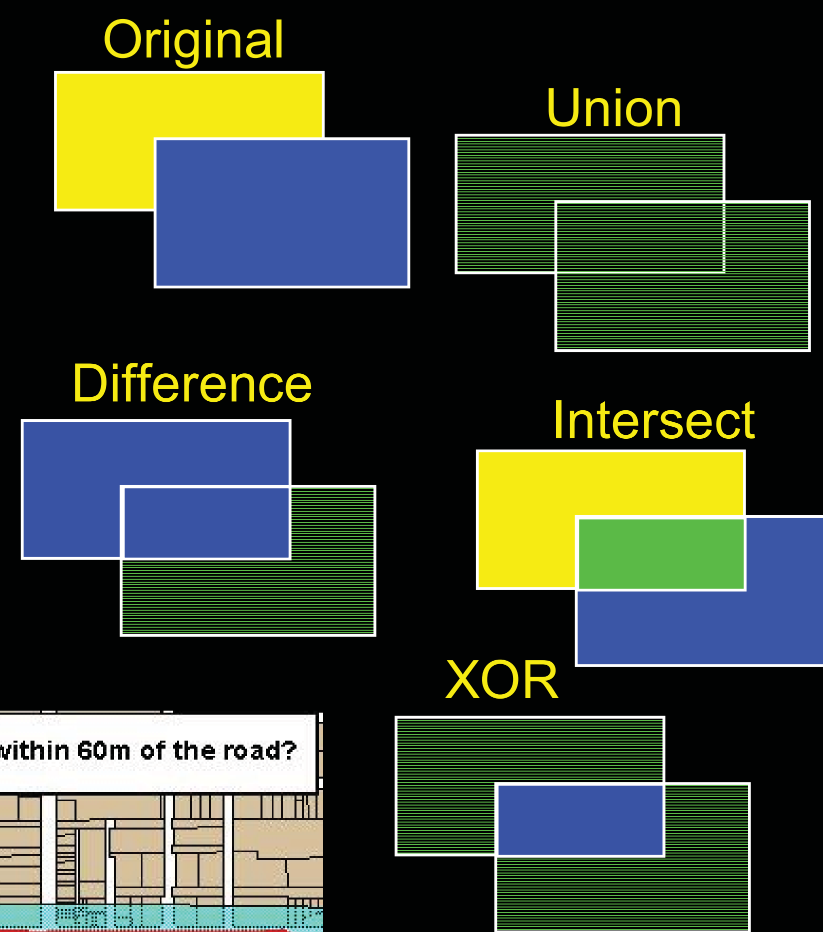


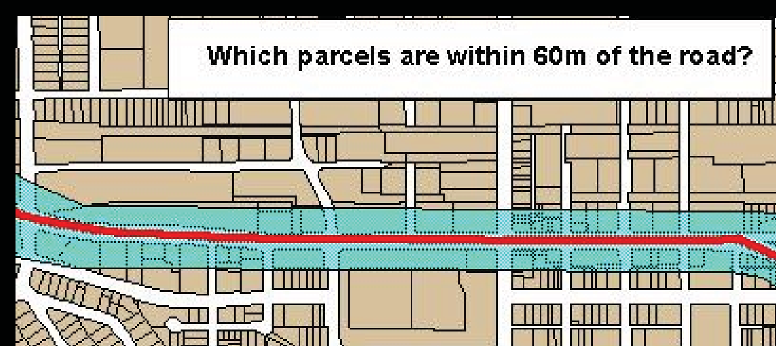
Road

ROAD_ID	NAME	SURFACE	LANES	LOCATION
1	Fisher Cir.	Asphalt	4	
2	Coop Ct.	Asphalt	2	
3	85Th St.	Asphalt	2	

Spatial Functions

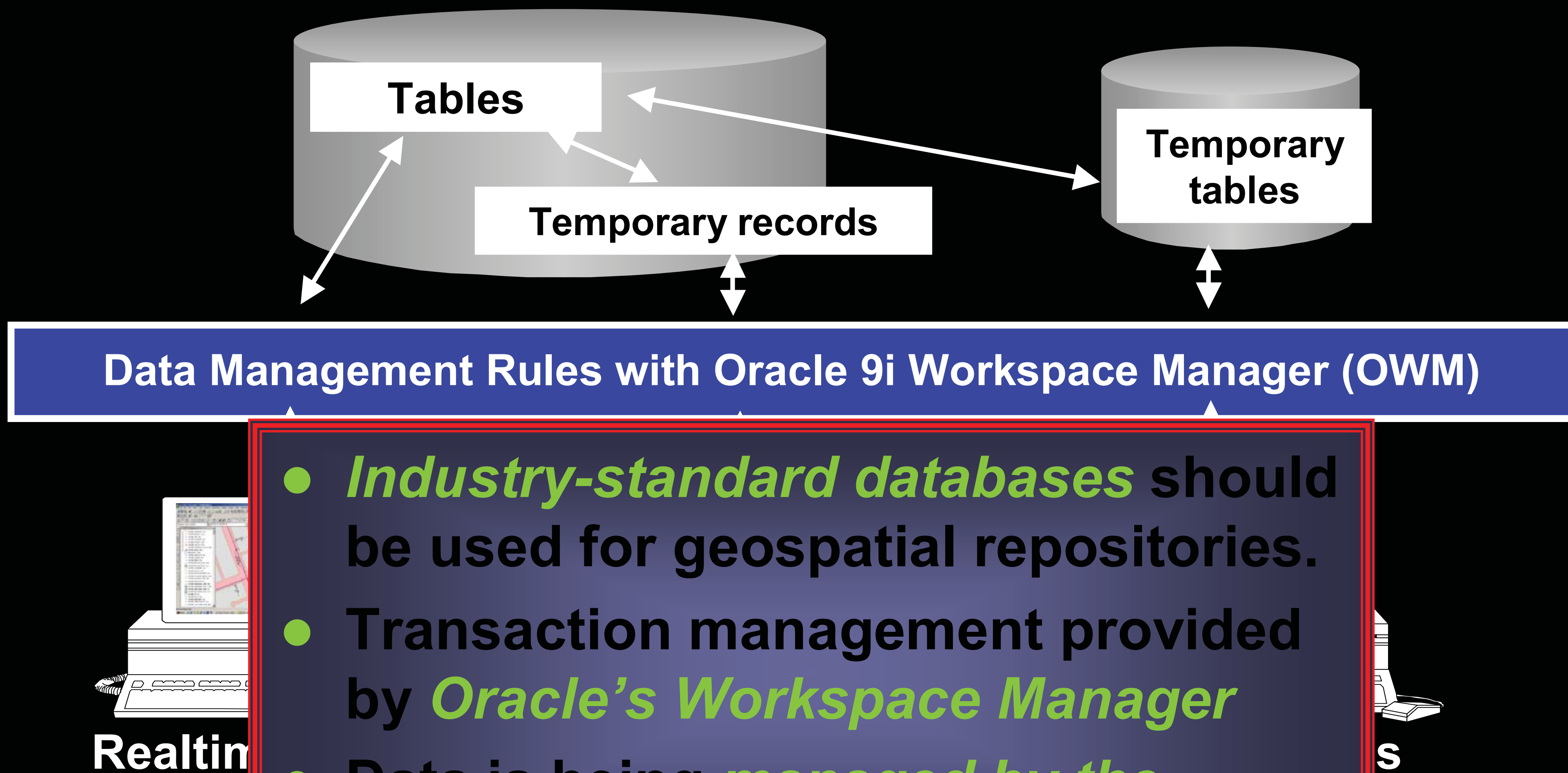
- Returns a geometry
  - Union
  - Difference
  - Intersect
  - XOR
  - Buffer
  - CenterPoint
  - ConvexHull
- Returns a number
  - LENGTH
  - AREA
  - Distance





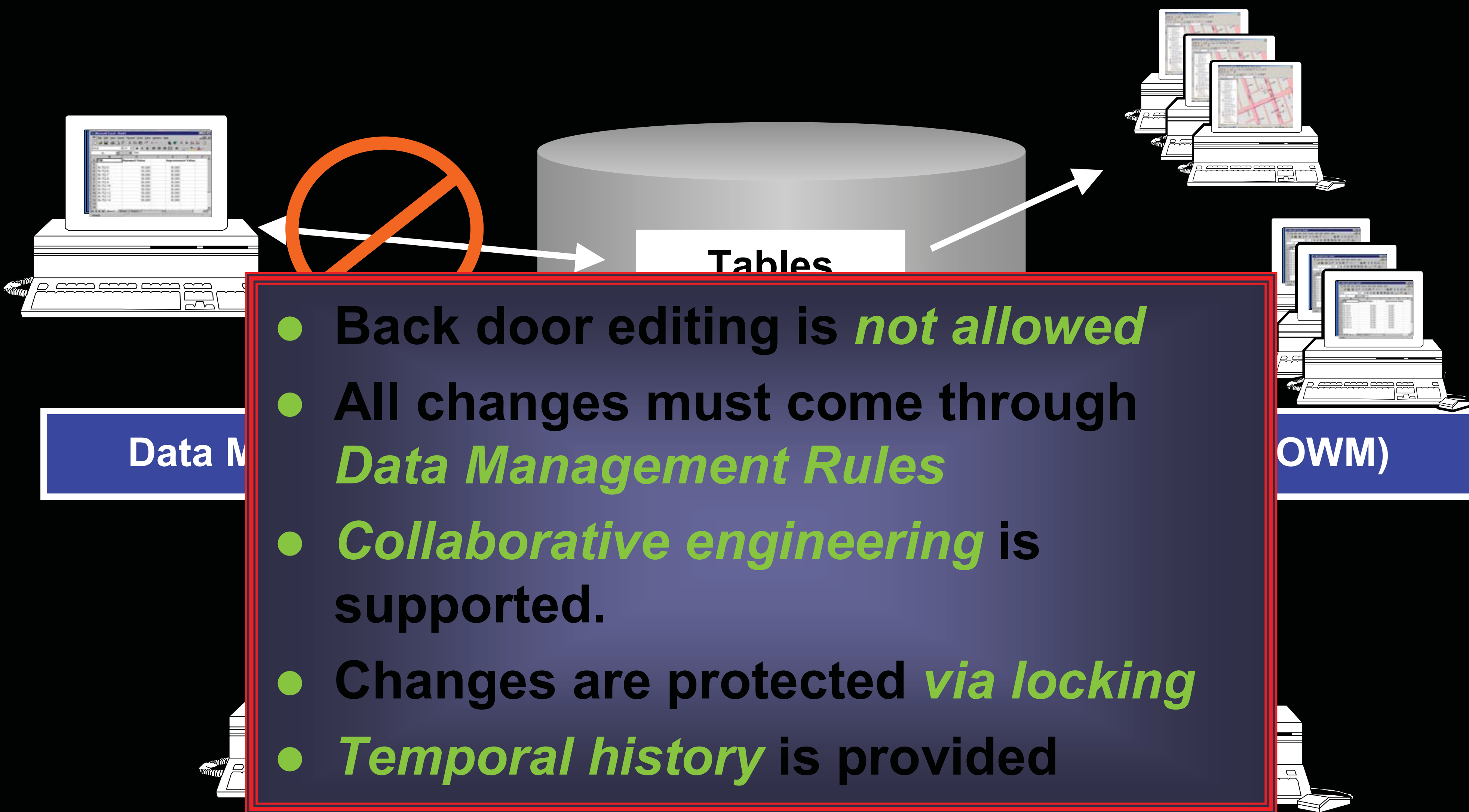
ORACLE

Workspace Manager: Change management



- *Industry-standard databases* should be used for geospatial repositories.
- Transaction management provided by *Oracle's Workspace Manager*
- Data is being *managed by the database*, not the application.

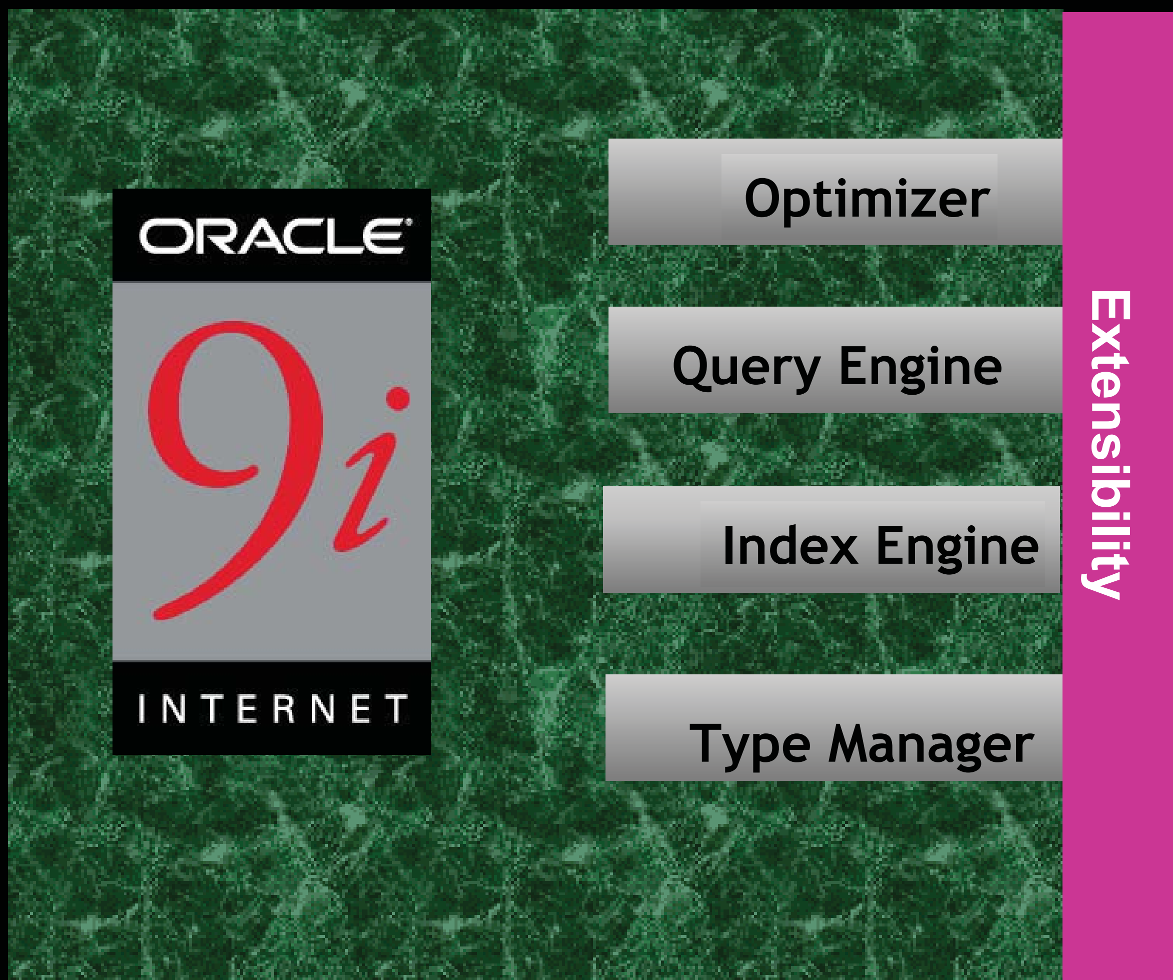
Workspace Manager – Data Management Rules



- Back door editing is ***not allowed***
- All changes must come through ***Data Management Rules***
- ***Collaborative engineering*** is supported.
- Changes are protected ***via locking***
- ***Temporal history*** is provided

ORACLE

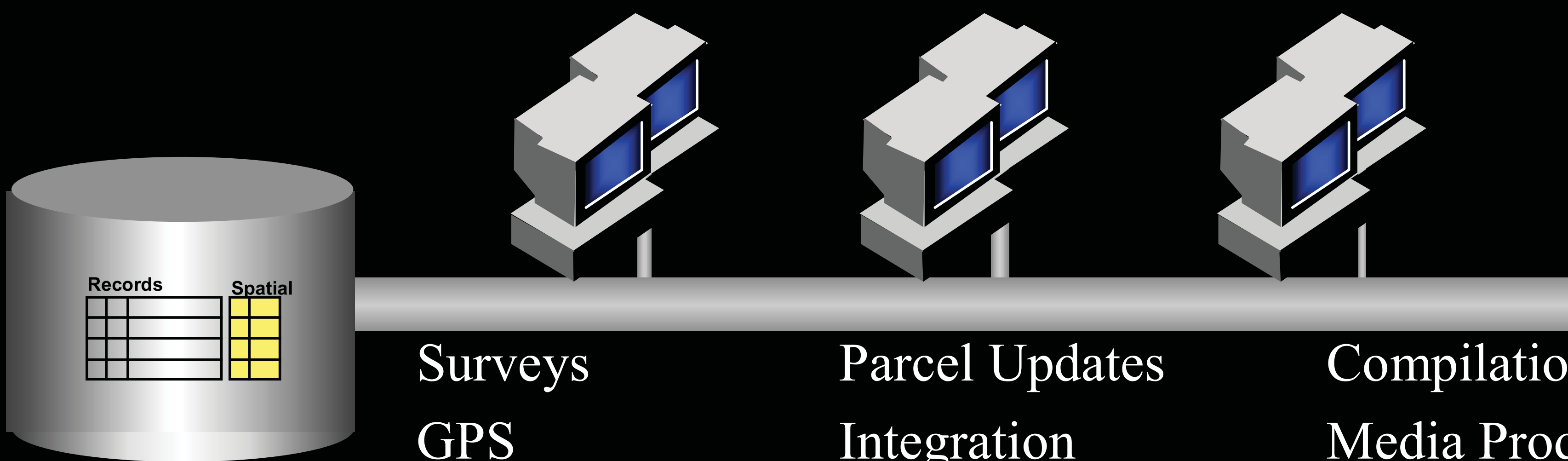
Extensible Database Framework



Oracle9*i* and Spatial Features

- Spatial Reference System
- Spatial Operators
- Long Transactions
- Linear Referencing
- Quadtree/R-tree index
- Parallel Index Create
- Java Classes
- Geodetic Support
- Spatial Aggregates

Spatial Platform for Integrated Enterprise Workflows



Enterprise Database & Applications

Data Collection

- Surveys
- GPS
- New Features
- Photogrammetry
- Online Updates

Production

- Parcel Updates
- Integration
- Long Transactions
- Versioning
- Topology Mgmt.
- Quality Control
- Security

Dissemination

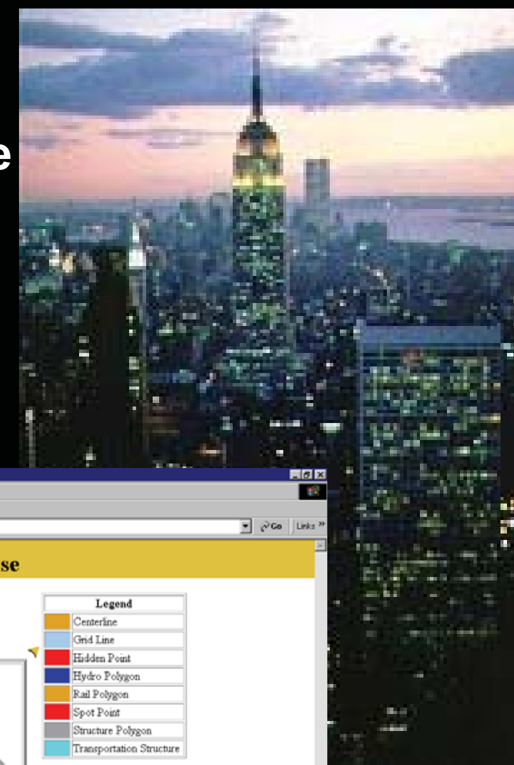
- Compilation
- Media Production
- Web Delivery
- Online Query
- Online updates
- Personalization
- Billing

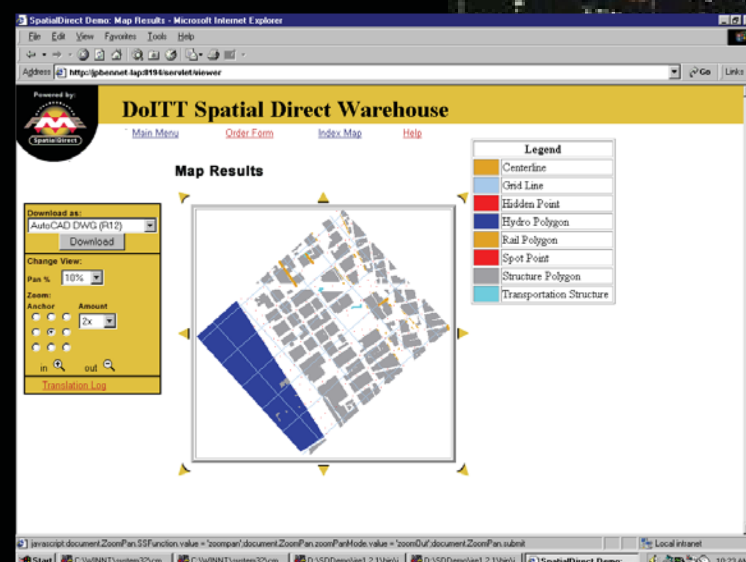
ORACLE

New York City Case Study

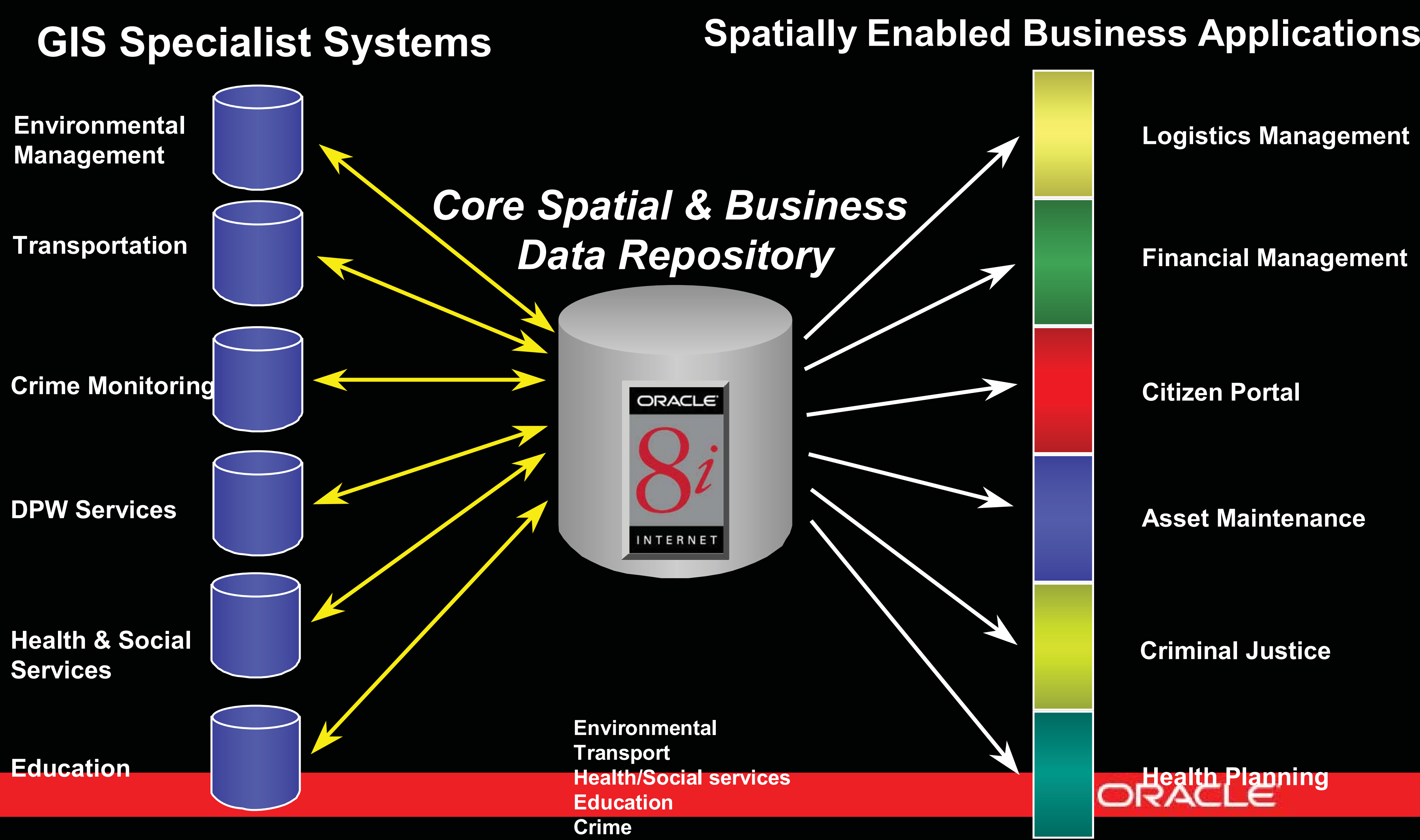
- NYC Standardized on 8iSpatial
- Open Architecture
- Largest, Most Accurate Municipal Spatial Data Warehouse in the world
- Leveraged Across Entire City/State
- Intregrate Multiple tools
  - ESRI
  - MapInfo
  - Smallworld

New York City
Spatial Data Warehouse





Integrated NYC Spatial Architecture



Working with Leading Partners

•

INTERGRAPH

MapInfo is desktop mapping

EngHOUSE

BENTLEY

ESRI The World Leader in GIS

Autodesk.

apic SA

TERRASOAR

FORMIDA SOFTWARE

SICAD® GEOMATICS

cadcorp

MAPQUEST

caris®

Atar INFORMATIC

PCI Geomatics Committed to GEO-Intelligence Solutions

GeoMicro, Inc.

Sedona GeoServices, Inc.

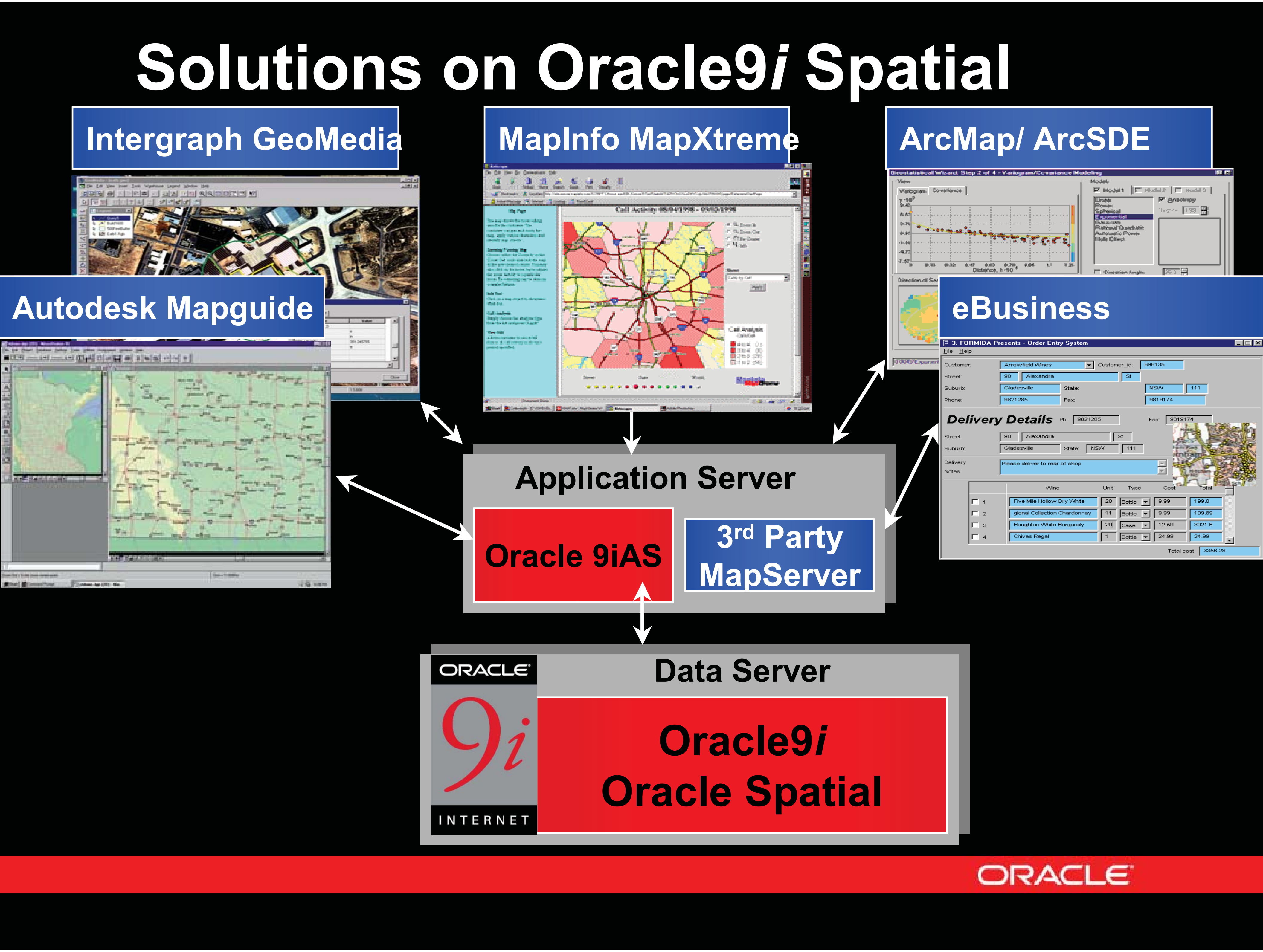
SAFE SOFTWARE INC.

ADVANCED VISUAL SYSTEMS INC.

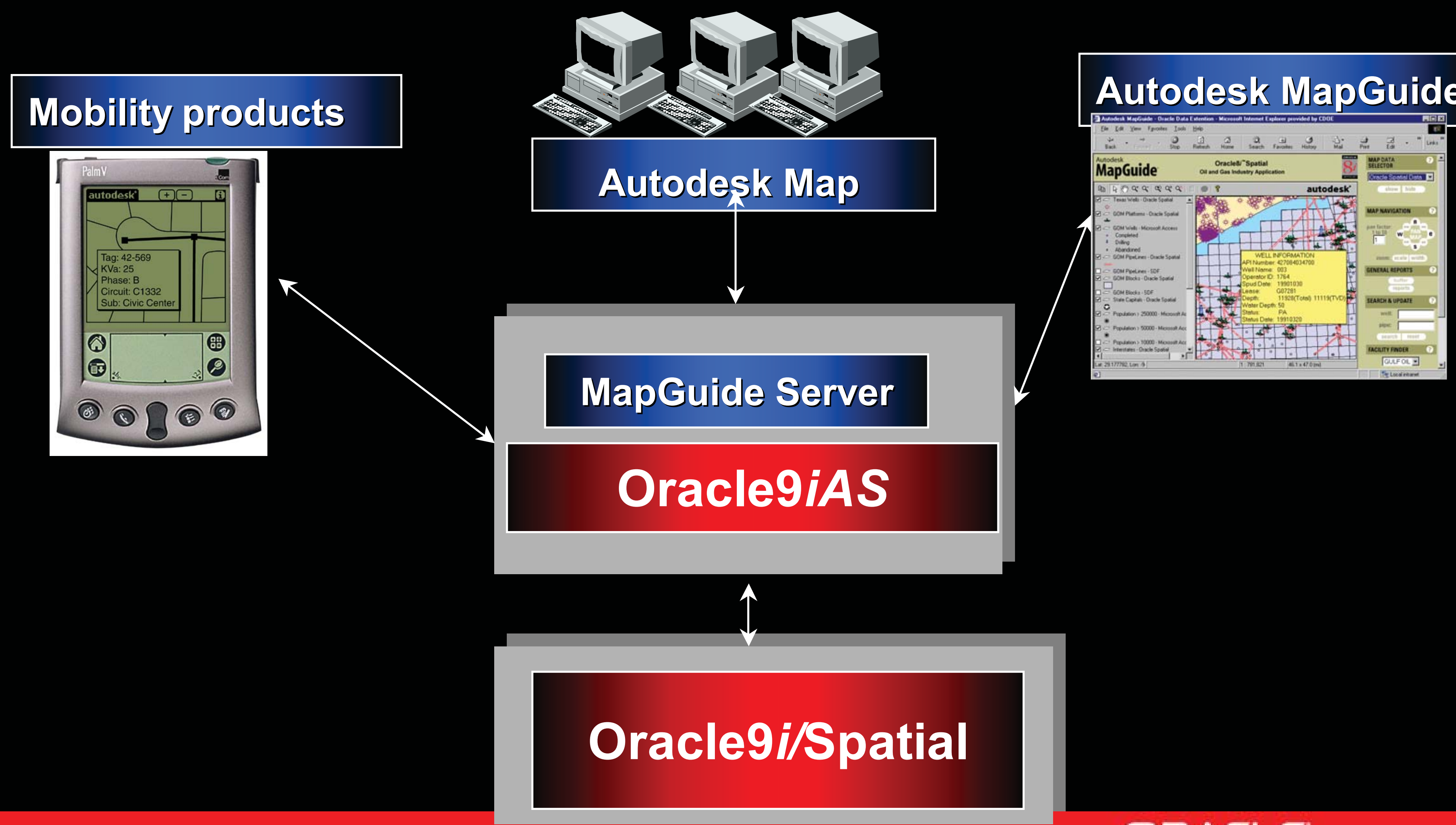
SMALLWORLD

Global Geomatics

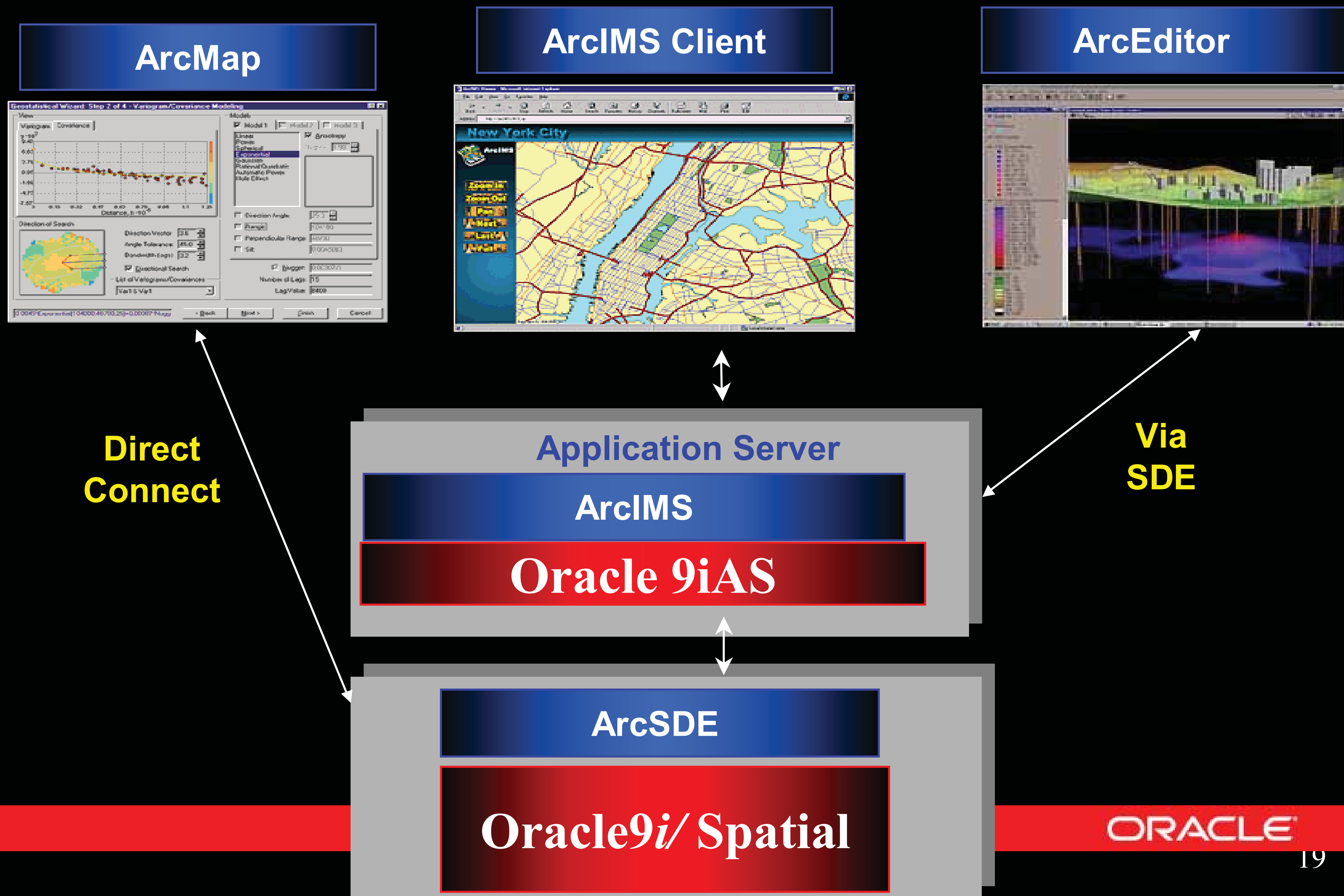
Qualitative Marketing Software, Inc.


Solutions on Oracle9i Spatial
Intergraph GeoMedia
MapInfo MapXtreme
ArcMap/ ArcSDE
Autodesk Mapguide
eBusiness
Application Server
Oracle 9iAS
3rd Party MapServer
Data Server
Oracle9i Oracle Spatial
ORACLE 9i INTERNET
ORACLE

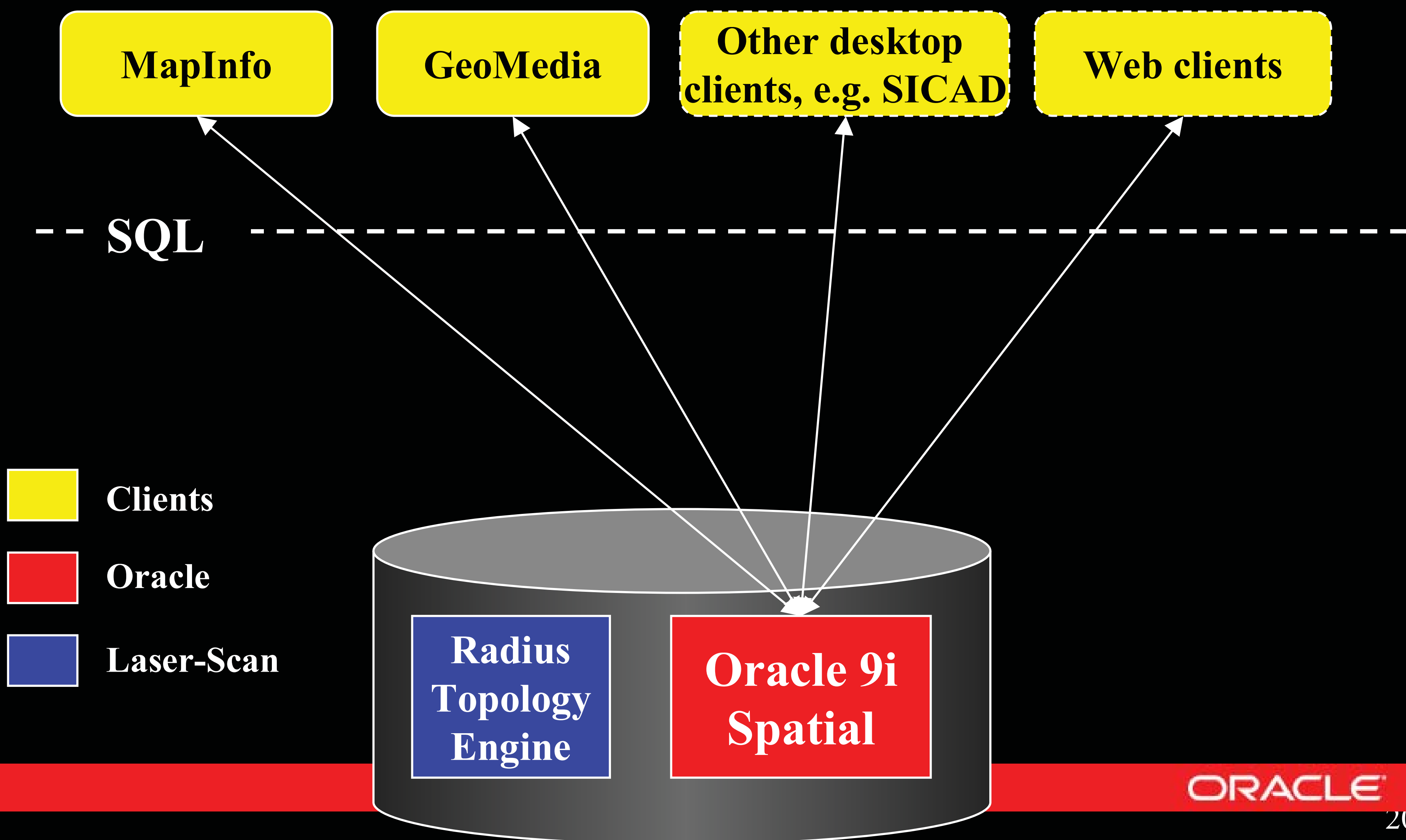
Autodesk & Oracle Platform



Oracle & ESRI Platform



Radius Topology Architecture



Spatial Standards Activities

- OGC:
  - Open Location Services
  - Geographic Markup Lanaguage (GML)
- ISO
- SQL92, SQLMM/Spatial

OpenGIS® PRINCIPAL MEMBER

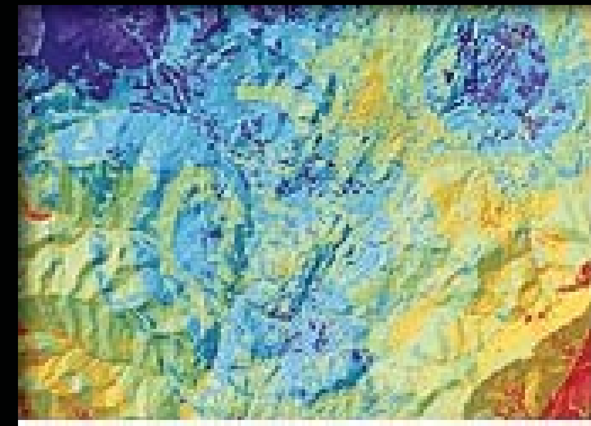
l.i.f. location inter-operability forum

ISO

On the Horizon… Oracle10*i*

- Raster Image Management
- Topology/Network Manager

GeoRaster



- New Raster Image & Grid Data Types
  - Satellite imagery, DOQs, Grids,
  - Open, general purpose 2D & 3D raster image management
  - Integration with 3rd party image analysis, GIS, compression
- Usage
  - Land management, Energy exploration, defense, satellite image portals, precision farming

ORACLE

Topology/Network Manager

- New Topology/Network Data Type and Schema
  - General purpose support and analysis of networks
  - Connectivity, adjacency and attribution at link level
  - 2D Planar topology for Land and Cadastre management
  - Conforms to OGC and ISO Topology specifications
- Usage
  - Land asset management
  - Network tracing (roadways, pipes)
  - Fleet, Supply Chain

ORACLE

Advantages of Managing Spatial Data in a Database

- **Same access methods (SQL)**
  - Select, insert, update, delete
- **Same RDBMS features available to GIS**
  - Replication, security, parallel, high availability...
- **Same utilities**
  - Load, import, export, backup
- **SAME DATABASE**
  - For spatial and non-spatial data

QUESTIONS & ANSWERS

http:// technet.oracle.com/products/spatial

ORACLE

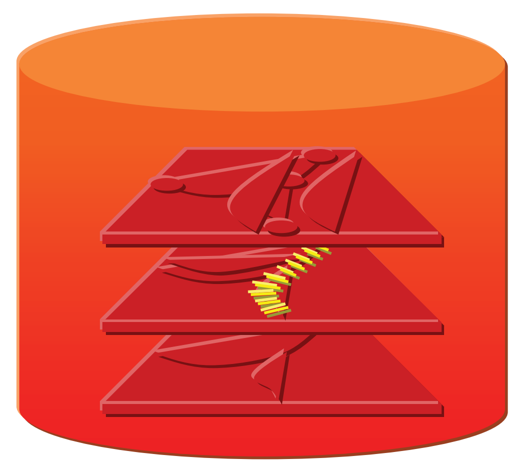
Next Generation Geospatial Databases

Dr David Maguire

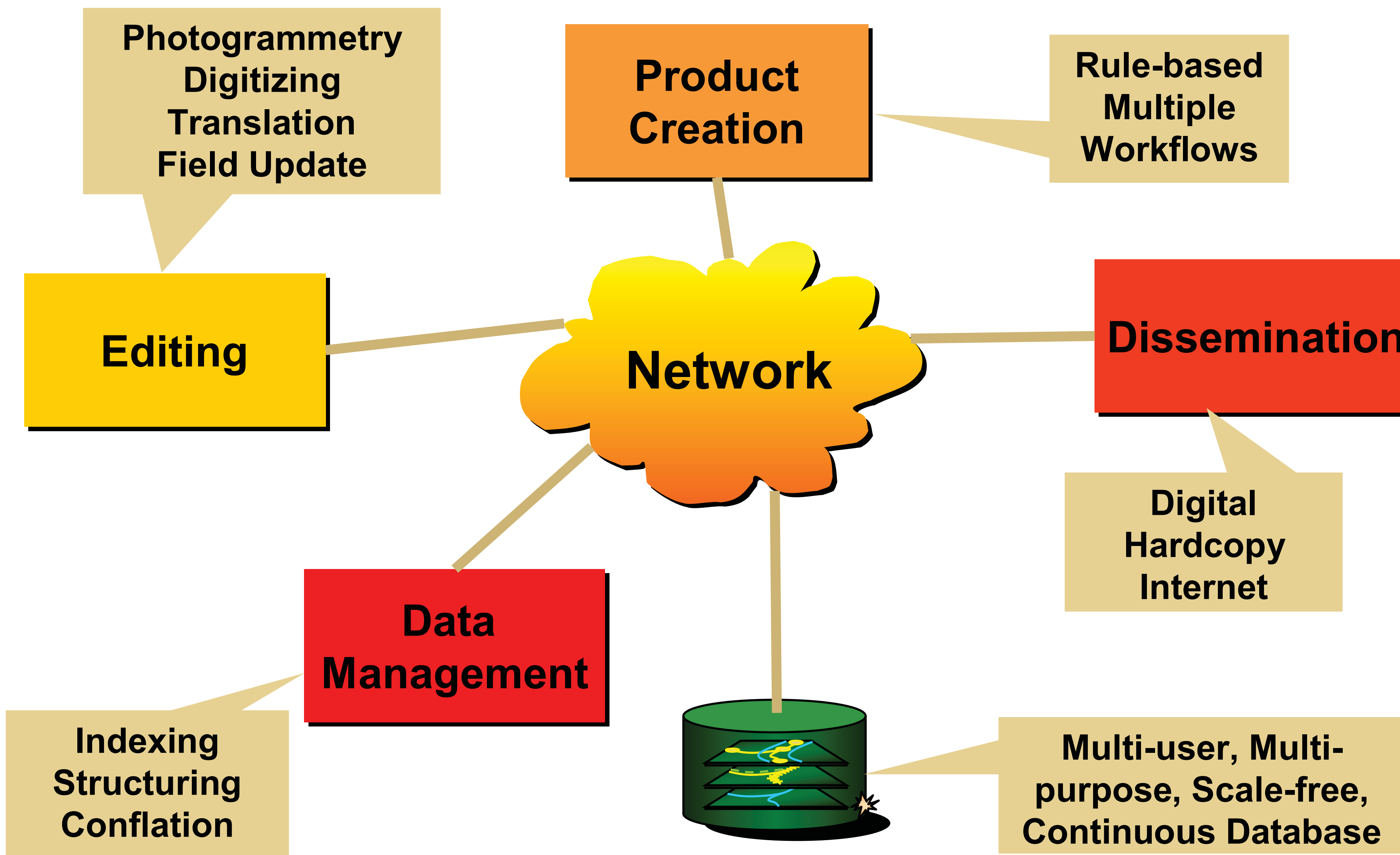
May 2002

Access to Government Data

- Spatial Databases
  - Now
  - Next
  - After Next
- Distributed GIS
  - Metadata
  - Portals
  - Web services



NMA Workflow



Database Approach

- All users share common database
- Continuous (joined up geography)
- DBMS gives …ilities
  - Integrity
  - Reliability
  - Flexibility
  - Scalability

Types of DBMS Model

- Hierarchical
- Network
- Relational - RDBMS
- Object-oriented - OODBMS
- Object-relational - ORDBMS

Table

Column = property

Row = object

Table = Object Class

Object Class with Geometry called Feature Class

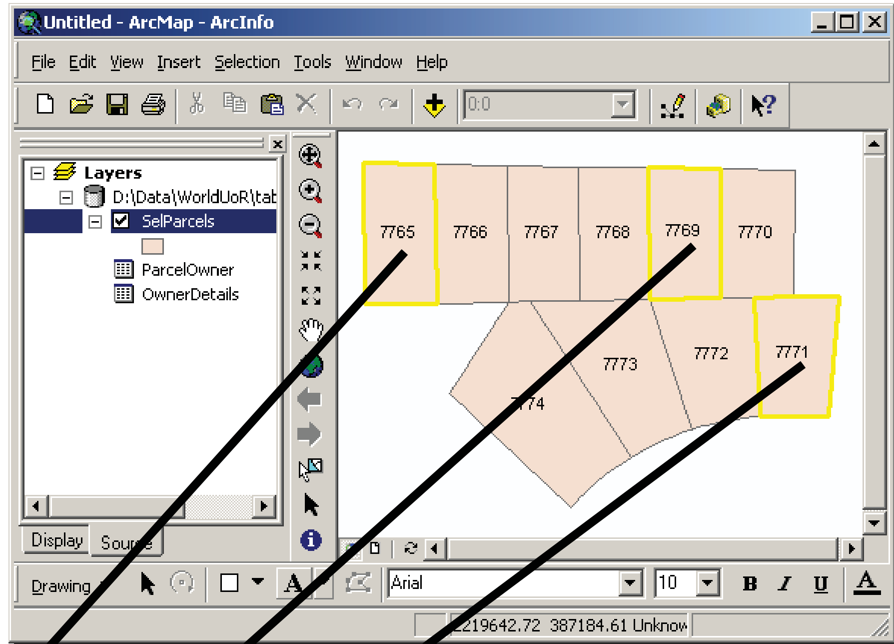
Attributes of STATES

FID	Shape*	AREA	STATE_NAME	STATE_FIPS
41	Polygon	51715.656	Alabama	01
49	Polygon	576556.687	Alaska	02
35	Polygon	113711.523	Arizona	04
45	Polygon	52912.797	Arkansas	05
23	Polygon	157774.187	California	06
30	Polygon	104099.109	Colorado	08
17	Polygon	4976.434	Connecticut	09
27	Polygon	2054.506	Delaware	10
26	Polygon	66.063	District of Columbia	11
47	Polygon	55815.051	Florida	12
43	Polygon	58629.195	Georgia	13
48	Polygon	6381.435	Hawaii	15
7	Polygon	83340.594	Idaho	16
25	Polygon	56297.953	Illinois	17
20	Polygon	36399.516	Indiana	18
12	Polygon	56257.219	Iowa	19
32	Polygon	82195.437	Kansas	20
31	Polygon	40318.777	Kentucky	21
46	Polygon	45835.898	Louisiana	22
2	Polygon	32161.664	Maine	23
29	Polygon	9739.753	Maryland	24
13	Polygon	8172.482	Massachusetts	25
50	Polygon	57898.367	Michigan	26
9	Polygon	84517.469	Minnesota	27
42	Polygon	47618.723	Mississippi	28
34	Polygon	69831.625	Missouri	29
1	Polygon	147236.031	Montana	30

Record: 0 Show: All Selected Records (0 out of 51 Selected.)

Relational Join

- One-One (1-1)
- One-Many (1-N)
- Many to Many (N-N)



Attributes of SelParcels

OBJECTID_1*	Shape*	OBJECTID	FID_	PARCELS_ID	Shape_Leng	Shape_Length
1	Polygon	140	11291	7765	359.372313	359.372313
2	Polygon	141	11292	7766	353.619020	353.619020
3	Polygon	142	11293	7767	350.879925	350.879924
4	Polygon	143	11294	7768	342.228455	342.228454
5	Polygon	144	11295	7769	343.753847	343.753847
6	Polygon	145	11296	7770	340.925943	340.925943
7	Polygon	174	11400	7771	333.647752	333.647752
8	Polygon	175	11401	7772	362.634496	362.634498
9	Polygon	178	11404	7773	431.353539	431.353538

Record: 1 Show: All Selected Records (3 out of 10 Selected.) Options

Table-Table Join

❶

Attributes of OwnerDetails

OBJECTID*	Owner	Zip	Street
1	Bill Smith	123	123 West Palm
2	Paul Star	867	235 Fern Avenue
3	Eric Clapton	334	The Mansion, Gate Lane
4	Peggy Taylor	124	4 Longhorn Plaza

Record: 3 Show: All Selected Records (1

❷

Attributes of ParcelOwner

OBJECTID*	Owner	Parcels_ID	Land_Value
1	Bill Smith	7764	100500
2	Bill Smith	7767	100650
3	Bill Smith	7768	100650
4	Paul Star	7769	100450
5	Eric Clapton	7770	100475
6	Paul Star	7771	100600
7	Bill Smith	7772	100400
8	Peggy Taylor	7773	100800
9	Eric Clapton	7774	110005
10	Peggy Taylor	7775	110000
11	Bill Smith	7766	100450
12	Paul Star	7765	100387

Record: 1 Show: All Selected Records (5 out of

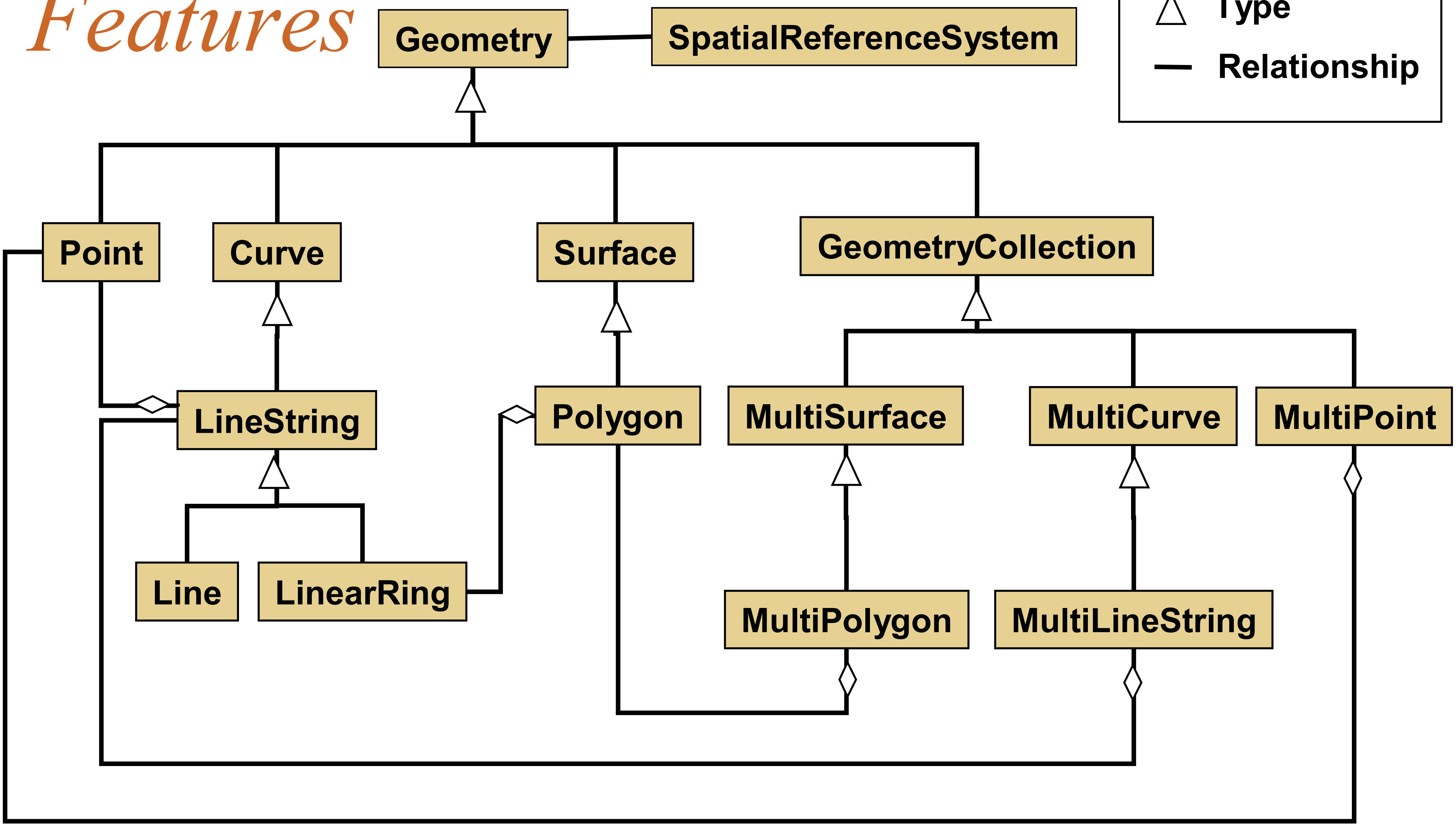
❸

Attributes of SelParcels

OBJECTID_1*	Shape*	OBJECTID	FID_	PARCELS_ID
1	Polygon	140	11291	7765
2	Polygon	141	11292	7766
3	Polygon	142	11293	7767
4	Polygon	143	11294	7768
5	Polygon	144	11295	7769
6	Polygon	145	11296	7770
7	Polygon	174	11400	7771
8	Polygon	175	11401	7772
9	Polygon	178	11404	7773

Record: 1 Show: All Selected Records (4 out of 10 Select

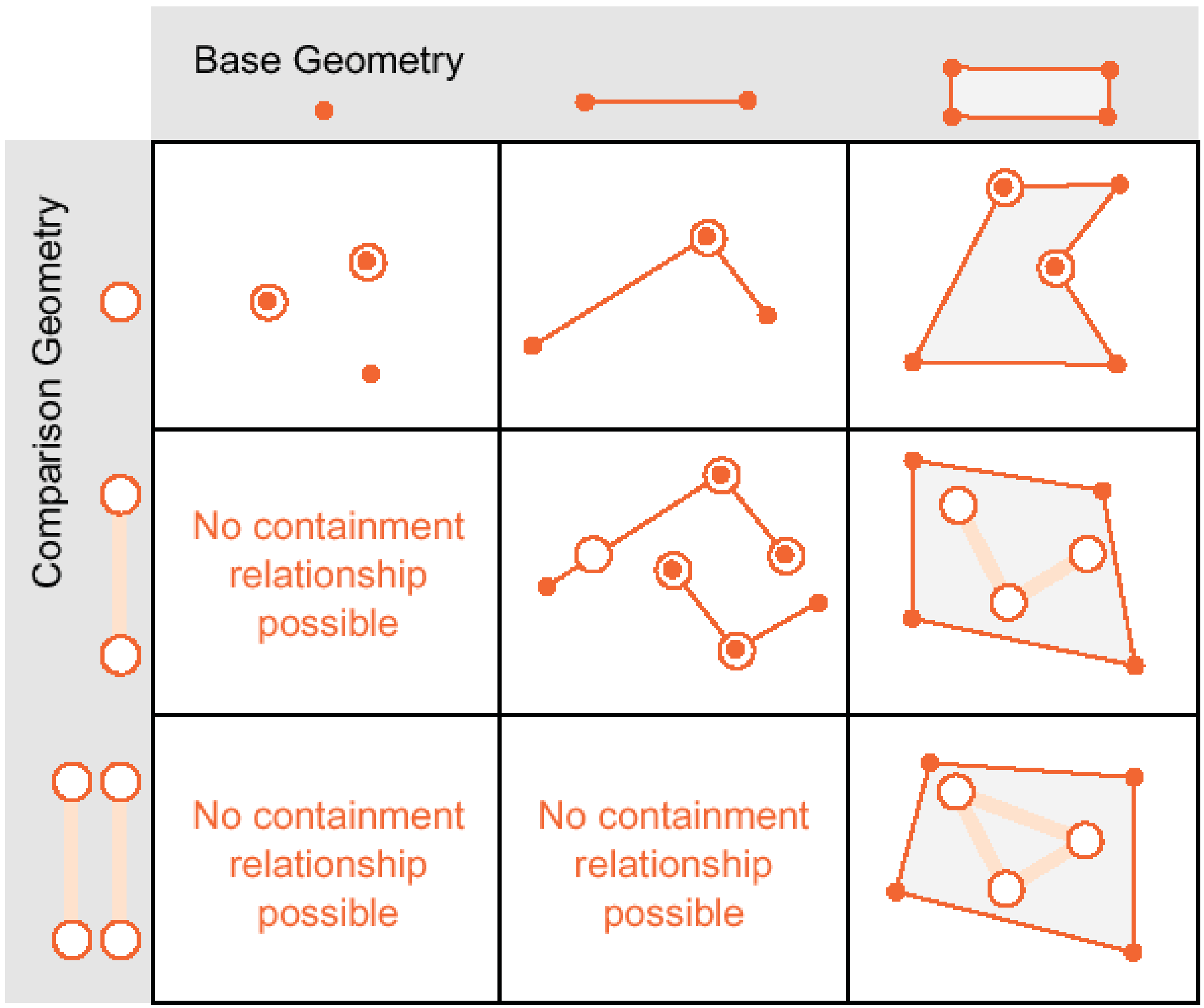
Spatial Types – OGC Simple Features



Spatial Relations

- Equals – same geometries
- Disjoint – geometries share common point
- Intersects – geometries intersect
- Touches – geometries intersect at common boundary
- Crosses – geometries overlap
- Within– geometry within
- Contains – geometry completely contains
- Overlaps – geometries of same dimension overlap
- Relate – intersection between interior, boundary or exterior

Contains Relation

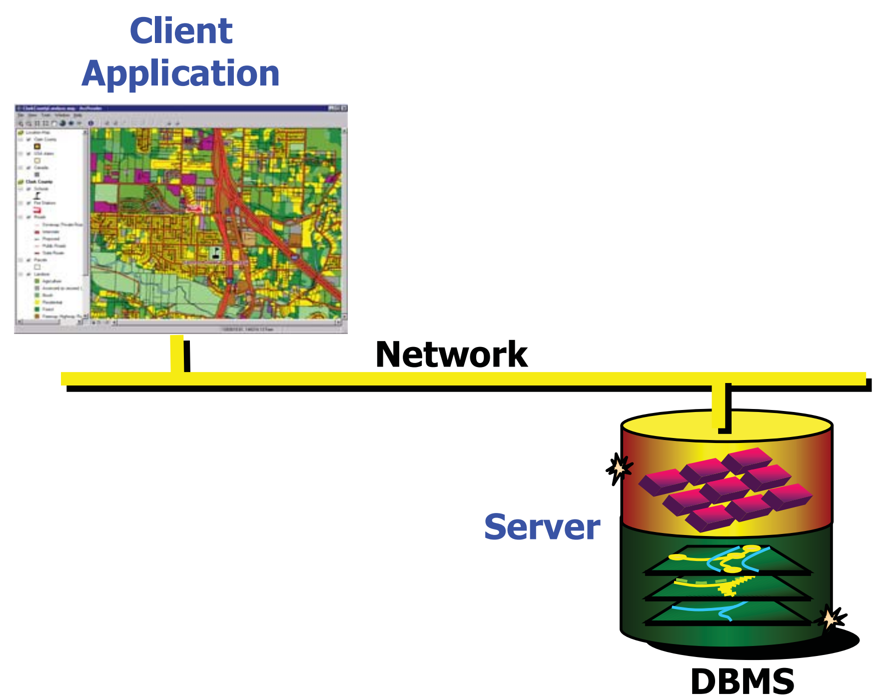


SQL Standardization

- SQL1 1986
- SQL2 1992
  - All shipping spatial database products = SQL2 + proprietary extensions
- SQL3 2002?
  - Adds support for Abstract Data Types (ADTs)
  - Part 3 deals with Spatial

Shipping Products

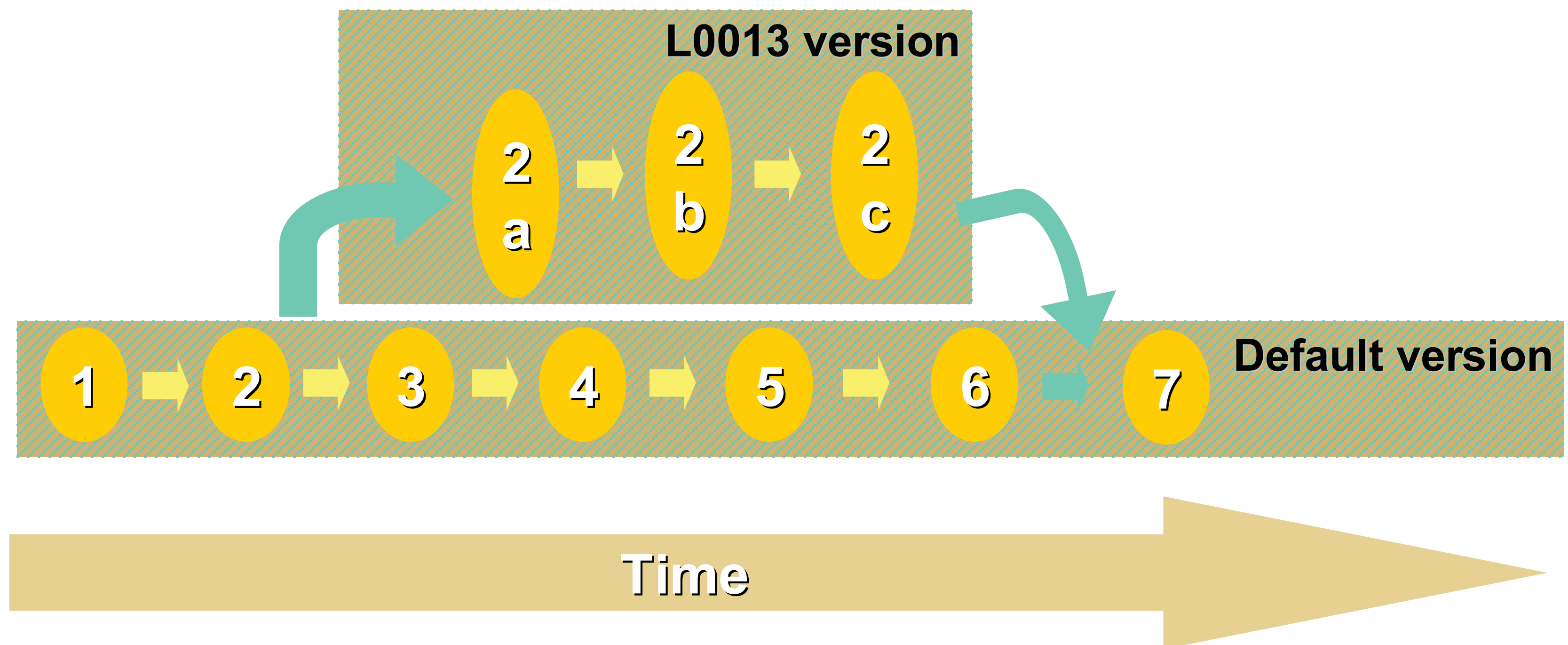
- O-R Extensions
  - IBM DB2 Extender
  - Informix Blade
  - Oracle Spatial
- Application Servers
  - Autodesk Design Server
  - ESRI ArcSDE
  - MapInfo Spatialware
- Client products
  - Numerous



Versioning

- Solves six key spatial database problems
  - Long transactions
  - Multi-user updates
  - Design alternatives
  - History
  - Distributed databases
  - Scalability

1. Long Transaction

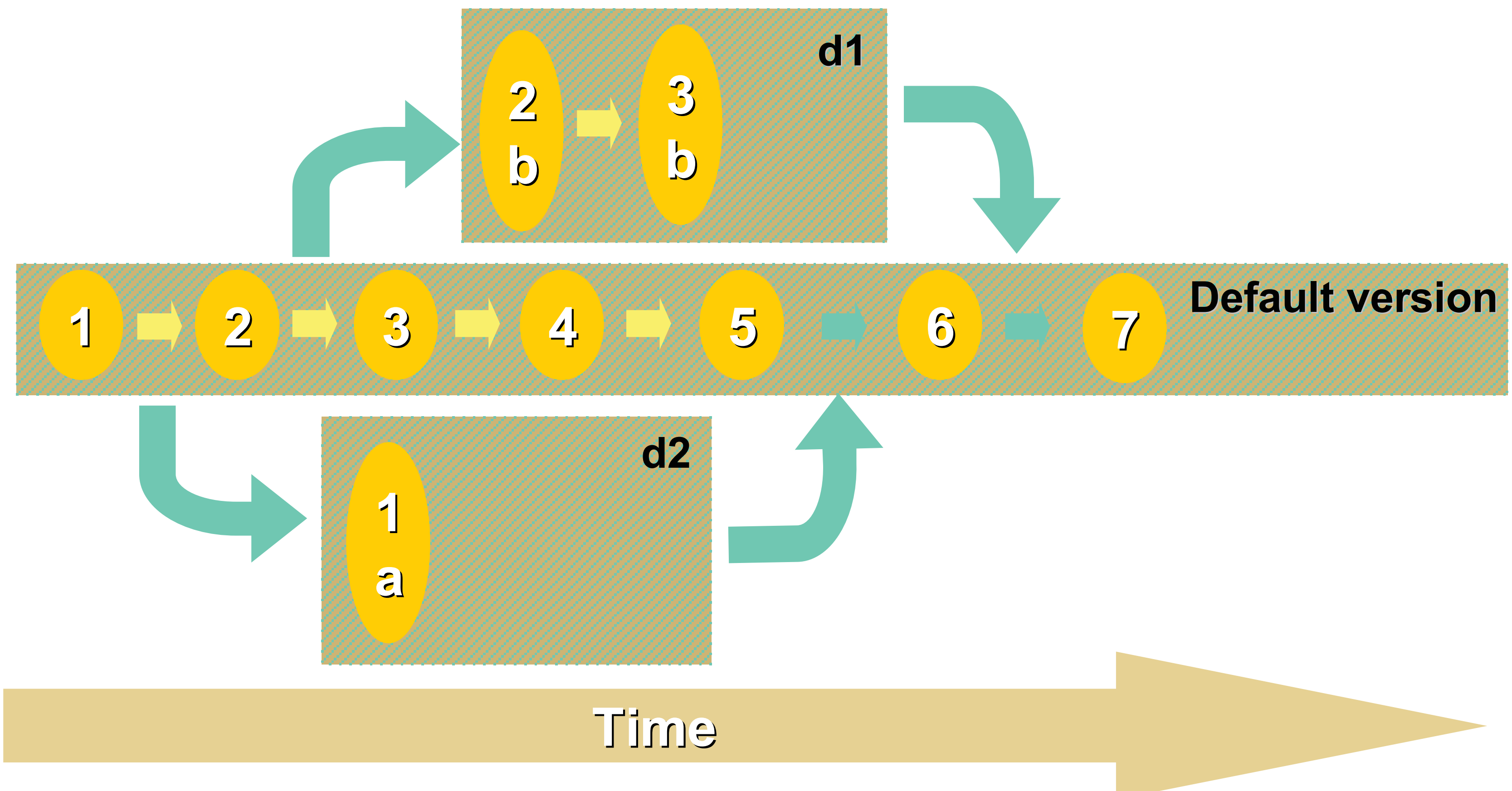


- Named version
  - Created / manipulated in an edit session
  - Edits independent of other versions
  - Changes stored in database

2. *Multi-user Updates (Concurrency)*

- Instead of locking, allow multiple versions created by single user or multiple users
- Each user believes they have full access to database
- Implement model for conflict detection and resolution
  - First to commit wins
  - Spatial / attribute conflicts resolved by GIS

3. Design Alternatives

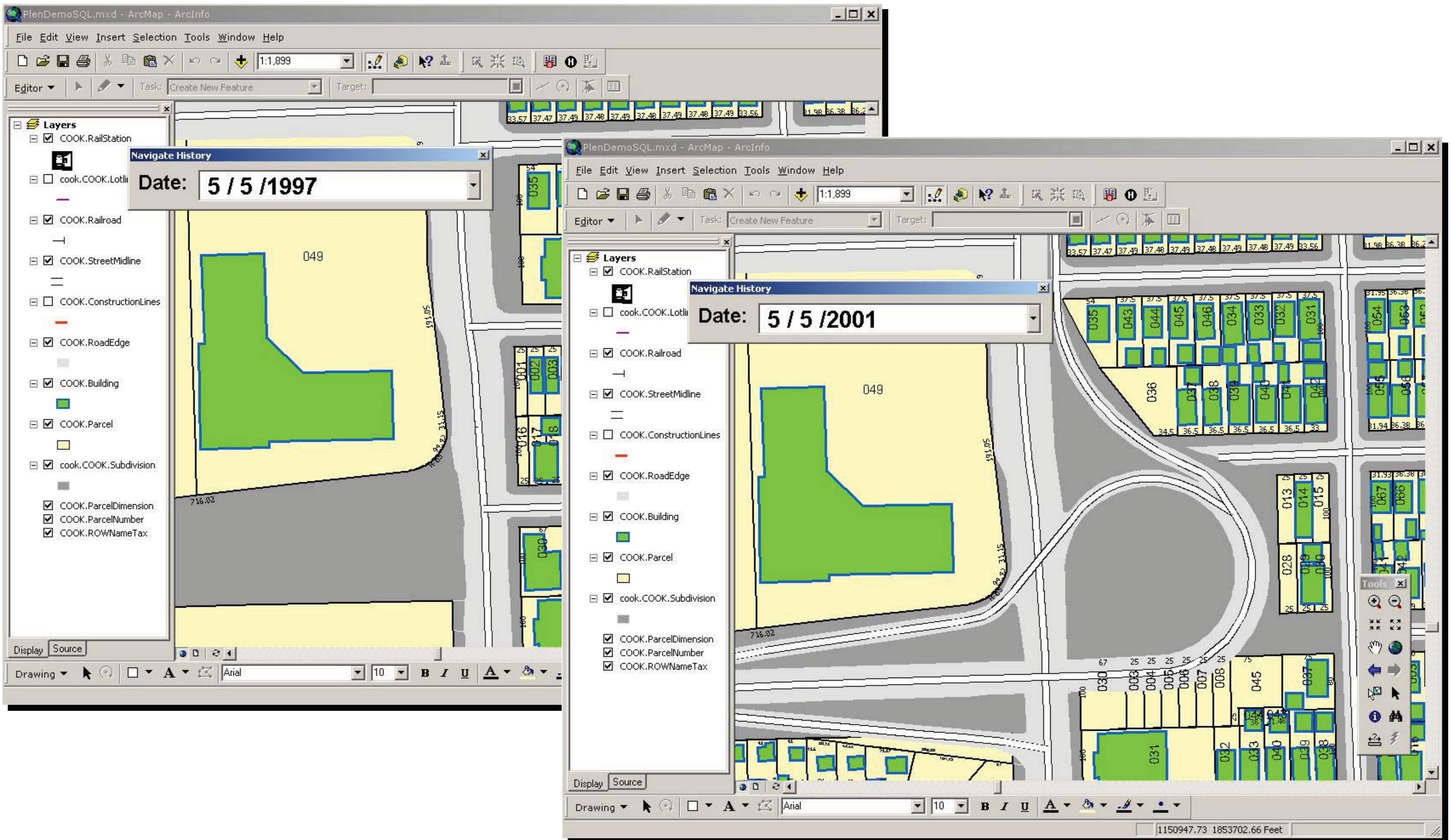


- Three alternative designs
- Three versions in database

4. History

- Retain historical state of database rows
- Maintain full integrity for all time points
- Associate information with change events
- Model what changed and how it changed

History

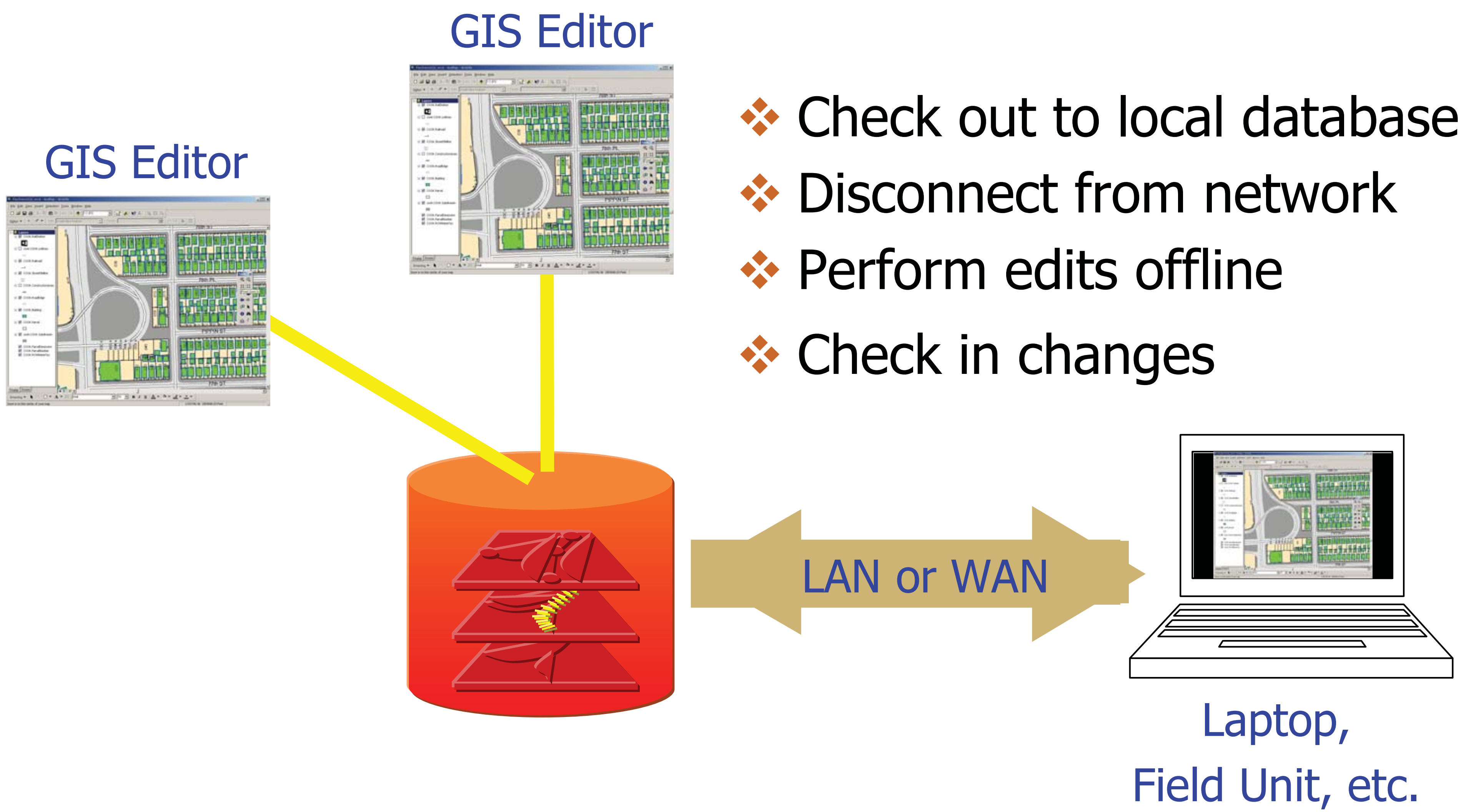

PlenDemoSQL.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
Editor
Task: Create New Feature
Target:
Layers
COOK.RailStation
cook.COOK.Lotli
COOK.Railroad
COOK.StreetMidline
COOK.ConstructionLines
COOK.RoadEdge
COOK.Building
COOK.Parcel
cook.COOK.Subdivision
COOK.ParcelDimension
COOK.ParcelNumber
COOK.ROWNameTax
Navigate History
Date: 5 / 5 /1997
049
Display
Source
Drawing
Arial
PlenDemoSQL.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
Editor
Task: Create New Feature
Target:
Layers
COOK.RailStation
cook.COOK.Lotli
COOK.Railroad
COOK.StreetMidline
COOK.ConstructionLines
COOK.RoadEdge
COOK.Building
COOK.Parcel
cook.COOK.Subdivision
COOK.ParcelDimension
COOK.ParcelNumber
COOK.ROWNameTax
Navigate History
Date: 5 / 5 /2001
049
Tools
Display
Source
Drawing
Arial
1150947.73 1853702.66 Feet

5. Distributed Databases

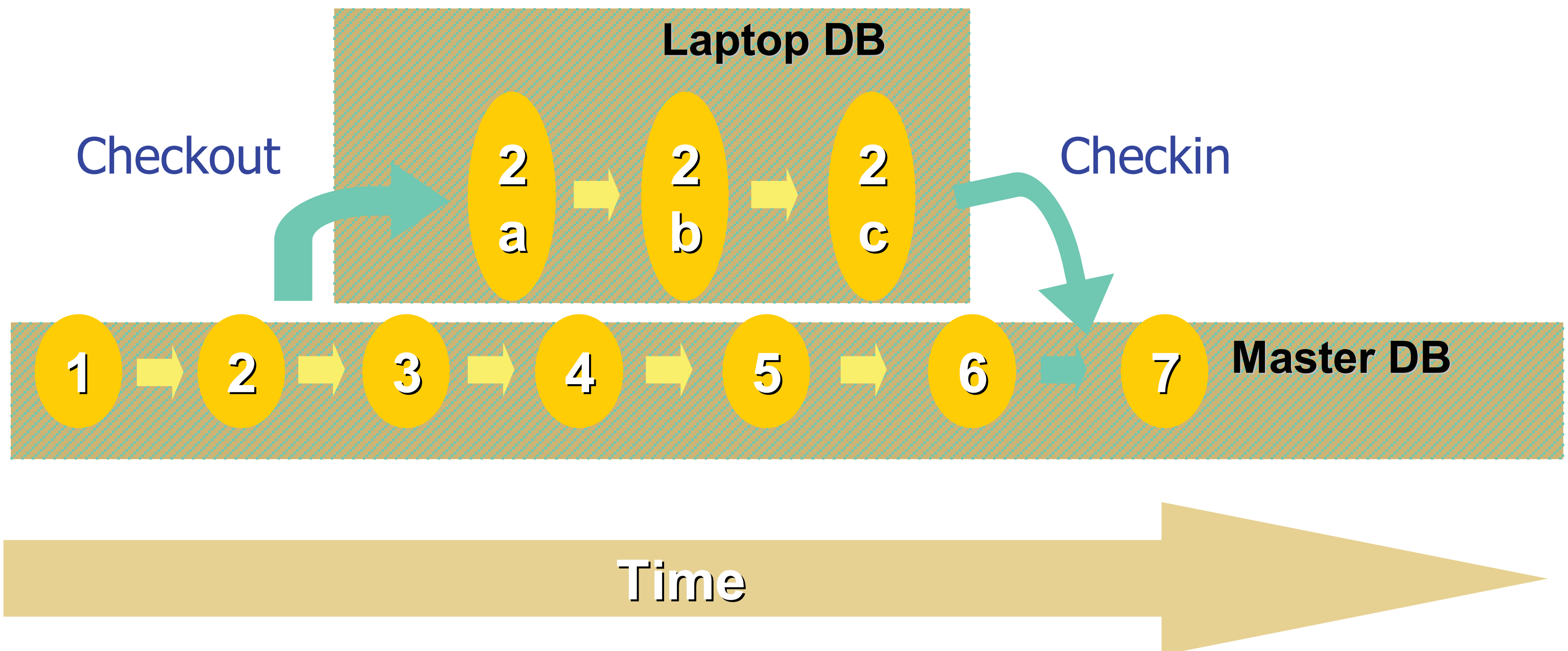
- Disconnected editing
  - Download database content to separate database system
  - Maintain database integrity (networks, topology, relationships, etc.)
  - Synchronize changes on reconnect
- Replication
  - Treat geographically distributed databases as though they were one (synchronization)

Disconnected Editing

- Check out to local database
- Disconnect from network
- Perform edits offline
- Check in changes



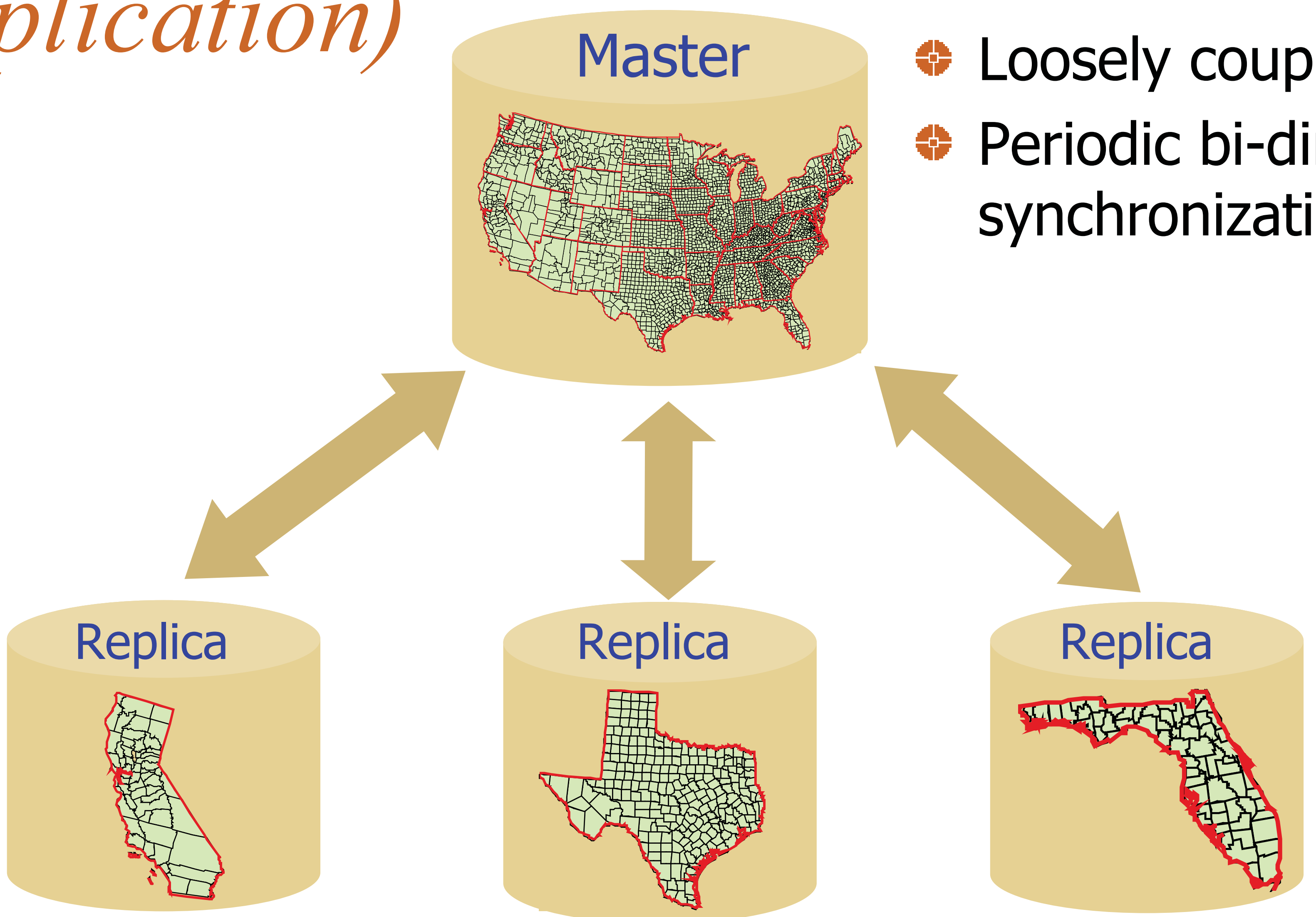
6. Distributed Versions



- Versions checked out to separate database by exporting/importing current master DB state
- Versions edited in parallel, accumulating change information
- Change information is imported (checkin) into master DB as a version and reconciled

6. Scalability (Distributed Replication)

- Loosely coupled
- Periodic bi-directional synchronization

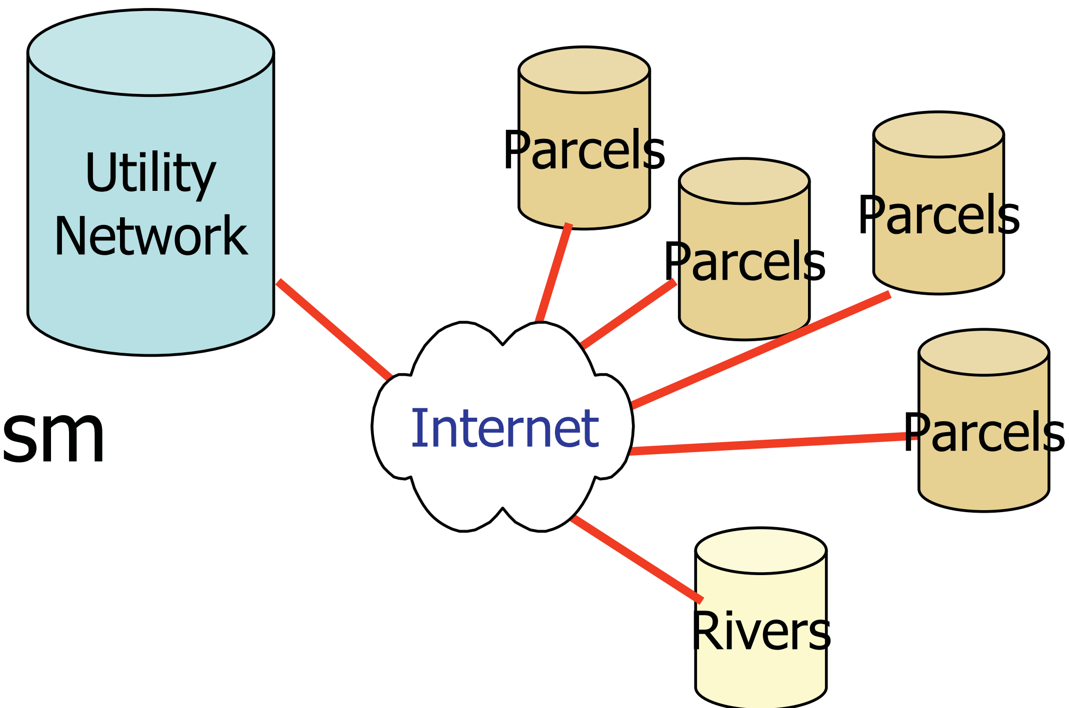


Distributed GIS Vision

- Widespread access to geographic data
  - Wired
  - Wireless
- Just in Time Delivery
  - Reduced cost of distribution
  - Intelligent objects from database
  - Up to date data
- Integrate data from multiple sources

Realizing this Vision Requires

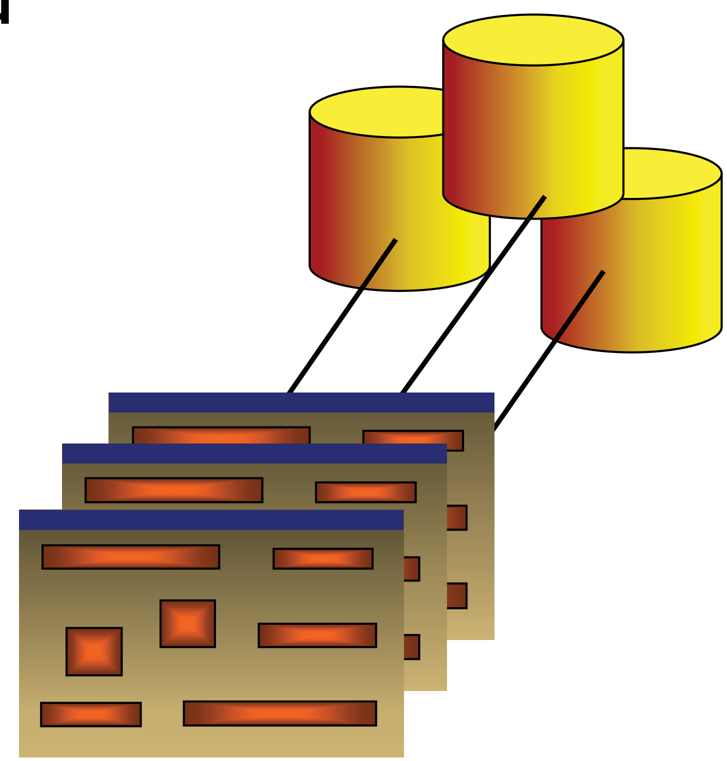
- Database-driven systems
  - Multi-user
  - Multi-purpose
  - Continuous
  - Secure
- Data access mechanism
  - Metadata catalog
  - Portal
- Distributed, network aware
  - Clients
  - Servers



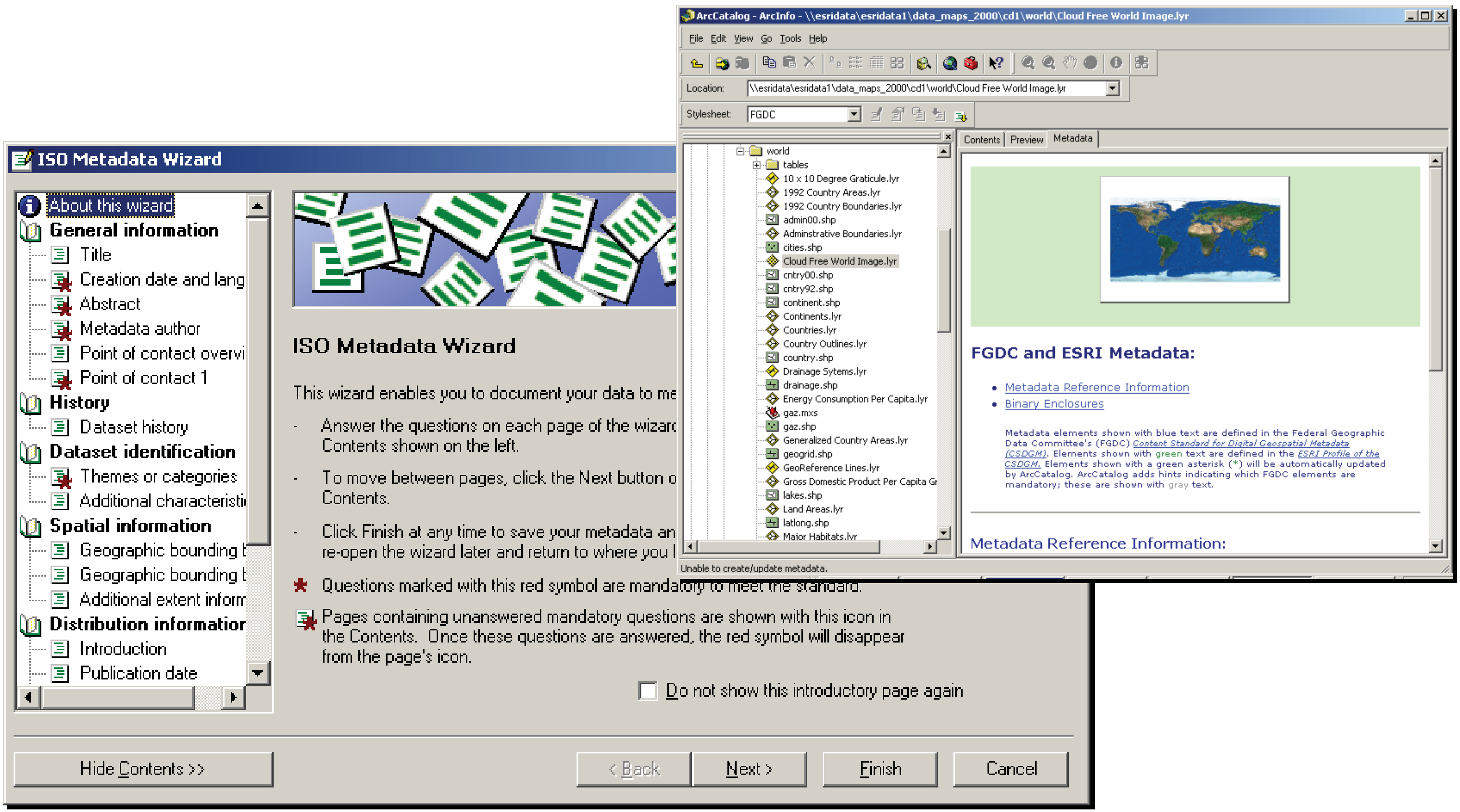
Metadata

- Data about data
  - Geographic area covered
  - Currency
  - Rules of use
  - Positional accuracy
  - Means of encoding
  - Datum
  - Map projection

Metadata is the Key to Access

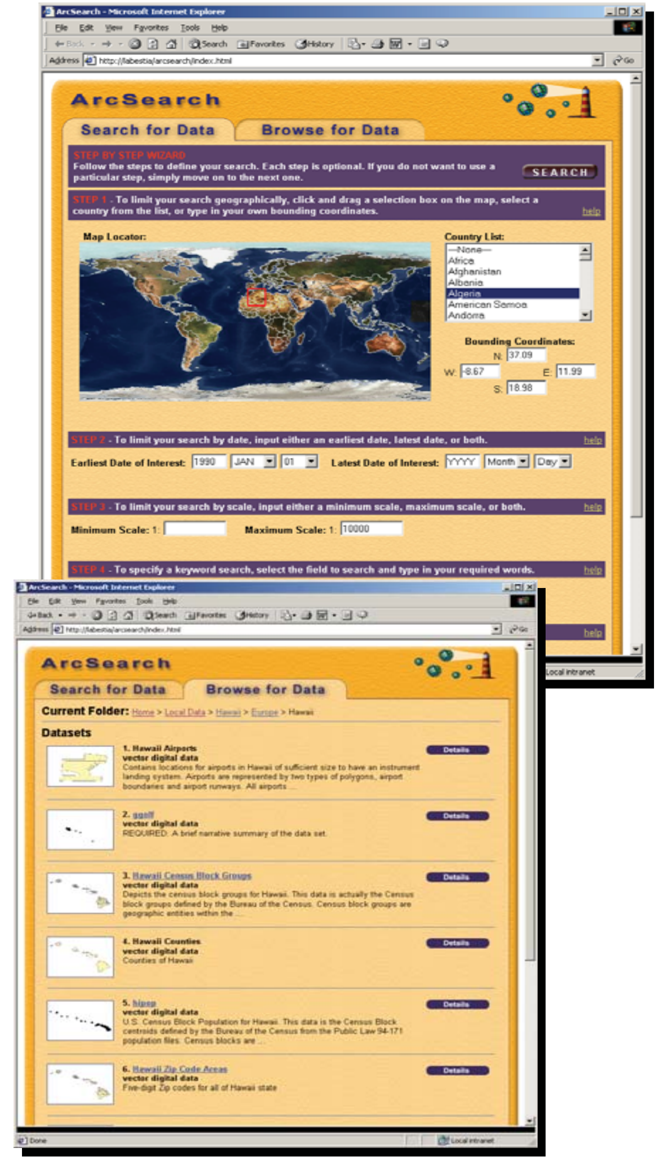


Create Metadata

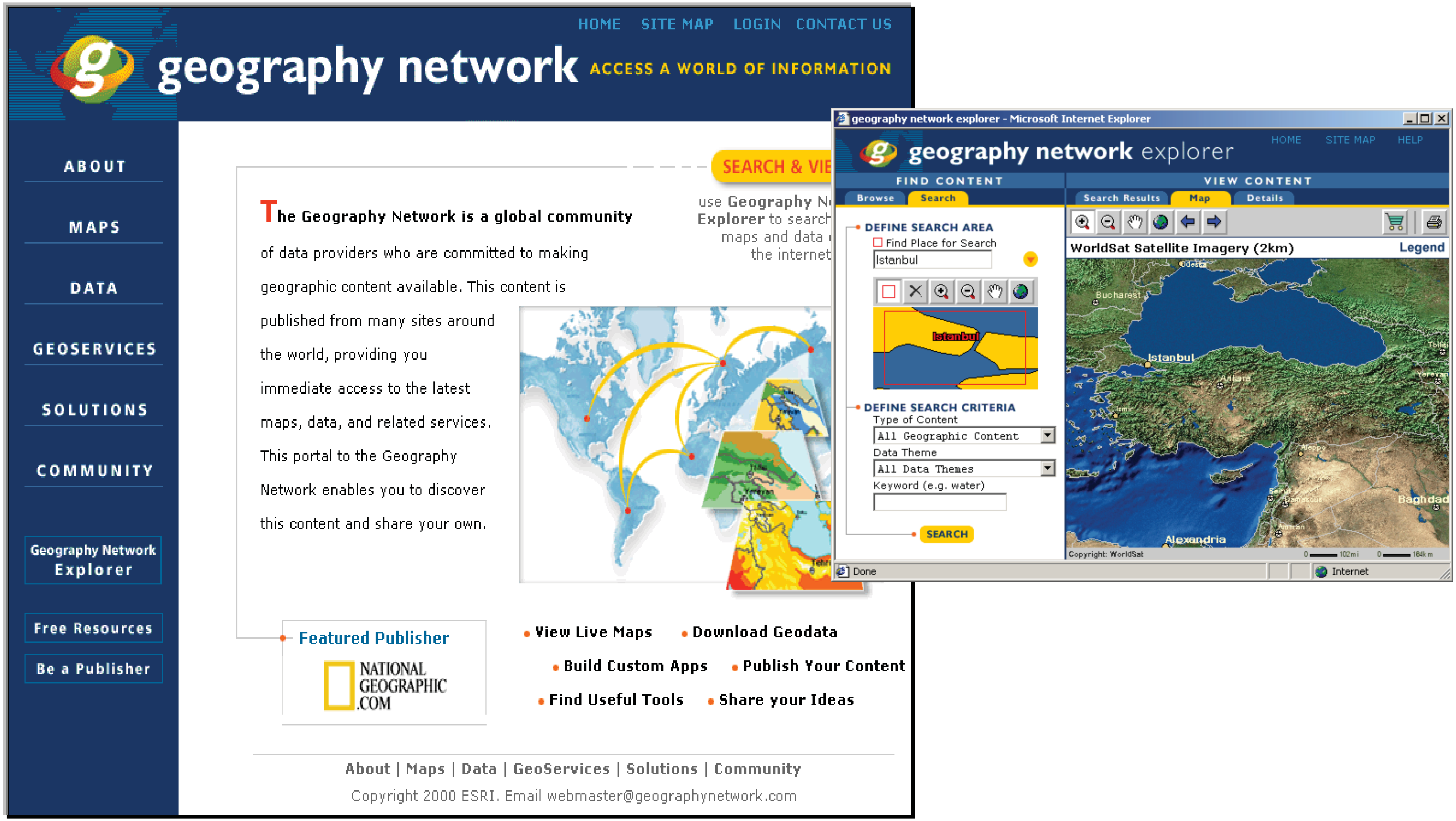


Server Metadata As Catalog

- Publish as metadata catalog
- Search via browser / desktop
- Access actual on-line data
- Easy to create and maintain
- Standards-based (ISO, CEN, Z39.50..)
- Store in database

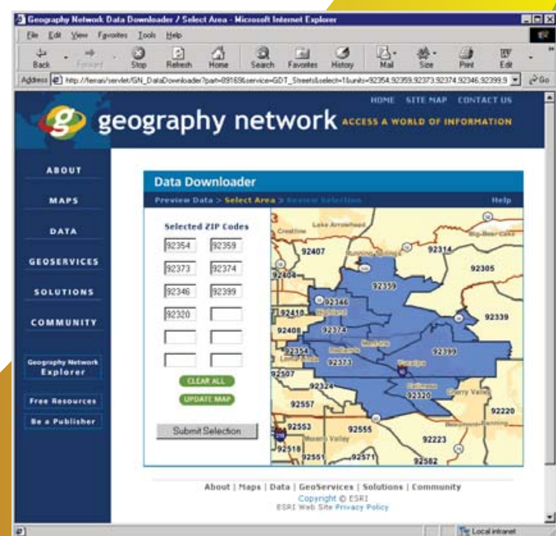


Portal - www.GeographyNetwork.com

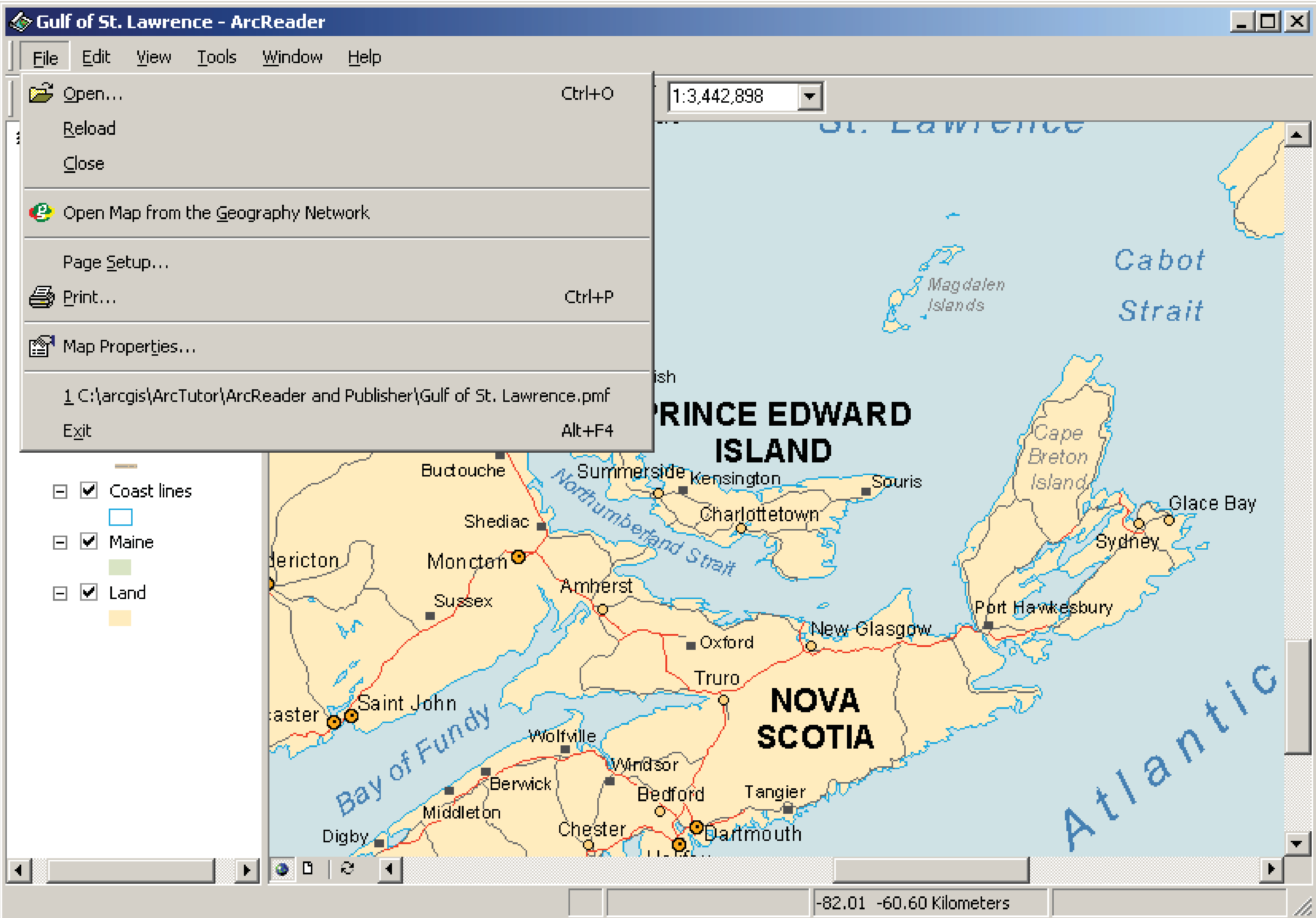


Geography Network Status

- 2,000+ Services
- 500+ Organizations
- 400,000+ Queries / day

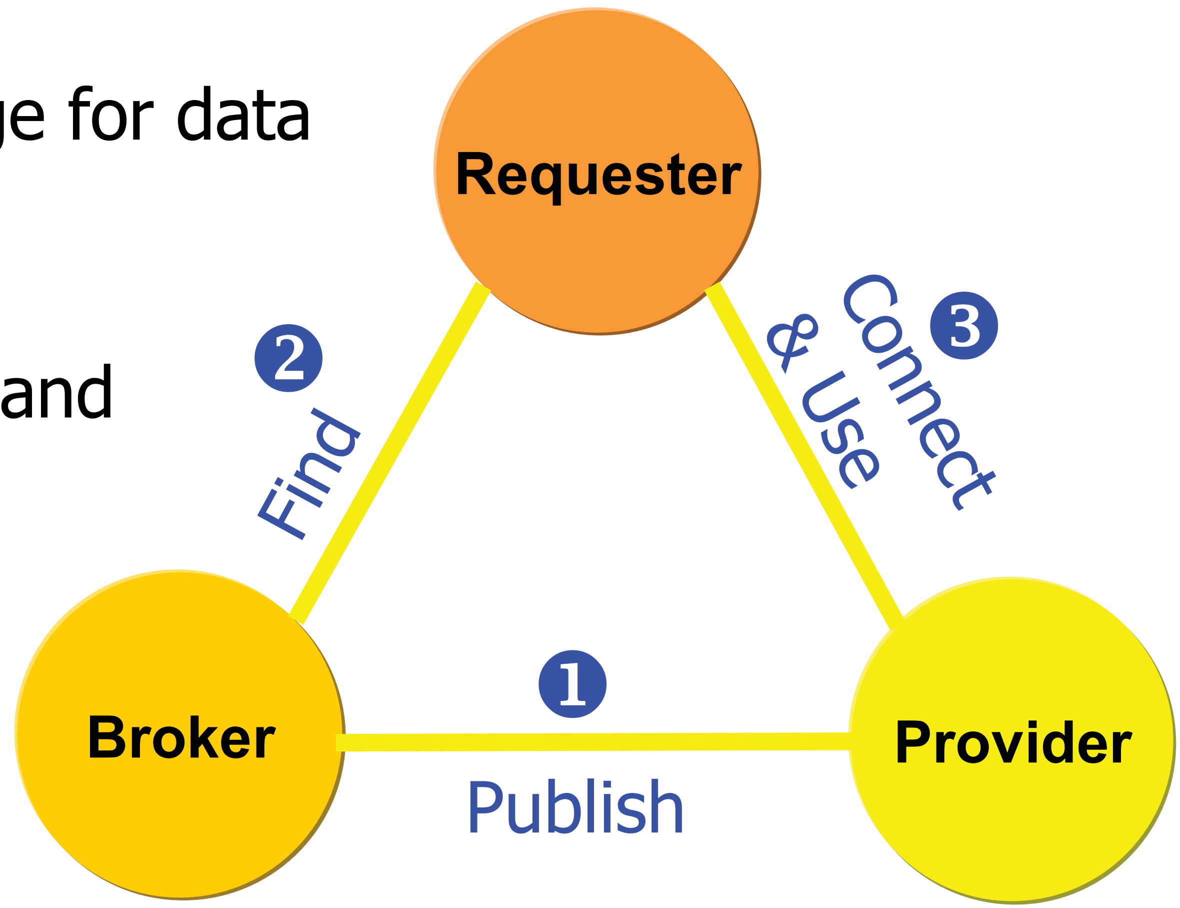


- GIS Integration
- e-commerce engine

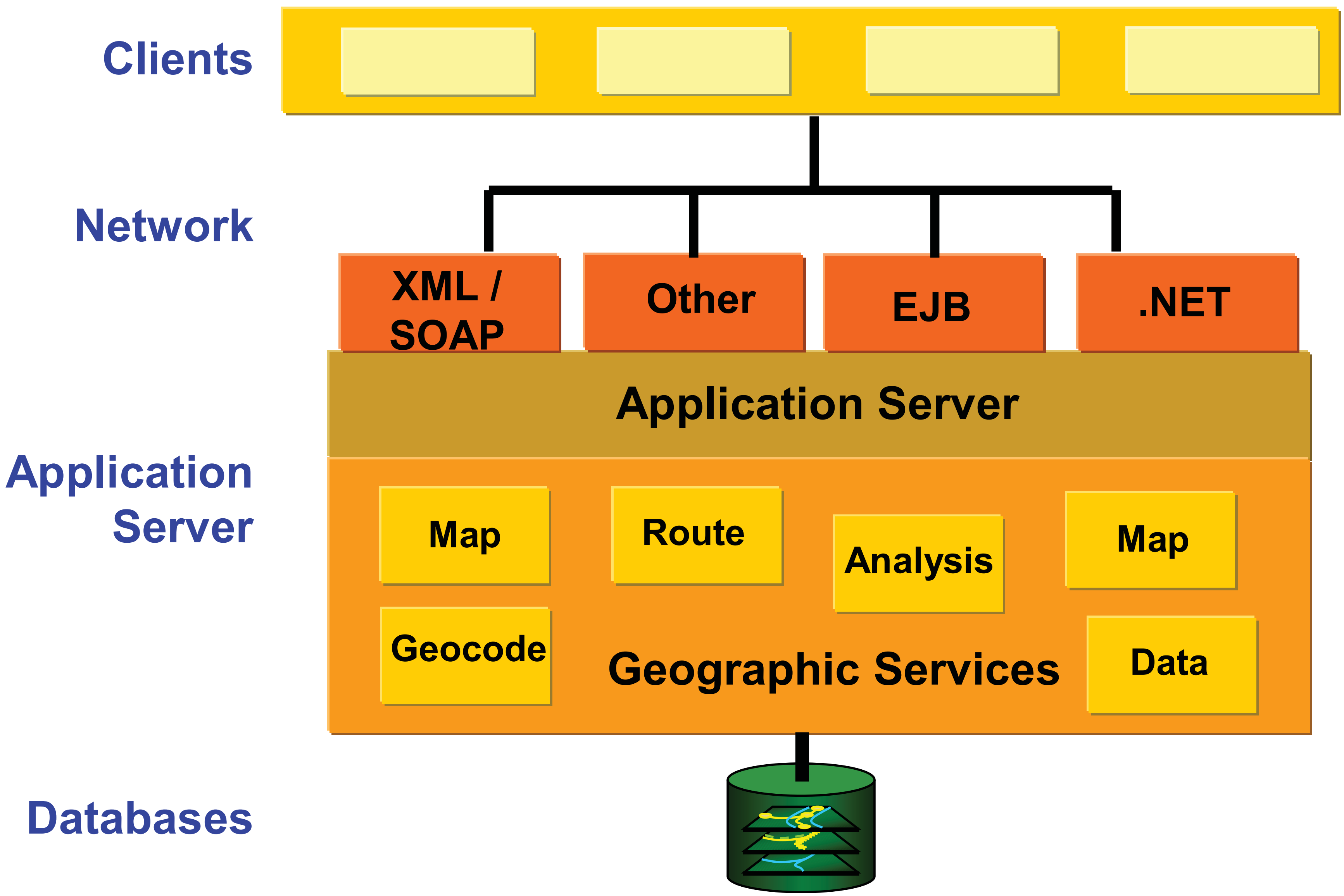

Gulf of St. Lawrence - ArcReader
File
Edit
View
Tools
Window
Help
Open... Ctrl+O
Reload
Close
Open Map from the Geography Network
Page Setup...
Print... Ctrl+P
Map Properties...
1 C:\arcgis\ArcTutor\ArcReader and Publisher\Gulf of St. Lawrence.pmf
Exit Alt+F4
1:3,442,898
Coast lines
Maine
Land
PRINCE EDWARD ISLAND
NOVA SCOTIA
Cabot Strait
Magdalen Islands
Cape Breton Island
Northumberland Strait
Bay of Fundy
Atlantic
Glace Bay
Sydney
Port Hawkesbury
Souris
Kensington
Summerside
Charlottetown
Buctouche
Shediac
Moncton
Amherst
Sussex
Saint John
New Glasgow
Oxford
Truro
Wolfville
Windsor
Bedford
Tangier
Berwick
Middleton
Digby
Chester
Dartmouth
-82.01 -60.60 Kilometers

Web Services

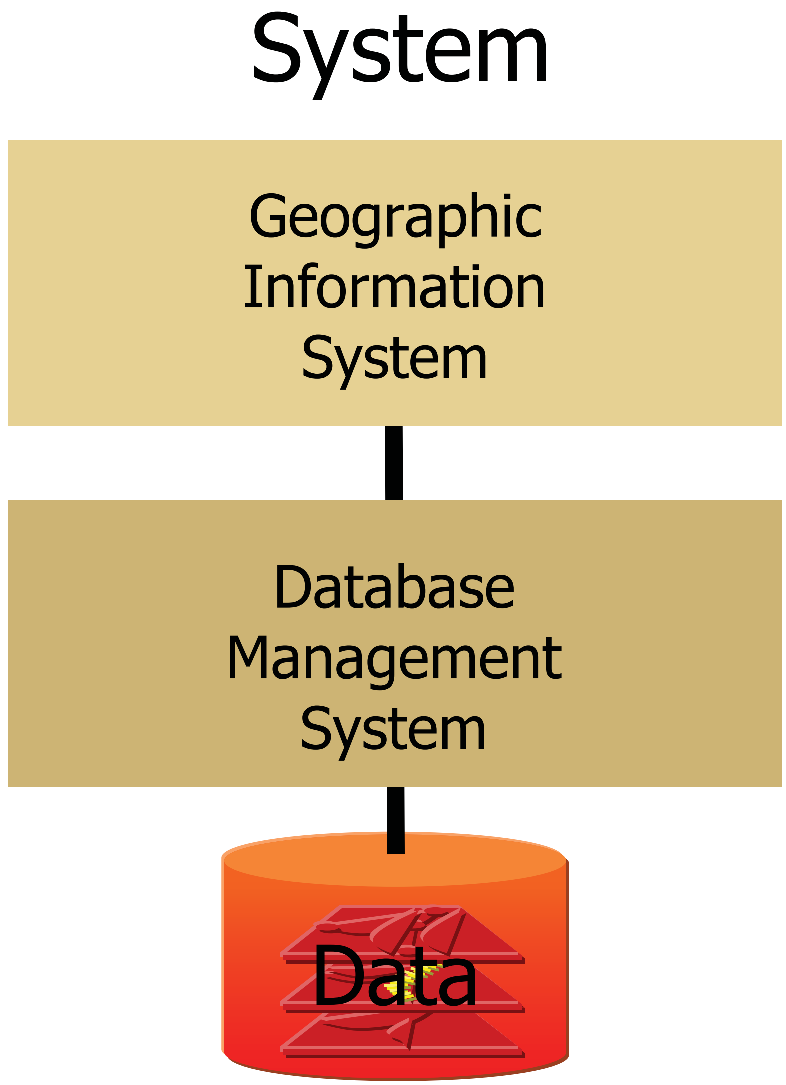
- XML
  - A common language for data and messages
- UDDI and WSDL
  - Mechanism to find and plug into services
- SOAP
  - XML protocol to control and use services



GIS as a Service



Role of DBMS



Task

- Data load
- Editing
- Visualization
- Mapping
- Analysis

- Storage
- Indexing
- Security
- Query

Thank You!

Cadastral needs and developments

Gerhard Muggenhuber, Austria

Email: geomugg@surfEU.at

International Affairs at
Federal Office of
Metrology and Surveying (BEV)

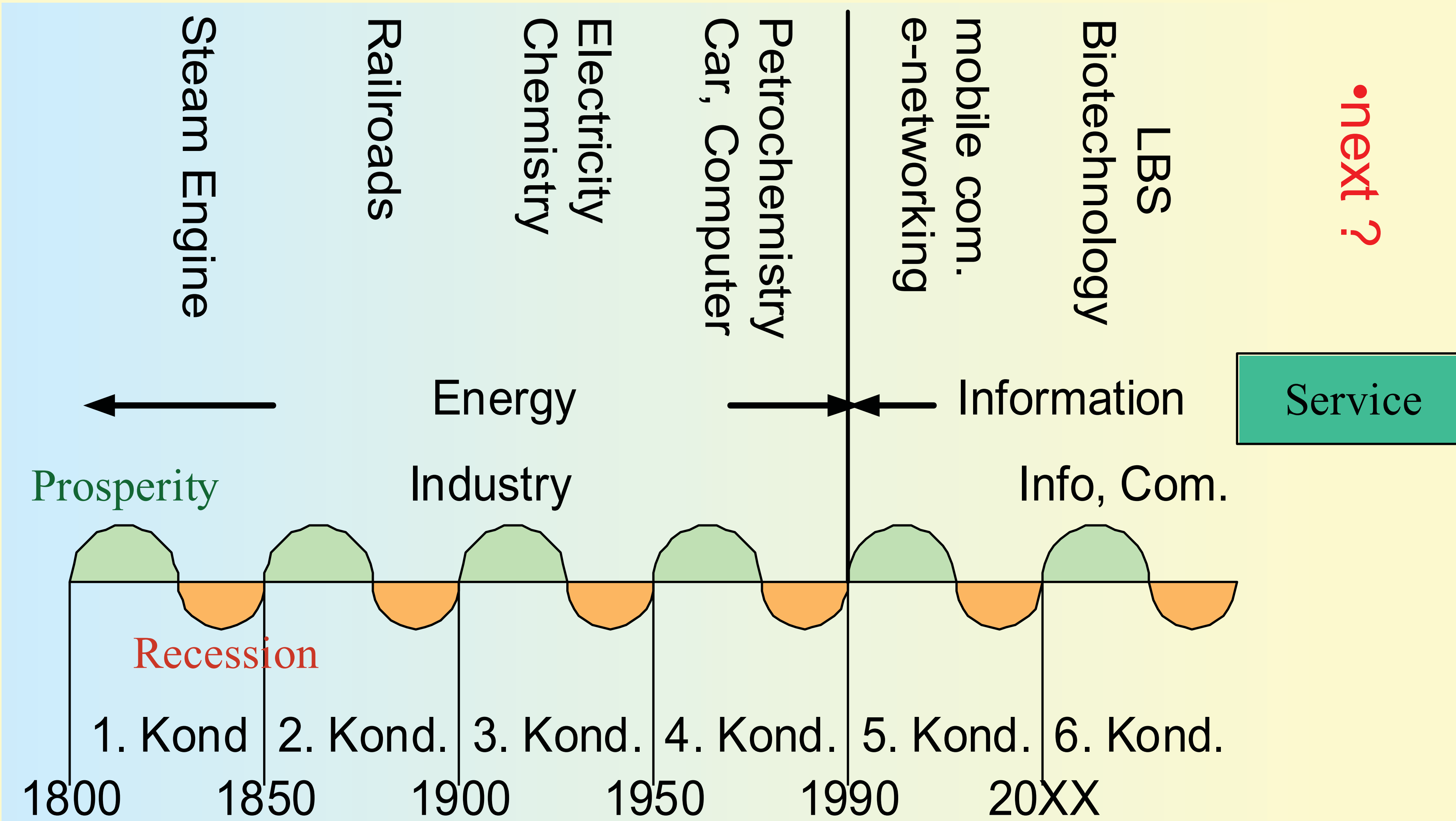
www.bev.gv.at

FIG-COM3
Spatial Information Management

http://fig3.boku.ac.at

Cadastre in a Service Society

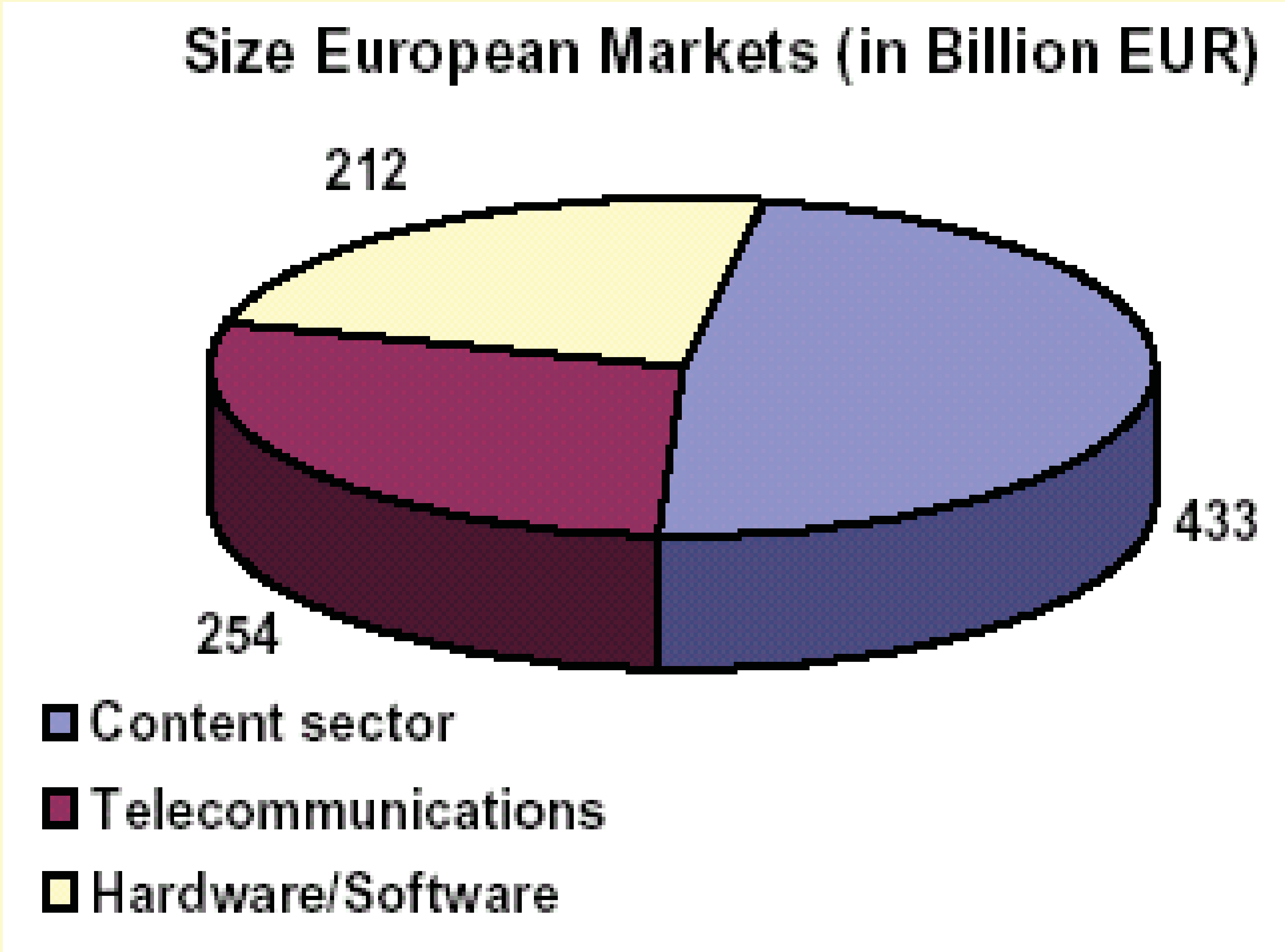
Cadastral concepts based on demand of Agricultural Society



Nach Nefiodow, L.A. (1997): Der 6. Kondratieff; Wege zur Produktivität und Vollbeschäftigung im Zeitalter der Information

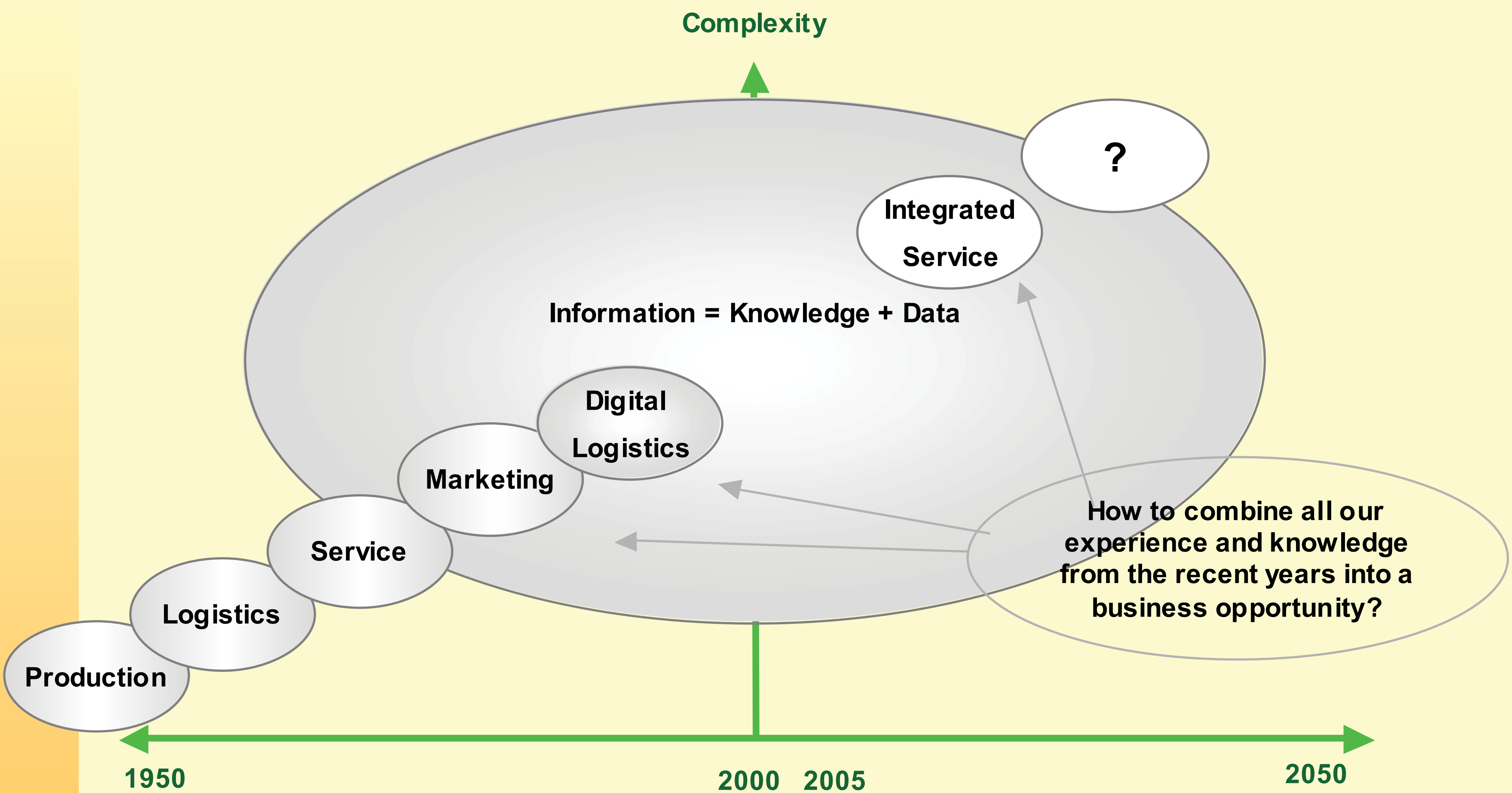
Increasing Information Market

eEurope 2002: Exploitation of Public Sector Information

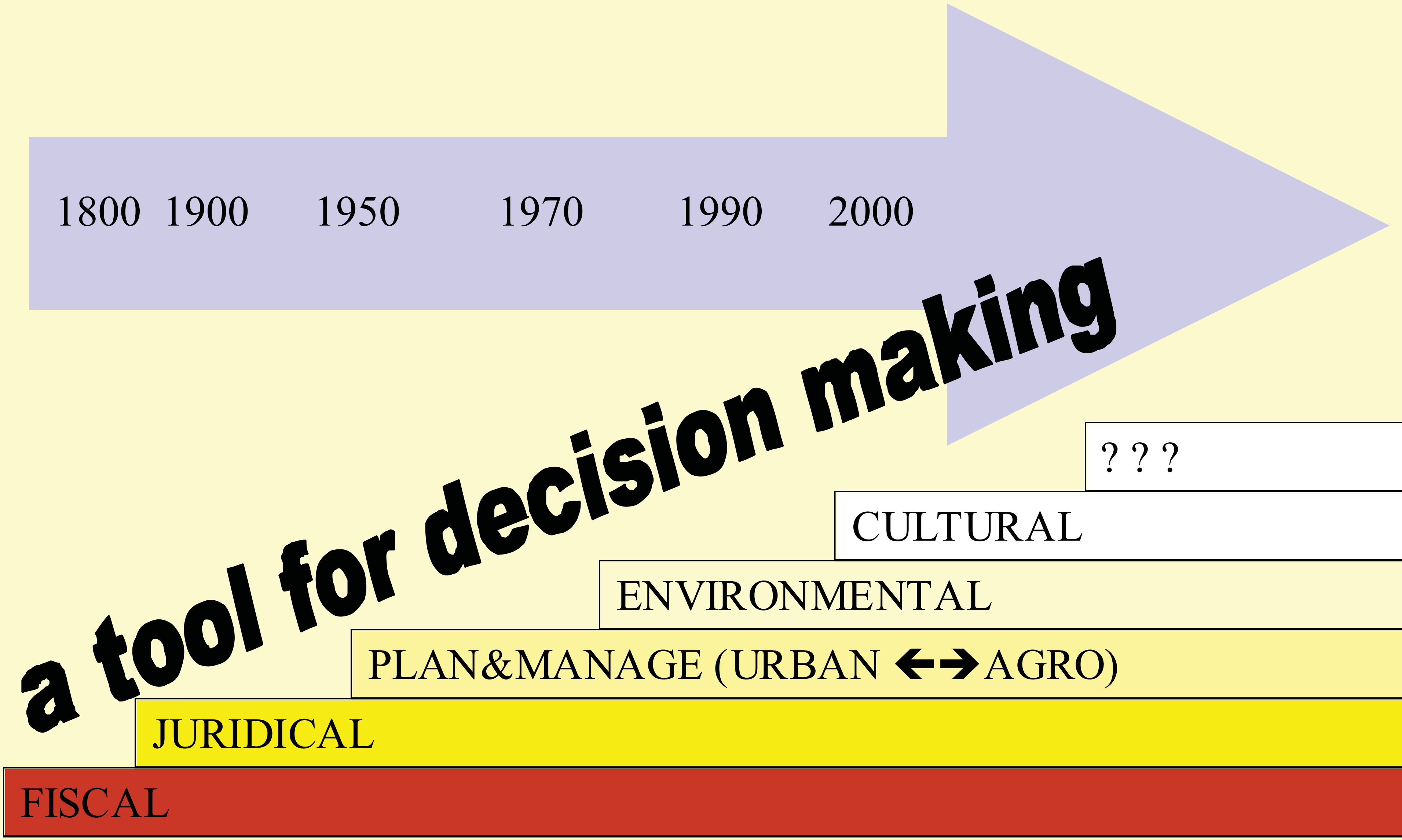


US the information market = 5 times of EU market.
Growth rate of 10 to 30 % in the last 6 years.
(EC-Study by Pira International, 2000).

Compete with the changing framework



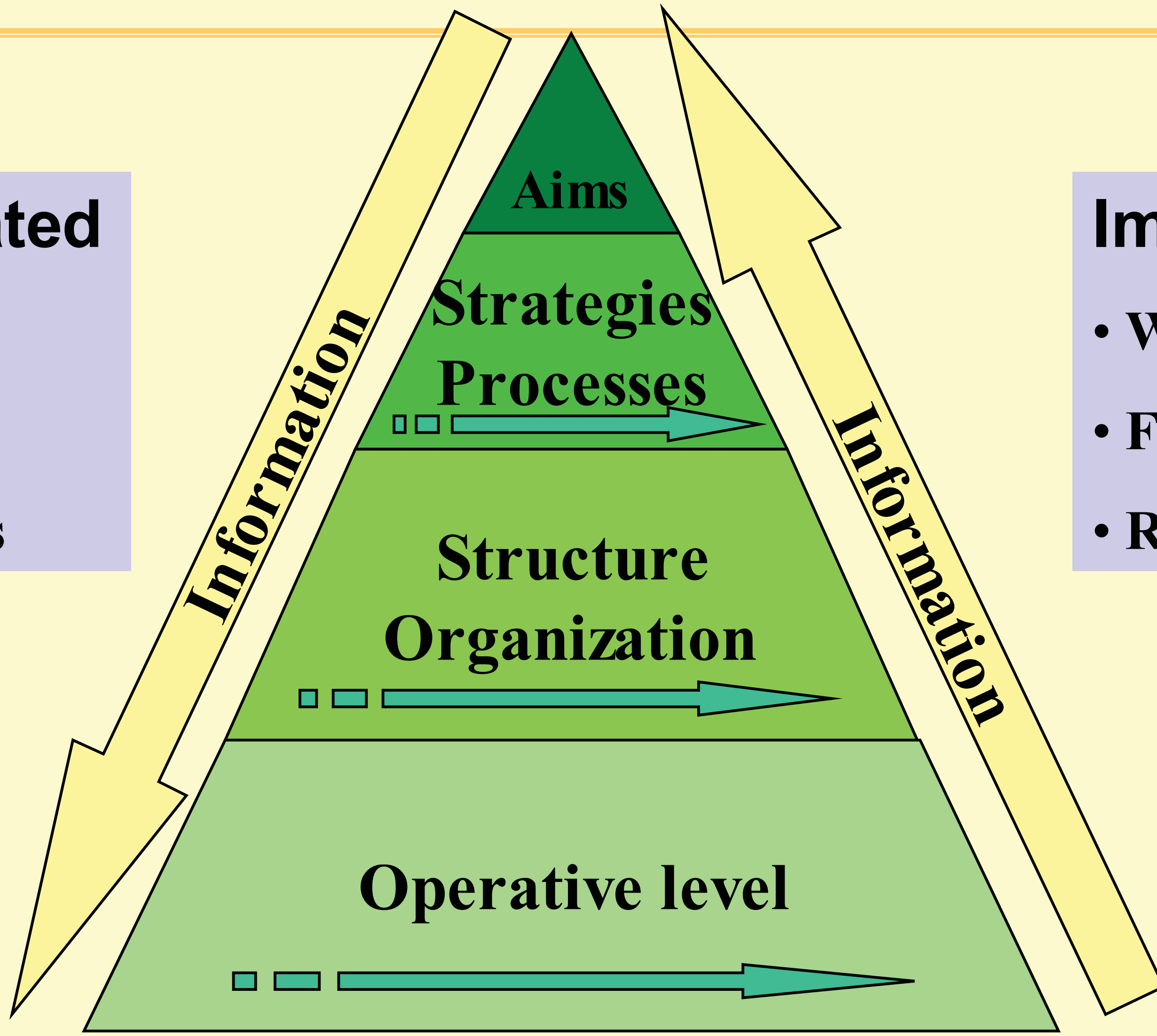
Integration of Services on Cadastre



Improving Cadastre

Coordinated

- Visions
- Aims
- Operations

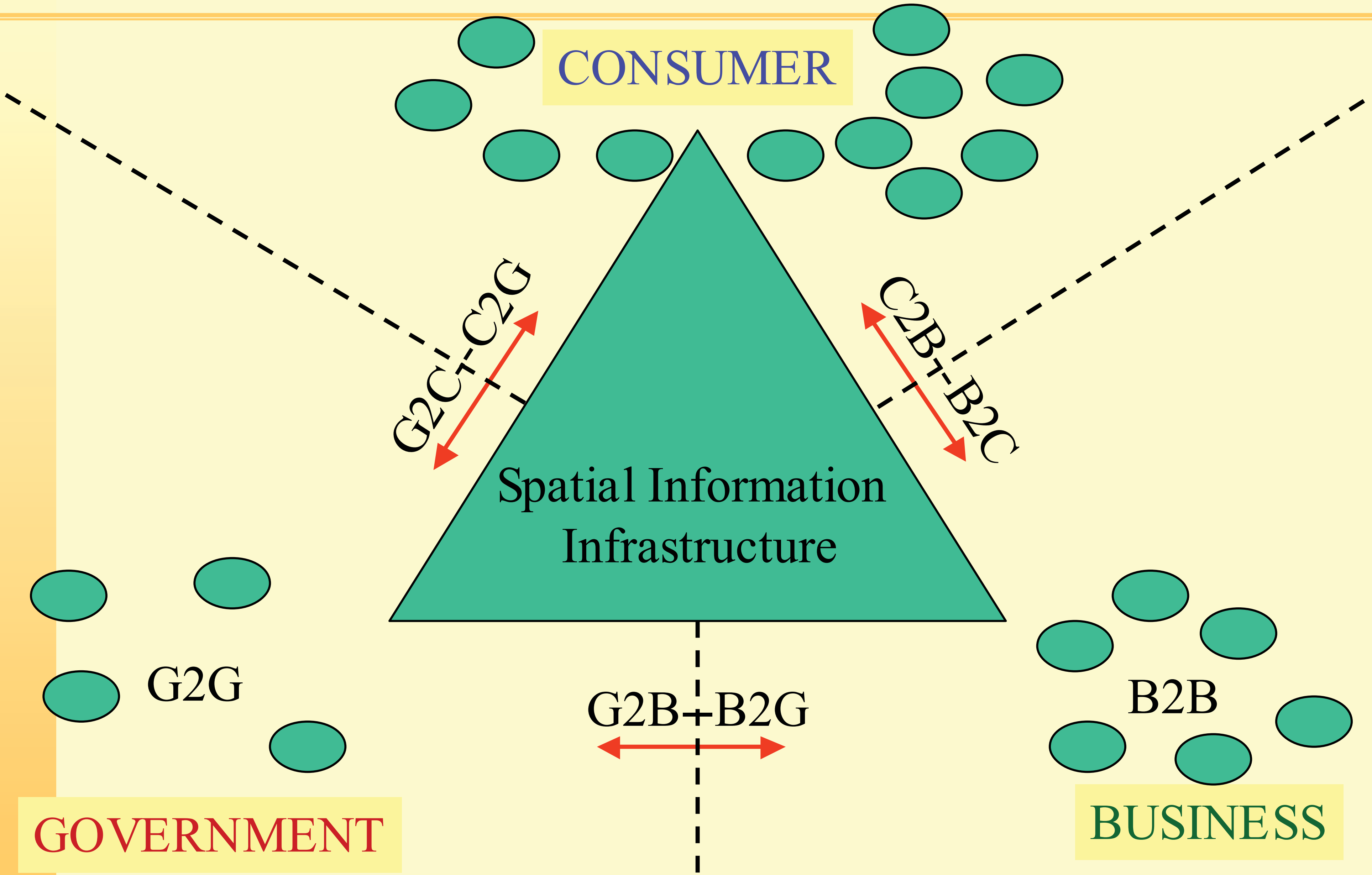


Improved

- Workflow
- Feedback
- Results

Shared meta-information for improved services

Cadastre as part of a Information Infrastructure



Land Administration (LA) and Geospatial Data

Land Administration in Europe

- Parameters
- Organizations involved
- Customers

Austria:

- BEV-Organization
- Spatial data structure
- Cadastre
- Landscape Information
- History

Users of Geospatial Data

European projects: IACS Customers ⇔ NMA

EULIS, INSPIRE , SABE

Global Geospatial data: GSDI, GMES, CIA, UN- Cartographic section

Key question within J.Delincé, JRC paper:

"Field Identification System in Agriculture":

How to get Mapping Agencies interested in the user needs?

Coordinated Geoinformation in Europe

European Council of Geodetic Surveyors (CLGE)

Geometer Europas (GE)

European Umbrella Org.for Geographic Info.(EUROGI)

EuroGeographics (CERCO+MEGRIN)

UN-ECE Working Party on Land Administration

urban data management society

Parameters for EU-Market

Product Market

Capital Market

europa

Service Market

Labour Market

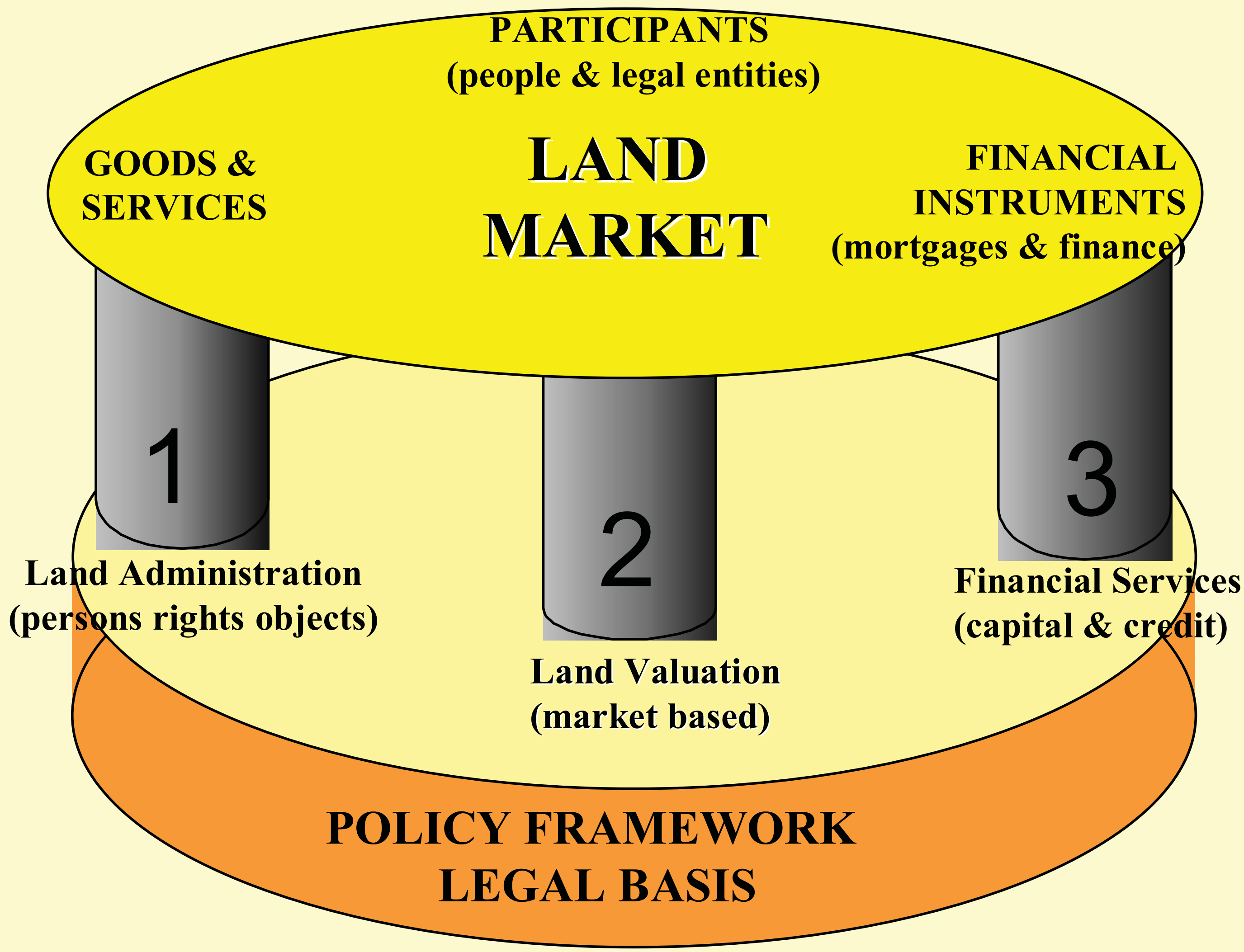
Land Market

Phare

OBNOVA

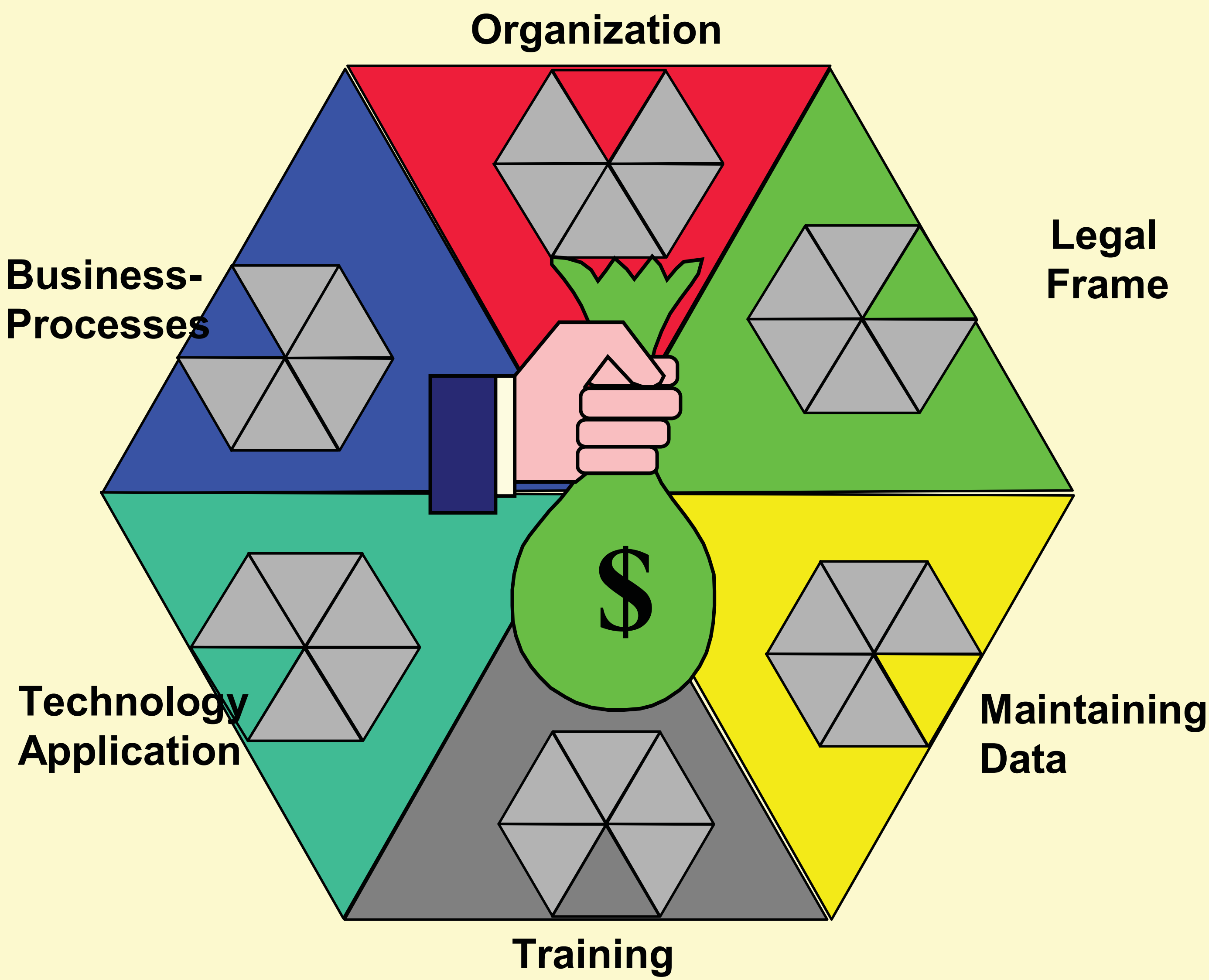
Tacis

Parameters for a Land Market



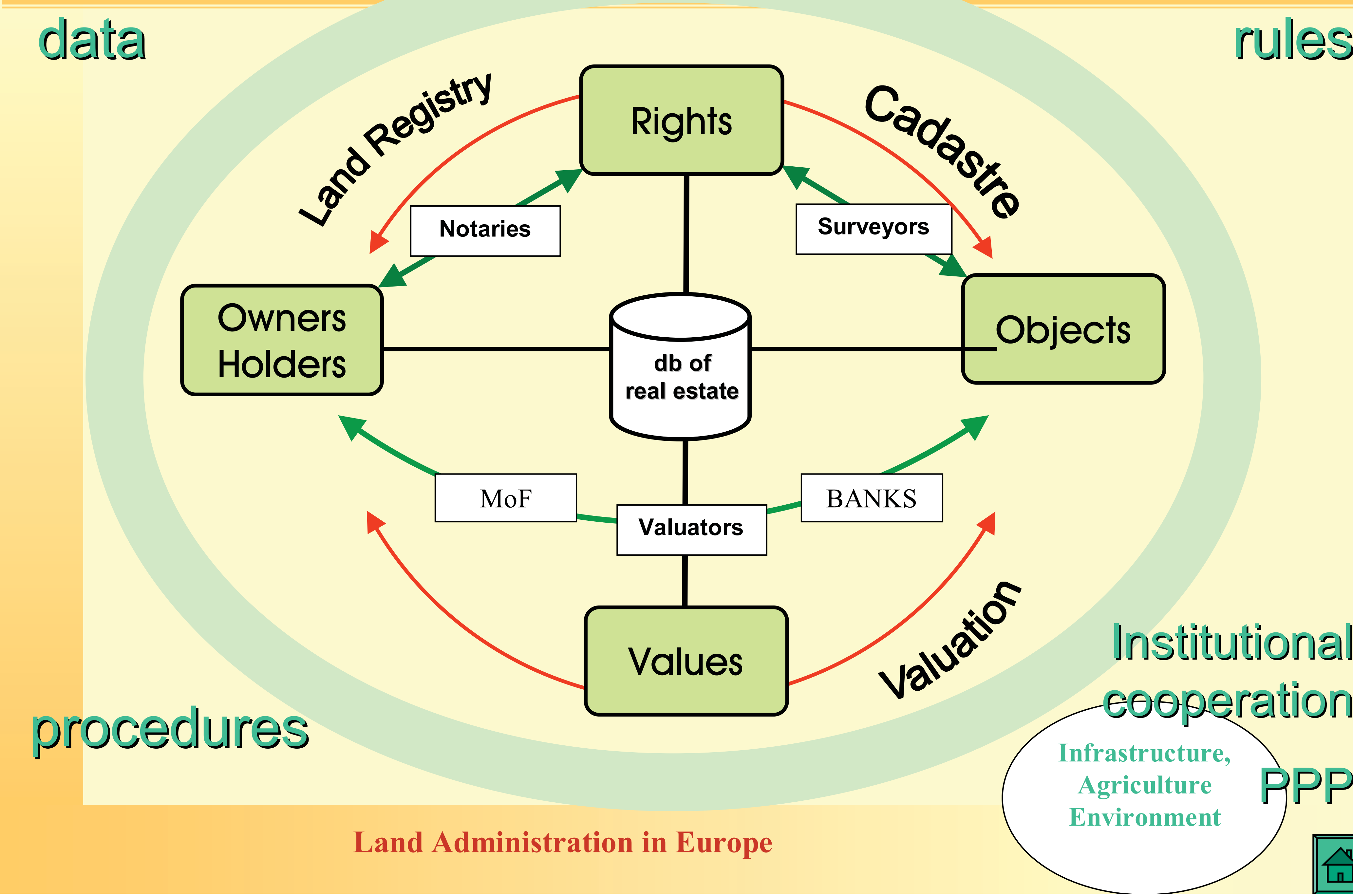
Parameters for Land Administration

Technology, legal-, economic frame, user orientation, skills, interdisciplinary cooperation

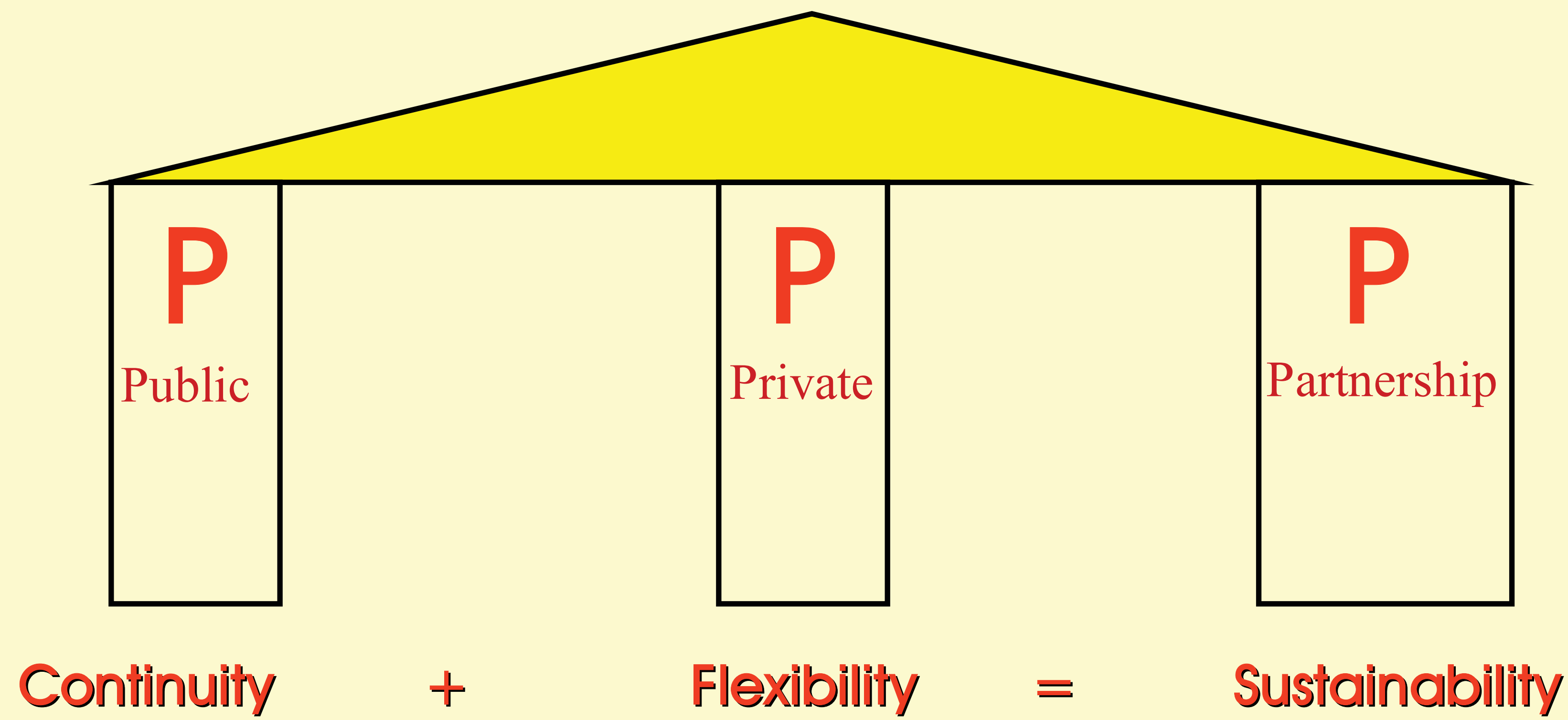


Land Administration:

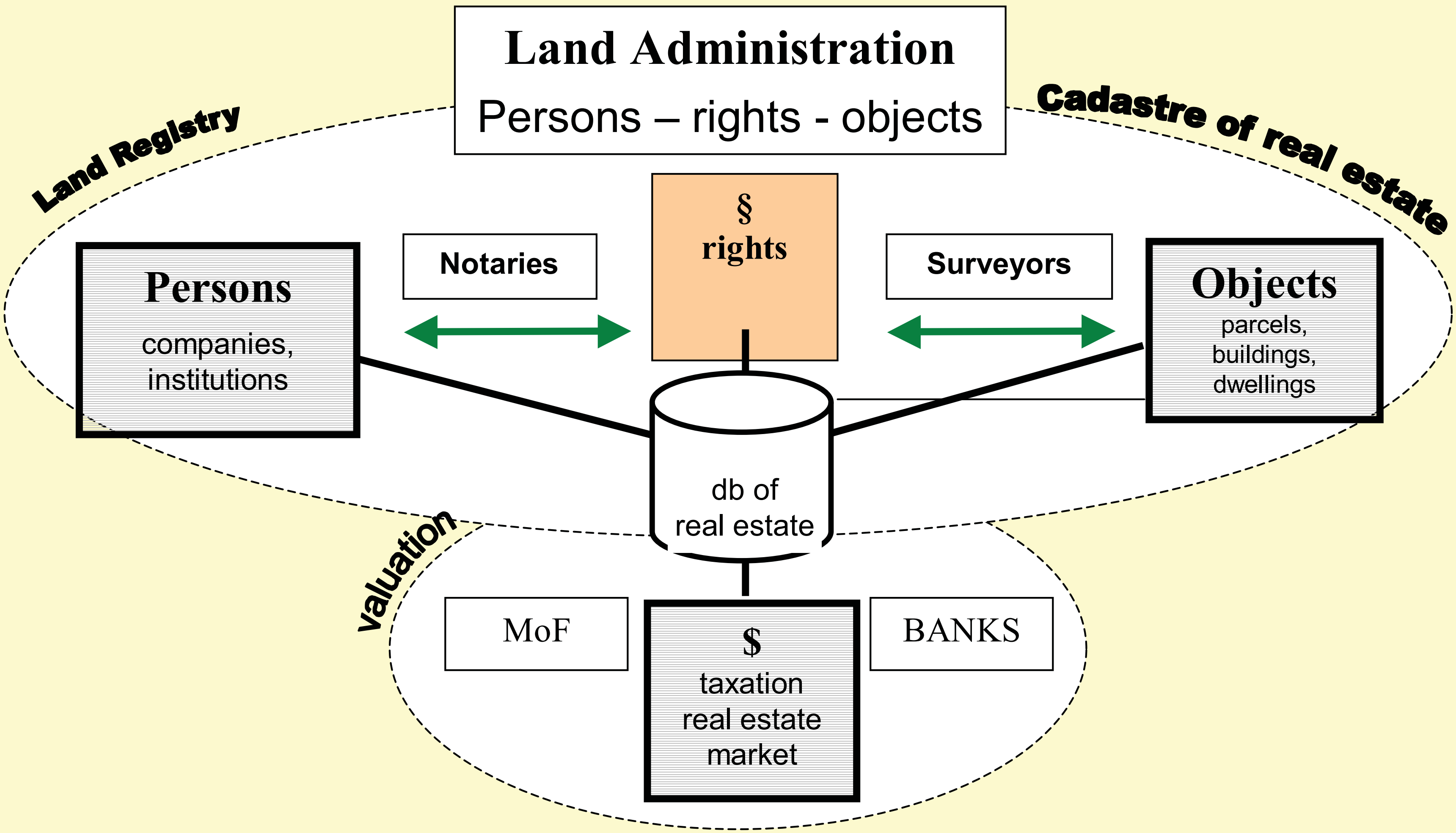
persons – rights – objects- value



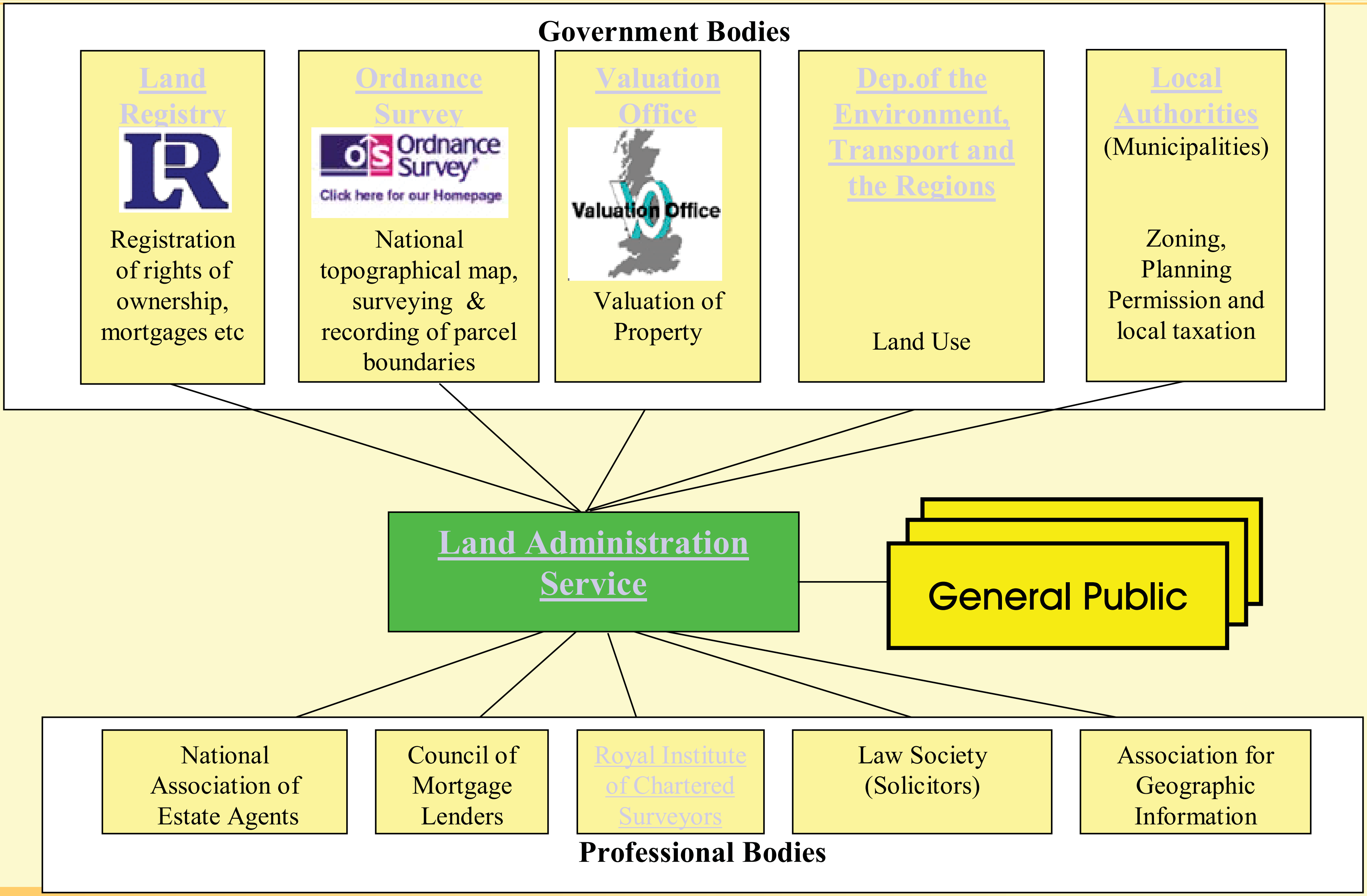
Interdisciplinary co-operation in Land Administration



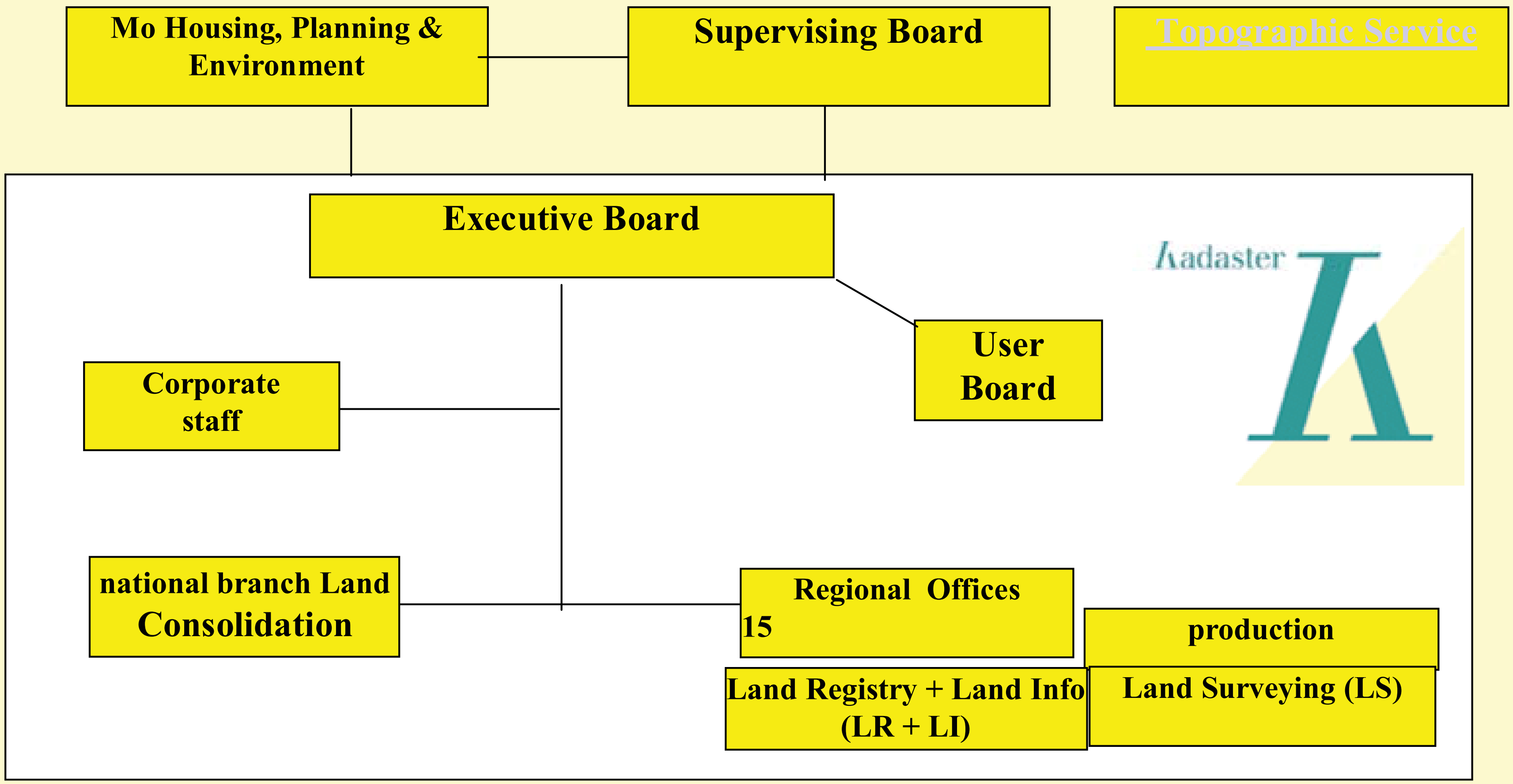
Land Administration



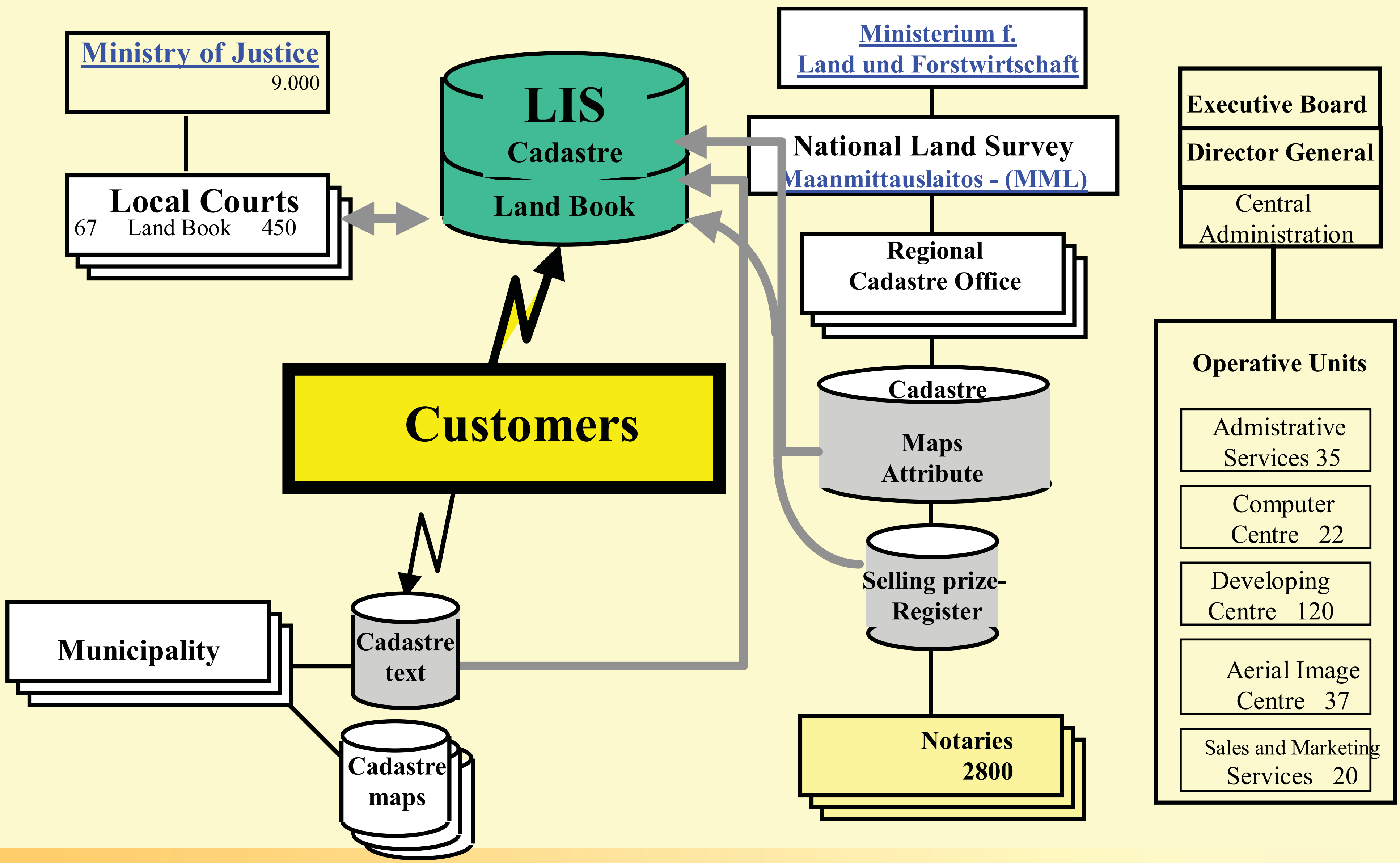
Organisation in UK



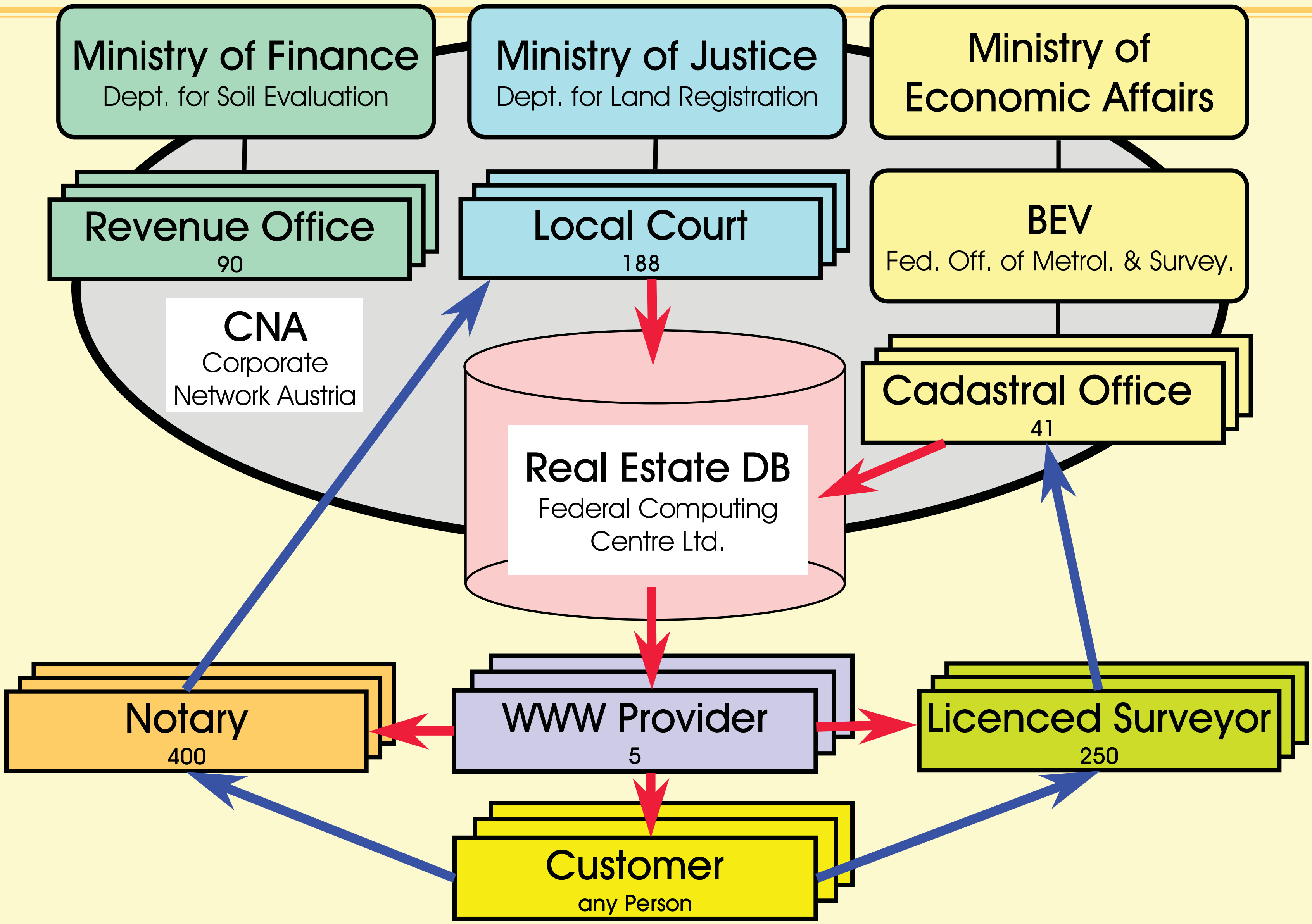
Organisation in NL



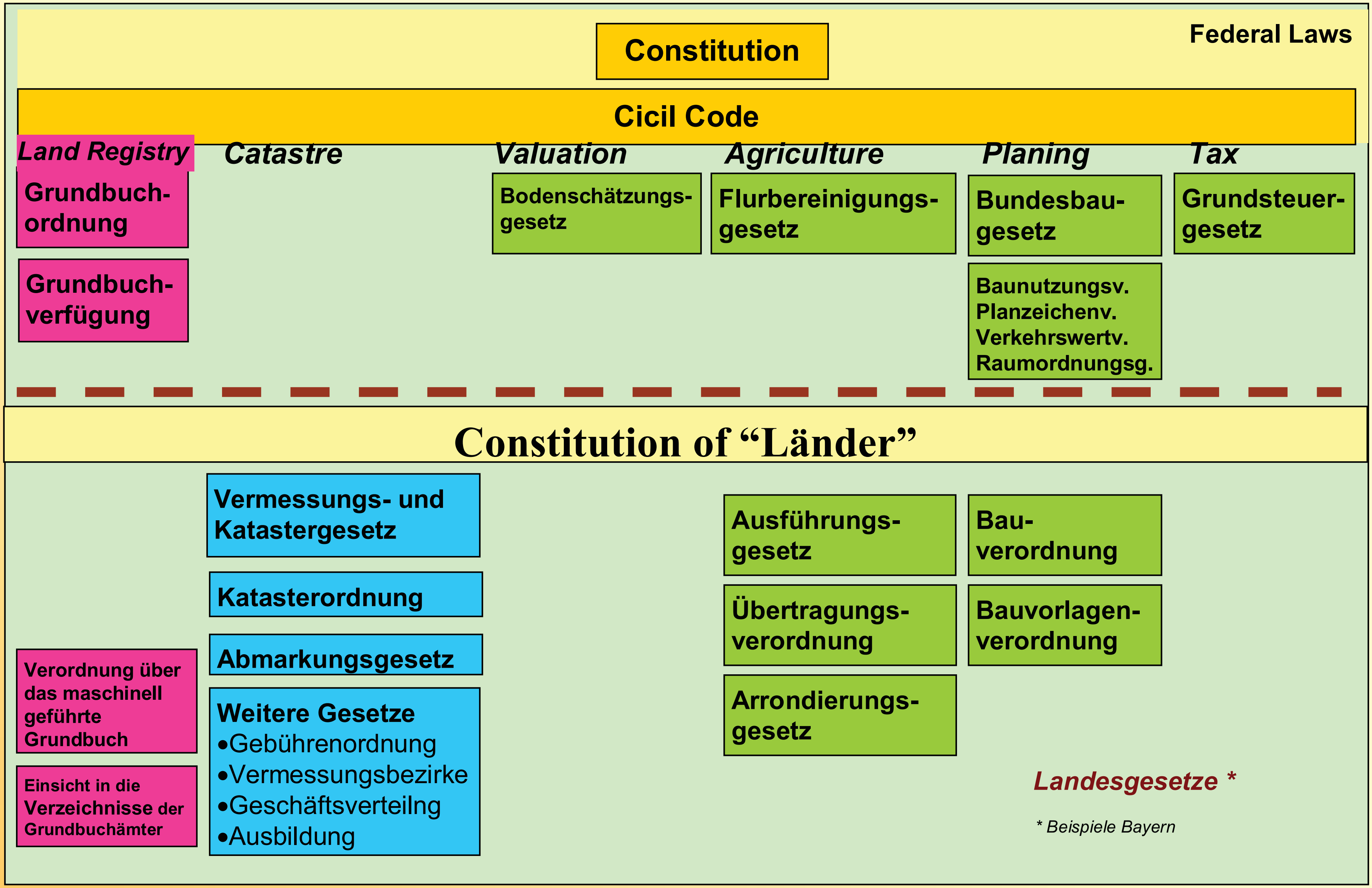
Organisation in Finland



Co-operation of private & public Sector



D: Legal Frame for Real Property



Coordinated Geoinformation in Europe

European Council of Geodetic Surveyors **(CLGE)**

Geometer Europas **(GE)**

European Umbrella Org.for Geographic Info.**(EUROGI)**

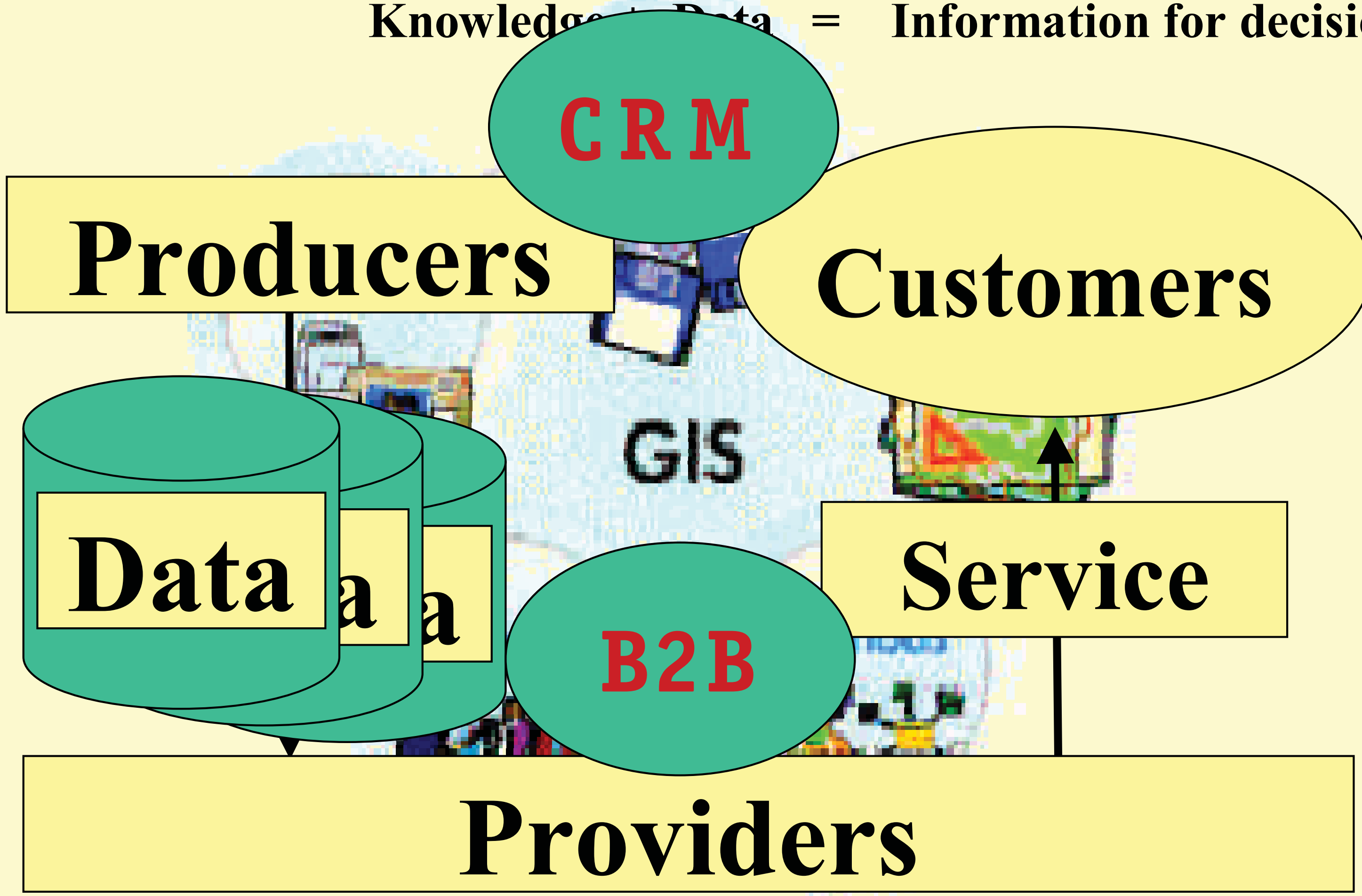
EuroGeographics **(CERCO+MEGRIN)**

UN-ECE Working Party on Land Administration

Access to Information

K+D=I

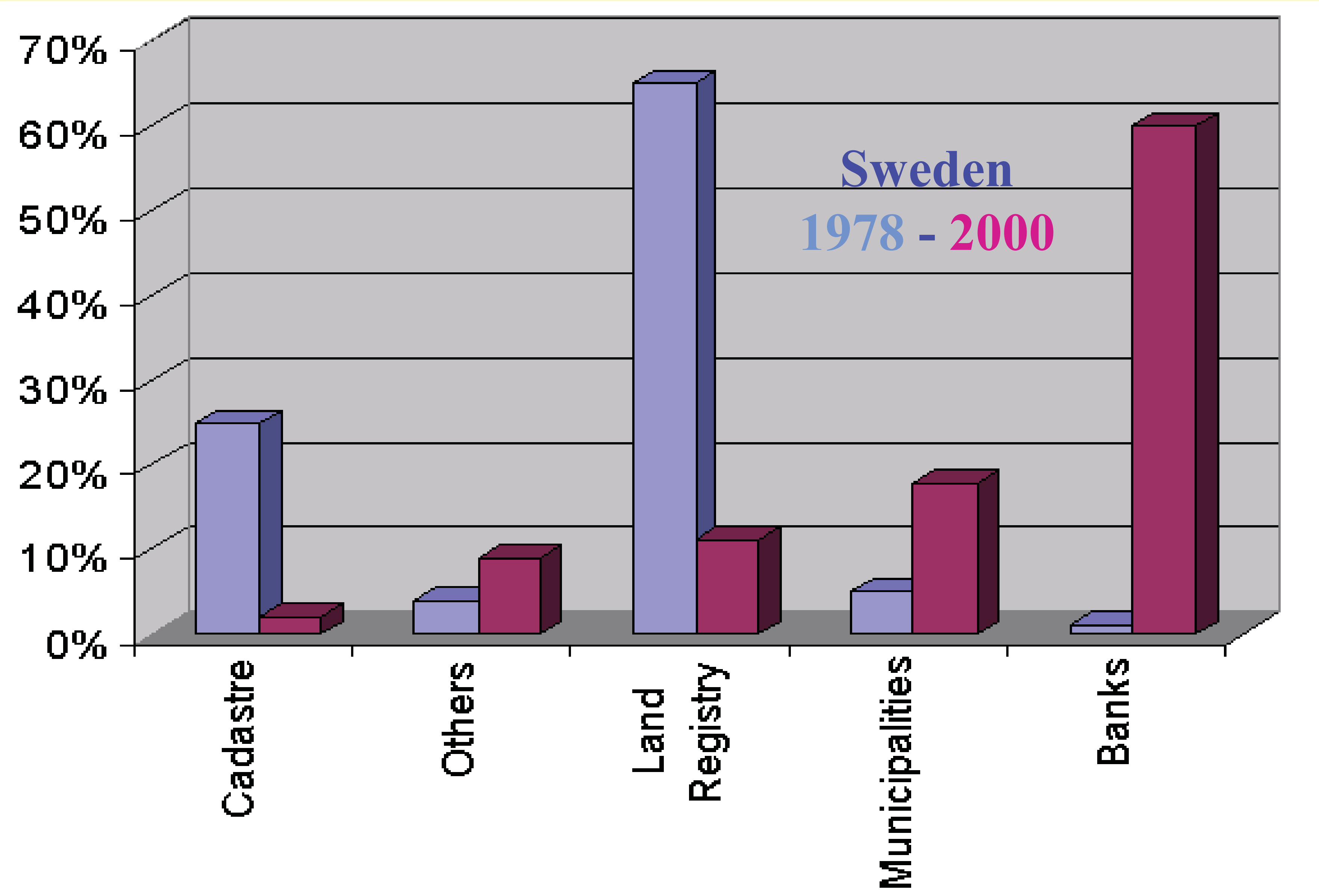
Knowledge + Data = Information for decision making



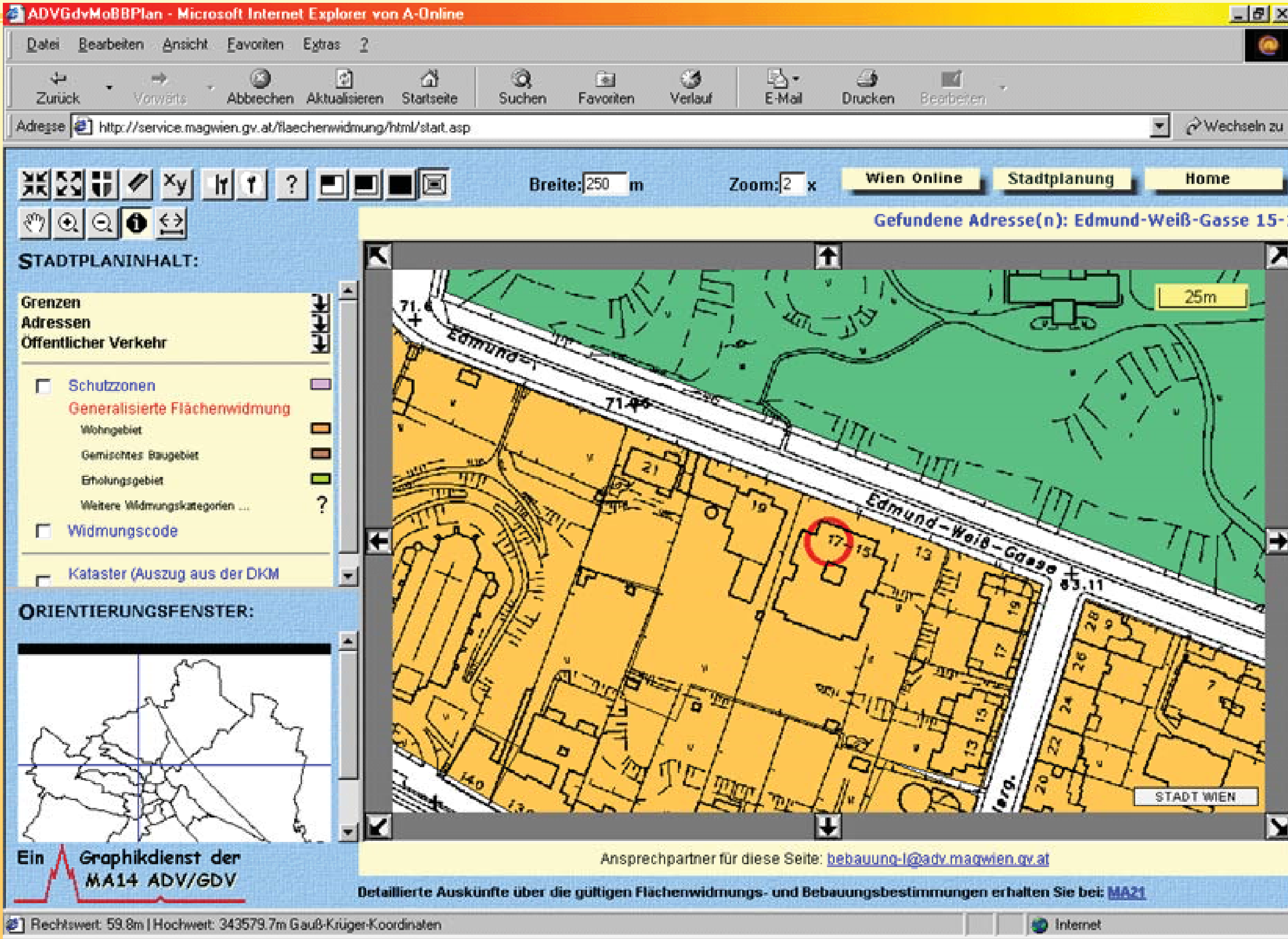
M-Commerce

Geo-Commerce

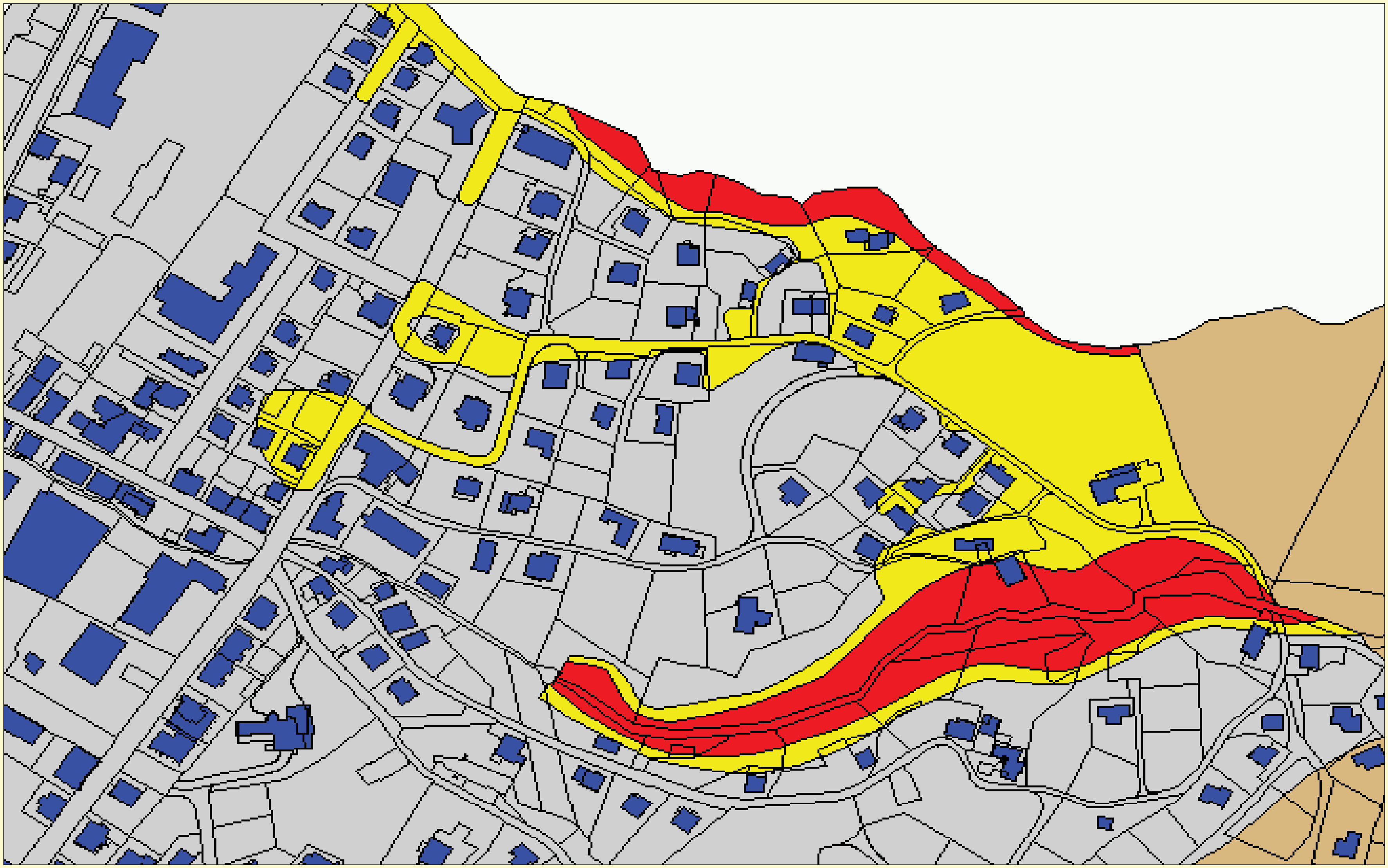
Online customers



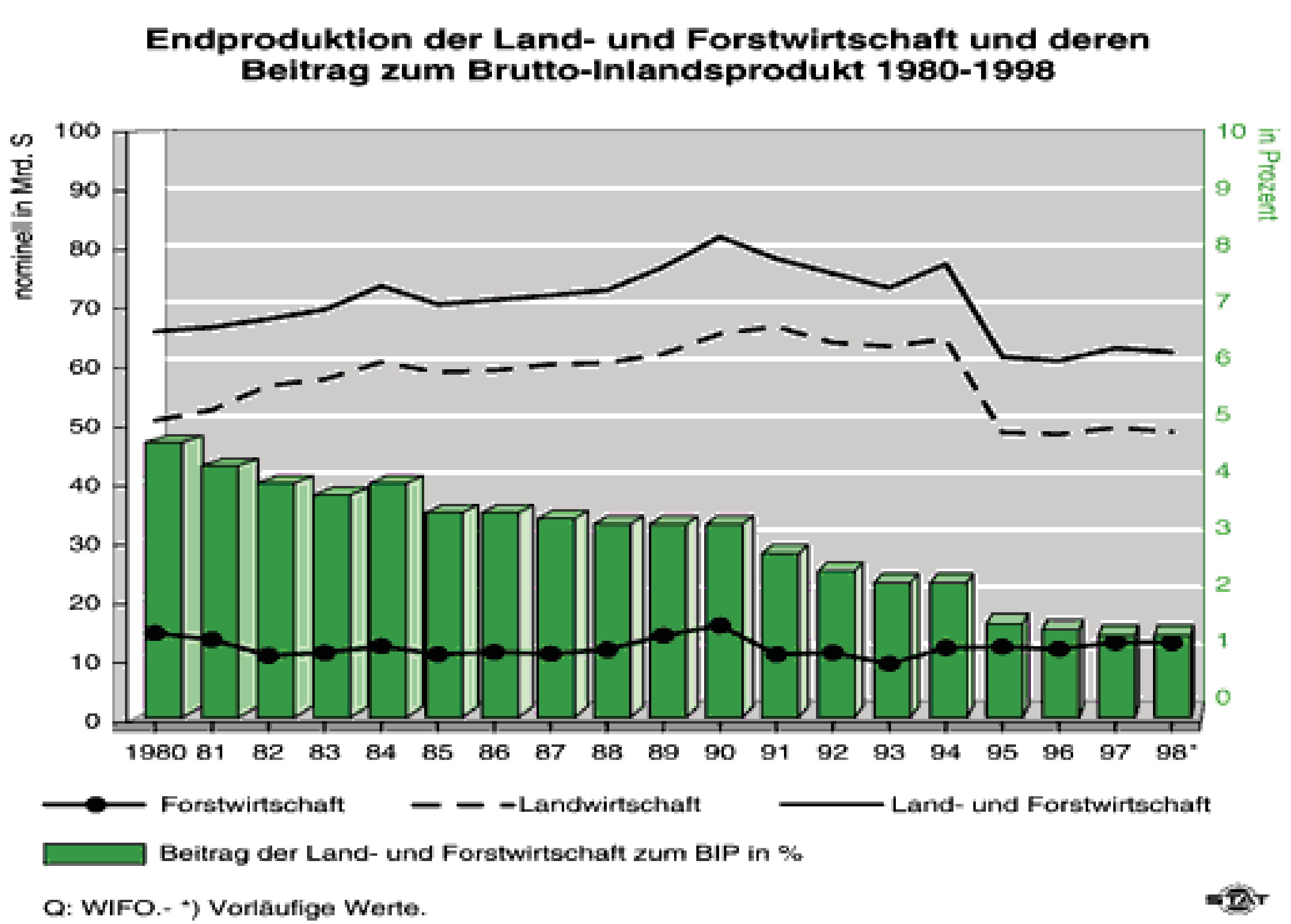
LA and General Planning

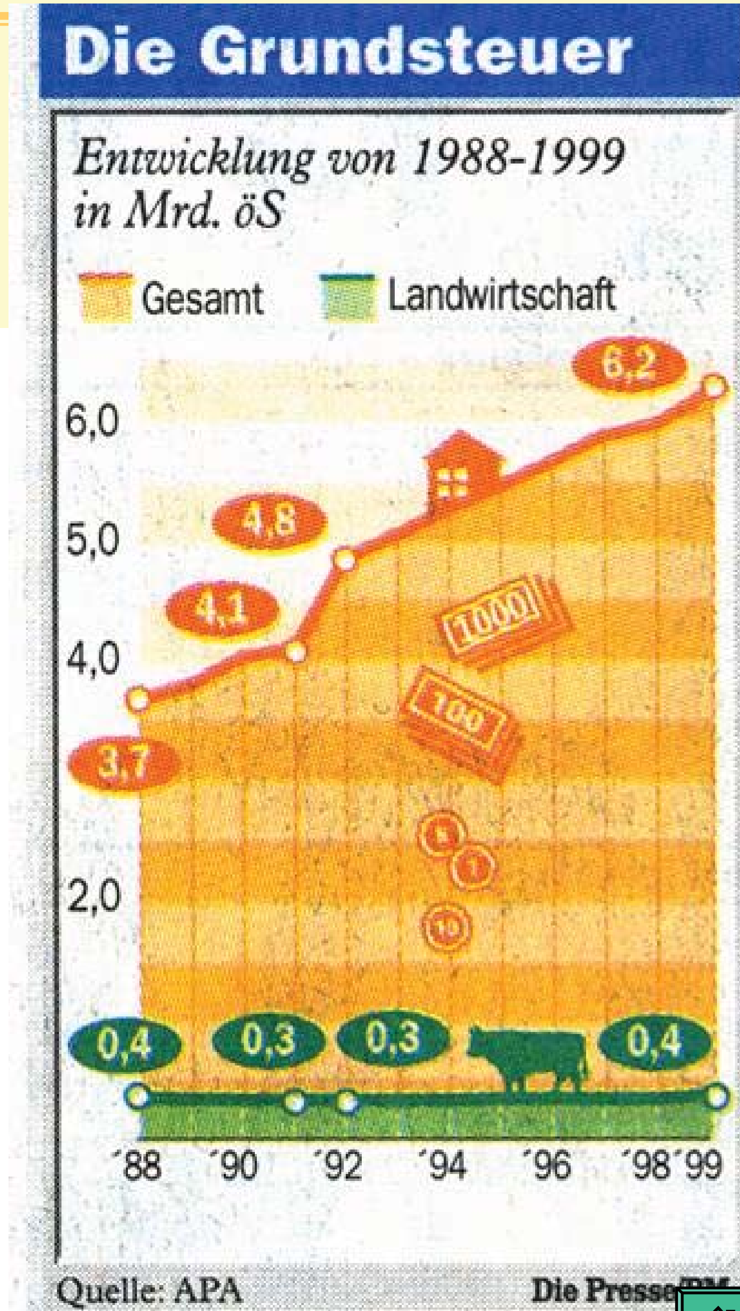


LA and Risk Management

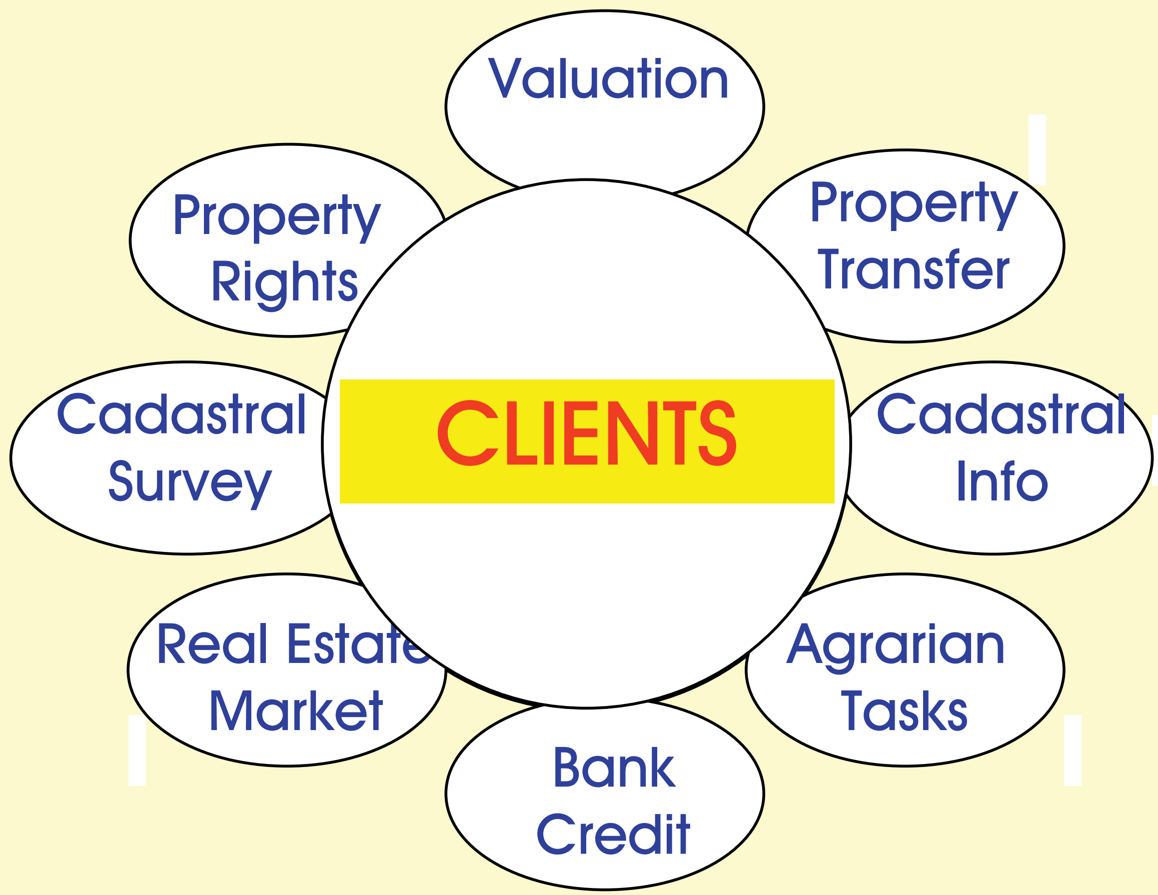


LA and Real Estate Tax





Applicant → Clients



PROVIDER + CONSUMER = PARTNER

solutions must fit to the consumers (citizens) need!

CONSUMERS

Service_1 | Service_n | Product_1 | Product_n

Spatial Information* Infrastructure

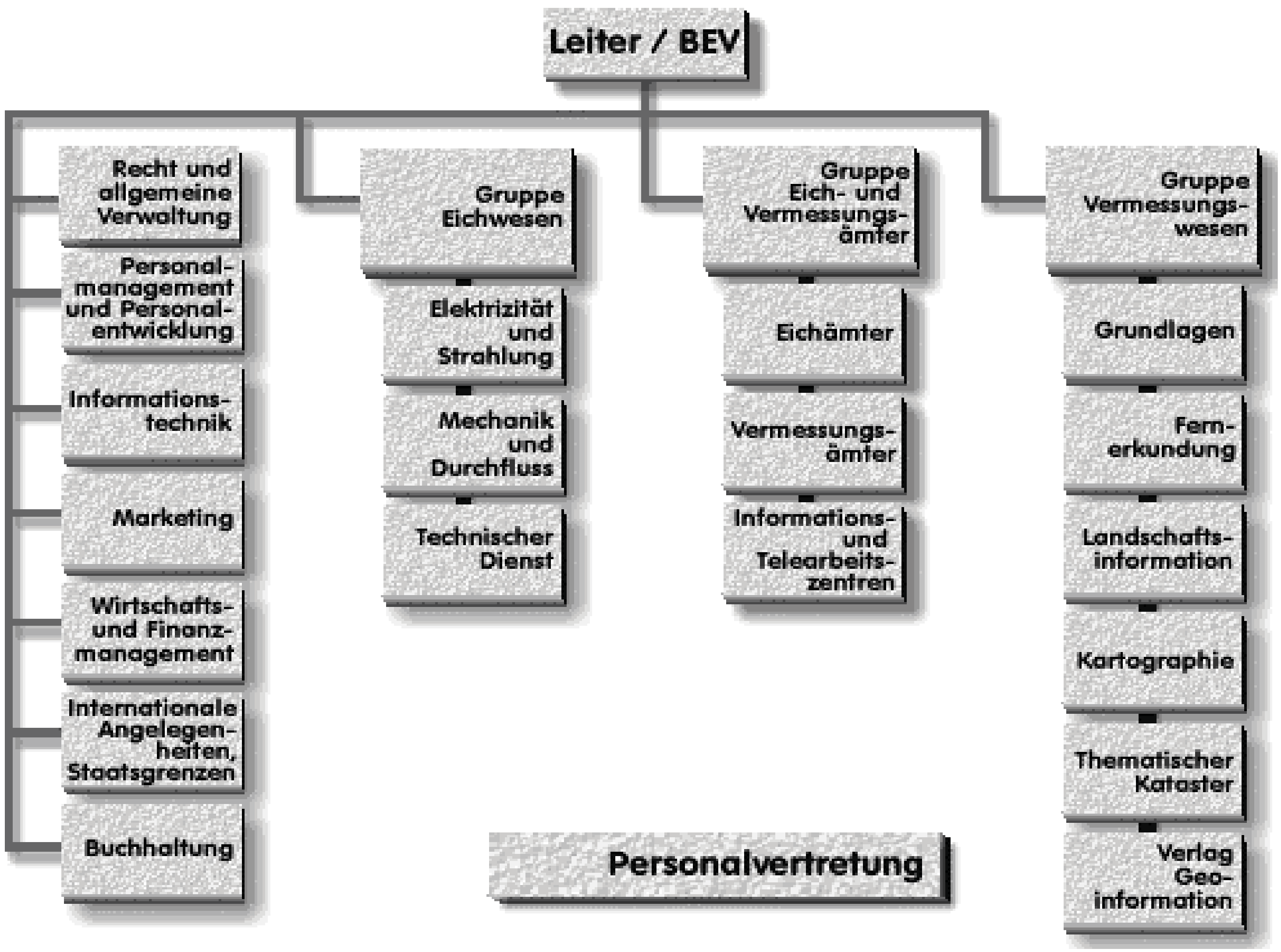
PROVIDERS

- Multidisciplinary approach
- Quality Management

Main Tasks of BEV

- Geodesy and control survey
- Providing and information delivery of base data for surveying and geoinformation
- Administration of the cadastre of real estate
- Production of the topographic maps of Austria of different scales
- Aerial photography for surveying including an image archive
- Fundamental and legal metrology
- National standards for the legal units of measurement

Structure of BEV

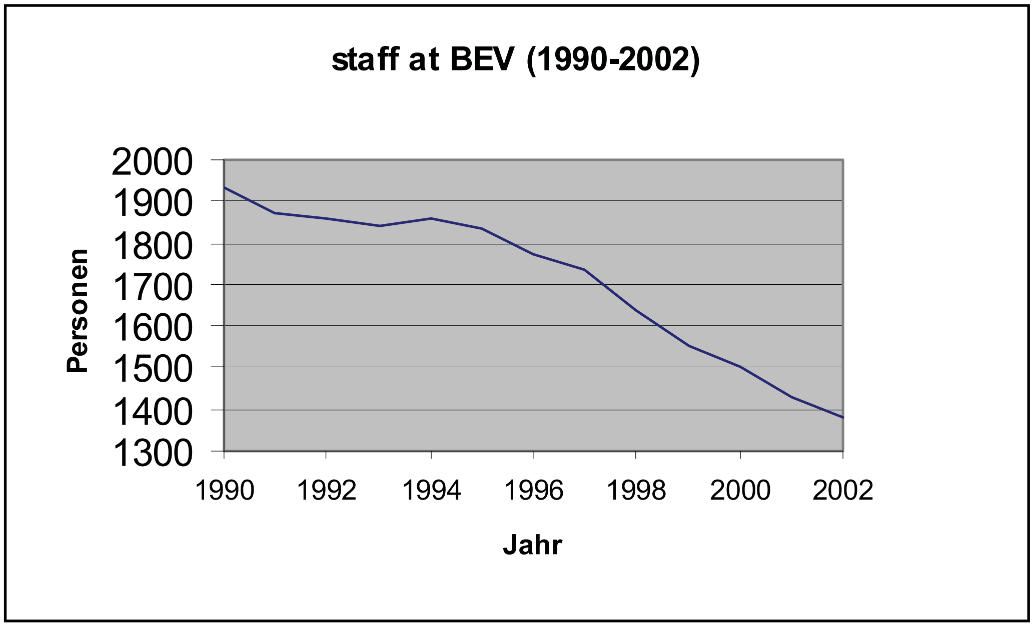


Budget, Personal Staff

- 1430 employees (2001)
- Annual budget:
  - 66 mill. € (surveying, cadastre, metrology)
  - 80 %. personal costs
  - 15 % permanent costs
  - 5 % investment
- Cadastre:
  - 730 employees in cadastre
  - 35 mill. € in cadastre
  - Cost proceeds: 11 %

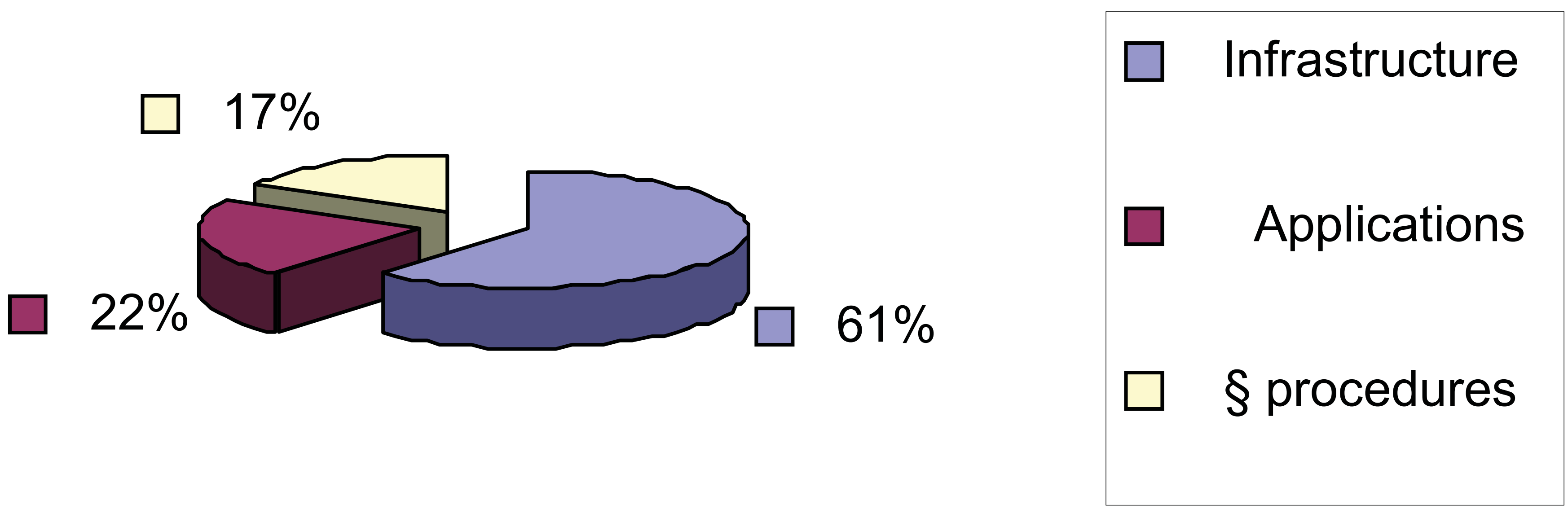
Challenges for the BEV

- Reduce Costs
- Reduce Staff
- Reduce tasks to kernel tasks
- Restructuring of organisation

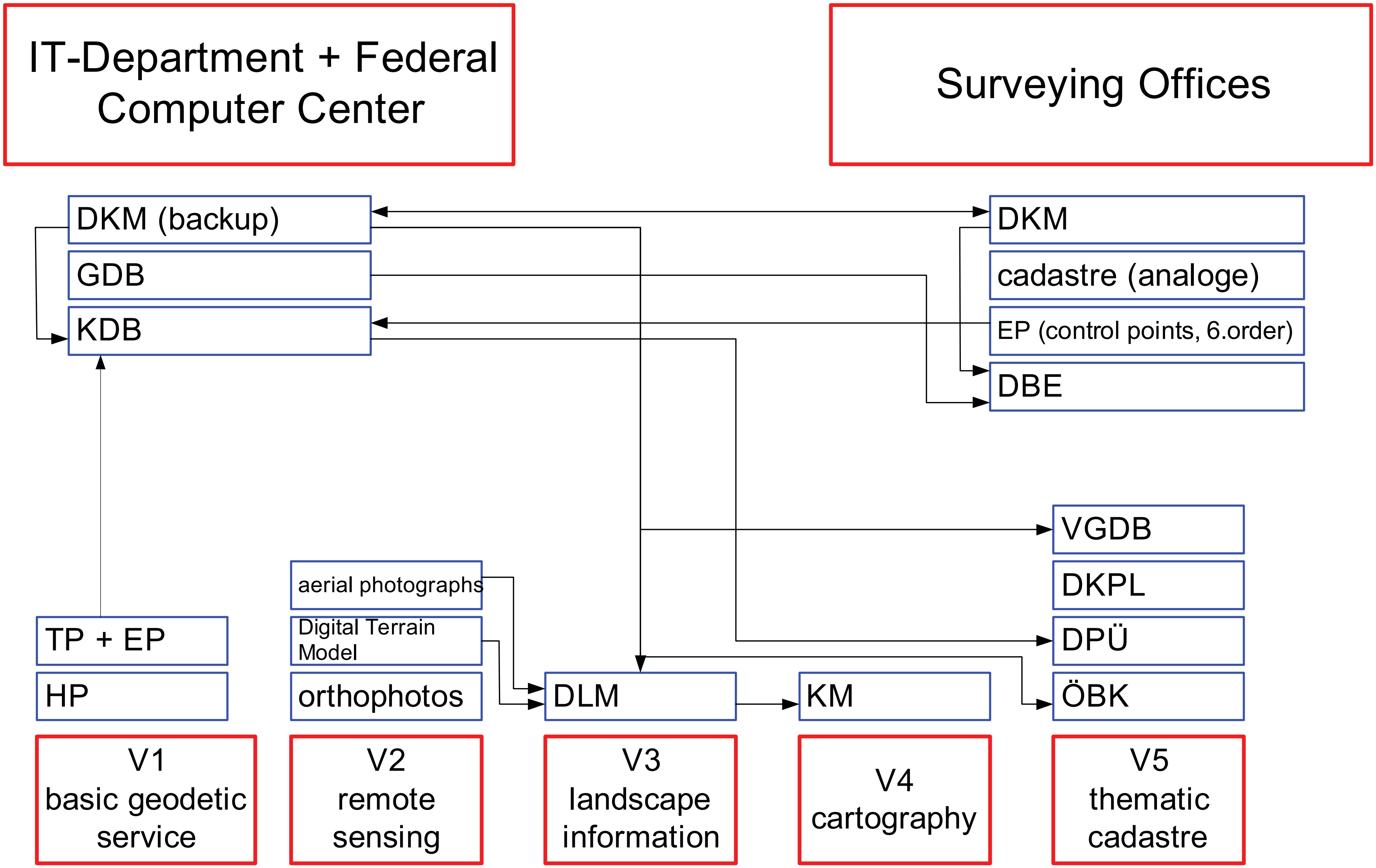


Status quo: BEV's economic output

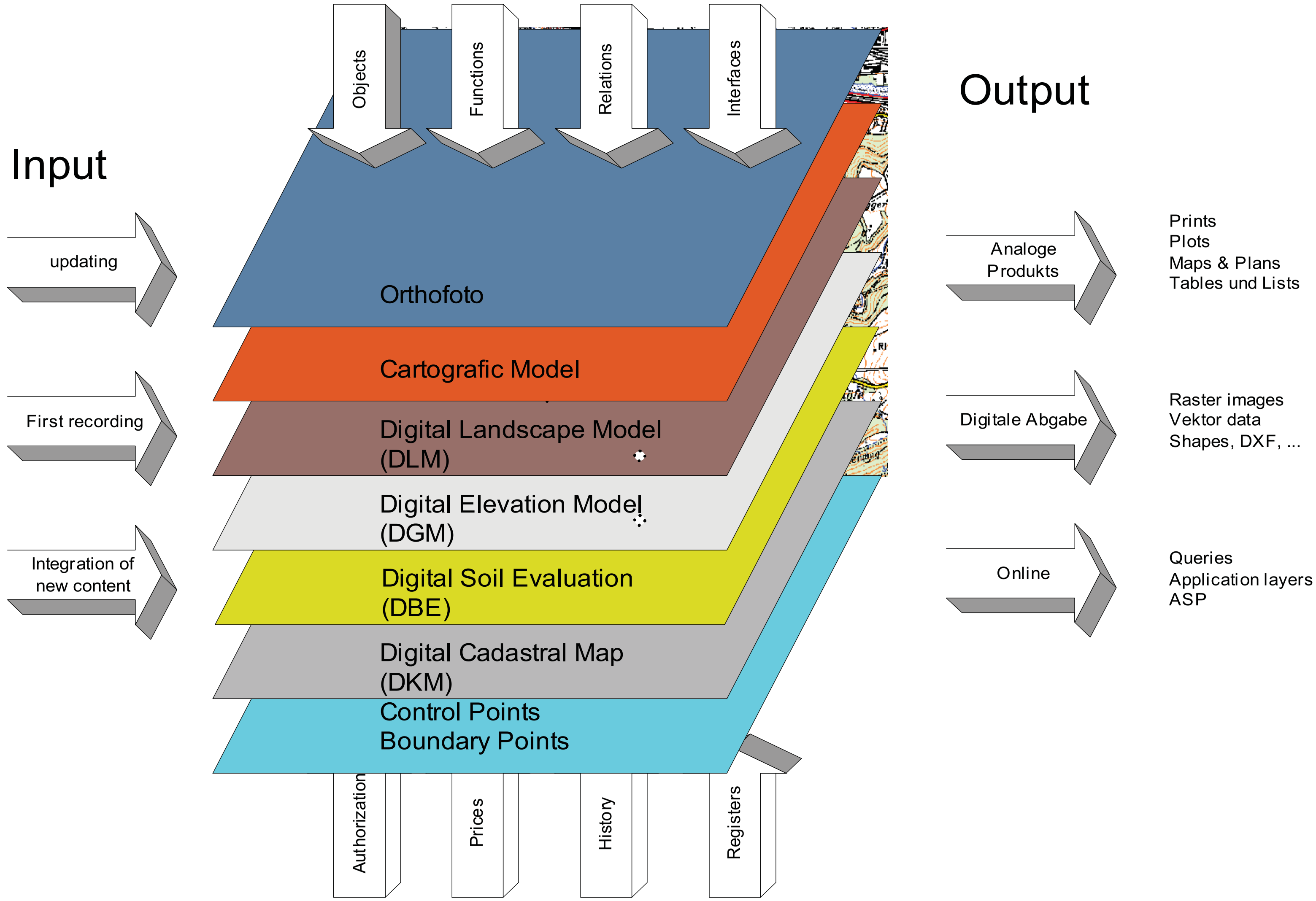
BEV 03/2001



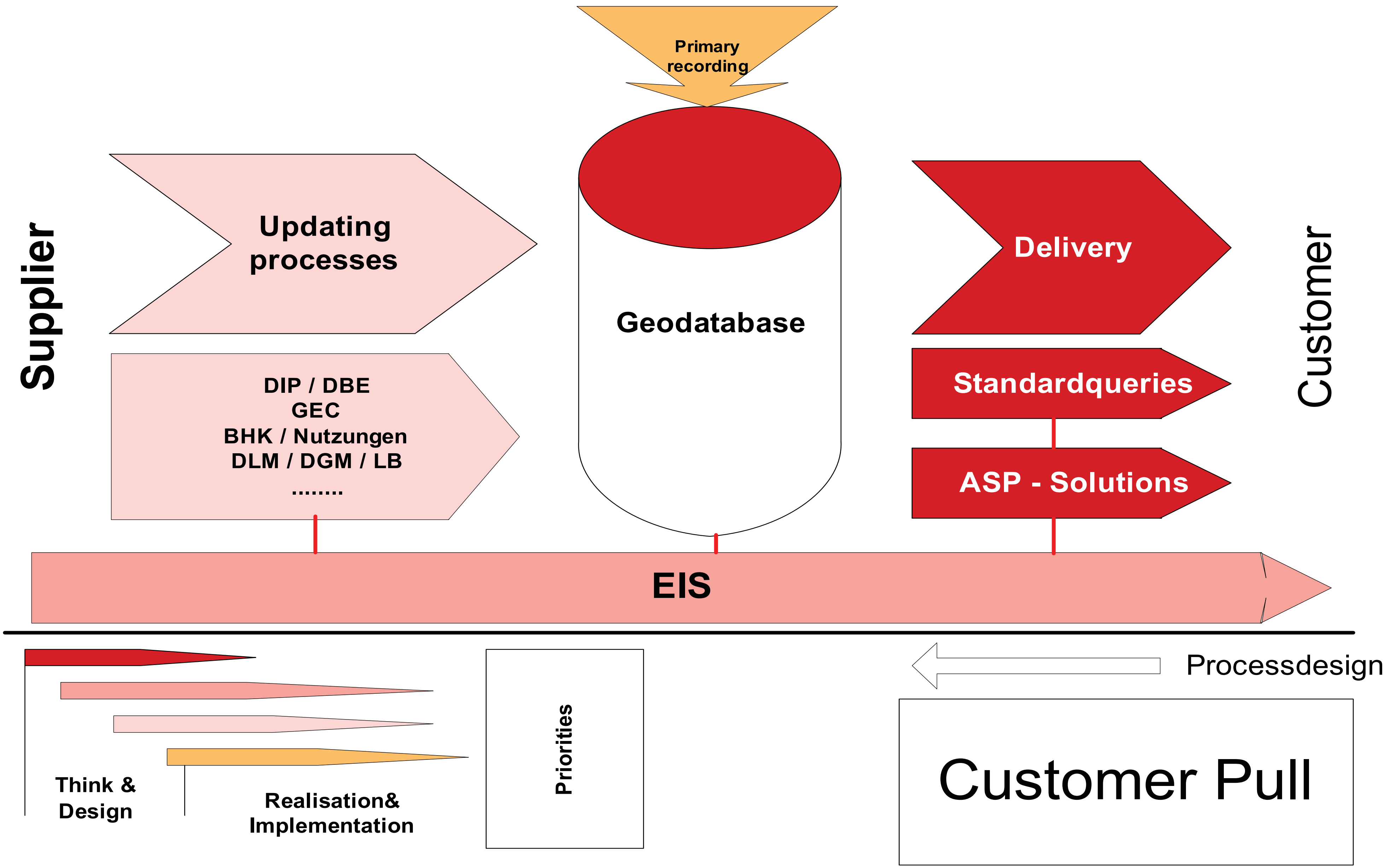
General Data structures



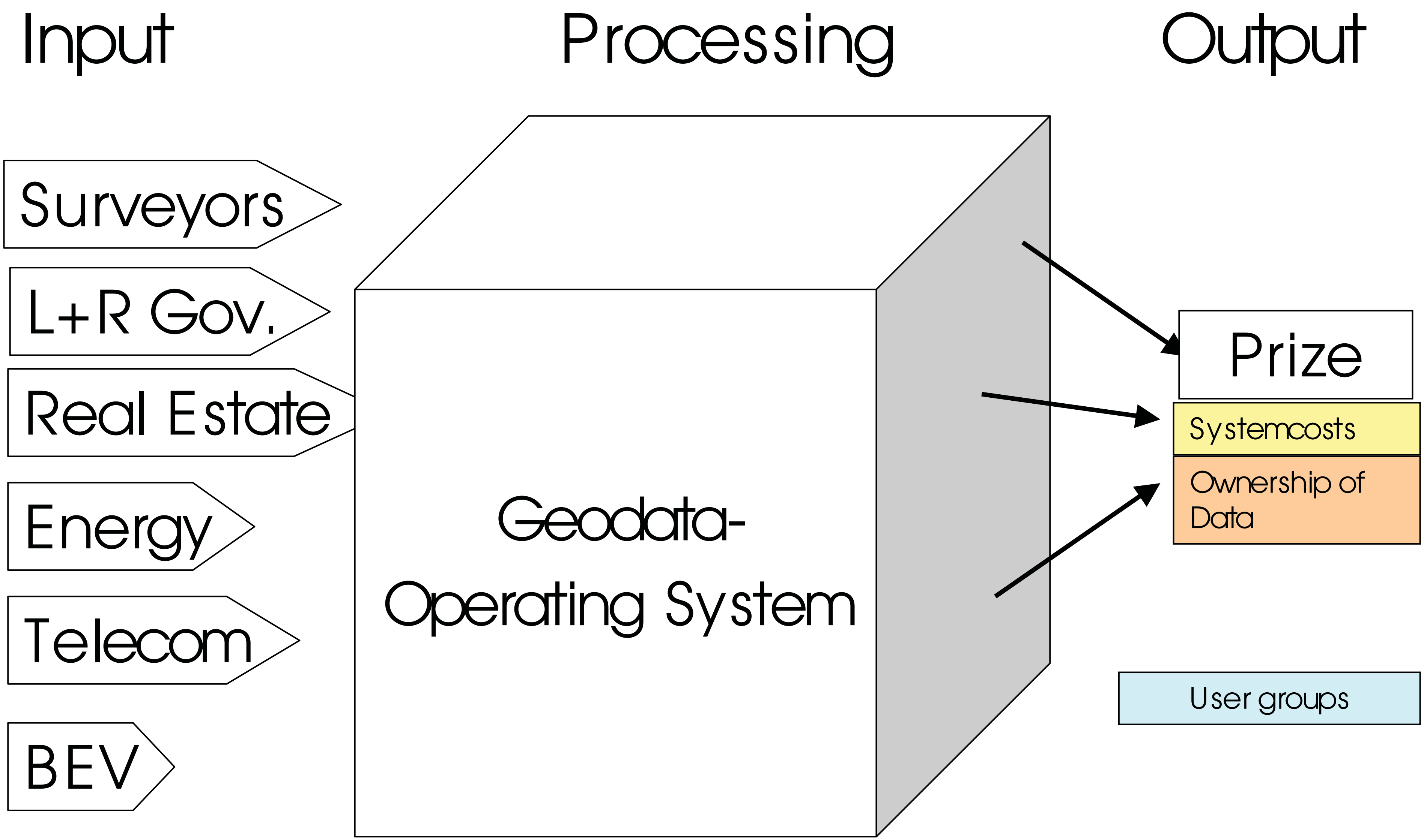
Geodata operating system



Geodatabase



Business Model: Geodata - Operating System

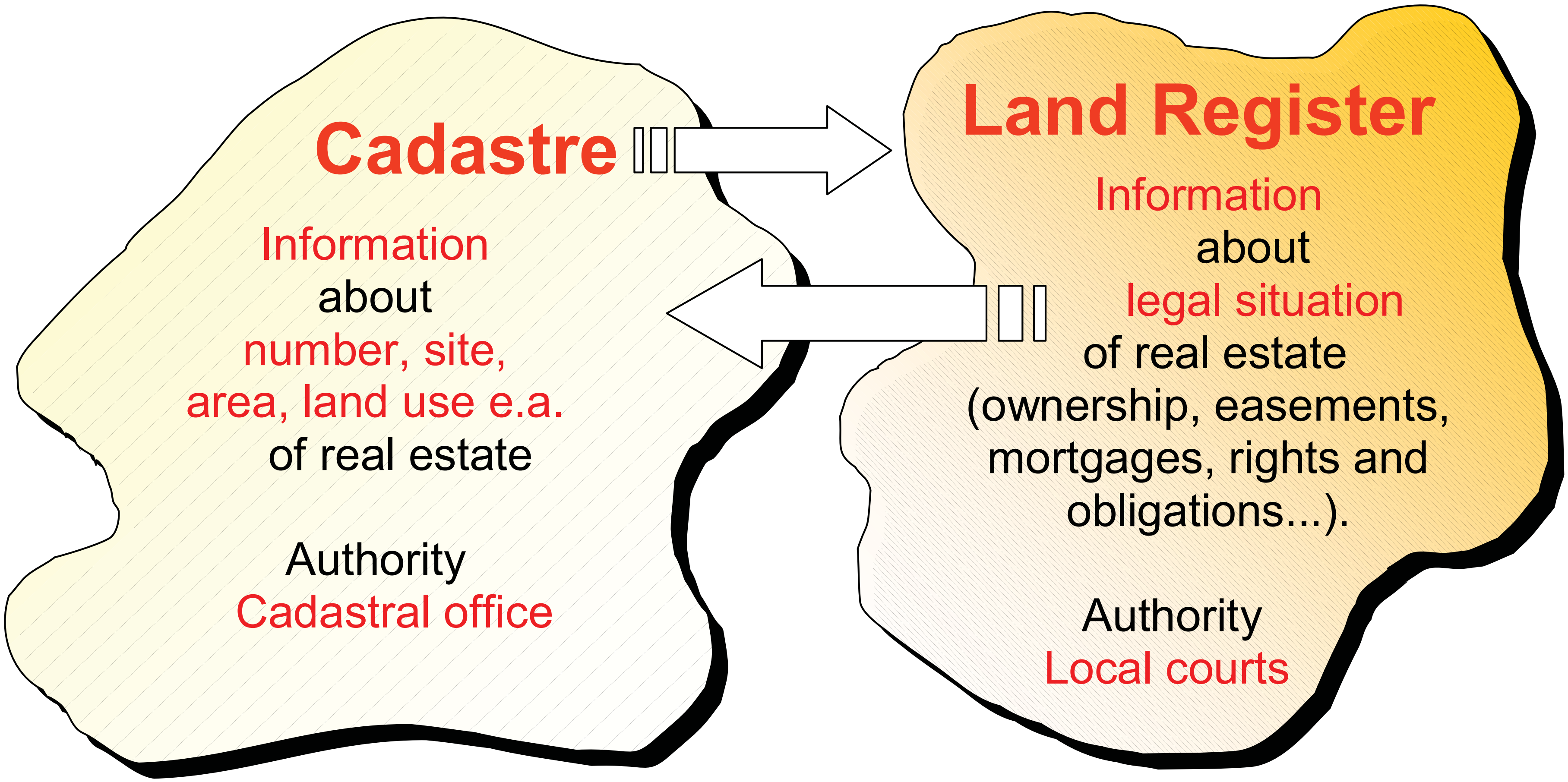


Austria

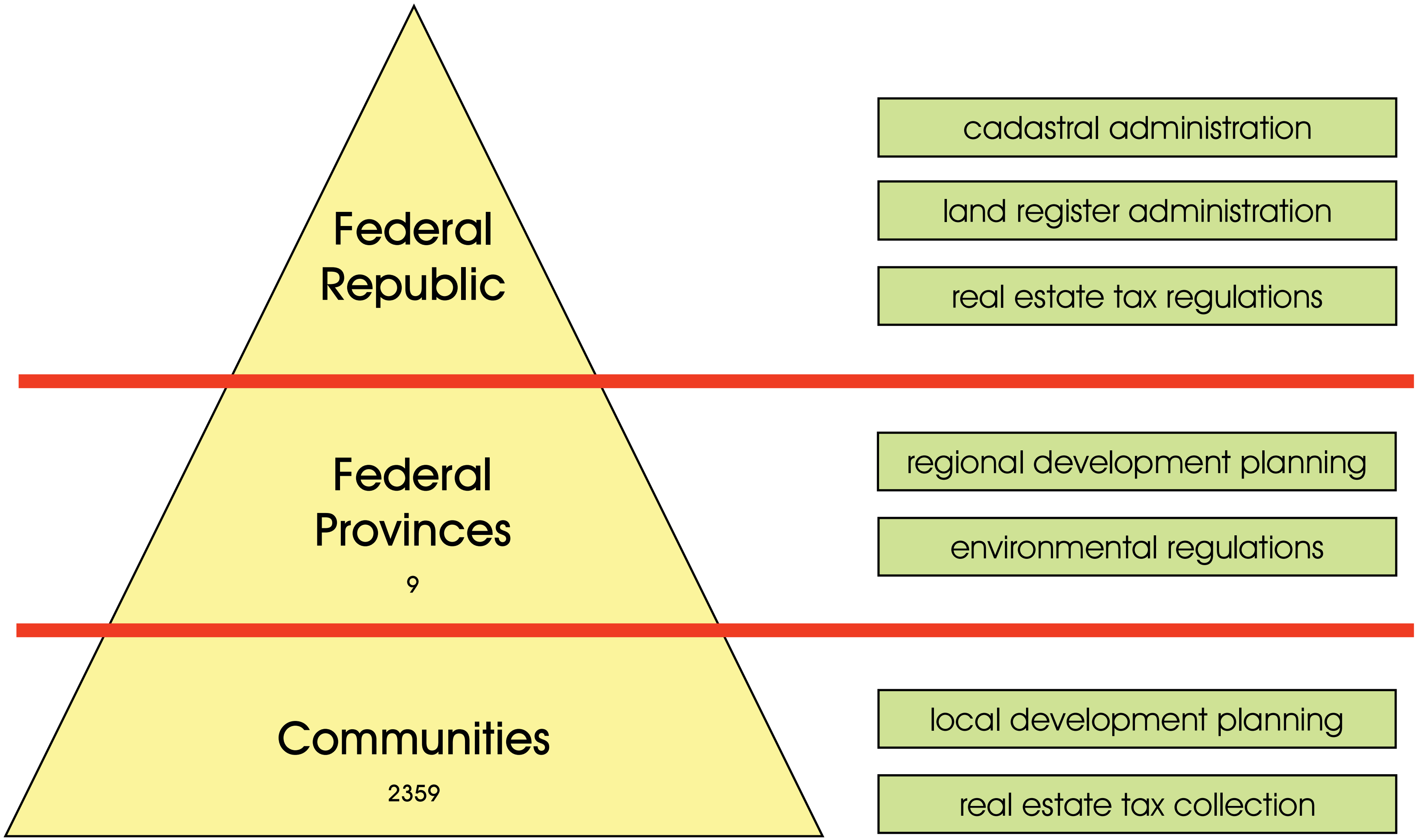
84.000 sqkm
8.1 mio inhabitants

11 mio parcels
41 Cadastral Offices
189 Local Courts

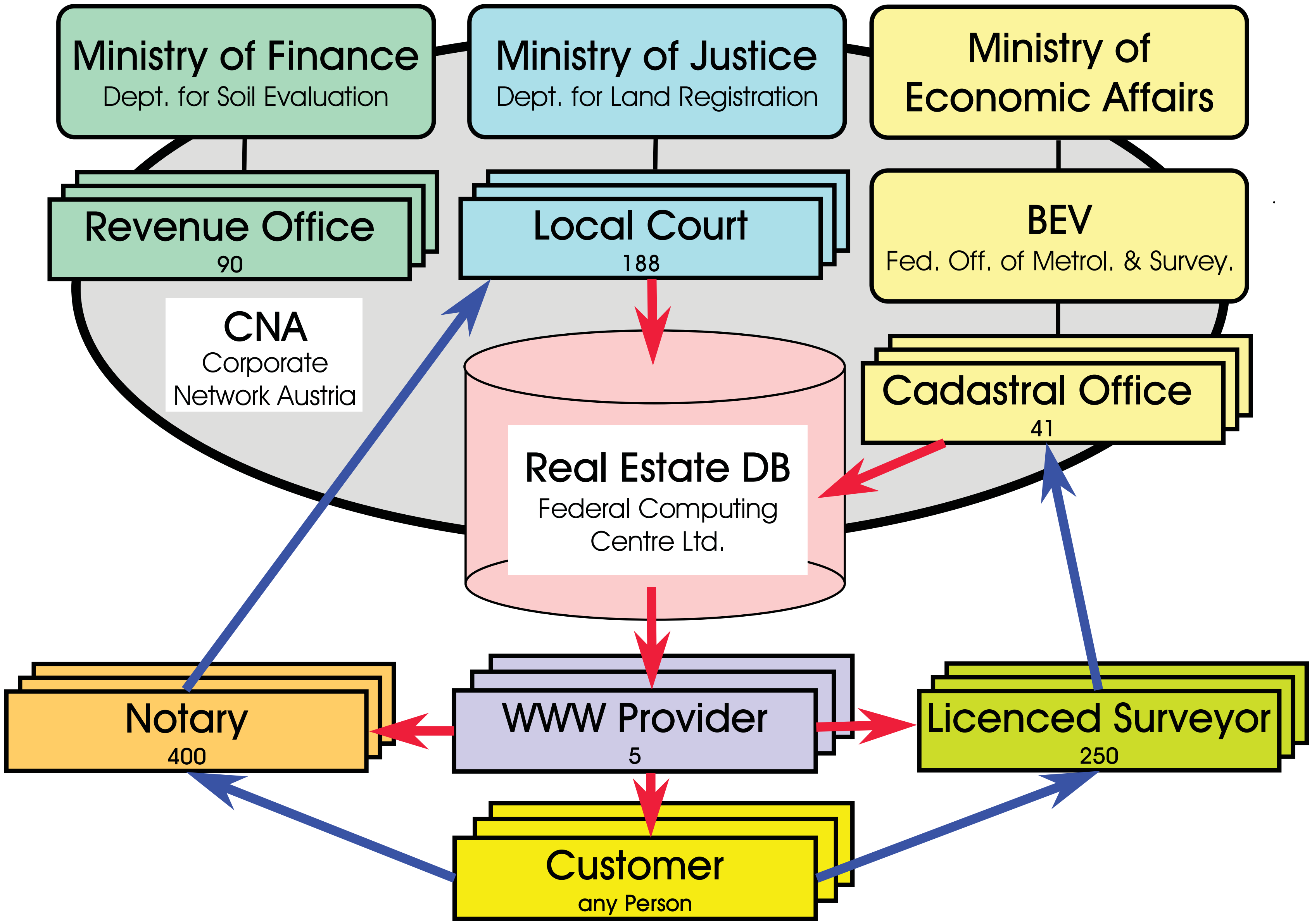
Austrian Land Registration System



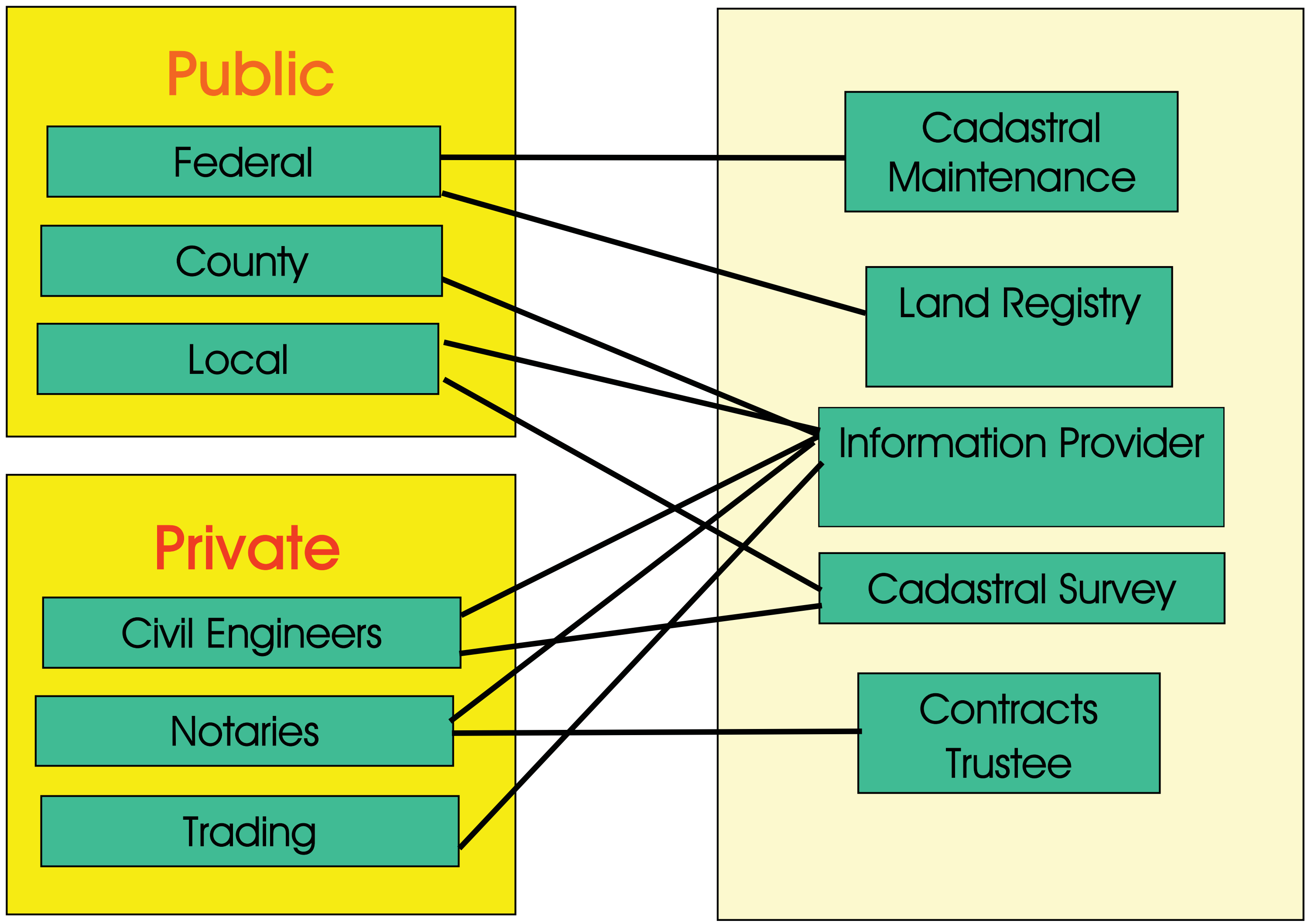
Public Land Administration



Co-operation of private & public Sector

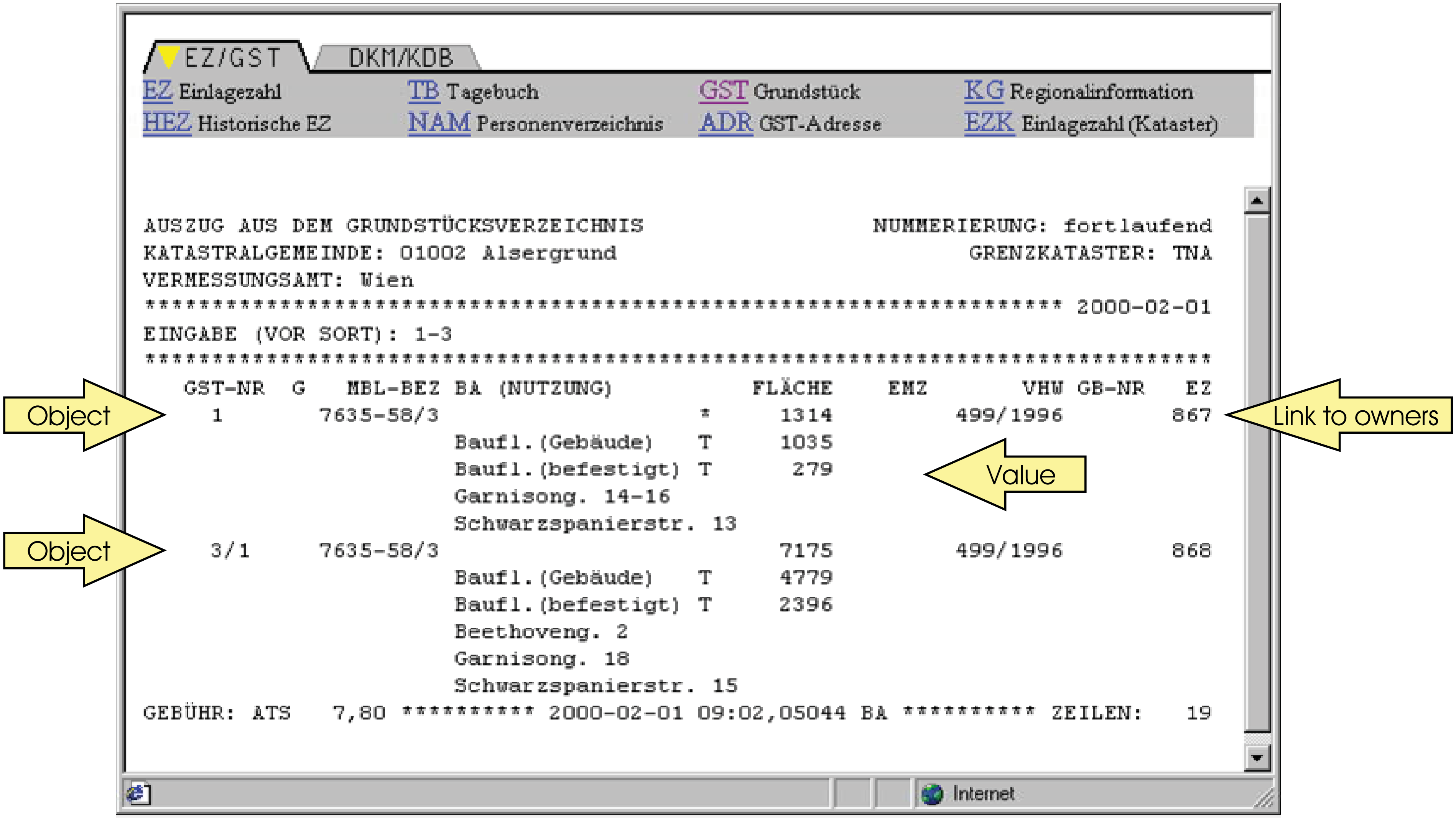


Public - Private Cooperation



tei Bearbeiten Ansicht Favoriten Extras ?
Zurück Suchen Favoriten Verlauf
sse C:\Eigene Dateien\Docint\Austria\BEV\Handbuch\handbuch.htm
Wechseln zu
Home
Suche
E-Mail
Handbuch für den Kataster
Rechtliche Grundlagen
Verfahrensabläufe im Vermessungsamt
Kommentare
Diskussionsforum
Gesetze
Verordnungen
Erlässe
» Einrichtung des Grenzkatasters
» Führung des Katasters
» Amtshandlungen
» Verwaltungsgrenzänderungen
» Erläuterungen
» Entscheidungen
» Beispiele
» Definitionen
» Abkürzungen
» Vorschläge und Anregungen sind erwünscht
(Geplant ist hier ein Diskussionsforum)
Arbeitsplatz

Real Estate Database - Cadastre Extract



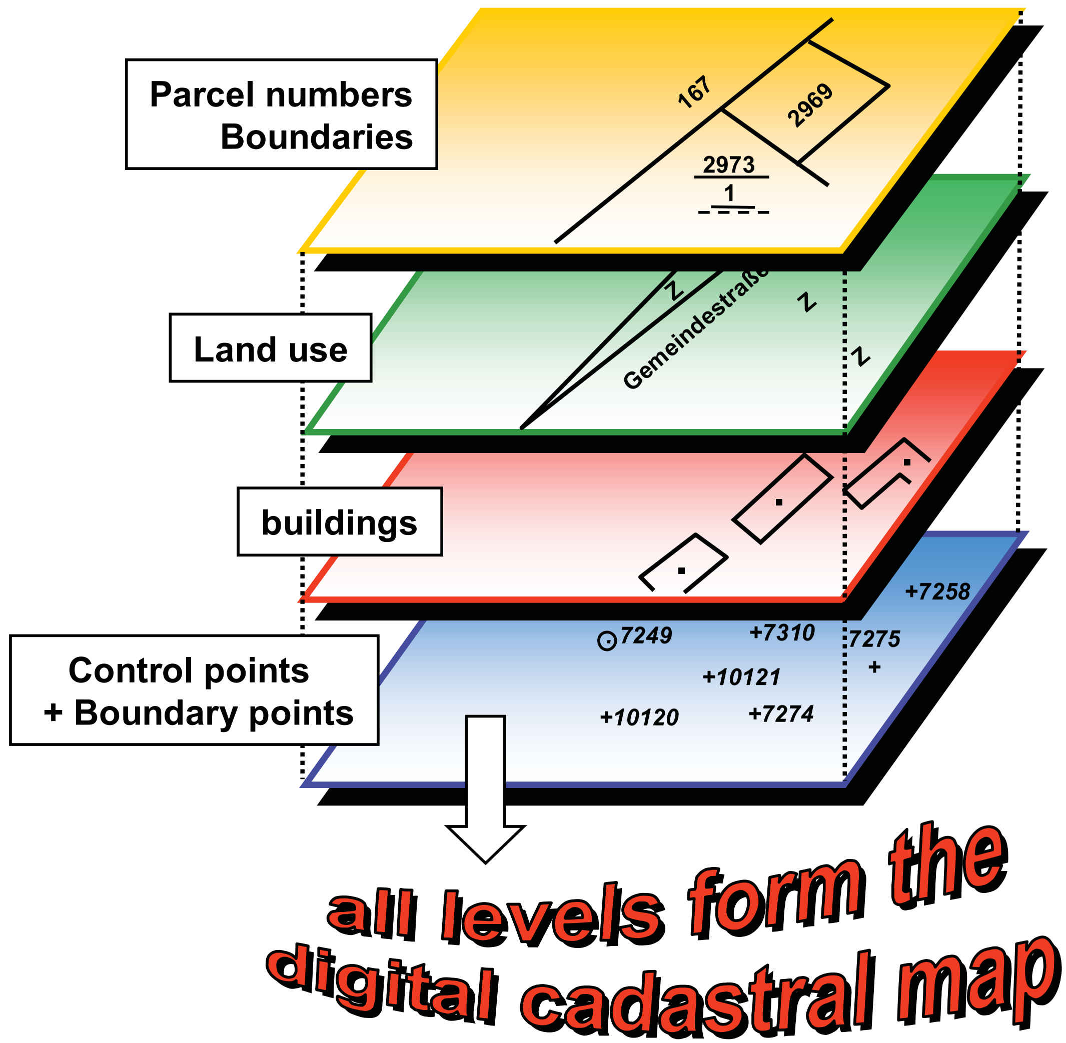

```
AUSZUG AUS DEM GRUNDSTÜCKSVERZEICHNIS                    NUMMERIERUNG: fortlaufend
KATASTRALGEMEINDE: 01002 Alsergrund                             GRENZKATASTER: TNA
VERMESSUNGSAMT: Wien
******************************************************************** 2000-02-01
EINGABE (VOR SORT): 1-3
*******************************************************************************
   GST-NR   G    MBL-BEZ BA (NUTZUNG)          FLÄCHE     EMZ        VHW GB-NR    EZ
     1         7635-58/3                   *     1314          499/1996          867
                        Baufl.(Gebäude)    T     1035
                        Baufl.(befestigt)  T      279
                        Garnisong. 14-16
                        Schwarzspanierstr. 13
    3/1        7635-58/3                         7175          499/1996          868
                        Baufl.(Gebäude)    T     4779
                        Baufl.(befestigt)  T     2396
                        Beethoveng. 2
                        Garnisong. 18
                        Schwarzspanierstr. 15
GEBÜHR: ATS   7,80 ********** 2000-02-01 09:02,05044 BA ********** ZEILEN:   19
```

Extract from the parcels register

```
GRUNDBUCH 01002 Alsergrund                                    EINLAGEZAHL   867
BEZIRKSGERICHT Josefstadt
****************************************************** ABFRAGEDATUM  2000-02-01
Letzte TZ  7206/1984
************************************* A1 *************************************
   GST-NR   G BA (NUTZUNG)           FLÄCHE  GST-ADRESSE
     1        GST-Fläche         *      1314
              Baufl.(Gebäude)           1035
              Baufl.(befestigt)          279  Garnisong. 14-16
                                              Schwarzspanierstr. 13
************************************* A2 *************************************
   1  a 8021/1976 Bauplatz (auf) Gst 1
   2  a 7206/1984 Denkmalschutz hins Haus auf GSt 1
************************************* B **************************************
   1 ANTEIL: 1/1
     Evangelischer Verein für Studentenheime
     ADR: Garnisong. 14-16   1090
      a 7922/1963 Kaufvertrag 1960-11-25 Eigentumsrecht
      b 9233/1963 Veräußerungsverbot
************************************* C **************************************
   1  a 9233/1963 Schuldschein 1963-12-06
          PFANDRECHT                                                 9,100.000,--
          1 % Z, 6 % VuZZ, NGS 910.000,-- für Stadt Wien
   2  a 9233/1963
          VERÄUSSERUNGSVERBOT gem WBFG 1954 für Stadt Wien
********************************** HINWEIS ***********************************
           Eintragungen ohne Währungsbezeichnung sind Beträge in ATS
GEBÜHR: ATS  11,70 ********** 2000-02-01 10:20,06664 BA ********** ZEILEN:   28
```

Internet

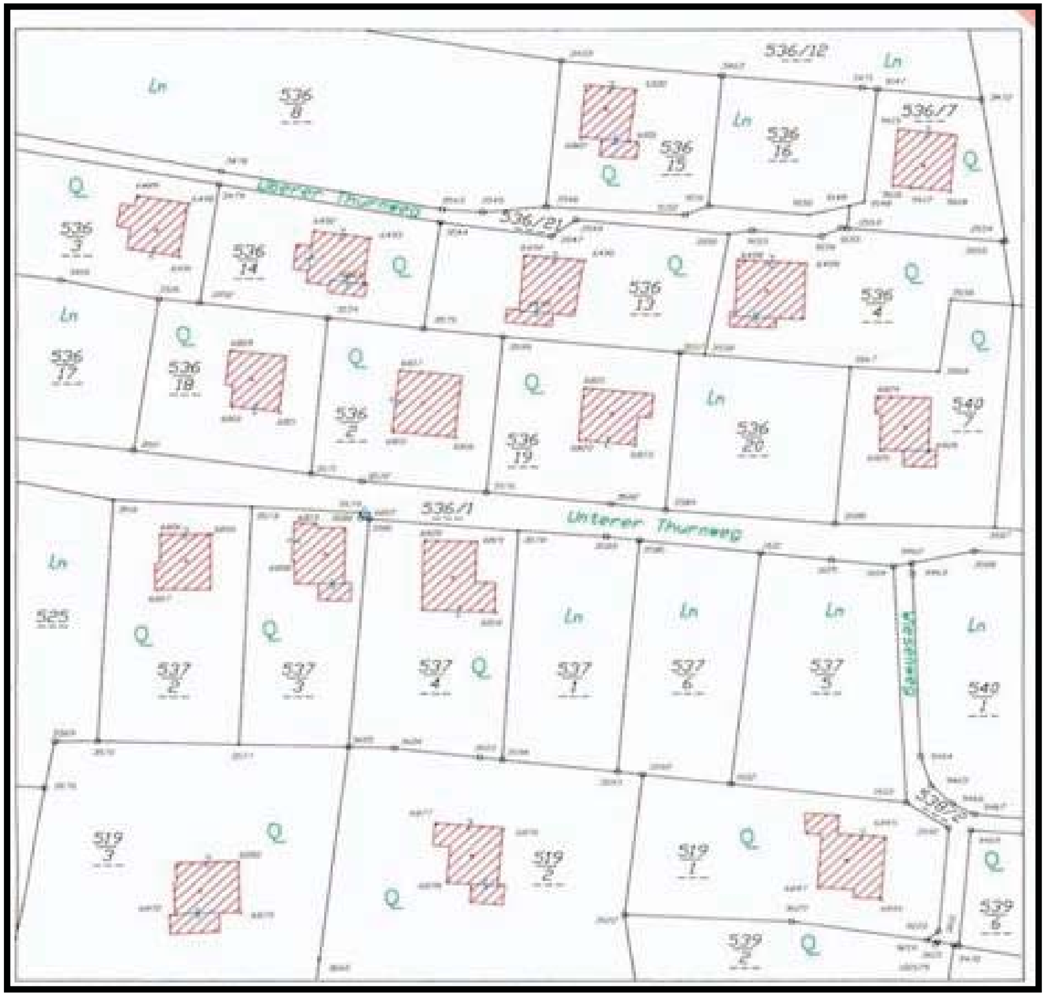
Digital cadastral map (DKM)



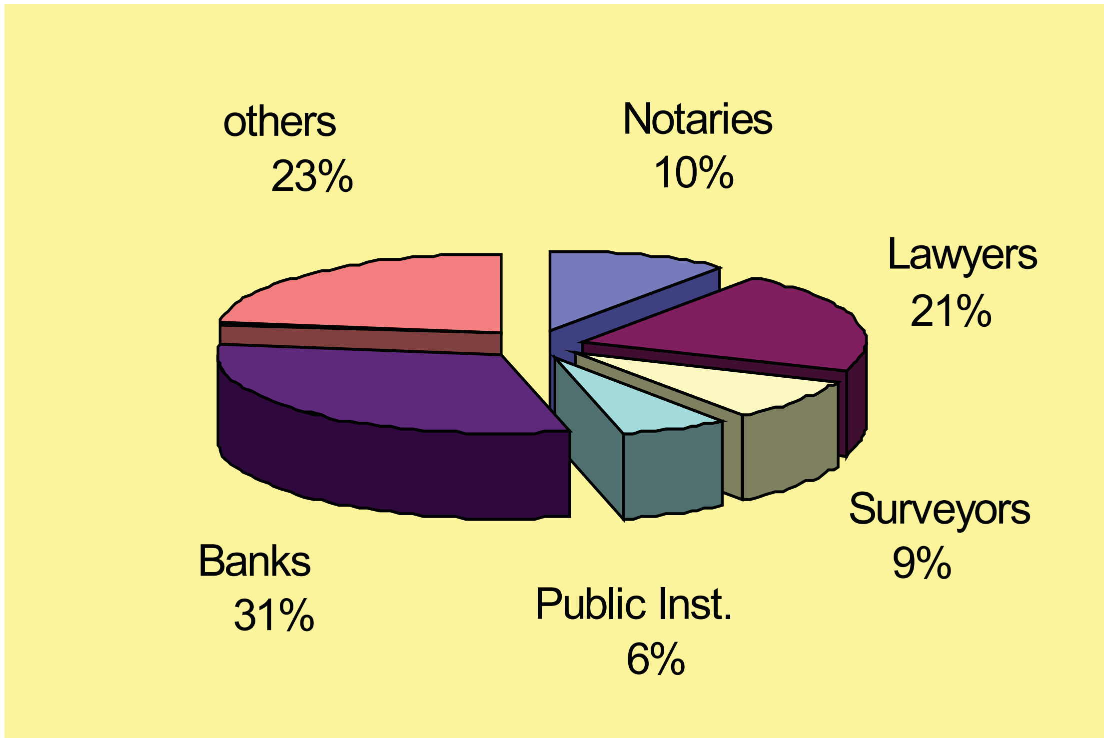
Output unreceptively to scale and sheet margins

linkable with other data (GIS)

Digital map sheet 1:1000



Customer service of the BEV



- Identify/Satisfy Clients
- Marketing + Awareness
- Increase Services

Integrated Administrative and Control System

- The main elements of the system are:
- **a) aid applications**; each farmer must submit every year an application indicating all agricultural parcels,
- **b) a computerised database**

c) an alphanumeric **identification system for agricultural parcels**, (areas to be located monitored over time, cross-checks and on-the-spot checks;

d) an integrated control system for administrative control and field inspections.

Total cost of IACS

The main categories are:

(a) implementation costs (only costs for EDP structures, control structures and experts are Community co- financed)

(b) running costs

Other implementation costs and the costs of running the IACS are borne by the Member States

Only 14 % of Germany's total implementation costs were reimbursed by the Community

View of the Agrarian Database (ADB

a) an **application** indicating all agricultural parcels

b) a computerised database

c) an alphanumeric **identification system for agricultural**

```
AUSZUG AUS DEM GRUNDSTÜCKSVERZEICHNIS / AGRAR-FLÄCHEN      BETRIEBS-NR.: 2312557
BEZIRKSBAUERNKAMMER Kirchdorf                    ANTRAGSDATEN VOM 1998-05-15
******************************************************** 1999-07-26, SEITE    1
FELDSTÜCK-NR:    6, OBERE HALS-WIESE, Grünland
    VERMESSUNGSAMT Steyr
    KATASTRALGEMEINDE 49403 Gleinkerau
   GST-NR  G   MBL-BEZ BA (NUTZUNG)            GSTVZ-FL    GENUTZTE-FL  RECHTSV.
   333                 Fläche lt. Antrag:       1 98 51       1 98 51   E   FEHL=L
             5329-44/4 Landw. genutzt           2 12 14
   337/1               Fläche lt. Antrag:       1 91 90       1 91 90   E   FEHL=L
             5329-44/2 Landw. genutzt           2 11 68
   339       5329-44/2 Landw. genutzt              4 18          4 18   E
   341/2               Fläche lt. Antrag:         88 95         88 95   E   FEHL=L
               5329-45 Landw. genutzt             72 63
                                   SUMME-FELDSTÜCK            4 83 54
******************************** BETRIEBSDATEN BESCHRÄNKT AUF OA. FELDSTÜCK ***
Erläuterung der Fehlercodes:
FEHL=L: Katasterfläche lt. Antragsteller ungleich GSTVZ-Fläche
************************************************************** FORMAT    1 A4 ***
```

BEV provided to the MoAF

- **Aerial photographs** infrared aerial images (1: 15000).

  GPS for camerapostion => fewer ground control-points
- **Interpretation of landuse**
- **Stereoplotting** boundaries of landuse were digitized
- **Improvement of the DTM**
- **Digitizing** and **updating** the cadastral maps
- Estimation of the **slope classes** of the parcels

Co-operation: MoAF and BEV in a common project :

- registration of 2.5 million **block parcels** (8 mill. cadastral parcels)
- installation of **numerical and graphical databases** connecting cadastral- and IACS information (ADB and DAM)
- **annual updates and crosschecks** of these databases to guaranty consistency
- **common working group**, defines annual support of BEV and the financial aspects

Next Generation Spatial Database Workshop – 2005

Keith Murray [OEEPE Commission 4]

22nd – 24th May 2002

Southampton

Ruth Williams

- Health & Safety

- Also your point of contact for:
  - Accommodation
  - Workshop Dinner
  - Transport

Ordnance Survey®

www.ordnancesurvey.co.uk

Next Generation Spatial Database Workshop – 2005

Welcome to Ordnance Survey
from
David Willey,
Deputy Chief Executive Ordnance Survey

New ways of working

Need new approaches

OEEPE

Next Generation Spatial Database Workshop – 2005

Aims

- A unique group - to share experiences
  - Issues we all face
  - Successful solutions
  - Things that have not worked so well
  - Plans for new databases
- Establish new contacts
- Four sessions – four perspectives
- Summary at the end – to be published

Programme: four sessions

- 22 May: Global Overview
- 23 May: National Mapping Agency view
- 23 May: Cadastre View
- 24 May: Technology provider view

Session Format

- Chairman
- Plenary – introduction of topic
- 2-3 presentations
- Questions

followed by

- Breakout – question (see programme)
- 2-3 groups ~ 1 hour
- All groups report back to plenary

Please enjoy the workshop

&

your time in Southampton

Session 1

Geospatial databases – the global drivers for change

Chair Peter Woodsford

Chair	**_Peter Woodsford_**		
14.30	Implications for Geographic Information providers in the Information Age - Keynote	Tony Davison, IBM	
15.30	From the outside to the inside & back again	Ed Parsons, Ordnance Survey	
16.00-16.30	Tea/Coffee		
16.30	Breakout Session	Subject: Top three issues facing NMA's and Cadastral agencies & their information	
17.30	Reporting back		
18.00	Close		

Session 2

The National Mapping Agency – What data, why, when, how?

Chair Keith Murray

Chair	***Keith Murray***		
09.00	Geospatial Database from Proprietary to Open	Fred Finch - OS Ireland	
09.45	A Geospatial Datawarehouse to provide cadastral and other data	Jens Ole Jensen – Kort & Matrikelstyrelsen	
10.30-11.00	Tea/Coffee		
11.00	Breakout Session	Subject: What architecture strategies (data & systems) are required to meet the needs of 2005?	
12.00	Report back		
12.30	Lunch		

Session 3

The Cadastre
New services, new directions

Chair Arbind Tuladhar

Chair	***Arbind Tuladhar***		
14.00	Geo-ICT technology push vs. Cadastral market pull	Peter van Oosterom (TU Delft) & Chrit Lemmen (Kadaster) and Rolf de Buy (ITC); Speaker: Arbind Tuladhar (ITC)	
14.45	Cadastral needs and developments	Gerhard Muggenhuber - Austrian Cadastre (BEV)	
15.30-16.00	Tea/Coffee		
16.00	Breakout Session	Subject: What kinds of data models will ensure data integration for future applications?	
17.00	Reporting back		
17.30	Close	Coach to hotel	
19.15		Coach hotel to SFC	
19.30	**Workshop Dinner**	**Southampton Football Club**	

Session 4

Database Technologies – What can they do for us? When?

Chair Ed Parsons

Chair	***Ed Parsons***		
09.00	Will your data service meet the customer needs in 2005?	Xavier Lopez, Oracle	
09.30	Building applications with government data – needs for the future.	David Maguire, ESRI	
10.00	Realising the benefits of object technology	Adrian Marriott, eXcelon	
10.30-11.00	Tea/Coffee		
11.00	Breakout Session	Subject: How will the database vendors solutions help me achieve my strategy?	
12.00	Reporting back		
12.30 – 13.00	Summing up & close		EP
13.00-14.00	Lunch		

Summary of workshop

- Database development will become a core activity
- Standard data model(s) ?
- Workflow management
- Cadastral / NMA database requirements are different
- Production/publication database needs
- Organisation issues > technology issues

Next Generation Spatial Database Workshop – 2005


Ordnance Survey®
www.ordnancesurvey.co.uk

From the outside to the inside & back again

Ed Parsons
CTO, Ordnance Survey

OEEPE Workshop, Southampton, May 2002

Ordnance Survey
www.ordnancesurvey.co.uk

Overview

- Where we are today..
- The “end to end” process
- Common points of pain
- The demands of OS MasterMap™
- The OS approach - Mayo

Where we are today

- The Internet changed everything ...
  - The amount digital information will double in volume over the next three years
  - Disparate information brought together through common standards e.g. XML
  - Software and data will be sold as a service
  - New pricing models of all information products

Where we are today

The challenges ahead ..

Geospatial data becomes embedded

Geospatial data needs to be real time

Attributes become more than geometry

Geospatial data becomes pervasive

Geospatial data is temporal

Geospatial data becomes disposable

Ordnance Survey®

www.ordnancesurvey.co.uk

Where we are today

- The end of the first generation digital mapping management systems
- Home built for single use
- CAD file structures
- Limited attribution
- Legacy hardware/software platforms

The "end to end" process

- Recognise a single process from the field to the map
- Includes ..
  - Survey
  - Data Capture
  - Quality Control
  - Data modelling
  - Cartography etc

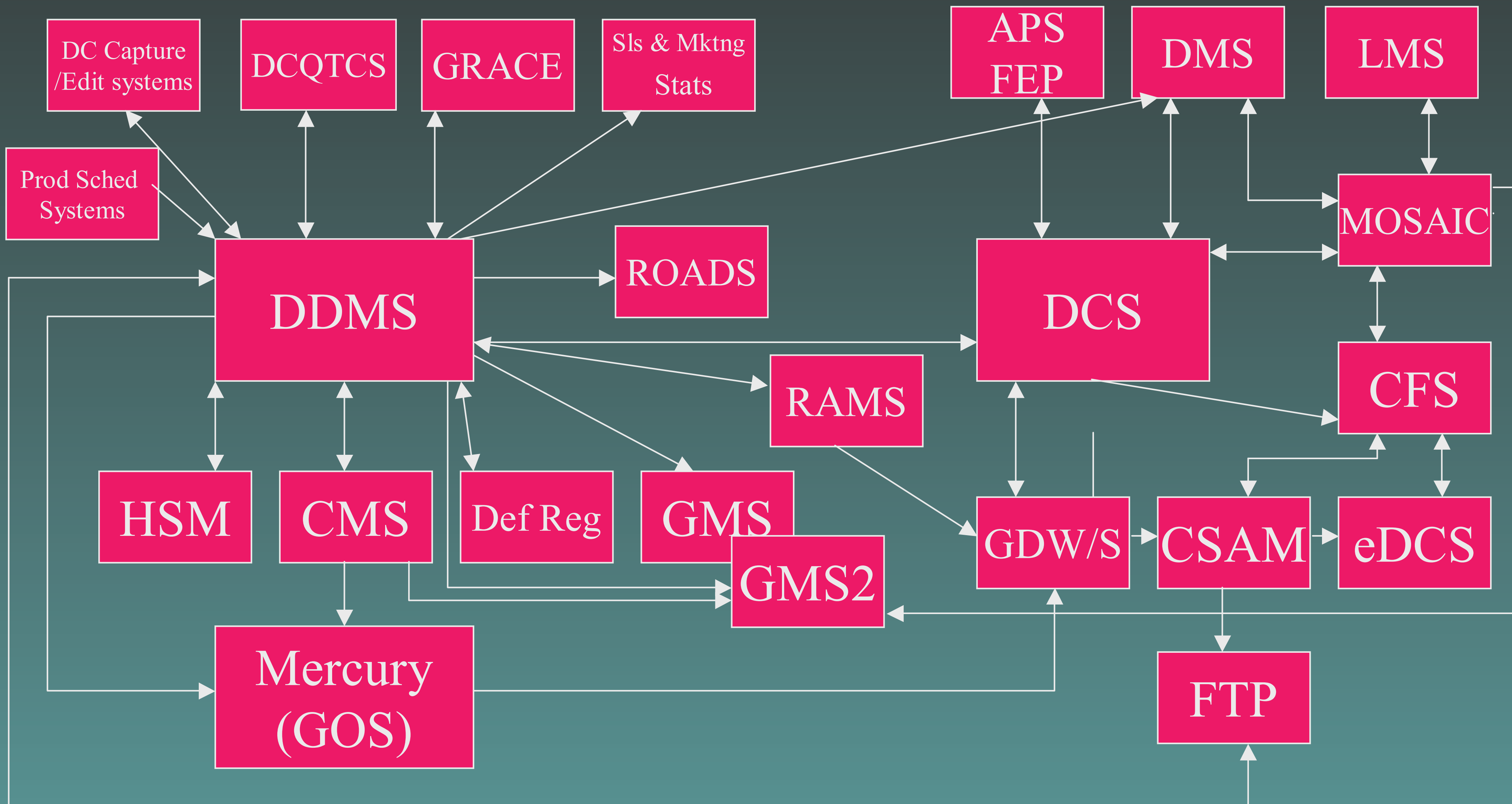
The “end to end” process

- Process needs support of a single integrated information system
  - Data Management
  - Workflow Management
  - Business rules / logic
- Separate data from representation
- Enterprise wide

Common points of pain !

- Existing systems lack agility
- Difficult to integrate/interoperate
- Very costly to maintain
- Failure of past knowledge management
- Bespoke developments !!

Common points of pain !



OS Current Geospatial database architecture

www.ordnancesurvey.co.uk

Common points of pain !

- Future systems need..
  - Seamless integration
  - Unified data management
  - Long transaction management
  - Structured data versioning
  - Robust geoprocessing
  - Flexible application configuration
- COTS based !

What is OS MasterMap™ ?

- Ordnance Survey's new DNF (Digital National Framework) based large-scale product and on-line service

Accessible

Definitive

Intelligent

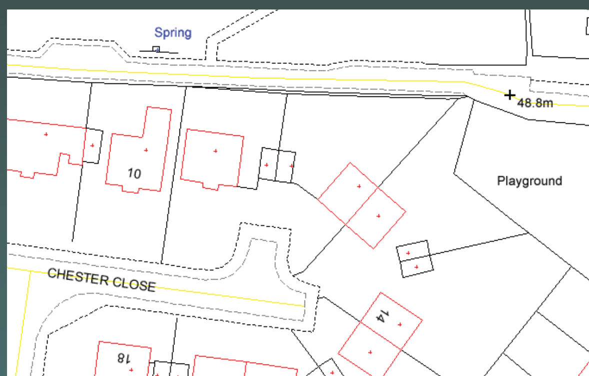
Ordnance Survey®

www.ordnancesurvey.co.uk

OS MasterMap™ features

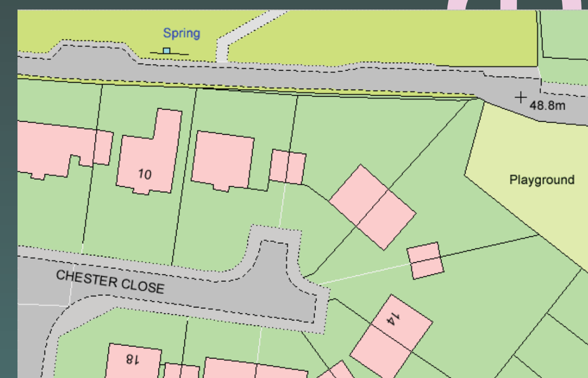
- A seamless polygon based database reflecting the real world
- Improved feature classification
- An on-line selection, quotation and delivery service
- Persistent Identifiers (TOIDs) for data exchange and sharing

Intelligent data ...



Topo96 Landline

Data restructure
Quality improvement
Spatial indexing



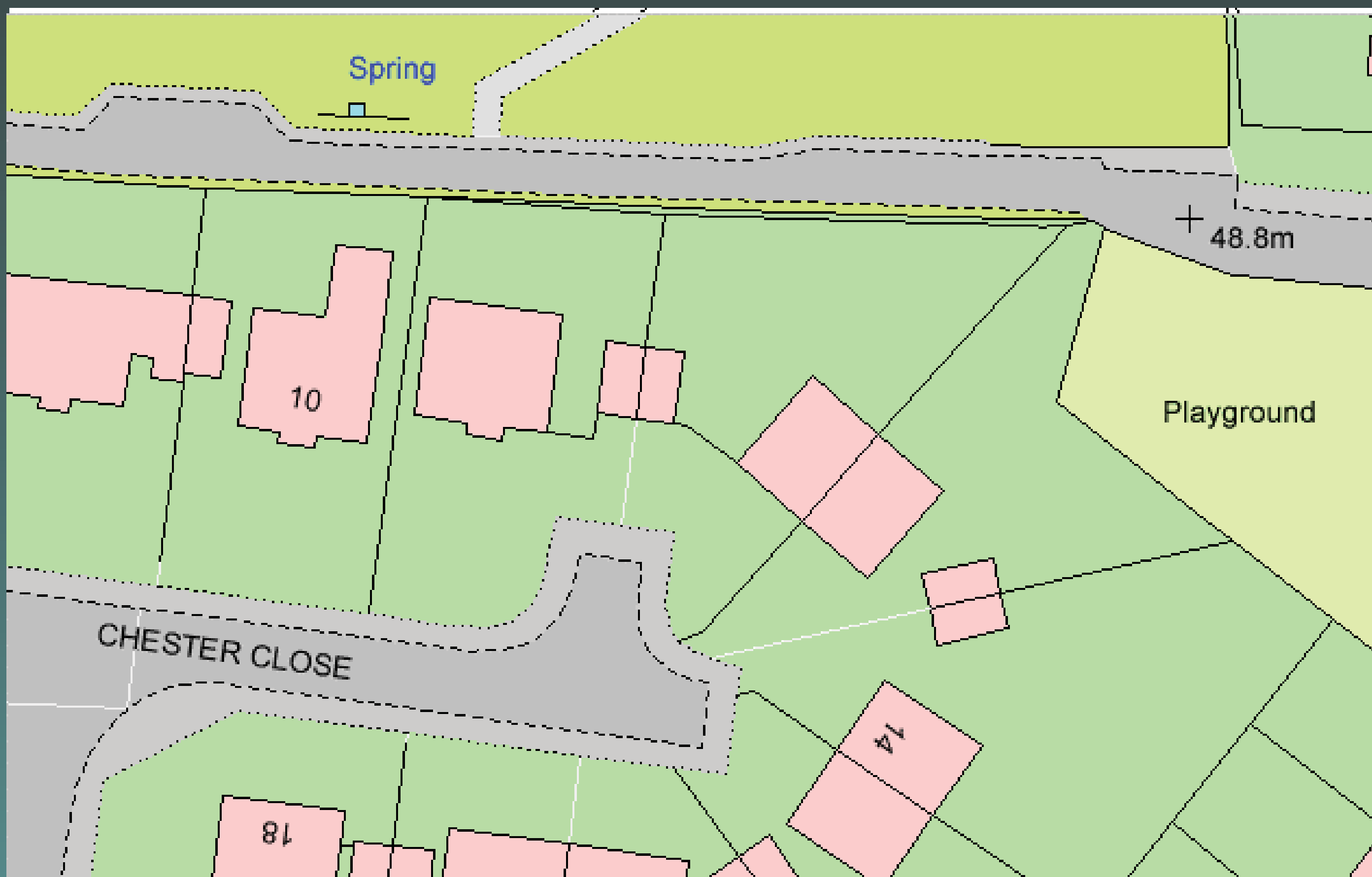
OS MasterMap Topo.

Database of over 400 million unique topographic objects

Ordnance Survey

www.ordnancesurvey.co.uk

Intelligent data ...



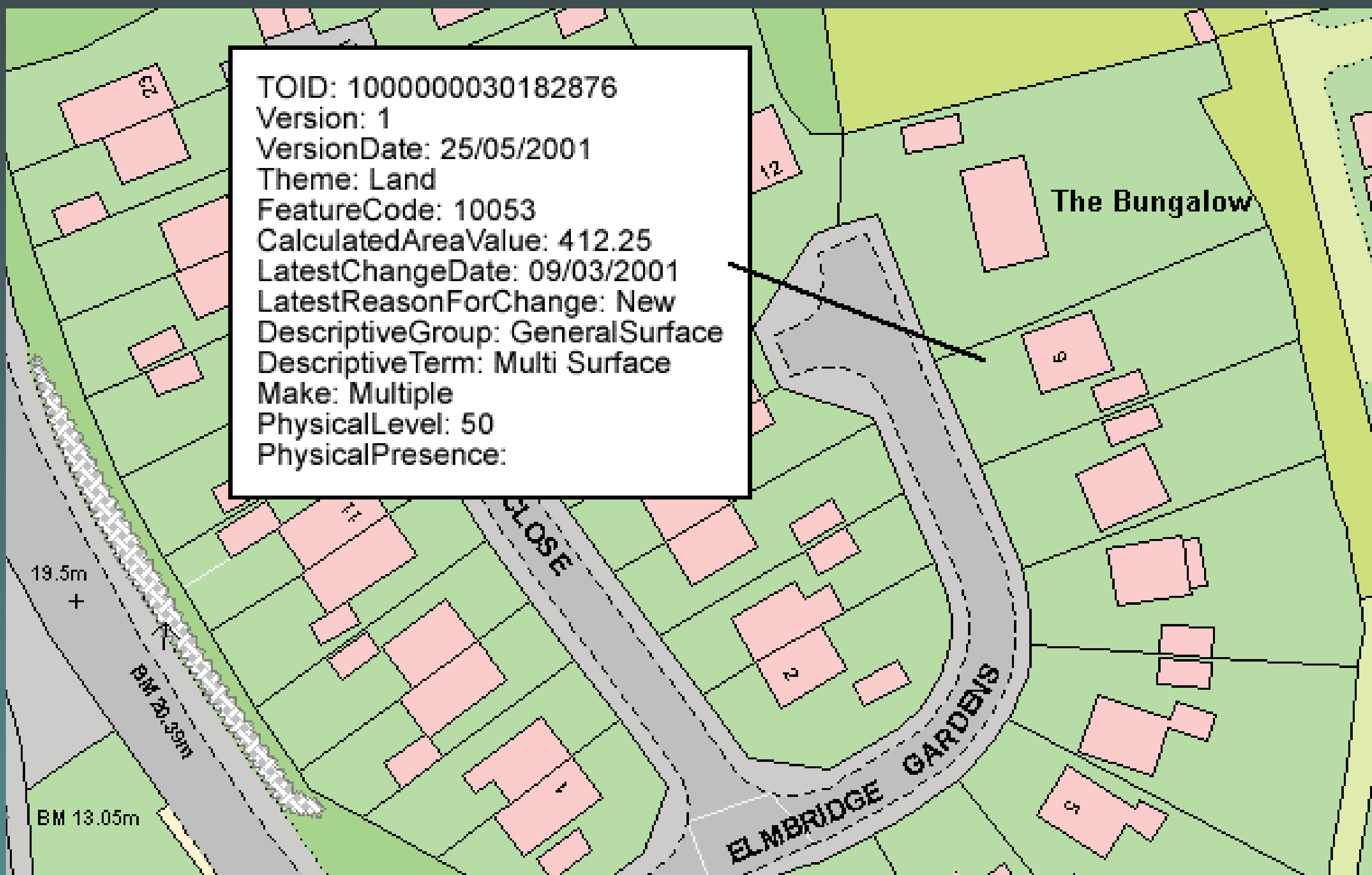
MasterMap Topo

Intelligent

Ordnance Survey®

www.ordnancesurvey.co.uk

Object attributes ...

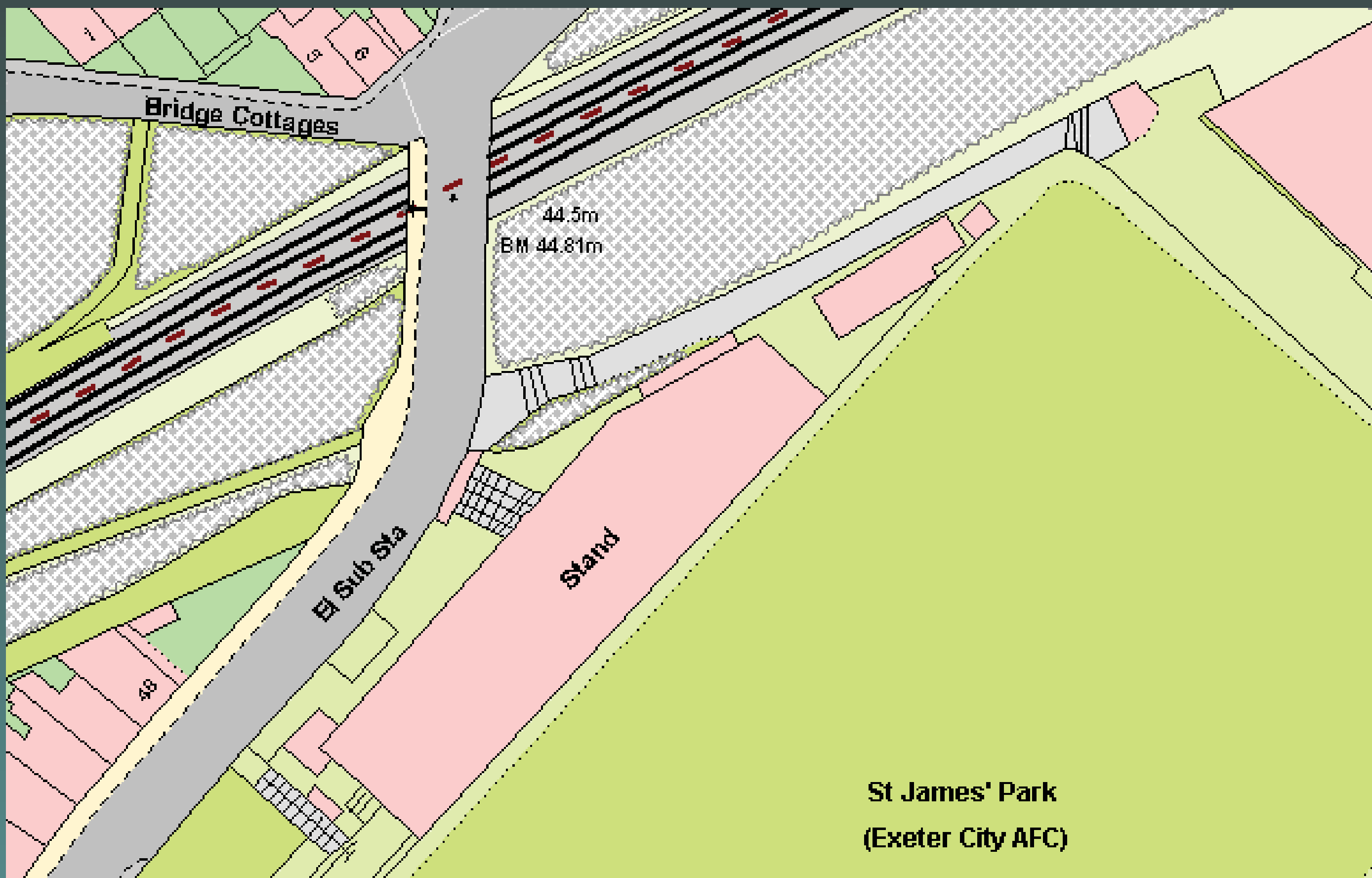


Definitive

Ordnance Survey

www.ordnancesurvey.co.uk

Seamless



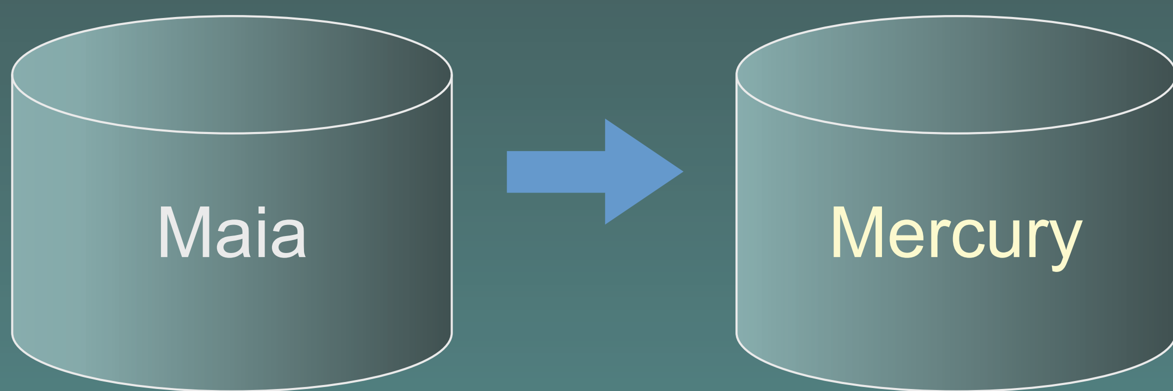
Accessible

The OS approach - Mayo

Shorts Mayo Composite
Shorts S21 Maia, and Mercury Flying boats.
Southampton 1939

The OS approach - Mayo

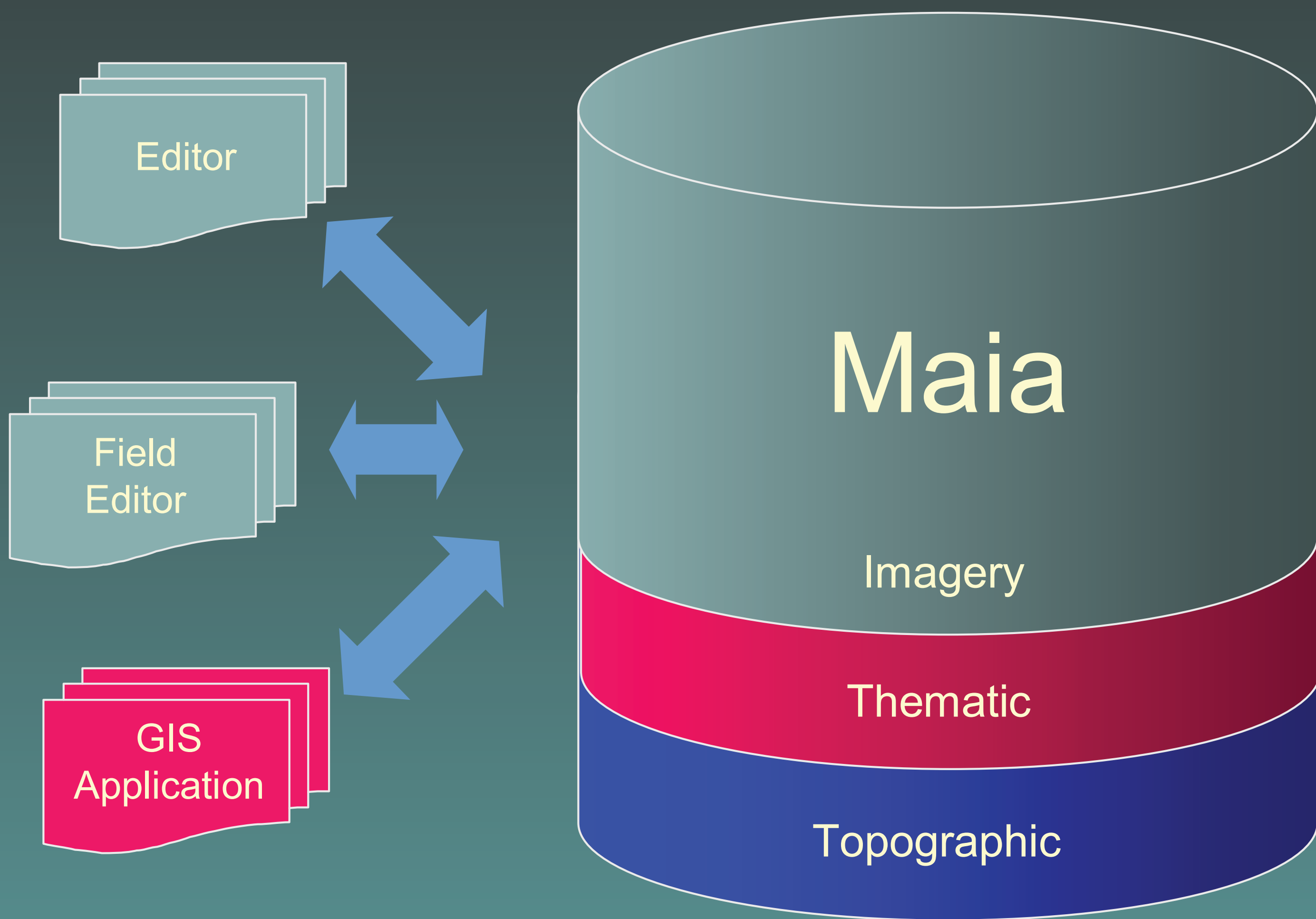
- A single Geospatial database architecture to support the production and publication needs of the Ordnance Survey



Internal,Transaction based system

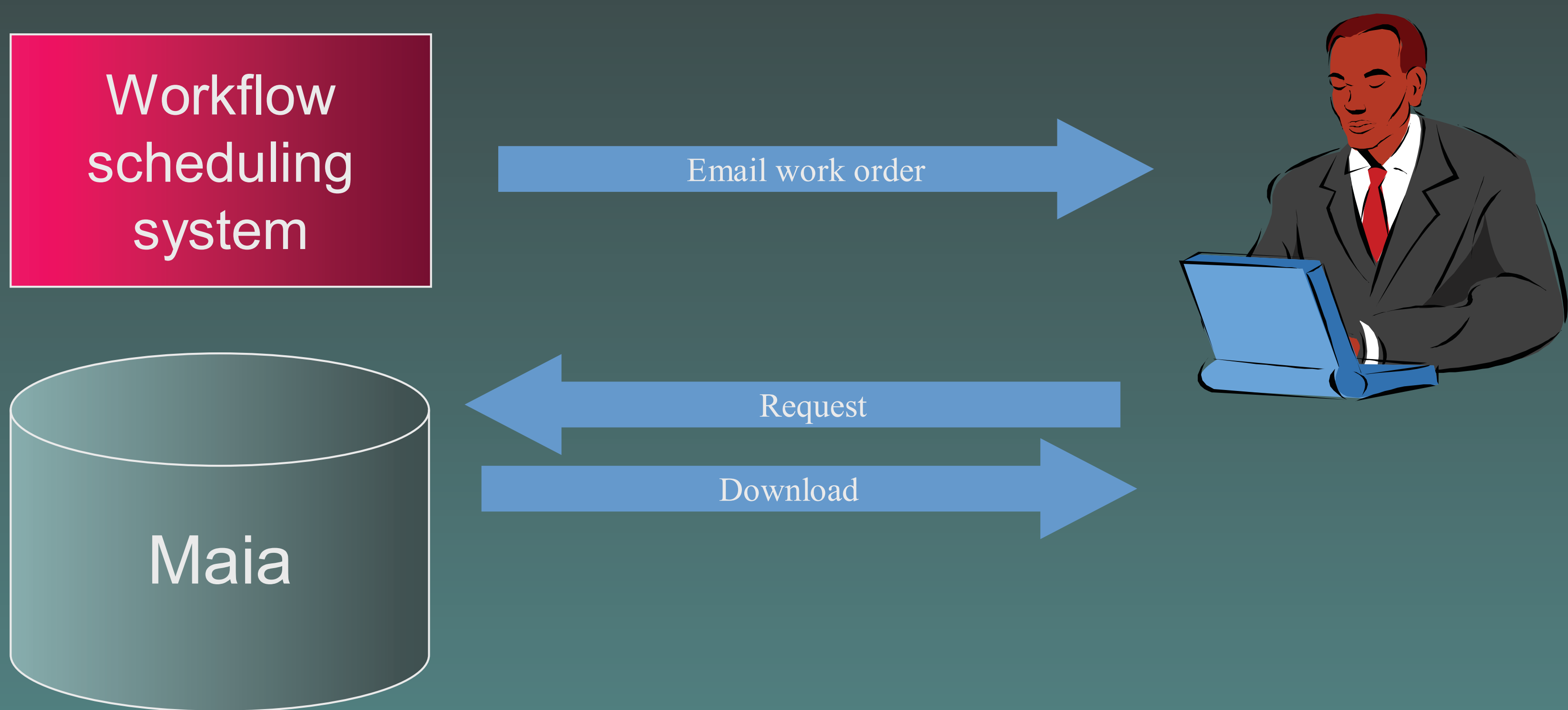
External, Publication system

From 50+ databases to one !



Ordnance Survey®
www.ordnancesurvey.co.uk

The data capture workflow

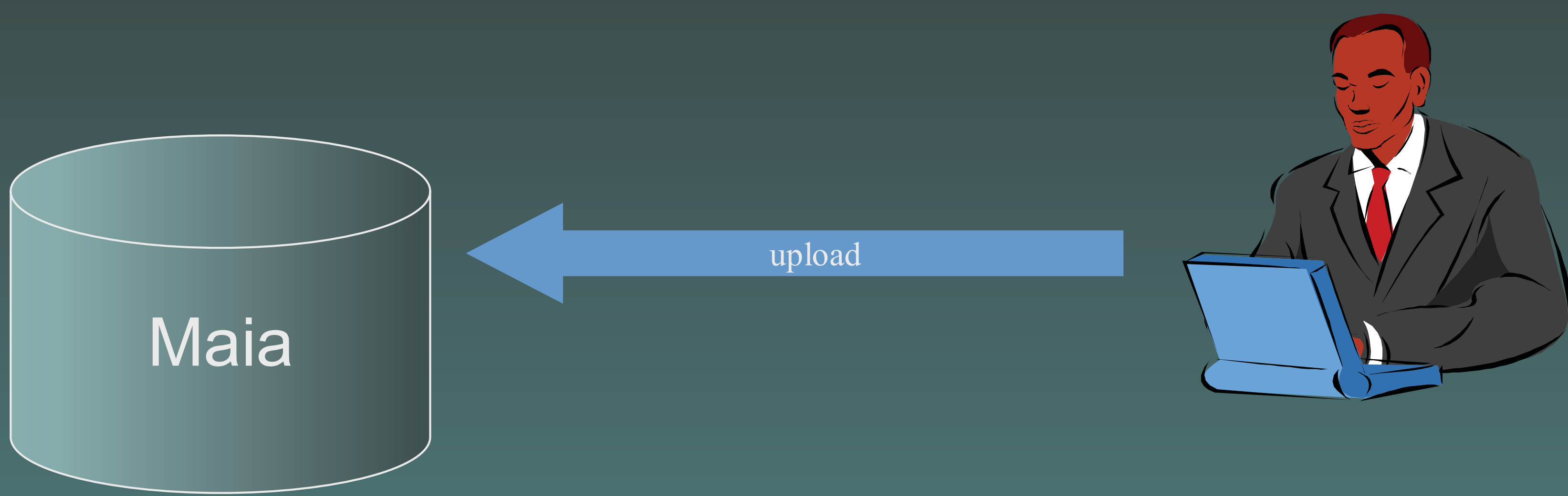


Surveyor browses and downloads working area over LAN or wired WAN

The data capture workflow

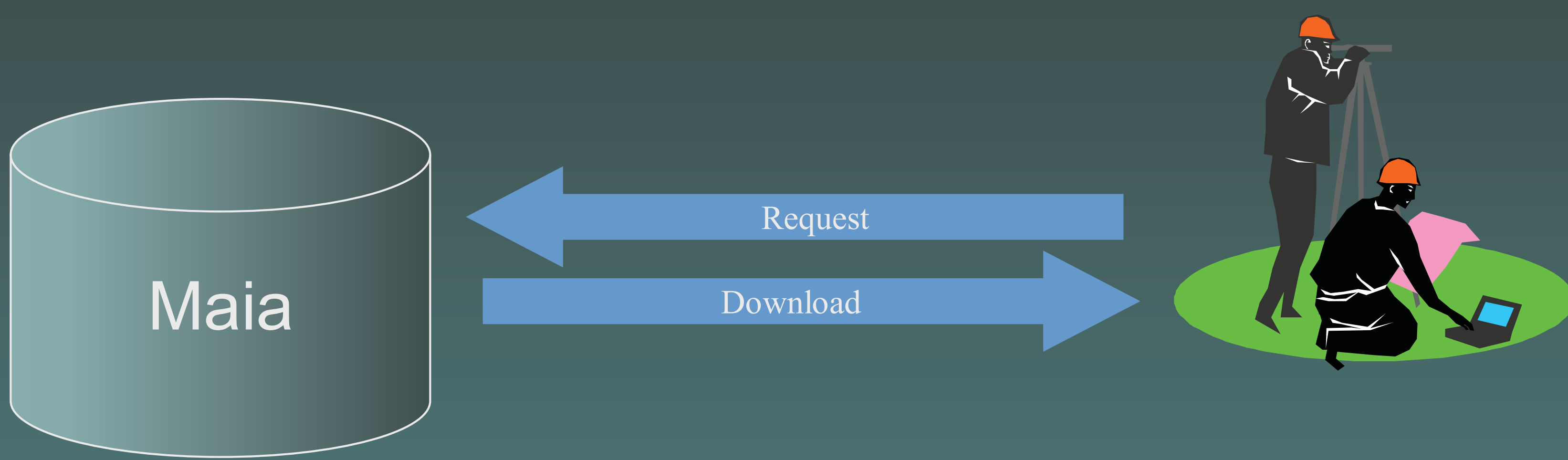
In the field works with local copy of data...

The data capture workflow



Back on the network, surveyor uploads changes only

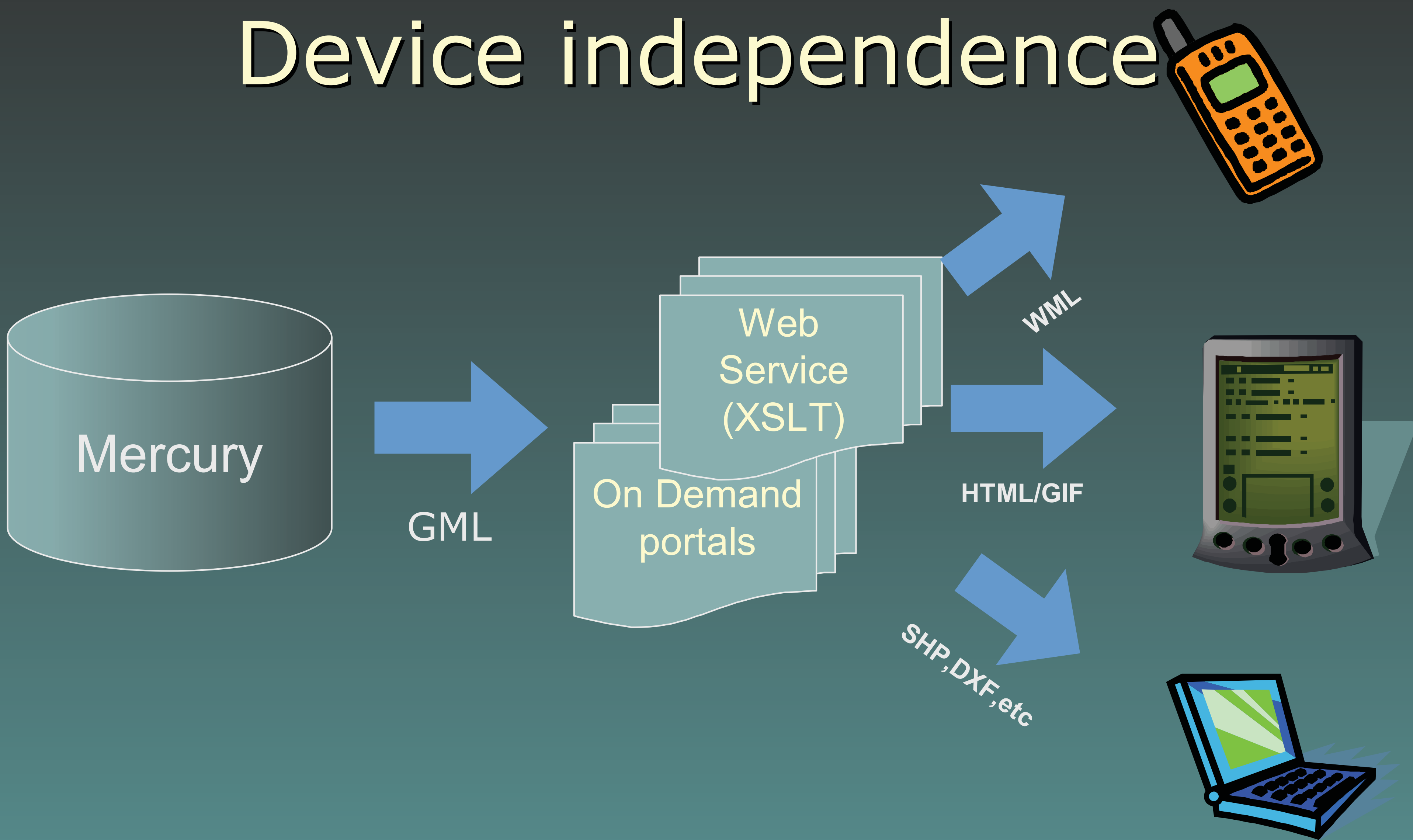
The data capture workflow



Data requested in the field across UMTS networks

Ordnance Survey®
www.ordnancesurvey.co.uk

Device independence



www.ordnancesurvey.co.uk

Some questions

- The role of middleware
- “Open” Geospatial databases
- Common data models
- Web services & federated databases

GML DEV DAYS

GML Developers' Conference 2002

Aim of the conference	Preregistration	Technical Exhibition
Target audience	Pricing	Accomodation
Location and Date	Request For Papers	Press
Conference Program	Workshops	Organizing Committee
Sponsoring		

Ordnance Survey · GeoConnections GéoConnexions Canada · galdos systems inc

Ordnance Survey · ORACLE · OGC Open GIS Consortium, Inc. · INTERGRAPH · Applied Sciences Continuing Education

Thank you

eparsons@ordsvy.gov.uk

Geo-ICT technology push vs. Cadastral market pull

Peter van Oosterom[1]
Christiaan Lemmen[2]
Rolf A. de By[3]
Arbind M. Tuladhar[4]

Key words: Cadastral Systems, OpenGIS Standards, DBMS, UML, GML, XML, MDA

ABSTRACT

This paper discusses the recent Geo-ICT developments, such as information system modelling standards, database technology, global positioning systems, Internet development, wireless communication and acceptance of geometry standards within general ICT tools and its uses on cadastral systems. Efficient design, development, testing and maintenance of cadastral systems allow the introduction of such systems within acceptable time and budgets. A basic condition for system development is analysis of user requirements. Those requirements can change in time, e.g. because of changes in legislation, governmental policy, new tasks for organisations or technology. It is therefore important to design generic and flexible information systems, e.g. to follow the (data) model driven architecture (MDA).

1. INTRODUCTION

Recent developments in Geo-Information and Communication Technology (Geo-ICT) have given tremendous push toward the development of cadastral systems and geo-spatial data infrastructures (GSDI). Both theoretical and practical developments in ICT such as the ubiquitous communication (Internet), data base management systems (DBMS), information system modelling such as Unified Modelling Language (UML), and global positioning systems will improve the quality, cost effectiveness, performance and maintainability of cadastral systems. Further, users and industry have accepted the standardisation efforts in the spatial area by the OpenGIS Consortium and the International Standards Organisation (e.g. the ISO T211 Geographic Information/Geomatics). This has resulted in the introduction of new (versions of) general ICT tools with spatial capabilities; e.g. eXtensible Mark-up Language/ Geography Mark-up Language (XML/GML), Java (with geo-libraries), object/relational Geo-DBMS including support of simple geographic features.

[1] Department of Geodesy, Section GIS technology, Delft University of Technlolgy, the Netherlands. email: oosterom@geo.tudelft.nl
[2] Netherlands Cadastre and Public Registers Agency, the Netherlands and International Institute for Geo-Information Sciences and Earth Observation (ITC), the Netherlands. Emails: lemmen@kadaster.nl or lemmen@itc.nl
[3] International Institute for Geo-Information Science and Earth Observation (ITC), The Netherlands. Email: deby@itc.nl
[4] International Institute for Geo-information Science and Earth Observation (ITC), The Netherlands. Email: tuladhar@itc.nl

It is the first time ever that such a set of worldwide-accepted standards and development tools are available (UML, XML, Geo-DBMS, OpenGIS standards). This creates new perspectives in both the development of new cadastral systems and in the improvement of or extension of existing cadastral systems. At the moment, the first Internet-GIS applications are already operational in a cadastral context. In the near future this will be extended to mobile GIS applications based on cadastral information (sometimes also called location-based services). Imagine the users of mobile phone or personal digital assistant (PDA) such as a civil servant of the municipality, a real estate broker, or a policeman, with their mobile using up-to-date cadastral information for their day-to-day tasks in the field: 'who is the owner of this building?', 'when was this building sold and what was the price?', etc.

On the cadastral market pull side, its new requirements that satisfy users have emerged due to the changes in government policies, legislation, emerging new tasks of the organisations and users, etc. A global overview of such user requirements related to cadastral systems is presented in Section 2 of this paper. In Section 3 an overview is given of recent Geo-ICT developments with a qualitative analysis on cadastral systems. The conclusions of this paper and a proposal on further development of *cadastral* OpenGIS standards are finally given in Section 4.

2. USER REQUIREMENTS

For an inventory of the general user requirements for cadastral systems, the *United Nations/Economic Commission for Europe* (UN/ECE) Land Administration Guidelines (*UN/ECE*, 1996) and UN/FIG the Bathurst Declaration (*FIG,* 1999) give some important and fundamental basic requirements. Furthermore, results from market surveys of the Netherlands Cadastre, have been used for inventory presented at the end of this section.

2.1 UN/ECE Land Administration Guidelines: user needs

The UN/ECE Land Administration Guidelines highlight the importance of addressing user requirements. Before altering an existing system or introducing a new one, it is essential that the requirements of those who will use or benefit from the system are clearly identified. Naylor (1996) relates this to the current market-oriented approach applied to land information. Products and services must certainly satisfy the user needs. The UN/ECE Guidelines state that a user can be anyone who is interested in land matters. A wide variety of user communities will need to be consulted in order to understand their requirements and the constraints under which they currently operate. The assessment of user needs should be made not only at the outset of the development of a new land administration system, but also *throughout its lifetime*. Questions need to be asked about the *categories of data* that will be required in the future. It may be an attractive idea to collect some types of data for some possible use in the future but if it is not necessary to do so at present, then few resources should be allocated for that purpose. A *step-by-step* approach may be more cost-effective.

Further requirements, which are recognised in the UN/ECE guidelines are:

- There will be a need for cooperation over who collects and coordinates data, what technology should be acquired so that *all components of the system are compatible*, how common standards and procedures can be developed, and other system-related decisions.

- *Data protection* has to be covered in a land administration system.
- Are *strata titles* (relating to the ownership of apartments, etc.) to be recognized? This subject has been discussed in a FIG workshop on 3D Cadastres (*Oosterom, P.J.M. van, J.E. Stoter, and E.M. Fendel,* 2001), organised in Delft, the Netherlands. The 3D representation of cadastral data is a typical example of a future need for certain areas in the world.
- The application of new technologies, such as GPS, should be assessed from an economic rather than a technical perspective. Provisions must also be made to accommodate future changes in the network that may occur as a result of technical improvements. These may affect *all coordinate-based systems*. If the coordinates are an essential component of the cadastral system, then the survey technique must be capable of producing these either *directly or indirectly*.
- A key component in any land administration system is the parcel identifier or *unique parcel reference number*. This acts as a link between the parcel itself and *all records related* to it. It facilitates data input and data exchange. Fiedler and Vargas (2001) recognise a technical requirement for cadastral data collection: the need to change the parcel identifier during the data collection process (first related to aerial photographs, later related to the administrative subdivision of the country).
- Orthophotomaps, rectified photomaps, or planimetric maps may be used depending on the user requirements, cost, and timing among other factors.
- *Redundancies* should be avoided.
- The management of an up-to-date land administration system inevitably involves the use of *modern information and communication technology*. It must be able to accommodate new user demands and to take advantage of *new technologies* as they become available. The technology adopted should be sufficiently flexible to meet *anticipated future needs* and to permit system growth and change. In this context, a framework for re-engineering land administration systems is given by Williamson and Ting (2001).
- When data collection starts, it is important that an *updating process* should be installed at the same time. See also Flores Silles, Javier (2001).
- Whilst more and more users require cadastral information that is frequently and quickly updated in real-time, the need to *secure data quality* should not be underestimated. One important aspect here is the management of topology integrated with geometry and other attributes (*Lemmen and Van Oosterom,* 2001).
- There are opportunities for greater cost-effectiveness in areas such as subcontracting work to the private sector; increasing cost recovery through higher fees, sales of information, and taxes; and by *linking* the existing land administration records with a wider range of land information. See also Bogaerts and Zevenbergen (2001).
- The public must understand and accept the level of information that is placed in the *public domain* or else people will find ways to avoid information appearing in the registers. See also Van der Molen (1999, 2000).

These requirements are of a more general nature. Some of them can also be interpreted as conditions for development of stable systems to run for a long time. Here it should be remembered that life-time of data is 50 years or more, of software 10 years or more and of hardware 3 years or more.

2.2 UN/FIG Bathurst Declaration

In the UN/FIG Bathurst Declaration, the importance of ICT for the development of land administration systems is underlined. Information technology will play an increasingly important role both in constructing the necessary infrastructure and in providing effective public access to information. Finally, there must be total commitment to the maintenance and upgrading of the land administration infrastructure. Some of the 'system development'-related recommendations of the Bathurst Declaration are:

- Encourage the *flow* of information relating to land and property between different government agencies and between these agencies and the public. Whilst access to data, its collection, custody and updating should be facilitated at a *local level*, the overall land information infrastructure should be recognised as belonging to a *national* uniform service to promote sharing within and between nations. See also Bogaerts and Zevenbergen (2001) and Williamson and Ting (2001).
- Recognise that good land administration can be achieved *incrementally* using relatively simple, inexpensive, *user-driven* systems that deliver what is most needed for sustainable development.
- Agencies should seek to develop multi-disciplinary, multi-national training courses in land administration and make these available at the local level through the use of modern information technology. Groot and Van der Molen (2000) present similar conclusions as a result of a Workshop on Capacity Building in Land Administration.
- In order to ensure sustainable development of territorial oceans claimed under UNCLOS (United Nations Convention on the Law of the Sea), the United Nations emphasise the need for claimant countries to develop their capability to support effective marine resource administration through the national spatial data infrastructure.

Williamson and Ting (2001) underline the importance of the Bathurst Declaration, as it establishes a strong link between land administration and sustainable development.

2.3 Business requirements: Observations from the Netherlands Cadastre

The information society has become prominently visible over the past few years (*Magis,* 1998). The use of information and communication technology (ICT) for *management, transactions and communication* is becoming increasingly popular. Customers are taking up a much more *directive* role. Organisations are becoming more dependent of each other and are in fact forced to openness (of systems) and exchange (of data). Developments such as *chain-orientation, digitisation* and *new technologies* are leading to the fading of physical product concepts. Information products are becoming flexible combinations of digital data components and additional facilities and services. To be able to operate as a supplier of information products in this changing environment in the long term, an organisation must understand the economic dynamics of information production. Technically, digital information products offer considerably more possibilities *for perfect reproduction and fast, inexpensive and easy distribution.* In addition, it is important to realise that a product does not have the same value to every customer and that as a consequence not every customer is prepared to pay the same price. A pricing policy based on customer-group differentiation is, in the Dutch situation, not feasible (principle for government operation). A pricing policy based on product differentiation is feasible. *Variation* in the product range is possible in many ways: by differentiation in access to the information, for instance (time, place, duration); or by differentiation in the actuality, completeness or extent of detail in the information; or by differentiation in the possibility for the user to download and store the information, to multiply it, print it, or in any possible way edit it. In addition, differentiation is possible in the speed of delivery, in user-friendliness, and in support. The variants can besides be used separately as well as in combination with each other. In view of the specific business characteristics, an information supplier should aim for *standards* (of distribution, exchange and usage) and *product flexibility*.

In the near future, customers want to have access to information *24 hours a day, 7 days a week, at home, in the office, and in the field*. They want to be served in a professional way, through user-friendly tools to information that is timely, up-to-date, reliable, complete, accurate, relevant, if necessary customised, *well-integrated with other relevant data sets of other suppliers*, good value for money by systems that are compatible with the customer's working procedures.

The Netherlands Cadastre's main customer group, the notaries, is becoming more cost-effective. They will have to differentiate their products, deliver their services faster and with higher quality, will have longer opening hours, will have to specialise, have to provide more and clearer information. Citizens want one-stop-shopping (integrated service delivery). And more and more they want to access the information through the Internet. *Electronic conveyancing techniques* such as electronic signatures, encryption, hash values, measures against bit-loss, are applied increasingly. Because the law has to provide legitimacy to the transmission, submission and registration of electric documents, changes in the (land) law are necessary. Expertise to define the new legal prescriptions concerning the authenticity of electronic documents, the certification authorities that are empowered to issue digital keys, is available now (*Van der Molen,* 2001). As land registers and cadastres play an increasing role in the knowledge regarding the legal status of land according to public law (the so-called public encumbrances) as a complement to the status according to private law, the submission and recording of government documents concerning government decisions on land with an effect on third parties, are within reach. This will contribute to the development of *e-government*. Modern *mobile computers* allow updating of (cadastral) maps during the field session and make geometric quality management possible in the field, so that detected errors can be investigated and rectified on the spot. Wireless data communication facilitates the transmission of work files of maps from the field to the office in order to establish an efficient work process. The recording process (throughput) will be improved through internal data communication offering a better integration between centralised and decentralised processes. Workflow management techniques will become applicable, which will have a positive impact on the management of daily fluctuating supply and demand, because an allocation of the workload is possible at the location where the work force is available that very moment. The integration of work processes allows for combining the benefits of centralised IT services and decentralised information management. Not the location but speed of access is important. On the output-side the strategic objective of making the accessibility of land information better, easier and cheaper, will be supported by data communication. A well-organised front office supported by an efficient back office provides a boost in customer-oriented services. Internet services can be applied here. This requires a reflection on opening hours, data quality, liability, data protection and copyright, privacy issues, and pricing policy. Establishing an e-commerce environment will require decisions on to which extent tailor-made land information products are offered, and how payment will be guaranteed. Land administration will become an important basis of establishing a GSDI.

Van der Molen (1999) concludes: users of cadastral information need clarity, simplicity and speed in the registration process. The information must be as complete as possible, reliable (which means ready when required by the users), and rapidly accessible. Finally, the system must be sustainable in order to keep the information up to date.

3. GEO-ICT DEVELOPMENT

Recent developments in Geo-ICT have now important roles for the development of cadastral systems and GSDI surrounding cadastral systems. The developments in ICT in general, and specifically the Geo-ICT can improve the quality, cost effectiveness, performance and maintainability of cadastral systems.

GISs are used within (local, regional, central) governments, utility, and other companies to support their primary business, which often depends heavily on spatially referenced data. Until recently the spatial data management was handled by GIS software outside the DBMS. As DBMSs are being spatially enabled, more and more GISs (Arc/Info, Geomedia, Smallworld, Geographic Microstation) are or will soon migrate towards an integrated architecture: all data (spatial and thematic) are stored in the DBMS. This marks an important step forward that took many years of awareness creation and subsequent system development. Many organisations are currently in the process of migrating towards this new architecture. This is a lot of work and will still take many years. The next step will be the creation of a common GSDI for related organisations; the so-called information communities. This can replace, in the long run, the exchange of copies of data sets between organisations. It requires good protocols, standardisation such as the OpenGIS (*Buehler, K. and L. McKee*, 1998) web mapping specification. But also the role of the Geo-DBMS becomes more important, because not a single organisation depends on it, but a whole community. The main use will be query-oriented (and less update-oriented, only the owner of the data is doing updates, others are only doing queries). This also means that Geo-data derivation (creating new spatial data sets out of existing ones) is an important component that uses advanced class of queries. Here, an essential component for realizing such query processes is the network infrastructure (bandwidth) itself.

In this section, we will first discuss the broader concept of the Geo-Data Infrastructure and relate this to cadastral systems in 3.1. Next, we describe the relevant OpenGIS standards in Section 3.2. This is followed by an analysis of the developments in modelling and structured information transfer (in 3.3). Section 3.4 shortly presents the developments in database technology. Finally, Section 3.5 introduces the concepts and technologies behind the location based services and again relate this to cadastral systems.

3.1 The Geo-Data Infrastructure

During the 5th Geo Spatial Data Infrastructure Conference (*Resolutions*, 5th GSDI Conference, 2001), the following definition of GSDI was agreed on by the GSDI Steering

Committee: “The Global Spatial Data Infrastructure supports ready global access to geographic information. This is achieved through the co-ordinated actions of nations and organisations that promote awareness and implementation of complimentary policies, common standards and effective mechanisms for the development and availability of interoperable digital geographic data and technologies to support decision making at all scales for multiple purposes. These actions encompass the policies, organisational remits, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure that those working at the global and regional scale are not impeded in meeting their objectives”. The processes of production, provision, use, maintenance, exchange and sharing of these data are complex in nature. Large data volumes have to be managed using database management technology and geographic information systems (GIS). The managers of geo-data are not only the custodians of cadastres, but also national mapping agencies, geological surveys, soil surveys, ministries, land use planning institutes and large municipalities.

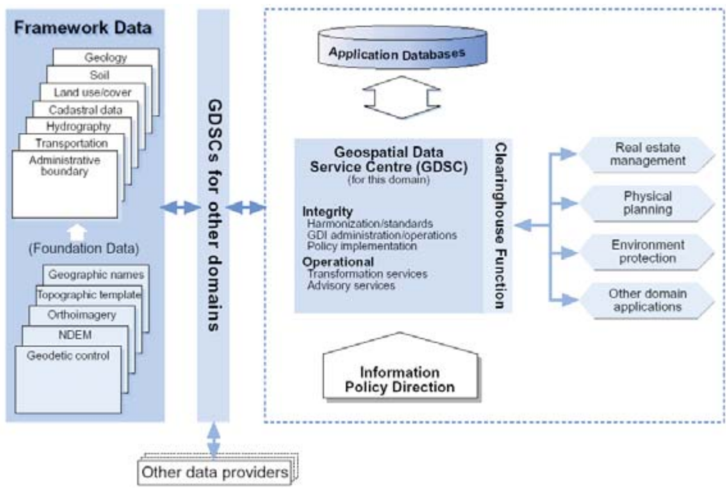


Figure 1: The foundation data for Geo Spatial Data Infra-structures in relation to the Geo Spatial Data Service Centre of the Environment and Physical Planning Domain (source: Groot and McLaughlin, 2000*).*

The success of the Internet in general has shown the power of an open infrastructure. The open standards and the decentralised architecture are responsible for the many free and non-free services. Besides the network infrastructure (wired and mobile), the GSDI (Figure 1) can be seen as composed of three important and quite different types of ingredients (*Van Oosterom et al,* 2000, *Groot and McLaughlin,* 2000):

- *Geo-data sets* in different domains. *Framework data sets* like cadastres, but also coverage data pertaining to soil, land use, hydrography, geology and transportation are all necessary tools of effective government. Framework data provide information on people and the land where they live and work. They supply information on the location of administrative boundaries and of objects like buildings and roads. They provide information on type of soil and pollution, ownership (land tenure), value and use of the land as well as geological information. Framework data/information can help governments to determine how to deal with land in their policies to combat poverty, to achieve sustainable settlement goals and to manage natural resources. A special sub set of geo-data are *foundation data*. This is the fundamental geographical reference for all other thematic application data. Foundation data concern Geodetic Control, National Digital Elevation Model, Ortho Imagery, the Topographic Template and Geographic Names. All these data sets should be well defined with respect to their data model, thematic contents, quality, accuracy, actuality, and so on.

- *Geo-Data Services* in general and the geo-DBMS specifically. The Geo-Data Service Centre (GDSC) harmonises/standardises all data for its application domain. It ensures they are described in a national metadata standard to facilitate the sharing of these resources by other potential users. The GDSC also enforces the information policies that control access, use and planning, in keeping with legislation and overall government policy. The geo-data sets are maintained in geo-DBMSs and are served to users from these geo-DBMSs via networks or traditional means. For these purposes, the DBMS has to support spatial data types and operators (simple analysis and selection-oriented queries), spatial clustering and indexing (for large data sets), and if possible support for advanced analysis (topology-based analysis). Also, temporal support is required in the form of some kind of future standard TSQL (*Snodgrass, R.T., I. Ahn and G. Ariav,* 1994) or in the meantime through some other (non-standardised) means: extension for a specific DBMS or explicitly in the application data model.
- *Interoperability standards* are required to enable the integration of the different data sets and to combine the geo-data processing services. In fact, different organisations and individuals using each other's geo-data sets in a digital environment can be regarded as parts of a distributed computing environment. One of the most obvious examples of this is an internet GIS retrieving *on-the-fly* data from different sources on the internet. To operate in such a heterogeneous world (different types of hardware, operating systems, geo-DBMSs, geo-data sets) interoperability standards at many levels are required.

All three ingredients have different aspects, which can be either technical or non-technical (organisational, financial, legal, etc.).

One of the most time-consuming tasks when implementing a GIS is obtaining geo-data. First, relevant data sets and sources have to be located, then these data sets have to be copied and converted into the local (DBMS of a) GIS. Some reasons why this process is so time-consuming are that it may be difficult to find the data, the data model of the source may be very different from the model implemented by the local system, the supported exchange formats of source and destination are different. To improve this situation, much effort is needed to create a GSDI, which should at least cover the following aspects: *consensus* on the geometric parts of the data model; support for both raster and vector data (including different spatial reference systems); a formal description of the geo-data sets (and geo-processes), that is, metadata standards, covering both the spatial and non-spatial aspects; *access to and query* the metadata and how the result of such a query is returned, this is called catalog services; selection of the geo-data; *format (and transfer)* of the resulting geo-data set. Instead of always copying data sets from one system to another, a new scheme becomes feasible once the above aspects are covered by implementations of geo-standards. This new scheme allows to keep the data at the source, which can then be used all over the world. At the client side, no data management is needed for data originating from other sources. This scheme also allows fair pricing of geo-data, because every time data from the source is used (possibly through a local cache) the user can be charged for this. Currently, in the *full data set copy* scheme the user has to pay for the whole data set, even if certain data (regions) are not used at all (in a certain period). The new *data at the source* scheme allows fairer pricing, both viewed from the vendor's and buyer's points of view.

An early pilot in this area started in 1996 and was called Geoshop (*Berg C. van den et al,* 1997*)*. The following partners were involved: the municipality of Almere, a cable TV company (Casema), and the Cadastre. It was based on the C++ Magma (server) and Java Lava (client) software developed by Professional GEO Systems (PGS); see Figure 2.

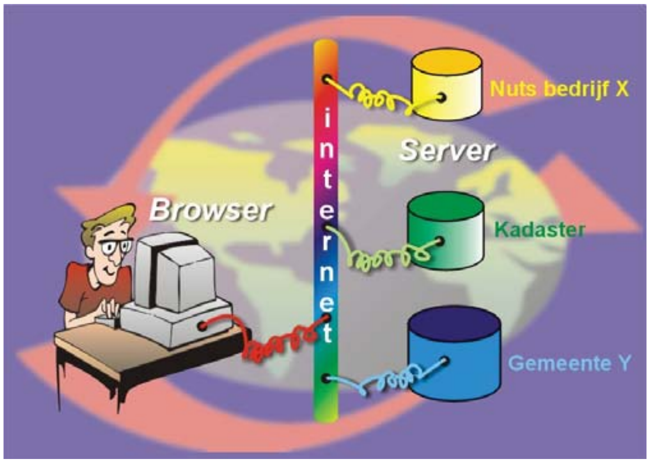


Figure 2: Geoshop example of Internet GIS.

Important characteristics are: access data form multiple and heterogeneous sources (Ingres, Informix, DXF files), raster and vector data at server and client side, client side software platform independent (Java platform). No attempt was made to implement the payment aspects as this was expected to be a more generic issue, also treated outside the GI community. Also, no standards for querying (meta) data and for returning the results were available at that time (1996).

3.2 OpenGIS Standards

The OpenGIS Consortium (OGC) and the official standardisation organisations (ISO and CEN) have addressed several aspects of the interoperating framework. In this section, we will focus on the OGC as they also (re-)use official ISO standards when appropriate. OGC has basically two levels of standards: abstract (comparable to 'official' standards) and implementation standards. Implementation standards describe the exact interfaces (protocols) of a (part of) an abstract standard in the context of a specific distributed computing platform. An overview of the OpenGIS domain can be found in the OpenGIS Guide (*Buehler, K. and L. McKee,* 1998). In Section 3.2.1, the feature geometry data model will be discussed. The next section covers the aspects of metadata and catalog services. Finally, Internet GIS standards are described in Section 3.2.3.

3.2.1 Feature Geometry

The first OpenGIS implementation standard was related to the feature geometry abstract specification. It was called Simple Feature Specification (SFS) and it standardised the basic spatial types and functions. The implementation specification for the SFS are described for three different platforms: SQL (*OpenGIS Consortium,* 1998), Corba, and OLE/COM. Currently, OpenGIS is revising/changing the feature geometry abstract specification to be consistent with the draft standard ISO TC 211 *Spatial schema* (*ISO TC 211/WG 2,* 1999): covering also 3D types, more geometric primitives (curve and surface types), and complex features (topology). What is still missing is the implementation specification of topology for specific platforms comparable to the implementation specification for simple features. What the abstract *feature geometry* and implementation *simple feature* specifications are for the vector model, are the abstract *earth imagery* and the implementation *grid coverage* (*OpenGIS Consortium,* 2000b) implementation specifications for the raster data model.

3.2.2 Metadata and catalog services

With respect to the contents and structure of the metadata, the OGC decided more or less to adopt the work by ISO TC 211 (*ISO TC 211/WG 3,* 1999). Much attention of the OGC has been paid to the related aspect of catalog services (*OpenGIS Consortium,* 2000a), describing how to access the metadata (and also data describing available computing services). This standard can and will be used in realising clearinghouses for geo-information all over the world; e.g., the new NCGI in the Netherlands (*Absil et al*, 1997) will be based on this OpenGIS standard.

3.2.3 Internet GIS

Strongly related to the previous standards, metadata and catalog services are the activities in the area of Internet GIS (or web-mapping). This may be seen as an interactive (and ultimate) form of interoperability as data from multiple sources can be retrieved and combined in the web-browser. However, instead of querying and receiving metadata, now the geo-data itself is queried and received. The OGC has created in this area:

- the implementation specification *Web Map Server Interface* (*OpenGIS Consortium,* 2000d) for the query aspects using three basic functions: GetCapabilities (what is available on the server), GetMap (raster images, graphic primitives or data) and GetFeature_info (fetch attributes), in addition the *Web Feature Server Interface* can be used for updating information and has functions for locking and committing transactions, and
- the recommendation *Geography Markup Language* (GML) (*OpenGIS Consortium,* 2000c) (simple features in XML with XSLT 'stylesheet' for presentation) for vector data transfer.

The query from a client uses the well-known structure of an Internet URL. The OGC has specified the names of the query parameters (e.g., BBOX, LAYERS, FORMAT) and the meaning and allowed values for these parameters. An example of a GetMap query:

```
http://b-map-co.com/servlets/mapservlet?WMTVER=0.9&REQUEST=map&
BBOX=-88.68815,30.284573,-87.48539,30.989218&
WIDTH=792&HEIGHT=464&SRS=4326&
LAYERS=AL+Highway,AL+Highway,AL+Highway&
STYLES=casing,interior,label&FORMAT=GIF&TRANSPARENT=TRUE
```

With respect to the different types of clients OGC has developed a model to compare these clients. See Figure 3: thin clients (which display only raster images JPEG and PNG), medium clients (with graphic primitives WebCGM and SVG) or thick clients (data in the form of simple features XML, that is GML (*OpenGIS Constortium,* 2000c), is processed at the client side).

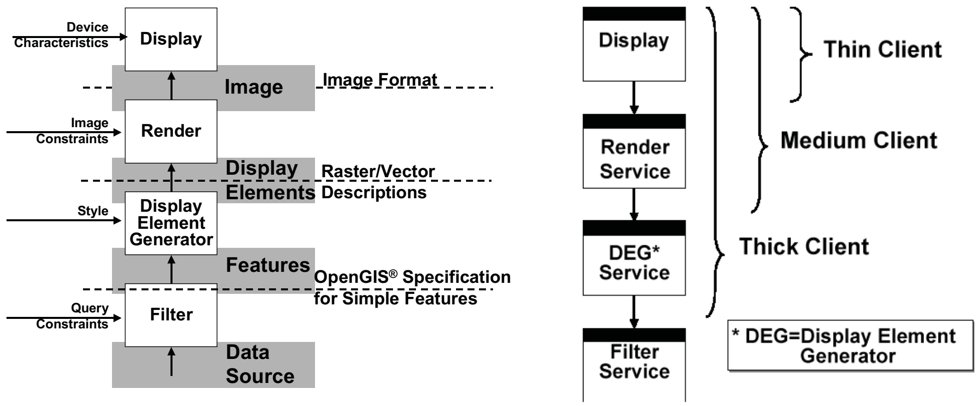


Figure 3: Different levels to exchange geo-information (OpenGIS Consortium, 2000d*).*

It is clear that when a certain process is not executed at the client side, the work has to be done at the server side. For example, generation of graphic primitives from GML data and/or convert graphic primitives into images (rendering). The FORMAT parameter in the query indicates what type of result should be sent back; in the example above a GIF image is requested. This could also have been a request for GML data. Below a fragment of a GML data sets is shown, taken from the domain of topographic mapping (*Vries, De et al,* 2001). The Topographic Service of the Netherlands has enriched a standard topographic map (1:10,000), which was converted into a GML (2.0) prototype by TU Delft. The format and structure resemble the well known HTML format where everything has a begin and an end tag: the whole object (tdn:SpoorbaanDeel), thematic attributes (tdn:begindatum) and geometric attributes (gml:Polygon).

```
<tdn:SpoorbaanDeel fid="TOP10.4200001">
 <tdn:top10_id>4200001</tdn:top10_id>
 <tdn:begindatum>06 Jul 2001 08:08:24</tdn:begindatum>
 ...
 <tdn:verkeersgebruik>Tram</tdn:verkeersgebruik>
 <tdn:aantal_sporen>1</tdn:aantal_sporen>
 <gml:geometryProperty>
  <gml:Polygon srsName="EPSG:7408">
   <gml:outerBoundaryIs>
    <gml:LinearRing>
     <gml:coordinates>
      191008.456,447232.635,0.0 190990.713,447236.938,0.0 190972.849,447239.952,0.0
      190955.904,447235.469,0.0 190940.491,447231.646,0.0 190923.831,447229.355,0.0
      190924.668,447229.093,0.0 190942.211,447223.787,0.0 190944.282,447224.343,0.0
      190957.890,447227.719,0.0 190973.223,447231.776,0.0 190989.103,447229.096,0.0
      191006.570,447224.861,0.0 191008.456,447232.635,0.0
     </gml:coordinates>
    </gml:LinearRing>
   </gml:outerBoundaryIs>
  </gml:Polygon>
 </gml:geometryProperty>
 <tdn:hoogteniveau>0</tdn:hoogteniveau>
</tdn:SpoorbaanDeel>
```

3.3 Unified Modelling Language and eXtensible Markup Language

For the several decades, the different modelling techniques have been employed in the design and development of information systems. Now, we have reached the stage of a world-wide acceptance of a standard modelling language: the Unified Modelling Language (UML) *Booch et al*, 1999). It supports a rich set of graphical notation describing classes, objects, activities, states, workflow, use case, components, nodes and the relationships among them. It provides significant benefits to system designers and organization by building rigorous, traceable and maintainable models, supporting development lifecycle. It is used for modelling both the data aspect (structural) and the functional aspect (behavioural) of information systems supporting both external and internal requirements. Thus, UML models can be used to describe and implement various components and their links of cadastre and land registration systems within the scope of management framework for developing information systems and their business processes (Tuladhar, 2002). There are three most important components in UML: use case diagram to capture external environments (user requirements and behaviours), activity diagram to show how use case can be realized, object model (system behaviours) showing interaction between actors and entity objects, and information model (commonly known as 'class diagram', which resembles the well-known entity-relationship diagrams). Since it is based on object-oriented technology, it can effectively be used to design a generic model. To this end, the simple and generic cadastral system can be developed on the top of Simple Feature Specification (refer to section 3.2.1) using UML.

An international consortium promoting the use of object-oriented information technology is maintaining the UML standard: the Object Management Group (OMG).

eXtensible Markup Language (XML) is the standard for the exchange of structured information and plays an important role in the Internet. Data models can be described within either a Document Type Description (DTD) or, more advanced an XML Schema document (*W3C, 2000a, W3C,* 2000b). XML documents must obey some basic rules: they must be 'well formed' (have corresponding begin and end tags) and 'valid'. An XML document is valid when its structure is conform to the definitions and declarations in the model (given in DTD or XML Schema). The XML, DTD and XML Schema standards are developed and maintained by the world-wide web consortium (W3C). XML Schema has a 'connection' to UML: the data part of the (UML) model can be transformed into the XML Schema (by hand or automatically). Many tools exist to manipulate the XML data and know how to interpret the models. The advantages of XML in general are that it is well readable by humans and machines (in contrast to binary formats), international (support of Unicode for non-western languages), methods to process XML documents are available (e.g. develop an XML Style Sheet Transformation, XSLT to convert a DLM into a DCM), extensible with own parts (using the 'XML Schema' language (*W3C,* 200a and *WC3,* 2000b)). Moreover it is very well supported by all kinds of software in the market (ranging from the web-browser to the DBMS).

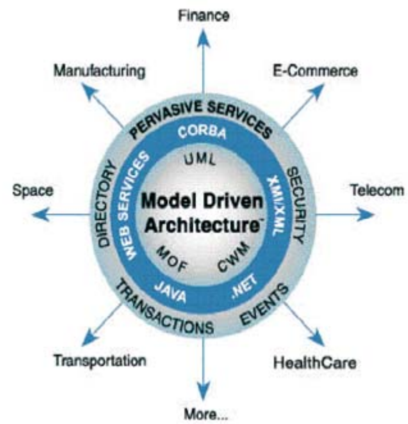


Figure 4: OMG's Model Driven Architecture (from Siegel, 2001*).*

As stated in the user requirements cadastral systems need to be generic and flexible because of the changing requirements over time. Flexible information systems are also one of the main motivations behind the model driven architecture (MDA) also promoted by the OMG (Siegel, 2001). The MDA is based on models of information systems (components) being described in UML. Other advantages of the MDA approach, specifically for today's highly networked, constantly changing systems environment, are: portability, cross-platform interoperability, platform independence, domain specificity, and productivity. Figure 4 shows the MDA in relationship with the different technologies being incorporated (including UML) and the relationship with the different domain specific models.

Now, returning to our specific domain, the spatial or geographic information of which the cadastral data form a sub set. There is a growing demand to distribute the geo-information in a more open transfer format. Geography Markup Language (GML) and its underlying technologies, OpenGIS specifications and OMG and W3C standards, have been introduced above. Now we will show some aspects in a little more detail. Two XML Schemas describe GML itself: the geometry schema and the feature schema. The UML schema belonging to the geometry schema is shown in Figure 5. With version 2.0 the OpenGIS Consortium decided only to use XML Schema to convey the structure of GML files (and not DTD anymore). The GML specification is basically a set of XML Schema documents with element declarations and type definitions plus a hierarchical structure for the relationships between types. In this way the specification offers a framework that can be used by organisations to make their specific XML application schemas for their GML implementations.

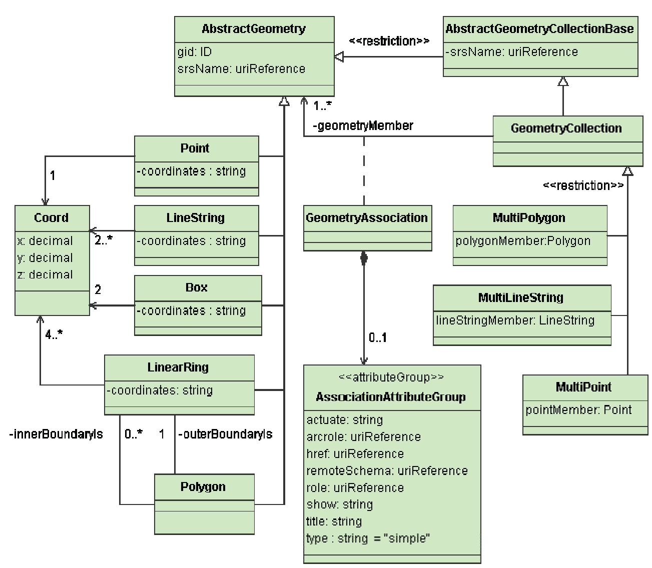


Figure 5: The UML schema of the GML Geometry model taken from (OGC, 2000c).

3.4 Developments in Database Technology

Both developments in hardware and software and in database technology will influence the future development of the Geographic Information Infrastructure. Current extensible DBMSs (*ASK-OpenIngres 1994, ESRI,* 1999) are very well capable of storing spatial data. Also simple queries like zooming in and zooming out can be handled efficiently. However, more complex operations, like map overlay, on the fly generalisation, enforcing correct topology during updates etc., are still not within reach of these systems. New developments in DBMS technology that can help to build a new generation of spatial DBMSs are for example Extensible (Object Relational) DBMSs, Object Oriented DBMSs, and Very Large Memory (VLM) Databases. The current status of the mainstream DBMSs with respect to handling geo-information can be found in the Appendix 2.

3.5 Location Based Services

Mobile information society is developing rapidly as mobile telecommunications moves from second (GSM) to third (UTMS) generation technology. The Internet and its services are coming to wireless devices. Location-based services (LBS) personal navigation are parts of mobile multimedia services. Personal navigation is a service concept in which advanced mobile telecommunications allow people to find out where they are, where they can find the products and services they need and how they can get to a destination. (*Rainio*, 2001).

Figure 6: The transmission senders and the positioning of the mobile phone.

The architecture of LBS consists of components from three different disciplines:

1. positioning of the mobile terminal, either based on the mobile phone network or on a positioning system such as GPS, GLONASS or Galileo;
2. wireless communication network, either based on GSM, GPRS or UTMS;
3. geo-information and geo-services based on GIS technology.

The three disciplines have one thing in common: in one way or the other the concept of location is important.

The supply of services is visible to the users as different service applications, in the background of which can be generic services and technologies, like data management, customer administration, data security, etc. Service applications can support all modes of transport (walking, skiing, vehicles such as the wheelchair, bicycle, motorbike, skidoo, car, taxi, bus, train, ship, plane, etc.). The services can include address, route, timetable, weather, accommodation, restaurant, and other guidance, traffic and travel services, as well as a description of any commercial and public services to be positioned. The nature of the information can be travelling directions, historical and cultural information, programme and event information, official regulations, etc. Safety is enhanced by the automatic transmission of the terminal device's location when making emergency calls. Positioning and the transmission of location data can also be applied with different kind of orders and when trying to locate a lost person, animal or object (*Rainio,* 2001). The use of mobile phones and personal digital assistants (PDAs) is growing rapidly. Wireless data transfer, mobile multimedia and mobile Internet are the main trends in mobile telecommunication. The third generation network, UMTS (Universal Mobile Telecommunication System) will provide 5-10 times more efficient data transfer links (300 kbit/s) than in conventional Internet modem use. Mobile data networks are also being joined by Wireless Local Area Networks (WLAN 11 Mbit/s) and other wireless data transfer systems (Bluetooth 770 kbit/s); although their coverage is limited to one hundred metres. One directional data transfer for digital television and radio offers up to 20 Mbit/s. This could be partly utilised for other data transfer.

4. PROPOSAL AND CONCLUSIONS

In this paper, we have first tried to get an overview of the general user requirements of cadastral systems based on the work done by the UN/ECE and the UN/FIG. This was augmented with user requirements from the Dutch Cadastre. One thing became very clear: cadastral systems are dynamic; they do have to change over time in order to support society in a sustainable manner. With these user requirements in the back of our mind we did analyse the recent Geo-ICT developments in a broader sense, that is, including as information system modelling standards, database technology, positioning systems, Internet development, wireless communication and acceptance of geometry standards within general ICT tools, on cadastral systems. It can be concluded from this analysis that the development and maintenance of the cadastral systems can benefit a lot from the new Geo-ICT and even completely new functions are now becoming possible (e.g. LBS).

However, in spite of the now available standards (for modelling UML), exchanging structured information (XML) and geo-information standards (OpenGIS Simple Features, Web Map servers, GML, etc.), there is still one important aspect missing. This is a standard and accepted base model for the cadastral domain. This should include both the spatial and administrative part and be based on the above mentioned core standards. Within the OpenGIS consortium there are several special interest groups (SIGs) working on generic domain models on which specific applications can be founded by assembling parts adhering to this domain model. The generic domain models itself are based on the core technology models (such as for geometry, time, meta data, etc.). The standardised cadastral domain model should be described in UML schemas and accepted by the proper international organisations (FIG, OpenGIS...). This will enable industry to develop products. And in turn this will enable cadastral organisations to buy these components and develop (and maintain) systems in an even more efficient way. More than two years ago the Technical Committee (TC) of the OpenGIS Consortium tried to set up a 'Land Title and Tenure SIG' without success. This in spite of several other successful domain SIGs within the OGC, such as Telecommunications, Defence and Intelligence, Disaster Management, Natural Resources and Environmental, etc. It is time to join forces between the FIG and the OpenGIS Consortium and start working a standard and accepted cadastral base model. This model can be used in (nearly) every country. Of course, on top of this cadastral base model, parts of the system may be added for specific situations in a certain country. That is, the model can be extended and adapted according to the theory of object-oriented systems.

Acknowledgements

The authors would like to thank Elfriede Fendel and Wilko Quak for their critical review of the draft version of this paper.

REFERENCES

Absil van de Kieft, I. and B. Kok, 1997, 'The development of a geo meta data service for the Netherlands'. In Third Joint European Conference & Exhibition on Geographical Information (JEC-GI'97)*,* pages 1165-1176, April 1997.

ASK-OpenIngres 1994: 'INGRES/Object Magemenent Extention', User's Guide, Release 6.5.

Berg C. van den , F. Tuijnman, T. Vijlbrief, C. Meijer, H. Uitermark, and P. van Oosterom, 1997: 'Multi-server internet GIS: Standardization and practical experiences'. In: International Conference and Workshop on Interoperating Geographic Information Systems, Santa Barbara, California, USA, December 3-4 and 5-6, 1997, interop'97, pages 365-378, Boston, 1997. Kluwer Academic Publishers.

Bogaerts, T. and J. Zevenbergen, 2001, 'Cadastral Systems – alternatives', Theme Issue 'Cadastral Systems', p. 325-337, Volume 25, number 4-5, 2001, Elsevier Science, New York.

Booch, Grady, James Rumbaugh, and Ivar Jacobson, 1999. The Unified Modeling Language. User Guide. Addison-Wesley, 1999.

Buehler, K. and L. McKee, 1998, 'The OpenGIS guide - Introduction to interoperable geoprocessing'. Technical Report Third edition, The Open GIS Consortium, Inc., June 1998.

ESRI, 1999, 'Managing Spatial Data: The ESRI Spatial Database Engine for Informix'.

ISO TC 211/WG 2, 1999, 'Geographic information - Spatial schema', Technical Report second draft of ISO 19107 (15046-7), International Organization for Standardization, November 1999.

ISO TC 211/WG 3, 1999, 'Geographic information - Meta data'. Technical Report draft of

ISO 19115 (15046-15), International Organization for Standardization, June 1999.

Fiedler, Theodor, and Vargas, Ronald, 2001, 'Rural Cadastre in Chuquisaca, Bolivia', GIM International, 2001-7.

FIG Publication No. 21, 1999 'The Bathurst Declaration on Land Administration for Sustainable Development', Bathurst, Australia, 1999.

Flores Silles, Javier, et. al., October 2000, 'Proposal of a maintenance model for cadastral systems in Bolivia, processes hand book', and: 'Foundation and references for the construction of the model for maintenance of the cadastral information system', published by Kadaster International, La Paz, Bolivia.

Groot, Richard and John McLaughlin, 2000, ‘GeoSpatial Data Infrastructures, concepts, cases and good practice’, Oxford University Press, Oxford, United Kingdom.

Groot, Richard and Paul van der Molen (ed.), 2001, ‘Workshop on Capacity Building in Land Administration for Developing Countries’, ITC, Enschede The Netherlands.

Lemmen, C., and P. van Oosterom, 2001, ‘Spatial Data Management on a very large cadastral database’, in: ‘Computers, Environment and Urban Systems’, Theme Issue ‘Cadastral Systems’, p. 509-528, Volume 25, number 4-5, 2001, Elsevier Science, New York.

Magis, M.C.D. RI, 1998, ‘Developing a customer-oriented Cadastre’, The Netherlands Public Registers and Cadastre Agency, UN/ECE Workshop on managing and developing effective land registration and cadastral services, London, United Kingdom.

Molen, Paul van der, 2001, ‘Datacommunication: a lifeline between land administration organisations and society’, FIG Working Week and Technical Conference, Seoul, Korea.

Molen, van der, P., 1999, ‘User demands and the application of new technologies’, ECE Workshop on Modern Cadastre and Land Registration Systems, Bonn, Germany.

Naylor, J, 1996, ‘Operations Management’, Pitman Publishing, London, United Kingdom.

Oosterom, P. van, W. Quak, T.P.M. Tijssen and E. Verbree, 2000, The architecture of the Geo-Information Infrastructure. In Proceedings of UDMS 2000 22nd Urban and Regional Data Management Symposium, Delft, the Netherlands, 11-15 September 2000, p. II.9-II.18.

Oosterom, P.J.M., J.E. Stoter, and E.M. Fendel (ed.) 2001, 3D Cadastres, Registration of properties in strata, Proceedings International Workshop on “3D Cadastres”, FIG, Delft, November 2001.

OpenGIS Consortium, Inc. 1998, ‘OpenGIS simple features specification for SQL’, Technical Report Revision 1.0.

OpenGIS Consortium, Inc, 2000a, ‘OpenGIS catalog interface implementation specification’ Technical Report version 1.1 (00-034), OGC, Draft.

OpenGIS Consortium, Inc., 2000b: ‘OpenGIS grid coverage specification’, Technical Report Revision 0.04 (00-019r), OGC.

OpenGIS Consortium, Inc., 2000c: ‘OpenGIS recommendation - Geography Markup Language (GML)’ Technical Report version 1.0 (00-029), OGC.

OpenGIS Consortium, Inc., 2000d: ‘OpenGIS web map server interface implementation specification’, Technical Report reversion 1.0.0 (00-028), OGC.

Rainio, Antti, 2001, 'Location Based Services: Location-based Services and Personal Navigation in Mobile Information Society', FIG Working Week, Seoul, Korea.

Resolutions, 5th GSDI Conference, May 21-24, 2001, Cartagena, Colombia.

Siegel, Jon and the OMG Staff Strategy Group (2001), Developing in OMG's Model-Driven Architecture. Object Management Group White Paper, November 2001 (rev 2.6).

Snodgrass, R.T., I. Ahn and G. Ariav, 1994, TSQL2 language specification. SIGMOD Record, 23(1):65-86.

Tuijnman, F. and P.J.M. van Oosterom, 1998, Geoshop, een multi-server internet-GIS (in Dutch). Geodesia, 40(4):165-174.

Tuladhar, A.M., 2002, "Developing a framework for cadastre and land registration systems in land administration organisations", presented paper on Joint session commissions 1 and 7, FIG XXII International congress, Washington, D.C. USA, April 19-26 2002

UN/ECE, 1996, *United Nations/EconomicCommission for Europe*, 'Land Administration Guidelines', Geneva, Switzerland.

Vries, Marian de, Theo Tijssen, Jantien Stoter, Wilko Quak, and Peter van Oosterom (2001), The GML prototype of the new TOP10vector object model, Technical Report GISt No. 9, TU Delft, Faculty CiTG, Department of Geodesy, December 2001.

Williamson, I. and L. Ting, 2001 'Land administration and cadastral trends – a frame work for re-engineering', in: 'Computers, Environment and Urban Systems', Theme Issue 'Cadastral Systems', p. 339-366, Volume 25, number 4-5, Elsevier Science, New York.

W3C, 2000a, XML Schema part 1: Structures and XML schema part. Technical report, World Wide Web Consortium, October 2000. Candidate Recommendation.

W3C, 2000b, XML Schema part 2: Datatypes. Technical report, World Wide Web Consortium, October 2000. Candidate Recommendation.

Appendix 1: RDBMS Spatial Data Management

This is a summary of a Product Survey published in GIM International, 2002-2 (Editor: Christiaan Lemmen).

Questions	***Answers by*** **Oracle** (Spatial Option), **Computer Associates** (Ingres II, Spatial Object Library), **IBM Informix** (Informix Dynamic Server) and **Sybase** (Adaptive Server Enterprise 12.5, with Spatial Query Server Option)
Supported spatial data types (to be applied direct in SQL)	
Spatial data-types (vector oriented) supported by your RDBMS	For products all spatial data types supported in Simple Feature Specification + circles, arcs, combinations of arcs and lines and rectangles
Spatial data-types (raster oriented) supported by your RDBMS	Not supported, only some 'output formats' like jpg, gif, bmp, png, geotiff
Supported spatial data operators (to be applied direct in SQL)	
Could you give an overview of spatial data-operators (vector oriented) supported by your RDBMS	A wide range of (OpenGIS-ISO/SQLMM compliant) spatial operators and functions is available in all products.
Could you give an overview of spatial data-operators (raster oriented) supported by your RDBMS	Not supported, except classical way of storage in Binary Large Objects
Spatial indexing	
Is there a specific spatial index supported by your RDBMS for fast data retrieval	Yes
If answer is 'Yes': which spatial index	Support for Quad-tree (most products) and R-tree (most products), sometimes a B-tree is used in a smart way
Spatial data clustering	
Is spatial data clustering supported for better access performance? Spatial data clustering means spatial data storage is organized in such way that spatial objects which are 'near' to each other on disk	Not really.
If yes: which method of spatial data clustering is supported	One product uses Hilbert functions
Support in topology	
Is there any support in storage of topologic relationships	Not supported
Is topologic structure management supported to support the realisation of a planar partition (e.g. cadastral parcels) or linear networks	Not supported
Do you support 'Clementini' and 'Egenhofer' operators	More and more products have support of more and more of these operators.
3D GIS support	
Can 3d GIS be supported by your RDBMS	Support depends on GIS tool
Can 3D coordinates be stored in your data types?	One product supports for 3D storing of lines/points/polygons. R-Tree supports 3D indexing, but no 3D operators. Linear Referencing can be 4D. Other products also support 3D representation but not for all data-types

Do you support spatio-temporal models? E.g.: can you support maintenance of history (reconstruction of the past) with functionality in your RDBMS?	Most products support spatio-temporal models
If yes: could you summarize (max. 100 words) how this is organized	One approach: it works by maintaining versioned tables, which implies a.o. that the version will always be part of the primary key. Workspace Manager can consolidate versions, by tracking changes to handed out copies. Other approach: since these are relational data-types, we can maintain a time attribute as another column in the table containing the spatial data.
Internet-GIS	
Can you support Internet-GIS applications with your RDBMS	Yes
VLM	
Very Large Memory (VLM) is not only important for spatial data management support. But because of the performance of a RDBMS in a VLM environment it is important to know if your RDBMS runs in VLM environment	Yes
Other	
Please give an overview of specific advantages or your product for management of large spatial data sets	A selection of replies: • Coordinate System Support, • Whole Earth indexing. Spatial data can be portioned • Performance • Ease of data management • Storage of Geodetic information
Could you give an overview of strategic partnerships within the GIS industry	Most RDBMS providers have strategic partner ships with GIS suppliers
Is your company a member of the Open GIS Consortium	Yes for one RDBMS supplier, No for others
For which specification is your RDBMS Open GIS compliant	One supplier replies: Several specifications, including: Simple feature specification, Open Location Services, GML 2.0 and WKT (Well Known Text). Other suppliers N/A

OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management

21-22 March 2002
ISTANBUL

(Editors: M.Orhan ALTAN, Hayati TASTAN)

Table of Contents

No	Content	Author(s)	Page
1	Opening Address	Dursun BAK	3
2	OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management (SDQM) -An Editorial Report	M.Orhan ALTAN Hayati TASTAN	4 - 9
3	Today's Priorities for Data Quality (Key note address)	Andrew U. Frank	10 - 12
4	Spatial Data Quality -Concepts And Iso Standards	Hayati TASTAN M.Orhan ALTAN	13 - 27
5	Implementing The Iso 19100 Standards in Denmark's National Datasets (Sea Charts, Top10dk And The Cadastral Map)	Jørgen Giversen	28 - 36
6	Implementing The ISO 19113 And ISO 19114 Standards in Norway's National Geodatabases	Øystein Andersen	37 - 38
7	Spatial Data Quality Management At Ordnance Survey	Andrew B. Smith	39 - 43
8	Quality Management of Atkis Data	Andreas Busch Felicitas Willrich	44 - 56
9	Spatial Data Quality: From Metadata To Quality Indicators and Contextual End-User Manual	Rodolphe Devillers Marc Gervais Yvan Bédard Robert Jeansoulin	57 - 67
10	Web Based Services For Data Quality Evaluation – A User-Oriented Approach	Anders Östman	68 - 69
11	Participants to the Workshop On Spatial Data Quality Management	-	70

Opening Address

Dursun BAK, Major-General

Commander of the General Command of Mapping,
National Mapping Agency of Turkey
and Head of the National Society of Photgrammetry
and Remote Sensing in Turkey

Ladies and Gentlemen,

I want to start with my thanks

- to the organization committee who bring together both the representatives of OEEPE and ISPRS which have considerable contributions to the spatial data quality with worldwide activities
- and to the representatives of the international scientific and application institutions
- as well as to all chairmen, speakers and participants contributed to this workshop.

Before taking about the subject of this workshop I want you know that I am sorry for that I will not be able to participate in the workshop completely because of the meetings that I have been attending since the 18^{th} March in Istanbul.

The importance of the technology is obvious for all disciplines of profession. Especially developments in Information Technology increase the importance of information and give the opportunity to store, manage and use the spatial information in computer environment, which was provided and used on paper maps before. Thus, the term map is replaced with digital geographic information, in other words “spatial data”.

As you might suppose, the most important component at the process of making geo-related decisions is spatial data. Furthermore, the accuracy of these decisions depends on the quality of spatial data used. Therefore, the spatial data quality is of great importance. Because the accuracy and reliability of decisions made by using the spatial data of poor quality will less.

Having this fact into consideration, people, institutions and organizations are making use of maps before and spatial data now require quality data whereas producers are trying to produce quality data in parallel with this requirement.

Being aware of the importance of the spatial data quality, General Command of Mapping, as the National Mapping Agency of Turkey gives a special consideration to this subject and makes every effort to actively participate in the activities carried out in the frame of his membership to the international scientific organizations such as ISO, ISPRS, OEEPE, ICA, etc which deal with spatial information.

Within this context, I believe this workshop organized by the General Command of Mapping and Istanbul Technical University in the view of OEEPE and ISPRS will make considerable contributions to the international studies carried out about spatial data quality. I thank again to all who contributed to this workshop and hope much success with my regards.

OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management (SDQM) - An Editorial Report

M. Orhan ALTAN (1), Hayati TASTAN (2)

(1) Chairman of the OEEPE WG on SDQM & Organizer of the Workshop on SDQM, Head of Division of Photogrammetry, Istanbul Technical University, Maslak, Istanbul.

(2) Co-chairman of the OEEPE WG on SDQM & Co-organizer of the Workshop on SDQM, Head of System Develoment Section, Inf. Sys. Dept., General Command of Mapping, Ankara.

1. HISTORY

In 1999, General Command of Mapping as the OEEPE National Member of Turkey, has made the Project Proposal "Developing a Spatial Data Quality Management System" in co-operatation with the Photogrammetry Division of the Istanbul Technical University. As being aware of the importance of the spatial data quality management, OEEPE accepted this proposal and setup a Working Group on Spatial Data Quality Management to lead the project. The project was started with an initial questionnaire to determine the possible participants to the project. But, unfortunately, the attraction to the project was too small. (Altan, 1999).

At the 97 the steering committee meeting in 2000, WG-SDQM has been tasked to organize a Phase 0 Workshop to identify key quality aspects to be studied in this SDQM project. The last two days of May 2001 was decided as the workshop date. Then taking the work to do, such as determination of the experts to invite, sending invitation emails or faxes, evaluating the replies, finalizing the agenda and sending the agenda to the OEEPE, into consideration a date shift to 1-2 October 2001 together with the following preliminary agenda have been proposed and accepted at the 98th OEEPE Steering Committee Meeting (Altan, 2000):

-1st Plenary Session (half day) on "Priorities for data quality (Keynote speech)" with a possible workshop breakout issue "What aspects of data quality are most important in an online world"

-2nd Plenary Session (half day) on "Managing Data Quality Experience and Perspective from National Mapping Agencies" with a possible workshop breakout issue "How do we distinguish customers' real needs for data quality from what they say they need".

- 3rd Plenary Session (half day) on "Managing Data Quality Experience and Perspective from National Mapping Agencies" (continued) with a possible workshop breakout issue "How do quality issues differ from those for paper maps"

- 4 the Plenary Session (half day) on "Industry perspectives" with a panel debate "How do system developers and international customers relate to issues of data quality"

According to the decision taken at the 98th OEEPE Steering Committee meeting, necessary attempts have been made to organize the workshop in October 2001. In this frame, pre-announcements and invitations covering the place, date and other details of the workshop have been sent to the regarding addresses. But the number of papers and registrations has not reached to a sufficient level. The reason was thought as being that the workshop preparation period overlaps with the summer season. Thus for being prepared in a much more longer time period a suggestion to postpone the workshop to 21-22 March 2002 and organize as a OEEPE/ISPRS Joint Workshop, as already been accepted by ISPRS Council before the OEEPE meeting, was accepted by the OEEPE Steering Committee, as well (Altan, 2001).

Finally, as the first step to start the OEEPE Project "Developing a Spatial Data Quality Management System", OEEPE WG-SDQM together with the ISPRS Commission IV has organized. In this organization, General Command of Mapping has prepared and hosted an OEEPE Workshop Web page (http://www.hgk.mil.tr/oeepe.htm) covering all the information on the workshop, which was linked from OEEPE and ISPRS Web Pages. The General Command of Mapping has also done preparation and mass production of the master CDROM covering the papers presented to the workshop. 5001 Euro from the Project budget have been used for the accommodations of the seven speakers excluding those from Turkey. The workshop has been realized as a joint workshop on spatial data quality management hosted by the General Command of Mapping (National Mapping Agency of Turkey) and Division of Photogrammetry of the Istanbul Technical University on 21-22 March 2002 in Istanbul. 25 people (16 from Turkey) have participated to the workshop.

2. WORKSHOP

At the opening address, Major General Dursun BAK, Commander of the General Command of Mapping (National Mapping Agency of Turkey and National Member Organization to OEEPE) and the Head of Turkish Society of Photogrammetry and Remote Sensing (TSPRS), has pointed out that the accuracy and reliability of decisions depend on the quality of spatial data used and thus spatial data quality management is an important issue to be focused on.

a. Oral Presentation Sessions

During the three oral presentation sessions, eight papers are presented whereas three breakout sessions are performed.

M.Orhan ALTAN, Chairman of the first session, Head of Davison of Photogrammetry of the Istanbul Technical Photogrammetry and Chairman of the OEEPE WG SDQM as well as Director of the ISPRS 2004 Congress, has briefed on the history of this OEEPE Project with the stress on the importance of the spatial data quality. First session consisted of two oral presentations and the first break

Andrew FRANK, the Key Speaker of the workshop, has made the first presentation of the workshop with the title "Today's Priorities for Data Quality". During his speech he pointed out that "Abstract assessment of data quality is not effective: different aspects of data quality affect different decisions differently, thus a general assessment is not possible." At the conclusion he said that "All errors in the data affect the decision, but some error affect decisions more than others and the users are interested in the dominant error - most errors in a data set have minimal effect on the decision." with a future work to develop "error models for all data types (relate to Scales of Measurements) and statistical propagation rules for discrete models" (Frank, 2002).

Hayati TASTAN, Co-chairman of the OEEPE WG SDQM, has made the second presentation of the workshop with the title "Spatial Data Quality - Concepts and Standards". He introduced the concepts for the spatial data quality management at both general- and specific level together with the general information about the ISO Technical Committee 211 as well as the purpose, scope and content of the related ISO standards (ISO 19913, ISO 19114, ISO 19115) (Tastan and Altan, 2002)

Costas ARMENAKIS, Head of the ISPRS Commission IV, has chaired the second session including two speakers and second breakout session.

Jorgen GIVERSEN, from National Survey and Cadastre - Denmark, presented the paper "Implementing the ISO 19100 Standards in Denmark's national Datasets (Sea charts, TOP10DK and the cadastral map)". He gave the first hand impressions together with problems they encountered during the implementation of the ISO 19113 and 19114 data quality standards in three major datasets in Denmark (Giversen, 2002).

Øystein ANDERSEN, from Department of Mapping Sciences - Agricultural University of Norway, presented the paper "Implementing the ISO 19113 and ISO 19114 Standards in Norway's National Geodatabases". He provided a short description on how ISO19113 and ISO19114 was implemented into a set of standards which has the purpose to ensure necessary quality of public geodata-bases in Norway and highlighted some of his experiences (Andersen, 2002).

The third session chaired by Oystein Andersen covered three oral paper presentations and the second breakout session.

Andrew SMITH, Head of Corporate Geospatial Data Management - Ordnance Survey, has presented "Spatial Data Quality Management at Ordnance Survey" describing the Ordnance Survey's experience with the process of spatial data quality management and offered some thoughts on future challenges (Smith, 2002).

Anreas Busch, from Federal Agency for Cartography and Geodesy, has presented "Quality Management of ATKIS Data" explaining a concept and their developments for automated quality control of the area-wide available topographic vector data set ATKIS® in Germany using images (Busch and Willrich, 2002).

Rudolph DEVILLERS, from Université Laval, Québec-Canada, has made the presentation "Spatial Data Quality: From Metadata to Quality Indicators and Contextual End User Manual". He described an approach that aims to reduce the risks of misuse of geospatial data by comparing data producer's specifications and data users needs and providing indicators describing data quality to users. He pointed out that a system, named Multidimensional User Manual (MUM), allows the management of geospatial data quality and the communication of the quality information using indicators that can be analyzed at different levels of detail (Devillers, et al., 2002)

Anders ÖSTMAN, from Luleå University of Technology - Sweden and member of the AGILE (Association of GI Laboratories in Europe), has presented "Web Bases Services for Data Quality Evaluation - A User Oriented Approach" based on open systems, as defined by the specifications from the Open GIS Consortium (OGC). He suggested that OEEPE and ISPRS consider establishing international activities within the field of spatial data quality, a co-operation with the AGILE Working Group on Spatial Data Usability is encouraged.

b. Breakout Sessions

At the first breakout session, the issue "What aspects of quality are most important in an online world" was discussed by two groups separately which were led by Andrew Smith and Andreas Bush. After discussions in different rooms the groups came together and the leaders of these groups presented the views of their groups on this issue. From the oral presentations, "Aspects of quality which are important in an online world" are listed as follow:

- Speed to access data
- Availability of up-to-date data, metadata and index of data
- Service
- Type of payment
- Scrutiny and tools to compare and evaluate
- Being closer to customers
- Query of data
- Security and Copyrights
- Interoperability
- Exchange formats
- Temporality
- Availability of software, online interfaces for WEB access
- Most important quality aspects (omissions, consistency, completeness)
- Feedback (direct & rapid) and corrective action
- Delivery (downloading)
- Selection capability
- Scale & Contents (heterogeneous)
- On-line order
- Marketing tool
- Information about possible use or limitations

At the second breakout session, the issue "How do we identify customers' real needs for data quality, by comparison with what they say they need? " is discussed again by the same two groups separately which were led by Andreas Östman and Rudolph Devillers. As at the first breakout session, after discussions in different rooms, the groups came together and the leaders of the groups presented the views of their groups. From the oral presentations, the answers to the issue of that breakout session are listed as follow:

- There are two types of customers: Professionals & Non-Professionals in Geographic Information.
- Customers' identification is not easy. That is, target customers first.
- We should better know their usage and applications of geographic information.
- We should provide sample data, for example some users want bigger lettering.
- We should identify what users want: maps or other forms of communication.
- Actions for visualization, spatial measurements and queries are not equal to information on quality.
- How many customers need information about quality.
- Instead of Quality Information, products should have brands what help to recognize the products and to know easier the quality needed.
- We should answer these questions first:

  - Do they know their needs?
  - How do the customers describe their needs?
  - How to measure the junction point between supply & demand curves for data quality?
- As a result we can identify customers' real needs by:

  - Feedback process of customers through questionnaires, pilot projects and education
  - A range of products with different acceptable quality levels for different needs
  - Measuring users' satisfaction
  - Negotiations.

At the third breakout session, the issue "How should we define an OEEPE Project on Spatial Data Quality Management?" is discussed by two groups together, which were led by Hayati Tastan. After discussions, agreed are on the following topics:

- Data sets (test & reference) should be common.
- Procedures for spatial data quality management should be user specified both for data quality model and data quality specs.
- Test and reference data sets may be composed of both imagery and vector data.
- Important points for this project are:

- Scale / resolution / level of data
- Format interoperability
- Metadata
- Two versions of dataset for different dates

3. CONCLUSION

This successful workshop has been the first step to start the OEEPE Project on SDQM with a number of valuable contributions from OEEPE and ISPRS member countries covering both theoretical and practical aspects of the spatial data quality management.The next step is supposed to be a new project proposal on spatial data quality management taking the important aspects of both oral presentations and breakout sessions and the suggestions of the steering committee into consideration. Although works on preparing this new project proposal has already started, after taking the suggestions of the OEEPE Science Committee and the approval for the continuation of the project by the OEEPE Steering Committee, it will be finalized and circulated by the end of June 2002.

Acknowledgments:

Thanks to OEEPE Steering Committee for the budget allocation to the workshop and thanks to all who made contributions to the workshop by making presentations, chairing sessions, leading breakout sessions as well as participating, organizing and hosting the workshop.

REFERENCES:

Altan,M.O., 1999, Progress Report on the SDQMS Project for the 97 th Meeting of the Steering Committee of the OEEPE.

Altan, M.O., 2000, Progress Report on the SDQMS Project for the 98 th Meeting of the Steering Committee of the OEEPE.

Altan, M.O., 2001, Progress Report on the SDQMS Project for the 99 th Meeting of the Steering Committee of the OEEPE.

Frank, A.U., 2002, Keynote Address: Today's Priorities for Data Quality, in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 2002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp. 1-3.

Tastan, H., M.O. Altan, 2002, Spatial Data Quality - Concepts and ISO Standards, in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 2002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp. 4-18.

Giversen, J., 2002, Implementing the ISO 19100 Standards in Denmark's national Datasets (Sea charts, TOP10DK and the cadastral map), in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 2002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp. 19-26.

Andersen, Ø., 2002, Implementing the ISO 19113 and ISO 19114 Standards in Norway's National Geodatabases, in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 2002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp. 58-59.

Smith, A., 2002, Spatial Data Quality management at Ordnance Survey, in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 2002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp.27-32.

Busch, A., F. Willrich, 2002, Quality Management of ATKIS Data, in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp.33-44.

Devillers, R., M. Gervais, Y, Bédard, R. Jeansoulin, 2002, Spatial Data Quality: From Metadata to Quality Indicators and Contextual End User Manual, in CD-ROM covering papers presented to the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 002, Istanbul, (Eds. M.O.Altan, H. Tastan), pp.45-55.

Today's Priorities for Data Quality

(Key note address)

Andrew U. Frank

Geoinformation

TU Vienna

Keywords: Data quality, spatial decisions, priorities for data quality.

1. OVERVIEW

- Information for decision making: Spatial decisions
- Closed loop semantics
- Assessment of quality with respect to decision quality
- Agent simulation for evaluation
- What are the next steps?

2. SEMANTICS IS OF IMPORTANCE

- The currently pressing questions of GIScience are all related to ontology and semantics:
  - Data quality
  - Value of data
  - Usability of data

3. DATA QUALITY RESEARCH

- Needs a fresh framework.
- The current (20 year old) concept of
  - Location-
  - Temporal-
  - Thematic-

  Quality is not effective.
- Standards based on these concepts are resisted by practitioners. They seem not to answer the right questions.

4. DATA QUALITY CAN ONLY BE ASSES WITH RESPECT TO A DECISION

- Abstract assessment of data quality is not effective:

- Different aspects of data quality affect different decisions differently. A general assessment is not possible.
- There seem to be some (limited number) of typical spatial decisions.

5. CLOSED LOOP SEMANTICS

- To define the semantics of data, the data has to be linked to the physical observation during the data collection.
- To decide about the usability of the data it has to be linked to the physical actions resulting from the decision.
- Connect the collection to the action changing the world: This closes the loop!

6. INFORMATION IS USEFUL IN DECISION MAKING

- It is only useful in decision-making.
- Often the decisions are chained and earlier decisions (e.g. about further data collection) must be related to the final physical action.

7. SPATIAL DECISIONS ARE ALLOCATION DECISIONS

- All spatial decisions are decisions about the allocation of some spatial objects to a specific use.
- This is not always immediately visible – how is a navigation decision an allocation?
- Most practical decisions are complex and consist of a combination of individual contributions to the decision.

8. THREE TYPES OF SPATIAL DECISIONS

- Spatial decisions are about the allocation of
  - Point
  - Line
  - Area

  objects.
- The determination of the shortest path is a 'allocation of line objects'.

9. EFFECT OF ERROR ON DECISION

- Errors affect decision. How much?
- It is sufficient to study a trivial algorithm, because any more effective algorithm must produce the same result (they are isomorphic).

10. TRIVIAL DECISION ALGORITHM

- Produce all possible solutions

- Evaluate their utility
- Select the solution with highest utility

11. HOW DOES ERROR AFFECT THE DECISION

- Excludes solutions
- Influences the utility and leads to the selection of a sub optimal solution.
- Decision-making is a discrete problem (remember diophantian equations?)

12. SIMULATION OFTEN NEEDED

- No closed formulae for the propagation of error.
- Consider correlation (specifically spatial autocorrelation)

13. EXAMPLE: SHORTEST PATH

- Allocation of a set of connected lines leading from a to b
- Optimal solution: shortest (minimal sum of length)
- Incomplete data or errors in topology make you miss a solution.
- Error in length affect the choice (hypothesis: effect cannot be more than the total sum of the errors of individual segment length)

14. WHAT IS ACHIEVED IN THIS FRAMEWORK?

- Linkage to value of data: Value of data is economically related to the assessment of the improvement of the decision by the information
- Value of quality can be assessed (for a specific decision situation).

15. CONCLUSIONS

- All errors in the data affect the decision, but some error affect decisions more than others.
- The users are interested in the dominant error – most errors in a data set have minimal effect on the decision.

16. FUTURE WORK

- Error models for all data types (relate to Scales of Measurements?)
- Statistical propagation rules for discrete models

Spatial Data Quality -Concepts And Iso Standards

Hayati TASTAN (1), M. Orhan ALTAN (2)

(1) Head of System Develoment Section, Inf. Sys. Dept., General Command of Mapping, Ankara.
(2) Head of Division of Photogrammetry, Istanbul Technical University, Maslak, Istanbul.

Abstract

This document introduces the concepts for the spatial data quality management at both general- and specific level together with the general information about the ISO Technical Committee 211 as well as the purpose, scope and content of the related ISO standards (ISO 19913, ISO 19114, ISO 19115).

Keywords: Spatial Data Quality, Quality Management, Quality Evaluation, ISO/TC 211.

1. INTRODUCTION

Spatial datasets are increasingly being shared, interchanged and used for purposes other than those intended by their producers. The opportunity for data users to select any spatial dataset is expanding. The value of data is directly related to its quality and therefore information about the quality of available spatial datasets is vital to the selection process. Spatial data users confront situations requiring different levels of data quality. Extremely accurate spatial data is required by some data users for certain needs and less accurate spatial data are sufficient for other needs. Information about the quality of spatial data is becoming a decisive factor in its utilisation. Technological advances allow the collection and use of geographic datasets whose quality can exceed that which is needed and requested by spatial data users.

The purpose of describing the quality of spatial data is to allow the selection of the spatial dataset best suited to application needs or requirements. Complete descriptions of the quality of a spatial dataset will encourage the sharing, interchange and use of appropriate spatial datasets. A spatial dataset can be viewed as a commodity or product. Information on the quality of spatial data allows a data producer or vendor to validate how well a dataset meets the criteria set out in its product specification and assists a data user in determining a product's ability to satisfy the requirements for their particular application.

2. CONCEPTS

2.1 General Concepts

The world of the spatial data quality contains a number of concepts which need being clarified. Even though the terms spatial and spatial data seem to be clear enough quality together with these terms brings new concepts, such as internal quality, external quality, quality check, quality evaluation, quality inspection, quality assurance, quality model, quality management, quality management system.

The **Quality** of a product or a service is defined as "the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs" (ISO 8402, 1994). In this definition, stated needs are the needs described explicitly in the product or service specifications while implied needs are those needs not explicitly specified because they are implied by the common understanding of a product or service (e.g. a knife must cut, a pencil must write).

This definition can be divided into an "external" component and an "internal" component when applied on spatial data. The **External Quality** (EQ) is the appropriateness of the specifications to the users needs : is our product well defined and does it correspond to an actual need among the community of users? The **Internal Quality** (IQ) is the respect of specifications: have we produced what was intended in the specifications. Quality is good if both internal and external quality are good ; quality may be poor if one or both components are poor. (Dassonville, 1999)

Quality Model (QM) is a set of quality parameters (such as accuracy, completeness, etc.) and their measures (such as, Circular Error, Standard Deviation, etc. to be used to measure the quality of a spatial data set and compare it to a reference dataset.

Quality Evaluation (QE)**, Quality Check** (QC) or **Quality Inspection** (QI) is the process of measuring the quality of spatial dataset and comparing it to reference dataset. This process measures how well a spatial dataset meets its specification. QE deals with internal quality. These measures are quality parameters among those that are described within standards (lineage, positional and semantic accuracy, completeness, consistency and temporal accuracy). To measure quality we may use software, visual examination or “ground truth”, we need to have a reference dataset.

Assessing and checking the quality of a product or a service only at the final stage of the production line (QE) are not enough. Specifying a Quality Model, defining the product specifications according the quality model, the quality of production processes of the spatial data set and procedures to achieve the quality objectives of the organisation in respect with the quality policy together with the definition of responsibilities and production interfaces comprise the **Quality Assurance** (QA)**.** In other words, QA is the process of defining the production process precisely and check that the producers have the necessary training and qualifications to do their job. QA involves the necessary actions to provide adequate reliance that a product or a service satisfy the given requirements. **Quality Control** (QC) may be broadly construed as “the science of discovering and controlling variations” by limiting them with the purpose of maintaining conformance of the product with the design specifications and user requirements. Thus, QC and QA may be used interchangeably.

Quality Management (QM) may be defined as a quality management approach of an organization, centred on quality, based on the participation of all its members and aiming at long-term success through customer satisfaction" (ISO 8402, 1994; .

A **Quality Management System (QMS)** can be defined as the managing structure, responsibilities, procedures, processes, and management resources to implement the principles and action lines needed to achieve the quality objectives of an organisation. A QMS can be seen as a complex system consisting of all the parts and components of an organisation dealing with the quality of processes and products.

The objectives a QMS can be summarized into five groups (CERCO WG on Quality, 1999):

- **A better management and a more effective organization**: This idea includes a strong intention of increasing benefits and productivity, by standing the amount of product inspection and cost investments in production control closely related to the level of users' satisfaction. It includes also explicitly a feasible way for solving organisational problems by defining clearly responsibilities and interfaces, for giving transparent rules applicable by every employee, for introducing continuing improvement as a regular part of day-to-day leadership. It is promoting a philosophy of measurement, calibration and accountability every time where it is required, when any risk may occurs ; costs are saved when quality inspection is positioned cautiously where there is a risk of non-quality. The goals are the reduction in overlapping-repeated- corrective work and unnecessary processes and wasted products, the protection of know-how and the reduction in training costs when staff changes.

- **An increase of workers' satisfaction and more commitment to the organisation :** A well implemented QMS gives not only more satisfaction to the management, but also to all the other employees. With a QMS they have a tool to demonstrate they have done 100% of the job. They also know to whom and where to give a direct feedback on (new) methodologies from there own experience. They will have a clear view on the demands of the customers, in terms of quantity as well as quality.

- **Better customer satisfaction:** The reason about users' satisfaction mentions the intention of fulfilling customers' needs. It gives a framework for having a consistent and improved approach to customers, and it provides a model for customers and partners, which perhaps sets the need of a mutual understanding (and therefore a common language for explaining needs and translating them into specification, but also for answering customers' claims). A QMS provides control and inspection proves to customers, giving more information on the products. The results of a QMS consist of a lot of quality proves produced all along the production steps, and giving trust to customers, before they use the product. The goal is the customer focus by actively reviewing customer needs through dialogue; making customers aware of new products and services; ensuring the organisation is aware of customer needs; corrective action when the product or service fails to meet expectations.

- **Improved quality of products, services and processes** : it means as much to define the product closest from the users' requirements (specification) than to reduce non-conformance and to track what happened during production ; it implies to develop a high control on processes, by describing and harmonising them, observing them, simplifying and optimising them. The knowledge on processes gives the possibility to detect errors, to manage them and to avoid repeating the same error. The goal is the continual improvement – of products, services, working environment, staff development, and management and production processes.

- **Promotion :** A QMS may also have to improve the corporate image of the organisation. It eases the cooperation among NMAs, by standardising rules inside the quality management field.

2.2 Specific Concepts

Specific concepts regarding spatial data quality management are listed in the ISO 19113. Some of them are presented at Table 1.

spatial quality	one or more characteristics of geographic data that describe the extent that it is fit for use
accuracy	the closeness of observations to true values or values accepted to be true
data quality element	component of the quality of a data set documenting quantitative information
data quality model	formed structure for identifying and assessing quality information
data quality overview element	component of the quality of a data set documenting non-quantitative information
data quality scope	identified collection of data for which quality information is repeated
data quality measure	type of test applied to a data quality scope
data quality evaluation procedure	operations used in applying and reporting a data quality measure
feature type	class of features with common characteristics
data quality result	value or set of values resulting from applying a data quality measure to a data quality scope or the outcome of evaluating the obtained value against a specified acceptable quality level
data quality unit	type of value that a data quality result is being reported in. Examples are meters, percentage, Boolean, etc
data quality metrics	result of a described quality measure consisting of value or set of values of a defined data quality unit
conformance	fulfilment by an implementation of all requirements specified
conformance quality level	threshold value or set of threshold values for data quality results by which a data producer determines how well a data set meets the product specification
data set	identifiable collection (A data set may be any grouping of data limited by some constraint such as spatial extent or feature; for purposes of quality evaluation, a data set may be as small as a single feature or feature attribute contained within a larger data set.)
data set series	collection of data set s sharing the same product specification data
direct evaluation method	method of evaluating the quality of a data set based on inspection of the items within the data set
data quality evaluation procedure	operation(s) used in applying and reporting a data quality measure
full inspection	inspection of every item in a data set
indirect evaluation method	method of evaluating the quality of a data set based on estimates derived from external knowledge such as data set usage, lineage, production method, and quality of source data used for the production
item	that which can be individually described or considered (An item may be any part of a data set , such as a feature, feature attribute, or combination of these)
population	group of items under consideration (e.g. All polygons in a data set, the name attribute of all roads in a certain geographic area)
sample	set of items drawn from and analyzed to estimate the characteristics of a population

Table 1. Specific concepts for spatial data quality management

3. ISO STANDARDS FOR GEOGRAPHIC INFORMATION

International Standardization Organization's Technical Committee 211 (ISO/TC 211) has been developing a family of standards on Geographic Information/Geomatics which are also cold as ISO 19100 series. 13 plenary meetings have been convened by this committee since 1994 as of today (March 21, 2002). The chairman if TC211 is Olaf Østensen from Norwegian Mapping Authority and the secretary is Bjørnhild Sæterøy from Norwegian Technology Centre. The committee has 30 active members and 20 observing members as shown at Table 2 and Table 3, respectively. As of the date October 26, 2001 (13th plenary meeting of ISO/TC 211,Adelaide, Australia), the scope of work is carried out by 9 working groups and consists of 37 standards where as 15 of them are almost ready as "Draft International Standard (DIS)" as shown at Table 4 (ISO/TC 211 Web Page, 2002a & 2002b).

This committee aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth. These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analyzing, accessing, presenting and transferring such data in digital/electronic form between different users, systems and locations. This work shall link to appropriate standards for information technology and data where possible, and provide a framework for the development of sector-specific applications using geographic data.

The objectives of these standards are

- increase the understanding and usage of geographic information,
- increase the availability, access, integration, and sharing of geographic information,
- promote the efficient, effective, and economic use of digital geographic information and associated hardware and software systems,
- contribute to a unified approach to addressing global ecological and humanitarian problems.

Australia	Denmark	Japan	Saudi Arabia	Switzerland
Austria	Finland	Rep. of Korea	South Africa	Thailand
Belgium	Germany	Malaysia	Spain	Turkey
Canada	Hungary	Morocco	Sweden	United Kingdom
China	Italy	New Zealand	Saudi Arabia	U.S.A
Czech Rep.	Jamaica	Norway	South Africa	Yugoslavia

Table 2. Active Members of the ISO/TC211

Argentina	Greece	Mauritius	Philippines	Tanzania
Colombia	Iceland	Netherlands	Poland	Ukraine
Cuba	India	Oman	Slovakia	Uruguay
France	Isl. Rep. of Iran	Pakistan	Slovenia	Zimbabwe

Table 3. Observing Members of the ISO/TC211

NO	NAME	STATUS	WG
19101	**Geographic information – Reference model**	**DIS**	**1**
19102	Geographic information – Overview	(Deleted)	1
19103	Geographic information –Conceptual schema language	CD	1
19104	Geographic information –Terminology	CD	1
19105	Geographic information – Conformance and testing	NWI	1
19106	**Geographic information – Profiles**	**DIS**	**5**
19107	**Geographic information – Spatial schema**	**DIS**	**2**
19108	**Geographic information – Temporal schema**	**DIS**	**2**
19109	**Geographic information – Rules for application schema**	**DIS**	**2**
19110	**Geographic information – Feature cataloguing methodology**	**DIS**	**3**
19111	**Geographic information – Spatial referencing by coordinates**	**DIS**	**3**
19112	**Geographic information – Spatial referencing by geographic identifiers**	**DIS**	**3**
19113	**Geographic information – Quality principles**	**DIS**	**3**
19114	**Geographic information – Quality evaluation procedures**	**DIS**	**3**
19115	**Geographic information – Metadata**	**DIS**	**3**
19116	Geographic information – Positioning services	CD	4
19117	**Geographic information – Portrayal**	**DIS**	**4**
19118	Geographic information – Encoding	CD	4
19119	**Geographic information – Services**	**DIS**	**4**
19120	Geographic information – Functional standards	TR	7
19121	Geographic information – Imagery and gridded data	TR	1
19122	Geographic information/Geomatics – Qualifications and Certification of Personnel	TR	7
19123	Geographic information – Schema for coverage geometry and functions	CD	2
19124	Geographic information – Imagery and gridded data components	NWI	6
19125-1	**Geographic information – Simple feature access - Part 1 : Common architecture**	**DIS**	**4**
19125-2	**Geographic information – Simple feature access - Part 2 : SQL option**	**DIS**	**4**
19125-3	Geographic information – Simple feature access - Part 3 : COM/OLE option	NWI	4
19126	Geographic information – Profile - FACC Data Dictionary	NWI	7
19127	Geographic information – Geodetic codes and parameters	WD	9
19128	Geographic information – Web Map server interface	NWI	4
19129	Geographic information – Imagery, gridded and coverage data framework	WD	6
19130	Geographic information – Sensor and data models for imagery and gridded data	WD	6
19131	Geographic information – Data product specifications	NWI	9
19132	Geographic information – Location based services possible standards	NWI	8
19133	Geographic information – Location based services tracking and navigation	NWI	8
19134	Geographic information – Multimodal location based services for routing and navigation	NWI	8
19135	Geographic information – Procedures for registration of geographical inf. Items	NWI	9

WG: Working Group; NWI: New Working Item; TR: Technical Report; WD: Working Draft; CD: Committee Draft; DIS: Draft International Standard; FDIS: Final Draft International Standard; IS : International Standard

Table 4. Scope of Work of the ISO/TC211

4. ISO STANDARS FOR SPATIAL DATA QUALITY

The task of ISO/TC 211 is carried out by 9 working groups that are organized in several geographic categories (Table 5). Each working group is responsible for developing specific parts of the ISO 19100 standards series (Table 4).

WG	Geographic Category
1	Framework and reference model
2	Geospatial models and operators
3	Geospatial data administration
4	Geospatial services
5	Profiles and functional standards
6	Imagery
7	Information communities
8	Location based services
9	Information management

Table 5. SISO/TC 211 Working Groups.

Regarding the ISO Standards for Spatial Data Quality, Working Group 9 that is dedicated to Geospatial Data Administration, is responsible for developing standards on Quality Principles (ISO 19113), Quality Evaluation Procedures (ISO 19114) and Metadata (ISO 19115). Metadata standard is partially related to spatial data quality. This standard is integral in that it provides the schema used to report the quality information.

ISO 9000 standard series are not directly related with spatial data quality, for they are general purpose standards developed from the basis of the need to control the quality of any type of production process. In other words, they concentrate on inspection of the ability to produce products not on the product inspection.

4.1 ISO 19113 (Geographic Information –Quality Principles)

Purpose of the Standard: A consistent suite of geographic information schemata allows geographic information to be integrated with information technology. The goal of this standard is to produce a subschema for the quality characteristics of geographic information. Quality information is essential to both the use and reuse of geographic information. A standardized conceptual schema for spatial data quality characteristics will increase the ability of geographic information created for one application to be properly evaluated for use in another application. The schema will be used by geographic information users add quality parameters to data being created and to evaluate the data received from other sources. Geographic information system and software developers will use the schema to provide applications that provide consistent methods of handling quality information. This standard should fit the quality assessment methods of the ISO 19114.

Scope of the Standard: This International Standard establishes the principles for describing the quality of spatial data and specifies components for reporting quality information. It also provides an approach to organizing information about spatial data quality. This International Standard is applicable to spatial data producers providing quality information to describe and evaluate how well a dataset meets its mapping of the universe of discourse as specified in the product specification, stated or implied, and to data users attempting to determine whether or not specific spatial data is of sufficient quality for their particular application. This International Standard should be considered by organizations involved in spatial data acquisition and purchase, in such a way that it makes it possible to fulfil the intentions of the product specification. This International Standard can additionally be used for defining application schemas and describing quality requirements. Although this International Standard is applicable to digital spatial data, its principles can be extended to many other forms of spatial data such as maps, charts and textual documents. This International Standard does not attempt to define a minimum acceptable level of quality for spatial data. (ISO/TC 211 Web Page -2)

Content of the Standard: This standard comprise the terms and definitions, data quality elements and data quality sub-elements, data quality descriptors for data quality sub-elements, data quality overview elements, identifying and reporting quantitative- and qualitative quality information. This standard also contains normative and informative annexes, such as "Abstract Test Suite", "Data Quality Concepts and Their Use" and examples defining data quality elements for different data sets such as "Digital Chart of the World", "Digital Terrain Map", "Land Use Dataset", "3D Road Network Database".

Data quality elements:

- completeness: presence and absence of features, their attributes and relationships;
- logical consistency: degree of adherence to logical rules of data structure, attribution and relationships
- positional accuracy: accuracy of the position of features;
- temporal accuracy: accuracy of the temporal attributes and temporal relationships of features
- thematic accuracy: accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships.

Data quality subelements

- completeness;
  - commission: excess data present in a dataset,
  - omission: data absent from a dataset.
- logical consistency;
  - conceptual consistency: adherence to rules of the conceptual schema,
  - domain consistency: adherence of values to the value domains,
  - format consistency: degree to which data is stored in accordance with the physical structure of the dataset,
  - topological consistency: correctness of the explicitly encoded topological characteristics of a dataset.
- positional accuracy;
  - absolute or external accuracy: closeness of reported coordinate values to values accepted as or being true,
  - relative or internal accuracy: closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true,
  - gridded data position accuracy: closeness of gridded data position values to values accepted as or being true.
- temporal accuracy;
  - accuracy of a time measurement: correctness of the temporal references of an item (reporting of error in time measurement),
  - temporal consistency: correctness of ordered events or sequences, if reported,
  - temporal validity: validity of data with respect to time.
- thematic accuracy

- classification correctness: comparison of the classes assigned to features or their attributes to a
- universe of discourse: (e.g. ground truth or reference dataset),
- non-quantitative attribute correctness: correctness of non-quantitative attributes,
- quantitative attribute accuracy: accuracy of quantitative attributes.

Descriptors of a data quality subelement:
Quality information shall be recorded for each applicable data quality subelement. The mechanism for completely recording information for a data quality subelement shall be the use of the seven descriptors of a data quality subelement:

- data quality scope
- data quality measure
- data quality evaluation procedure
- data quality result
- data quality value type
- data quality value unit
- data quality date

Data quality overview elements:

- Purpose: Purpose shall describe the rationale for creating a dataset and contain information about its intended use.
- Usage: Usage shall describe the application(s) for which a dataset has been used. Usage describes uses of the dataset by the data producer or by other, distinct, data users.
- Lineage: Lineage shall describe the history of a dataset and, in as much as is known, recount the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form. Lineage may contain two unique components:
  - source information shall describe the parentage of a dataset,
  - process step or history information shall describe a record of events or transformations in the life of a dataset, including the process used to maintain the dataset whether continuous or periodic, and the lead time.

4.2 ISO 19114 (Geographic Information –Quality Evaluation Procedures)

Purpose of the Standard: Consistent methods of reporting the quality of geographic information will not be enough to assure consistent evaluation of data set quality. The quality information reported for a geographic information data set will also depend on a consistent application of standardized methods for measuring the quality of geographic information. The results of one method of measuring quality may not be readily comparable to another although each is valid. A standardized set of evaluation criteria and procedures will guarantee that the relative quality of one data set versus another can be determined. For the most part, this standard will be used by geographic information users when they create data of when they evaluate data from other sources. Geographic information system and software developers may also use this standard to build tools for carrying out quality procedures within their application software.

Scope of the Standard: This International Standard provides a framework of procedures for determining and evaluating quality that is applicable to digital geographic datasets, consistent with the data quality principles defined in ISO 19113. This International Standard also establishes a framework for evaluating and reporting data quality results either as part of data quality metadata only or also as a quality evaluation report. This International Standard is applicable to data producers when providing quality information on how well a dataset conforms to the product specification and to data users attempting to determine whether or not the dataset contains data of sufficient quality to be fit for use in their particular applications. Although this International Standard is applicable to all types of digital geographic data, its principles can be extended to many other forms of geographic data such as maps, charts and textual documents.

Content of the Standard: This standard comprise the flow and steps of the process for evaluating spatial data quality together with direct and indirect evaluation methods. This standard also contains examples for evaluating, sampling, aggregating and reporting data quality information.

Spatial Data Quality Evaluation Steps:

Selecting Data Quality Element : Using the product specification, the data producer identifies the data quality element or data quality elements, which have to be evaluated to determine conformance to the specification. Using the user requirements, the data user selects the data quality element or data quality elements that have to be examined to determine how well the data set meets user requirements. Data quality elements are (as stated before) completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy.

Selecting Data Quality Sub-element: Using the product specification, the data producer identifies the data quality sub-element or data quality sub-elements, which have to be evaluated to test for conformance to the specification. Using the user requirements, the data user selects the data quality element sub-element to be evaluated in the test for how well the data set's quality meets the user requirements. Data quality sub-elements are (as stated before) horizontal accuracy, vertical accuracy for positional accuracy; classification (feature & attribute) accuracy and quantitative attribute accuracy for thematic accuracy; temporal accuracy, temporal consistency and temporal validity for temporal accuracy, feature completeness, attribute completeness and spatial completeness for completeness; domain, format and topological consistency for logical consistency.

Selecting Data Quality Scope: The data producer shall determine the scope of the quality evaluation on each sub-element that is necessary to test for conformance to the specification and meet the intended product purpose. Each data quality sub-element may have a different scope or multiple scopes depending upon the product specification. Using the user requirements, the data user defines the scope of the quality evaluation necessary to test for how well the data set's quality meets the user requirements.

One data quality scope is provided for each applicable data quality sub-element. A data quality scope can be a **data set series** to which a data set belongs, **the data set**, or an **identified reporting group**. The product specification and data quality overview elements are used to determine a data quality scope for each applicable data quality sub-element. If a data quality scope cannot be identified, the data quality scope shall be the data set.

Quality can vary within a data set. Multiple data quality scopes may be provided for each applicable data quality sub-element to more completely report quality information. A data quality scope is adequately identified. The following can be used to identify a data quality scope:

- the level, such as a data set series to which a data set belongs, the data set, or a reporting group,
- the types of items (lists of feature types and feature attributes) or specific items (lists of feature instances and attribute values),
- the geographic extent,
- the temporal extent, including the time frame of reference and accuracy of the time frame.

Selecting Data Quality Measure: One data quality measure is provided for each data quality scope. A single data quality measure might be insufficient for fully evaluating a data quality scope and providing a measure of quality for all possible utilization of a data set. A combination of data quality measures can give useful information. Multiple data quality measures may be provided for a data quality scope.

Choosing Data Quality Evaluation Method: For each data quality measure selected, the producer or the user chooses the quality evaluation method to be used. There are two quality evaluation methods, direct and indirect. The direct method is accomplished by sampling or full inspection of the data in the data set; the indirect method is accomplished by evaluation of data quality information from sources other than the data in the data set. Choice of quality evaluation method is determined by the data producer and the data user.

- Direct evaluation methods: There are two types of direct evaluation methods, full inspection evaluation method and sampling evaluation method. Full inspection evaluation method involves testing 100 percent of the items in a population to determine a quality result. Sampling evaluation method involves testing only a sample of all the items in a population to determine a quality result. Inspection by sampling and full inspection evaluation methods may be accomplished by either automated or non-automated means. For example some types of errors, such as feature consistency, attribute consistency and attribute value consistency can be detected by a computer program automatically (Tastan, 1999). For inspection by sampling, ISO 2859 (Inspection by attributes) and ISO 3951 (Inspection by variables) can be utilized. Steps for inspection by sampling are:
    - Items are defined. An item is that which can be individually described or considered. An item may be any part of a data set, such as a feature, feature attribute, or combination of these.
    - The data quality scope is divided into statistical homogeneous lots. Homogeneity may be evaluated based on the source data of production, production system (hardware, software, skill of operator), complexity and density of features. A lot is the minimum unit to which the result of quality evaluation is attached. If the lot does not pass inspection, the all items in the lot may be discarded or reproduced. In this sense, the definition of a lot is strongly related with the production process itself. For example, for a 1:25 topographic database populated from cartographic sources, a lot can be the coverage of a 1:250 K map sheet (i.e. 96 sheets per lot).
    - Lots are divided into sampling units. Sampling unit area is a minimum geographical area in the model world for which the inspection is conducted for all items belonging to the geographical unit. Number of items in a lot is considered as lot size. For the example above, the coverage of a 1:25 K map sheet can be a sampling unit.
    - by simple random sampling for inspection, sampling units are selected from each lot. The number of these units should be taken from the ISO 8422 (Sequential sampling plans for inspection by attributes) and ISO 8423 (Sequential sampling plans for inspection by variables). For the example above, 10 out of 96 map sheets may be selected.
    - All items, which belong to the selected sampling units, are inspected.
    - If the number of non-conforming items reaches a specified rejection number determined by AQL or LQ, the lot is not accepted. Since the inspection is by attribute, ISO 2859 applies. If the average and variance of inspected values do not satisfy limiting conditions determined by AQL, the lot is not accepted. Since the inspection is by variable, ISO 3951 applies.
    - If all the lots are accepted, the data quality scope is accepted.

- Indirect evaluation methods: Indirect evaluation methods are based on estimates of data quality measure values from sources other than the data items of the data set. The variety of sources includes, but is not limited to, metadata, knowledge of the data set's purpose, data set lineage documentation, history of uses made of the data set, and quality reports on the data used to produce the data set. Knowledge of the production process and errors that may have been introduced or detected during production is useful.

Specifying Conformance Quality Level: Producer or user specifies the conformance quality level for each data quality measure such that it establishes conformance or non-conformance to the product specification or user requirement.

Determining Data Quality Values: Data quality values are computed by applying the chosen quality evaluation method to the related sub-element scope.

Assessing Conformance to Product Specification: Data quality values are compared to the specified conformance quality levels for each selected data quality sub-element. Assessment results are either conforming or non-conforming (i.e. acceptable or non-acceptable).

Aggregating Quality Evaluation Results: The quality of a data set may be expressed by an aggregated quality result. This may require combining quality results from data quality evaluations based on differing data quality elements or data quality sub-elements, each result with perhaps different meaning than the others. The purpose of such an aggregation, even of dissimilar data quality results, is to provide a single measure of data set quality. A data set may be deemed to be of an acceptable aggregate quality even though one or more data quality elements or data quality sub-elements fail acceptance. The aggregate data set quality (ADQ) may be evaluated by several techniques:

- **100% pass/fail method for Aggregate Data Quality (ADQ):**

$$ADQ = v_1 * v_2 * v_3 * \ldots * v_n$$

where v_i is the data quality evaluation result of each data quality subelement scope (pass:1, fail :0) and n is the number of quality sub-element measured. If ADQ = 1, then the overall data set quality is deemed to be fully conforming. If ADQ = 0, then it is deemed non-conforming.

- **Waited pass/fail method for ADQ :**

$$ADQ = v_1 * w_1 + v_2 * w_2 + v_3 * w_3 + \ldots + v_n * w_n$$

Where w_i is the weight (0.0-1.0) of the sub-element, which is based on the significance to the purpose of the product. The technique does provide a magnitude value indicating how close a data set is to full conformance as measured.

- **Minimum/Maximum Value method for ADQ:**

$$ADQ = MAX(v_i, i = 1 \ldots n)$$

$$ADQ = MIN(v_i, i = 1 \ldots n)$$

The technique does provides a magnitude value indicating how close a data set is to full conformance as measured, but only in terms of the sub-element represented by the maximum or minimum.

Reporting Quality Evaluation Results: One data quality result is provided for each data quality measure. The data quality result shall be either:

- the value or set of values obtained from applying a data quality measure to a data quality scope, or
- the outcome of evaluating the value or set of values obtained from applying a data quality measure to a data quality scope against a specified acceptable quality level. This type of data quality result is referred to in this part of the International Standard as pass-fail.

A data quality result contains the following quality indicators:

- data quality scope,
- data quality measure,
- data quality evaluation procedure,
- data quality result,
- data quality value type,
- data quality date.

These results are reported as metadata or additional data quality report. Identifiable data quality overview elements are also stated in the quality reports.

4.3 ISO 19114 (Geographic Information –Metadata)

Purpose of the Standard: A consistent suite of geographic information schemata allows geographic information to be integrated with information technology. The goal of this work item is to produce a schema for geographic information metadata. Metadata includes information about the currency, accuracy, data content and attributes, sources, prices, coverage, and suitability for a particular use. Data describing a data set is becoming ever more important for locating and accessing information of all kinds. A standardized conceptual schema for geographic information metadata will increase the ability of geographic information created for one application to be found and properly evaluated for use in another application. The schema will be used by geographic information users to add metadata in a consistent and verifiable form to data being created and to evaluate quickly and accurately the data being selected from other sources. Geographic information system and software developers will use the schema to provide applications that provide consistent methods of handling metadata.

Scope of the Standard: This International Standard defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. This International Standard is applicable to:

- the cataloguing of datasets, clearinghouse activities, and the full description of datasets;
- geographic datasets, dataset series, and individual geographic features and feature properties.

This International Standard defines:

- mandatory and conditional metadata sections, metadata entities, and metadata elements;
- the minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data);
- optional metadata elements – to allow for a more extensive standard description of geographic data, if required;
- a method for extending metadata to fit specialized needs.

Content of the Standard: This standard comprise UML (Unified Modelling Language), metadata applications together with annexes covering metadata schemas, data dictionary for geographic metadata, metadata extensions and profiles, abstract test suite, comprehensive dataset metadata application profile, dataset metadata profile (XML DTD), metadata extension methodology, metadata implementation, hierarchical levels of metadata, implementation examples, multilingual support for free text metadata element.

5. CONCLUSIONS

The ISO standards on Spatial Data Quality may be used

- both by spatial data producers who are providing quality information to describe and assess how well their spatial data set meets its mapping of the universe of discourse as specified in the spatial data product specification
- and by spatial data users who are attempting to determine whether or not specific spatial data is of sufficient quality for their particular application and for describing their quality requirements.

On one hand, the subject "Spatial Data Quality Management" has long been considered an attractive and hot topic at GIS symposia and workshops. On the other hand, the user community of spatial data quality composed of both spatial data producers and users continue to ask when ISO standards will be ready to be used and bear fruit in the form of usable tools that can be applied in practical life (Tastan, 1997; Tastan, 1999; Tastan & Altan, 1999; Tastan & Altan, 2000). In respond to this question two progressive steps might be suggested:

- Transforming ISO Standards for Spatial Data Quality into useful and practical software application tools
- Improving these tools so that they can be used in conjunction with COST GIS software packages trough an international open spatial data format (e.g. Geography Markup Language – GML v3.0).

REFERENCES:

CERCO WG on Quality,2000, Handbook for Implementing a Quality Management System in a National Mapping Agency.

Dassonville,L., F.Vauglin, A.Jakobsson, C.Luzzet (1999). Quality management, data quality and users, metadata for geographical information, *In Proceedings of the International Symposium on Spatial Data Quality '99*, pp. 133-143.

ISO 8402, 1994, Quality management and quality assurance - Vocabulary.

ISO/TC 211 Web Page, 2002a, http://www.isotc211.org/pow.htm (last updated 2002-02-16), accessed 21 February 2002.

ISO/TC 211 Web Page, 2002b, http://www.isotc211.org/scope.htm (Last updated 2002-01-31), accessed 21 February 2002.

ISO / DIS 19113, 2001, Geographic Information - Quality principles.

ISO / DIS 19114, 2001, Geographic Information - Quality evaluation procedures.

ISO / DIS 19115,2001, Geographic Information - Metadata.

ISO 2859, Inspection by Attributes.

ISO 3951, Inspection by Variables.

Tastan, H., 1997, An Automatic Quality Control System for Digital Elevation Models, *In: Proceedings of the Second Turkish-German Joint Geodetic Days (Eds. M.Orhan Altan, Lothar Gründig),* pp.677-686, Berlin.

Tastan, H., 1999, Data Quality in Geographic Information Systems (in Turkish), PhD. Thesis (Supervisor: Prof.Dr.M.Orhan Altan), Istanbul technical University, Division of Geodesy and Photogrammetry Engineering, Istanbul.

Tastan, H. and Altan, M.O., 1999, "Spatial Data Quality", *In: Proceedings of the Third Turkish-German Joint Geodetic Days (Eds.M.Orhan Altan, Lothar Gründig)*, Istanbul, June 1-4, Istanbul, pp.15-30.

Tastan, H. and Altan, M.O., 2000, Automatic Checking of Feature and Attribute Consistency of a Spatial Database, *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B4/3, Com. IV, pp.1066-1073, Amsterdam.

Implementing The Iso 19100 Standards in Denmark's National Datasets (Sea Charts, Top10dk And The Cadastral Map)

Jørgen Giversen

National Survey and Cadastre - Denmark
Rentemestervej 8, DK-2400 Copenhagen NV

Abstract

The National Survey and Cadastre – Denmark, has decided to investigate the **possibilities for implementing** the **ISO 19113 and 19114** data quality standards in our three major datasets (The Sea charts, The TOP10DK and **The Cadastral Map**). A group of internal experts has come to the conclusion, that it is possible to use the standards as a common way of describing data quality. We have also taken the first steps in implementing the ISO 19113 and 19114 data quality standards in the Cadastral Map system. I will try to give you some **first hands impression** on that.

We have also encountered a **lot of problems**, when trying to read, use and implement the standards. Especially on understanding the text in the standards and the different terminology in different chapters. We have also been discussing the **test procedures** (sampling methods etc.) At the end I will try to give some information on where we are going in the **next step**.

1. INTRODUCTION

In this paper I will try to explain what the National Survey and Cadastre – Denmark is doing to implement a common set of data quality standards to give a common way of describing data quality for the 3 major Danish spatial datasets. First of all KMS (Danish for National Survey and Cadastre – Denmark) has not implemented what I am going to describe in the next chapters. We are, at the moment, in the middle of a total organisational restructuring and refinancing process. Because of that, the standardisation process has been halted for some time. I hope that it will come alive in the very near future.

The KMS has the responsibility for updating, develop, maintain and assure data quality for the 3 major digital Danish spatial datasets. The TOP 10DK (topographic dataset in 1:10.000 for Denmark). The digital Nautical see charts. The digital Cadastral map.

2. BACKGROUND

In 1997/98 the board of directors decided to form a group to be in charge of all kinds of standardization projects. Among their responsibilities was the standardization of data quality descriptions.

2.1 The standardization group

The standardization group formed in 2000 a group of experts from all the offices that has the responsibilities for the 3 major spatial data sets in KMS. The group also consisted of some experts for our IT-office and some standardization experts from other offices. The purpose for the group was to investigate the possibilities to use a common set of standards that describes the data quality.

2.1 Group of experts

The group of experts were formed to investigate the ISO 19100 set of standards. The group had to find out if the ISO quality standards could be used to give a common quality description for our 3 major data sets. The conclusion was that we could use the ISO 19100 set of standards for that purpose

3. IMPLEMENTING THE “LIGHT” VERSION OR THE “FULL” VERSION?

One of the main discussions in the group of experts was how could we possible convince the top management that it would be a good idea to start implementing the ISO 19100 set of quality standards in the production and in the data models for the data sets?

One of the first things that we discovered was that we had to use a lot of effort and resources to implement the standards as is described in the documentation. We had to find a way were we could make some shortcuts. But basically we concentrate on the following issues, as you all probably know.

The quality can be put into 2 categories the quantitative components and the qualitative components. The quantitative components are the direct measurable, either by internal tests, for instance topology, or by external tests against a reference data set. The qualitative components are not directly measurable, but attaches to the origin etc. of the data set.

The quantitative components consists of:

- Completeness
    - Omission
    - Commission
- Thematic accuracy
    - Classification correctness
    - Quantitative attribute correctness
    - Non-quantitative attribute correctness
- Spatial accuracy.
    - Absolute external accuracy
    - Relative internal accuracy
    - Gridded data position accuracy
- Temporal accuracy
    - Accuracy of a time measurement
    - Temporal consistency
    - Temporal validity
- Logical consistency.
    - Conceptual consistency
    - Domain consistency

- Format consistency
- Topological consistency

The qualitative component consists of:

- Origin of the data set.
- Purpose for the data set.
- Usage – For what purpose is the data set designed.

For a more detailed definition and explanation of each components please look in the documentation of ISO 19113 and ISO 19114.

3.1 The full version

The full version was a way that we thought would be a very huge task to overcome. By the full version we mean that we take all aspect of the standards, both all the mandatory issues and all the non-mandatory issues under consideration.

3.2 The light version

The light version was discussed because it would be more suitable for KMS compared to the full version. The light version consists of many different shortcuts. First of all we only take the mandatory stuff, and then we make some adoptions to make it fit better into the 3 different data sets data models.

4. EXPERIENCES FROM IMPLEMENTING THE ISO 19113 AND 19114 IN THE CADASTRAL MAP SYSTEM, SO FAR.

4.1 The Cadastral map approach

The first data set that we looked at was the Cadastral Map, because we decided that we should start here to do the deeper investigation.

4.2 Categories

When we looked at the data set, we decided to make a categorization of all the features. The categories are as follows:

- Fixed points
- Administrative boarders
- Cosmetic information (for different scales)
- Cadastral identification
- Forest registration
- Topographical information
- Other cadastral information
- Boundaries
- Boundary points

4.3 Sheets

After we had made the categorization we looked at each category to find out how it would fit into this schema.

Scope	Boundaries		
Data quality element	**Data quality subelement**	**Relevant ?**	**Course**
Completeness	Commission	Yes	All boundaries has to come from the real world
	Omission	No	All boundaries has to come from the real world
Logical consistency	Conceptual consistency	Yes	Not mandatory to check
	Domain consistency	Yes	Here is a great possibility for human errors
	Format consistency	No	Not mandatory to check
	Topological consistency	Yes	In spite of the harmonization process still actual
Positional accuracy	Absolute or external accuracy	Yes	The map has to be shown with other geometry
	Relative or internal accuracy	Yes	Some boundaries has to be straight
	Gridded data position accuracy	No	Not in the map
Temporal accuracy	Accuracy of a time measurement	No	The map will always be up to date
	Temporal consistency	No	The map will always be up to date
	Temporal validity	No	The map will always be up to date
Thematic accuracy	Classification correctness	Yes	The boundary has to be the right feature class
	Non-quantitative attribute correctness	No	Not mandatory
	Quantitative attribute accuracy	No	Not mandatory

Table 1. Shows if it is relevant to test a certain category the different data quality elements

After we had investigated each category, we started to look at each feature in each category with some of the following results. Here are some examples.

Data quality components		
Scope		Boundary, Boundary for roads etc.
Bounding		
	Bounding Coordinates	The hole country except Københavns and Frederiksberg municipality and Christians island
	Reference System	System 1934 / 1945
Data quality element		5 – Thematic accuracy

Data quality sub elements			2 – Non-quantitative attribute correctness
Measure			
	Description		Pass – Fail
	Identification		50202
	Method		
		Type	Internal
		Description	Use computer aided checking to investigate if the boundaries have the right attribute information's. Supplied with some sampling tests against the archives
	Result		
		Value type	
		Value	
		Unit	
		Description	
Date			
Conformance Result			
Pass-Fail			

Table 2. The figure shows the test report for the feature type Boundary

Data quality components			
Scope			Boundary points
Bounding			
	Bounding Coordinates		The hole country except Københavns and Frederiksberg municipality and Christians island
	Reference System		System 1934 / 1945
Data quality element			1- Completeness
Data quality sub elements			2 - Omission
Measure			
	Description		Pass-Fail
	Identification		10209
	Method		
		Type	Internal
		Description	Visual tests and computer aided tests, combined with material from the archives.
	Result		
		Value type	
		Value	
		Unit	
		Description	
Date			
Conformance Result			
Pass-Fail			

Table 3. The figure shows the test report for the feature type Boundary points

Data quality components			
Scope			Parcel identification, parcel number.
Bounding			
	Bounding coordinates		The hole country except Københavns and Frederiksberg municipality and Christians island
	Reference System		System 1934 / 1945
Data quality element			1 - Completeness
Data quality sub elements			1 - Commission
Measure			
	Description		Pass-Fail
	Identification		101010
	Method		
		Type	Internal
		Description	A computer aided test that finds if there is more than one parcel number for each parcel. (very seldom)
	Result		
		Value type	
		Value	
		Unit	
		Description	
Date			
Conformance Result			
Pass-Fail			

Table 4. The figure shows the test report for the feature type Parcel number.

As you probably noticed we haven't filled in the results of the test, the reason for that is very simple, we have not performed any tests yet :-) I will explain later.

5. PROBLEMS THAT WE HAVE ENCOUNTERED

Well first of all, we found out that if we want to implement the ISO quality standards in the data sets, we have to allocate a lot of resources to this purpose. We also discovered a lot of difficulties described below.

5.1 Understanding the text

We found that text in the standards is very difficult to read and understand. Even for experts it is quite difficult to understand all details in the documentation. We also found out that we cannot use the documents on every person in the staff, because the documents are very complex. If we want to implement the standards we need translate some of the documentation, or write some document that describes the concepts in a more public language.

5.2 Understanding the concepts

The concepts of the documents are in some cases very difficult to understand. The reason as I can se it is that there is used different expressions for the same issues in the different documents. That's sometimes very confusing, when you read it.

5.3 Defining test suits, spatial statistics etc.

In the ISO quality documents (19113 and 19114) the test procedures is very short described and the documents points to another ISO standard concerning general statistics (ISO2859), but when you work with geographical data you have to consider aspects such as time and space when you develop strategies for the test procedures.

In the forth-coming work you have to develop strategies for test procedures for geographical data, among more detailed issues such as:

- How to take representative area from Top10DK where you have all feature types and they are equally represented?
- How to take samples from the Cadastral Map; Does it at all make sense to use sampling in this data set, and what features are relevant to test?
- How do you take a representative sample from the Sea Charts?

There is no doubt that the problems concerning the sampling methods are very different in the 3 different data sets. It demands a great knowledge to statistic sampling methods in general, but you also need a great knowledge about the object types and the data that are represented in the different data sets. I will recommend that a group with the proper knowledge are formed to develop the strategies for the test procedures.

5.4 Reporting internal and external?

We discovered that we had to take care of the reporting process. Quality information that are attached to the data sets can be divided into 2 groups. Information's that are stored and used internal in KMS, and the information that are used by the customers. Well it is so that the to groups of information is dependent on each other, because one of them is a subset of the other.

The quality can be divided into 3 levels:

1. On top you have a description that covers the whole data set. For instance it can be, information about the qualitative components.
2. On the next level you take the descriptions on object class level. Quality information on this level could for instance be how many percent of the lakes are defined in the dataset compared to ground truth. Defined by the metadata standard (ISO 19115).
3. And on the final level you have all quality information on each object and its attributes. For instance the geometrical accuracy on each point in the object. Defined by the quality standards (ISO 19113 and 19114).

The external users have to have access to quality information on all 3 levels. The use of the 3 different levels of information is required on different times in the users use of the data set. The information that are provided in level 2 also has to be available on level 3, and the same is given for level 1 to level 2.

The information that are provided in level 1 are targeted against the users who are searching for data sets that satisfy their needs in their current situation. That mean that this information are targeted against metadata services such as the. http://www.geodatainfo.dk/.

The second levels of data quality information are those information's the users expects to be delivered to day.

The third level of information's is targeted to both internal and external users. It will probably be that type of information that we will deliver in a couple of years. We may expect that it will not be all of our users that demand this kind of information; some of the external users may not use to deliver that information in their data sets.

6. WHERE ARE WE GOING IN THE NEXT STEP?

There are a lot of problems that has to be solved. If KMS is serious about describing data quality according to the ISO 19100 standards such as the ISO 19113 and 19114. Below I will explain what tasks are essential for that work

- KMS has to gain knowledge about sampling methods in spatial data sets. This knowledge is present in demark, so we have to either by the knowledge or educate the personnel that needs the knowledge.
- The work with data quality has to be integrated with the product development and it also has to be integrated with development of new production and maintenance systems for the data sets.
- The implementation and use of data quality standards have to have focus by the top management. One thing is for sure it will demand a lot of resources to implement and use the quality standards. The increased resources will be used in the following areas:
    - Read and understand the "heavy" documents.
    - There is not very much international experience to lean on, because of that, everything is "learning by doing", everything has to be done from the bottom and up.
    - You have to understand the concepts, in spite of the inconsequent terminologies in the standards.
    - You have to develop test procedures to describe the data quality. Perhaps the test procedures have to be developed separately for each data set.
    - Data has to be tested and described according to the standard. A lot of resources will be used to educate personnel to develop the test procedures etc. Even if the staff is well educated, the first tests will take longer, because the personnel haven't got routines yet.
    - We have to form a group internal in the organisation were we can discuss problems that we have encountered in the tests the work with the standard. Such a group is very important in the start of the implementation of the standards, because of much uncertainty and lack of experience and knowledge
    - Tools have to be developed for the tests.
- Methods have to be developed to describe the data quality, the methods and tools that have to be developed, can hopefully be used in all 3 data sets. But because of the fact that the 3 data sets are different of nature there has to be developed sub tools for each data set.

By implementing the quality standards for all KMS major data sets, the data will be described in such a manor that our customers/users exactly knows what they get/buy. Which we think is a major product enhancement. It also lowers the risk for misunderstanding what the user actually can use the product to.

But we also has to have in mind that the standards are not final and are certainly not very user friendly, which demand that the standards has to be revised within a couple of years. The result of a revision will of course have an effect on the procedures and tools that KMS has developed to describe the data quality of the data sets.

The board of director decided in the middle of 2001 to continue the data quality work. The responsibility was placed the cadastral and topographic division, but at the moment the work halted until a new organisation is formed.

7. CONCLUSIONS AND REFERENCES

The conclusion is, that we can use the ISO 19100 data quality standards to describe the data quality in the 3 major digital Danish spatial datasets. The TOP 10DK (topographic dataset in 1:10.000 for Denmark), the digital Nautical see charts and the digital Cadastral map. There is a lot of problems that we have to solve, but with a lot of effort and top management focus we can have a common data quality description for all 3 data sets. Witch will benefit both KMS and our customers in the long term.

ACKNOWLEDGEMENTS :

Jan Hjelmager, KMS
Jan Thorsen, KMS
Bettina Andersen, KMS

REFERENCES:

ISO/TC211 **N 1034** Text for DIS 19113, Geographic information - Quality principles, as sent to ISO Central Secretariat for issuing as Draft International Standard

ISO/TC 211 **N 1030 T**ext for DIS 19114 submission to ISO as Draft International Standard.

ISO/TC 211 **N 1024 T**ext for DIS 19115 submission to ISO as Draft International Standard.

ISO 2859-1 Sampling procedures for inspection by attributes

Report on implementing ISO data quality standards in KMS (Internal paper from KMS not public avaible)

Implementing The ISO 19113 And ISO 19114 Standards in Norway's National Geodatabases

Øystein Andersen

Department of Mapping Sciences
Agricultural University of Norway
P.O. Box 5034
1432 Aas NLH
Norway

Abstract

This document provides a short description on how ISO19113 and ISO19114 was implemented into a set of standards which has the purpose to ensure necessary quality of public geodata-bases in Norway. Some experiences are highlighted.

Keywords: Quality of geodata-bases.

1. INTRODUCTION

In Norway, a set of national standards has been issued since 1999, with the purpose of improving and ensuring the quality of national geodata-bases.

2. THE SET OF STANDARDS

The following standards have been issued:

Networks	Staking out of buildings and controlling them	Surveying and mapping of property borders	Maps and geodata	Quality control of geodata	Product specification for largescale databases
Quality control of surveying, mapping and geodata (The Geodata-standard)					

Some of the standards are detailed in the following.

Experience 1
The ISO19113 and ISO19114 standards are made for those who are making national (or regional standards). They are not suited for being directly used as national standards.

2.1 Quality control of surveying, mapping and geodata (The Geodata-standard)

This standard provides the foundation for all the other standards, and is based on ISO19113 Geographic information – Quality principles.

The standard defines the quality model (quality elements, subelements and measures). A subdivision of the area of a community into 4 different quality classes is defined. Data quality management of geodata and geodata-bases is defined on an overview level. Necessary formulaes are defined. Plans for the work with geodata-bases are specified. Examples are included.

2.2 Networks

This standard mainly describes how to control the quality of geodetic networks. A special feature is that requirements are put onto angles and lengths, not onto coordinates of points.

2.3 Staking out of buildings and controlling them

This standard has two purposes: Describing quality control of the staking out of buildings, and how to measure newly built objects. The newly built objects create a need for updating the databases. The principle is laid down that it is the builder the should carry trhe cost of utdating the databases.

2.4 Surveying and mapping of property borders

2.5 Maps and geodata

How to do the mapping, including photogrammetric mapping.

2.6 Quality control of geodata

How to check the accuracy of digital maps.

Experience 1
The ISO19114 standard is not good on the topic of geographical sampling quality control.

2.7 Product specification for largescale databases

The specifications and requirements are found here.

Spatial Data Quality Management At Ordnance Survey

Andrew B. Smith MA, MAppSci, MBA, ARICS

Head of Corporate Geospatial Data Management
Ordnance Survey

Romsey Road, Southampton, United Kingdom, SO16 4GU

Abstract

This paper describes Ordnance Survey's experience with the process of spatial data quality management and offers some thoughts on future challenges.

Keywords: Ordnance Survey, geospatial data, data quality, data management

1. INTRODUCTION

Ordnance Survey is the national mapping agency of Great Britain. It employs 1800 staff and has an annual turnover of £100 million (160 million Euros), including an agreement with government for services that are considered to be in the national interest. The data quality management process can be described diagrammatically by the following circle of activities (Fig.1):

Figure 1. The data quality process.

2. CUSTOMER NEEDS

At Ordnance Survey, data quality management starts, as it always should, with the identification of customers' changing needs. Regular customer satisfaction surveys and complaint handling procedures feed in any sources of dissatisfaction with existing products and services. Account managers maintain a constant dialogue with major customers and business partners, which serves to identify both where existing needs are not being fully met and how those needs are likely to change in the future. Technology tracking, which is part of our Research function, keeps a watch on developing technologies that may provide new market opportunities. Database business managers in our marketing areas synthesize all this information and try to define how the spatial data that underpins our various product and service offerings needs to be developed and enhanced.

This may sound straightforward, but in reality it is still difficult to be sure we are going to be able to meet the real needs of our present and future customers. In order to be there when they need us, we often have to try and anticipate their requirements several years in advance, when they are poorly understood and ill defined. Either over- or under-specifying data quality requirements can be very expensive.
The developments that are currently generating most interest are the location-based services, which telecommunications companies are hoping to offer in a few years, using third-generation mobile devices. If the more exciting forecasts are to be believed, one will soon be able to summon a taxi at the press of a single button and then be interactively guided to book your train or air ticket, hotel room, theatre seat and evening meal all whilst on the move.

Less futuristic but also very challenging are the "joined up government" services being driven, in Britain, by the "Modernising Government" initiative. An example is the National Land Information System (NLIS), which aim to speed up and simplify tasks such as house sale and purchase or the notification of a change of address.

The data requirements to support applications such as these will be very challenging. In days gone by we published new editions of paper maps very few years. As this became inadequate for customers' needs, we introduced means of producing intermediate versions via microfilm and then as a dyeline copy of the surveyor's master document. Web-based technology will provide virtually real-time access and demand a similar level of data currency for some elements, particularly where there are safety and liability issues. Data will need to be seamlessly networked across a range of public and private organisations, not only within individual countries but also internationally. Organisations maintaining data, both spatial and non-spatial, will need to co-operate on technical standards and business arrangements as never before. Most customers will be unaware of the sources of the data they are making use of, but if they are to derive the benefits they expect someone will need to have solved the thorny issues of data incompatibility and quality assurance. Partnerships and joint ventures will be essential. Few organisations have money to waste in collecting or updating information, which others are already maintaining or could do so more efficiently.

3. STRATEGY/PLANS

Commitment of senior management to a strategy of data quality improvement and maintenance is essential for business success to be sustainable. It is only too easy to be so intent on meeting today's needs that future opportunities are overlooked or have to take a low priority when funding is allocated. Top management should have their eyes at least partly on the horizon, balancing the needs of strategic investment with the funding of today's operations. Strategies are notoriously difficult documents to write. They tend to be dull and unimaginative and to be instantly despatched to gather dust in everyone's bottom drawer. If they are too general they may say nothing useful; if they are too specific they may be out of date as soon as they are issued. Ordnance Survey's corporate geospatial information strategy is currently embodied in a set of presentation slides, which can be continuously updated but which present a lively view of the "big picture", as its author refers to it.

Having periodically obtained the endorsement of top management for this evolving picture, the strategists need to work closely with the planners, to convert strategy into workable programmes and operational plans covering at least the next year to 18 months. Such programmes identify the goals to be achieved, the dates of any key milestones, which will be responsible for what and, critically, how the various programmes must fit together. It is important to work from the whole to the part, leaving the fine detail of method and resource allocations to internal project planning.

Data quality programmes must inevitably compete for funds in the periodic budgeting round with building projects, human resource development proposals and other ideas of various sorts. To help gain approval, these all need business cases prepared in a common currency, assessing business benefit against cost in financial terms as far as possible. This is where top management's commitment to data quality is really tested and the groundwork of lobbying and informing should pay dividends.

Timing is critical. Ideas for data improvement and new product development must pass through a funnel, with distinct gateways to sift out the non-runners. In order to ensure that market opportunities can be grasped at the right time, research work, concept testing, prototype development and creation of automated routines all have to be done early enough. It is only too easy to finish up using expensive manual editing to improve data because there was insufficient time to allow an automated routine to be devised.

4. STANDARDS

In the past, Ordnance Survey put a lot of effort into understanding and influencing the development of a range of international standards relating to geographic information. In the early days of digital mapping in Great Britain, Ordnance Survey was a prime mover behind making the National Transfer Format (NTF) into a British Standard (BS 7567). NTF Level 2 is still important for the supply of Land-Line, our flagship data product, along with DXF for CAD applications. We have resisted supplying data in a wider range of proprietary formats owing to the complexity of ordering and supply options, which multiple versions of these entail.

Today we take a more pragmatic stance over international standards. Data for the Digital National framework (DNF) release later this year will be in GML (Geography Mark-up Language) format, developed from the more general XML by the Open GIS Consortium (OGC). We are keeping closely in touch with the evolution of GML. Otherwise, a watching brief is maintained on the standard-making activities of ISO (specifically ISO/TC/211), CEN, the British Standards Institute (BSI) and internet standards bodies such as W3C and IETF.

5. SPECIFICATIONS

Ordnance Survey's specifications for surveying and cartographic depiction were historically enshrined in a set a text manuals affectionately known as "the modules", "the Red Book" and "the Biscuit Book"! Their modern equivalents are a set of quality system documents, created according to ISO 9000 principles, which are regularly reviewed and which can be accessed via the department's intranet. For each corporate data set there is a suite of documents covering some or all of the following:

- Overview;
- Glossary;
- Data model(s);
- Data structures;
- Data classification (and meta-data);
- Data consistency/integrity;
- Capture/improvement and maintenance;
- Data quality;
- Maintenance of geometry and attributes;
- Data presentation; and
- Detail catalogue.

These highly detailed documents are essential to the maintenance of national consistency and to enabling data to be updated and handled by multiple software systems. Changes to specifications to meet marketing requirements or system changes have to be formally controlled. However, every change in specification brings a problem of inconsistency. Only rarely can immediate retrospective action be justified. It takes many years for some changes to be fully implemented and in most cases 100% consistency is never achieved.

Specifications define an ideal, which in practice may not be attainable. Geospatial data in which all the elements are 100% complete and matching their real world state at every moment is as yet a goal which no organisation has been able to achieve. So tolerances and acceptable quality levels need to be defined in respect of each element of quality:

- Completeness;
- Currency;
- Logical consistency;
- Attribute accuracy; and
- Positional accuracy.

For instance, for currency and completeness Ordnance Survey has an overall "agency performance monitor" that 99.5% of all major features that have been in existence for six months or more will be recorded and available in its database at any point in time. Other tolerances are set for every process and agreed with both external contractors and internal work areas before quality assurance begins.

6. TESTING AND ACCEPTANCE

The data implications of changes to any system that handles corporate geo-spatial data have to be carefully evaluated. Changes to editing and validation systems must be made in sympathy if a sudden rash of validation failures is to be avoided. Testing tools need to be constantly re-configured to cope with new data formats and specification changes.

Data which is output from new and developing systems and processes must be tested for conformity to specification before formal acceptance notices, which permit new systems or enhancements to be implemented, can be issued. A radical development such as Ordnance Survey's current programme to deliver the Digital National Framework requires a substantial programme of data testing as pre-launch versions of several complex-, interacting systems evolve. Feedback and close co-operation between data testers and system developers is essential, but it is also vital that the testing retains its independence and has the confidence and backing of higher authority if any internal dispute should arise.
Every movement of data between systems involves a validation check to ensure that data integrity has not been compromised. Validation and access failures need to be analysed and errors rectified, sometimes a time-consuming and difficult task requiring a high level of skill and experience. Validation checks have to be tight, in order to trap failures reliably. The temptation to cut corners in the interests of expediency is often very strong, but should be resisted as it invariably creates more problems than it solves.

7. ASSURANCE

Quality assurance differs from quality control in that it is a complete and independent check applied to a sample of data taken from a wider population, whereas quality control should be applied to all data as an integral part of all stages of an editing or manipulative process.

Until recently, currency and positional accuracy have been regarded as by far the most important elements of geospatial data quality assurance. We test currency and completeness by a quarterly audit of a randomly generated sample of 600 digital mapping units (DMUs) from across the country. These are subjected to a complete local perambulation on the ground. Feedback is supplied to field operations managers on age profiles of unsurveyed change and the quality of change intelligence monitoring.

Positional accuracy is tested by a combination of field and photogrammetric methods, using the latest GPS equipment and digital photogrammetric workstations. The challenge is to keep ahead of advances in the methods that are being used for routine production work. Accuracy testing must be to a higher order, rather than simply repeating what was done in the first place.

We have recently embarked on a major programme of positional accuracy improvement, shifting the co-ordinates of features in some cases by as much as 6 to 8 metres. Whilst essential to making our data fit for

modern quality requirements, this process is a severe challenge for us, for our customers and for their system suppliers. It has required intensive consultation and collaboration. We are providing files of link vectors that relate the new positions of features to the old. These will enable customers to move their own data in sympathy where for instance a pipeline or cable has been located in relation to the edge of a road.
The other elements of data quality, particularly logical consistency and attribute accuracy, are becoming increasingly important for the applications which are making ever greater demands on data quality. Checks of these aspects now form essential parts of the quality assurance process.

To give some idea of the size of task which can be required, whereas Great Britain is covered by only 204 sheets of the 1:50 000 scale Landranger maps, there are some 230 000 DMUs in the main topographic database. These were originally digitised from maps maintained at the three "basic scales", 1:1 250, 1:2 500 and 1:10 000. A quality improvement programme, such as we have undertaken over the last six months, requires each of these map units to be extracted from the databank, edited, validated and returned to the databank. Up to a thousand people, including 250 staff in-house and some in India, Indonesia and Eastern Europe, have been working on this programme. A sample from each source has been subjected to rigorous quality assurance, including both automated tests and visual checks. The size of the sample diminishes as confidence in the output grows.

A challenge for the future is working out how best to carry out quality assurance when editing and customer supply operate in real time from the core database and there is no possibility of putting work into a "quarantine" area on its return from an updating process. The answer must be to increase the extent to which quality control is built into editing software, so that only valid changes are permitted and the user is alerted immediately rather than the data having to be fixed when it fails a subsequent, off-line validation check.
A further aspect of quality assurance involves testing all new and revised versions of digital products and delivery systems at the request of product and service managers. This enables them to launch their new product or service confident that they will not be causing their customers problems and giving rise to a barrage of complaints.

8. ADVICE AND COMMUNICATIONS

Feedback to those involved in updating or improving geospatial data, both internal and external, needs to be quick and professional, operating in an atmosphere of co-operative assistance. A confrontational approach is of benefit to no one. It is in everyone's interests that standards are met as quickly and efficiently as possible. Although, as was mentioned earlier, there is a distinction between quality assurance and control, in practice the two tend to merge, with quality assurers providing advice on quality control, which subsequently improves confidence in the delivered results and ensures that acceptable quality levels are met more quickly. New processes, both in-house and by contractors, are audited at an appropriate stage, to ensure that they are capable of achieving acceptable quality levels.

Despite all our best efforts to test products before they are delivered, customer complaints do still occur. Sometimes these relate to genuine omissions, errors or non-conformances, but many complaints and enquiries result from misunderstandings about what we are aiming to provide. We still have work to do to make our data quality statements more meaningful for our customers. Most complaints and enquiries require some form of investigation and decisions on corrective and preventative action.

9. CONCLUSION

With the feedback from customer complaints, helping to define future needs, the circle is completed.
This paper has attempted to describe Ordnance Survey's experience with the process of spatial data quality management. There are some formidable challenges ahead as we strive to meet ever-growing demands created by new applications and advances in technology.

Quality Management of Atkis Data

Andreas Busch (1), Felicitas Willrich (2)

(1) Federal Agency for Cartography and Geodesy
Richard-Strauss-Allee 11, D-60598 Frankfurt, Germany
busch@ifag.de

(2) Institute for Photogrammetry and GeoInformation, University of Hannover
Nienburger Str. 1, D-30167 Hannover, Germany
willrich@ipi.uni-hannover.de

Abstract

Describing the quality of digital geodata in a geodatabase is required for many applications. As well customers as users want to know how good the data are and if the data are up to date. We present a concept and our developments for automated quality control of the area-wide available topographic vector data set ATKIS® in Germany using images. The automation comprises automatic cartographic feature extraction and comparison with ATKIS, which both are triggered by additional knowledge derived from the existing scene description. To reach an operational solution the system is designed as an automated system, which admits user interaction to perform a final check of the fully automatically derived quality description of the data.

Keywords: quality control, road extraction, semi-automatic procedure, user interaction, integration of image processing and GIS

1. INTRODUCTION

Describing the quality of digital geodata in a geodatabase is required for many applications because tasks like environmental planning, documentation and analysis highly depend on the quality of the input data used for it. Additionally a quality description in terms of geometric and thematic accuracy and completeness is a prerequisite for identifying areas of interest where an updating has to be performed.

For checking the quality of existing geodata and for updating the data an efficient quality management system is required to ensure that a production process of geodata delivers the desired quality. Such a system has to cover a chain of processes for quality control that guaranty to keep the specified quality. Each step in quality control checks one aspect of quality specifications defined for the underlying data model. Thus, quality management should be the first step in data processing, data analysis, maintenance or homogenisation of different data sets to ensure a well-defined result in any of these processing tasks.

A major task of the Bundesamt für Kartographie und Geodäsie (BKG, Federal Agency for Cartography and Geodesy) consists in providing the geodata of the ATKIS project on the territory of the Federal Republic of Germany. The acronym ATKIS® stands for the Authoritative Topographic-Cartographic Information System for Germany. ATKIS is a trademark of the Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV).

Components (Fig. 1) of ATKIS are object-based digital landscape models (DLM), digital topographic maps (DTK) in vector and raster formats and standardised orthophotos (DOP). The object-based digital landscape models encompass several resolutions and the digital topographic maps of different scales are derived by a transformation to map geometry and symbol add-ons.

The ATKIS DLMBasis, i.e. the ATKIS data of the highest resolution or of the largest scale (cf. Fig. 1), are produced by the federal states (Länder) of Germany based on different data for the acquisition and are delivered to the BKG, where they are stored in a database at the Geodata Centre (GDC) of the BKG (Endrulis 2000). Since these data that are supplied by 16 surveying authorities are delivered to customers on the one hand and are used to derive data of smaller scales within the BKG on the other hand, a system for quality management of the ATKIS data is essential.

This paper presents the concept of quality management of the ATKIS DLMBasis as it is proposed at the BKG. Parts of it already are performed in an operational way within the daily production process. To solve the complete process chain in an efficient way BKG has initiated a common project with the University of Hannover to develop a system for automated quality control of ATKIS DLMBasis using digital orthoimages which is also described in this paper.

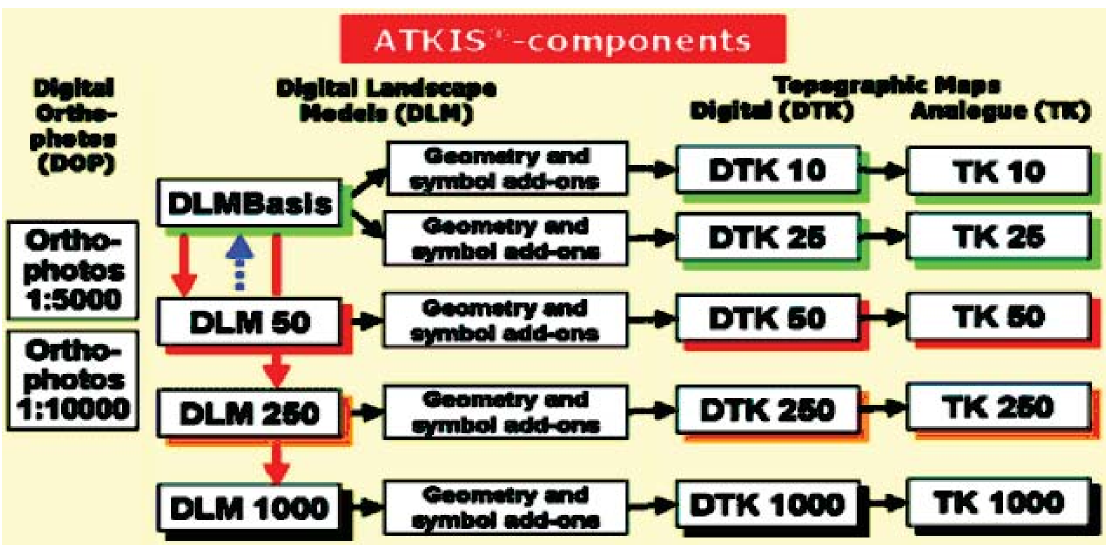


Figure 1. The components of ATKIS (cf. AdV 2000)

2. QUALITY OF GEOTOPOGRAPHIC DATA

2.1 Quality Measures

For rating the quality of geodata we need a certain set of measures, which give us expressive, comprehensive and useful criteria. That is the reason why quality measures are part of norms or specifications from e.g. ISO, CEN or the OpenGIS Consortium. We want to start with a coarse subdivision of quality measures into two categories, which due to the following arguments are important for practical applications:

1. Quality measures that concern consistency with the data model,
2. Quality measures that concern consistency of data and reality within the scope of the model.

A complete check of the first category can be performed automatically using solely the ATKIS vector data and functionality and routines within a database or GIS without any additional data. This inspection can be done exhaustively, i.e. the whole area covered by the data can be checked. On the other hand the comparison of data and reality is much more expensive. Performing it for the whole area requires nearly the same order of effort as the initial acquisition of the data.

Joos (2000) suggests a system of four criteria, which are conceptually independent or orthogonal, namely completeness, correctness, consistency, and accuracy.
The quality elements of ISO 19113

- completeness
- logical consistency
- positional accuracy
- temporal accuracy
- thematic accuracy

are strongly related to the criteria above. The ISO elements are refined by so-called subelements. So completeness comprises missing objects, i.e. omissions, and the complementary event, where a data object does not exist in reality, i.e. commission. Within ISO 19113 correctness is defined as a subelement of thematic accuracy. Temporal accuracy measures whether a dataset is up-to-date. Some quality elements of ISO differentiate between features and attributes. The quality elements of ISO are listed in Table 1 together with additional information.

2.2 Concept of Quality Management of the ATKIS DLMBasis at the BKG

Since the data of the ATKIS DLMBasis are delivered by 16 federal states, it is the task of the BKG to join them to one homogeneous set of data. This includes establishing logical and geometrical consistency at the borders of the different data sets. All incoming data sets and update files from the surveying authorities of the federal states have to undergo the quality control. Table 1 shows the current status of quality control at the BKG within the framework of the quality elements of ISO 19113. Automatic routines that test conformity of the data sets with the model, have been implemented at the Geodata Centre (GDC). These routines perform an exhaustive check, which includes the data quality element logical consistency and some other subelements of ISO 19113. Because spatial omissions concern the question, whether the whole area is covered by ATKIS objects without any gaps, they are revealed without any comparison of the data to the real world. In case of attribute completeness both commissions and omissions can be checked, because the model defines which attributes must be present for a given type of object and which not. Additionally for temporal consistency some plausibility checks are applied. All these tests concern quality measures belonging to the first category, i.e. quality measures that concern consistency with the data model. If once implemented, they perform on the full coverage of the data requiring computation power and time only. Unlike these automatic tests the comparison of the data to the real world will always require manpower. In Table 1 this is indicated by the term automated. Automated procedures consist of automatic steps that are started by an operator and give back a result that requires further processing by the operator.

Any error that is detected during the quality control is reported to the respective federal state. Since the federal states are producers of the data of the ATKIS DLMBasis, they are responsible for the appropriate amendment of the data. When the errors have been corrected, the updated datasets from the surveying authorities of the federal states are delivered to the BKG again where they are stored in the database at the GDC. This procedure guarantees that there exists one unique dataset of the ATKIS DLMBasis only.

Quality element	Quality subelement	Coverage	Status
Completeness	Omission (Feature completeness)	sample	automated
	Commission (Feature completeness)	sample	automated
	Omission (Attribute completeness)	full	automatic
	Commission (Attribute completeness)	full	automatic
	Omission (Spatial completeness)	full	automatic
	Commission (Spatial completeness)	—	no
Logical consistency	Conceptual consistency	full	automatic
	Domain consistency	full	automatic
	Format consistency	full	automatic
	Topological consistency	full	automatic
Positional accuracy	Absolute or external accuracy	sample	automated
	Relative or internal accuracy	sample	automated
	Gridded data position accuracy	—	no
Temporal accuracy	Accuracy of time measurement	—	no
	Temporal consistency	full	automatic, checked partially
	Temporal validity	sample	automated
Thematic accuracy	Classification correctness	sample	automated
	Non-quantitative attribute correctness	sample	automated
	Quantitative attribute accuracy	sample	automated

Table 1. Quality elements from ISO and their current status for the quality control at the BKG

2.3 Conformity of Data and Reality

A complete comparison of data and reality requires a lot of effort and cost, but it furnishes all the update information for the data. Update processes of digital geotopographic databases have changed in comparison to the periodic update cycles of analog maps. There is no technical problem in altering a specific feature or attribute as soon as the information about its change is available. As a consequence revision of geoinformation turns from periodic to continuous procedures. This requires that the information about changes in the real world has to be available continuously, too, if it is of relevance to the geotopographic data. Since nearly all of the changes in the real world are man-made, information concerning the changes is available very early, usually already during the phase of planning. Therefore the surveying authorities of Germany are forcing topographical information management to gather information about changes and make it available in time for the update of ATKIS. The flow of information from the authorities that cause the changes plays an important role in topographical information management.

This paper focuses on a method for quality control of the ATKIS DLMBasis that compares the data with reality using sensor data from an independent source. Thus, we look at quality control as an independent procedure to rate the quality of geodata by sample and to detect deficiencies within the chain of production. It depends on the type of sensor which features and attributes can be verified. Our main interest concerns objects where most changes arise and that are important, namely the road network and built-up area. Currently we are using only digital orthophotos as sensor data. If the orthophoto is of recent date, it is an up-to-date reference of reality and can be used to assess temporal accuracy, too. Orthophotos and photogrammetry are very suitable tools to determine the positional accuracy of features and geometric attributes like the width of a road. Nevertheless, there will be always features and attributes that are not detectable by the sensor data. In these cases quality control has to be based on the topographic information management.

Figure 2. How to check conformity between data and reality is performed

3. SYSTEM OVERVIEW & COMPONENTS

In chapter 2.2 we gave an overview over the quality management of the ATKIS DLMBasis at the BKG. The data quality concerning consistency with the data model is already checked fully automatically in an operational way. For efficiently checking the consistency of data and reality within the scope of the model, BKG has initiated a pilot R&D project together with the University of Hannover to develop a system for automated quality control by comparing ATKIS to imagery covering the scene.

The main idea followed with the developments is to check the quality of the ATKIS DLMBasis by extracting features from black and white orthoimages and comparing the extracted information to the DLM. To increase the efficiency of the quality control, extraction and comparison should be performed fully automatically.

3.1 System Overview

Automatic feature extraction from aerial images has been a major activity of international research in photogrammetry and computer vision during the last decades (e.g. Förstner et al. 1999, Baltsavias et al. 2001). Although there is many success in cartographic feature extraction experiences have shown that algorithms particularly give good results if applied to well-defined application areas. The reason is that all approaches need additional knowledge to be involved by using appropriate models, which can more easily be formulated for restricted situations. Walter (1999) for instance developed a system that supports the operator in quality control of area and line objects in ATKIS by automatically extracting land cover classes by multi-spectral classification from satellite imagery and comparing it to the corresponding ATKIS objects. He uses knowledge derived from the existing GIS for defining training sets for a supervised classification. Those ATKIS objects that show a high probability of differences between the extracted object classes are indicated to be presumed changes and can be visualized for further interactive analysis by the operator.

Knowledge based systems have proven to be suitable framework for representing knowledge about the objects and exploiting it during the recognition process. Liedtke et al. (2001) present a system for knowledge-based image interpretation which models structural dependencies by semantic networks with concept nodes and edges describing the relations between the nodes by steering the top-down and bottom-up analysis. The system is designed to use holistic methods for feature primitive extraction attached to nodes on different semantic levels. Examples for land use interpretation are given using ortho images, laser DEM and an initial segmentation derived from vector data given by a GIS.

We present an automated system prototype for knowledge-based quality control, which is designed to combine fully-automatic analysis with interactive pre-processing by an operator to reach an optimal workflow. We admit a final user interaction to transfer the knowledge based image interpretation techniques to an operational solution for practical applications as in general one can not expect any automatic image analysis tool to lead to 100% reliability which however is needed for operational systems (cf. Lang and Förstner 1996, Gülch 2000). The fully automatic part attains to focus the interaction and thus reduces the amount of interaction by an operator, which is the time consuming part in the quality control process chain.

The knowledge we use is partially derived from the existing geodata that is from ATKIS and coded in rules. In future it will be implemented in the knowledge-based system presented in Liedtke et al (2001). Although in general the system is designed to handle all object types of ATKIS we presently are focussing on those objects for which the highest up-to-dateness is required, and test it for roads.

The system development is embedded in a broader concept of a knowledge-based workstation, which provides functionality from photogrammetry, GIS, and cartography for the acquisition, and maintenance of geoinformation. A major goal of this concept is to integrate several components performing different tasks within the framework of a knowledge-based system.

3.2 System Components

The system is designed as a knowledge-based photogrammetric-cartographic workstation, which provides functionality from knowledge-based photogrammetric image analysis and cartography for the production of geoinformation. It consists of three major parts: a. the GIS component, b. the photogrammetric component and c. the knowledge-based component:

a. *The GIS component:* The GIS component of the system is based on the GIS ArcInfo 8 and runs with the desktop version under Windows. It is used for automatic pre-processing of the ATKIS data, as an interface to the database and to the image processing system, for interactive post-processing of automatically derived results and generally spoken as an user interface and for the overlay of aerial images and ATKIS data.

b. *The photogrammetric component:* The photogrammetric component running under Linux comprises the automatic cartographic feature extraction modules and the comparison with the original vector data leading to quality measures. Both tasks are triggered by the GIS data being a valuable source of additional knowledge. The underlying result of the feature extraction steps as well as the automatically derived quality of the ATKIS objects are stored in exchange files. They are transferred to the GIS component and are used to support the operator during the interactive final check and during geometric corrections.

c. *The knowledge-based component:* This part of the system is responsible for making pre-knowledge from the GIS available and transfer it in a suitable way to the photogrammetric component, that is to the object extraction, comparison and evaluation algorithms. As the link to the photogrammetric component is very close it is also running under Linux. Pre-knowledge especially is used for defining regions of interest, for selecting the appropriate algorithms, for parameter control of the road extraction and evaluation of deviations between ATKIS and extracted features. Additionally it is helpful for steering the complete automatic workflow.

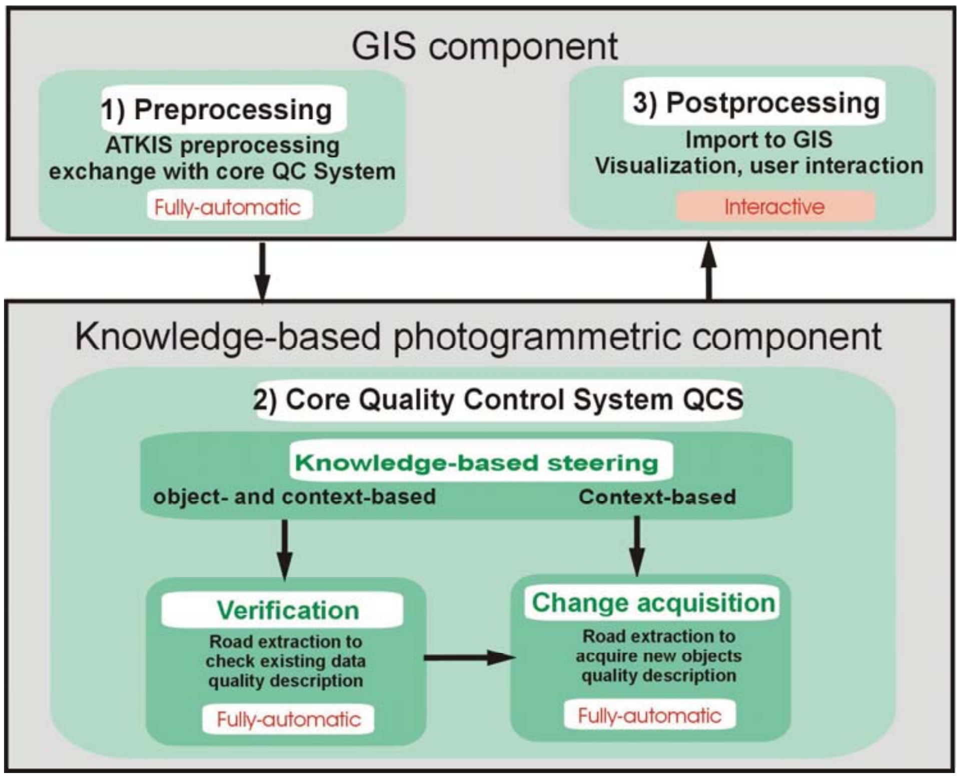


Figure 3. The components of the system for quality control

The result of the automatic knowledge-based reasoning is imported to the GIS component so that it is available to the operator. Quality measures are delivered as attributes and are used for an appropriate visualisation. During the final interactive check of the results the operator focuses on roads that were indicated as not verified and roads where the photogrammetric component indicated an uncertain situation. Thus he only has to handle a reduced number of objects contained in the GIS. Missing objects, geometric differences that exceed the tolerances given by the ATKIS data model and wrong values of attributes are recorded and forwarded for the correction of the database.

4. PROCEDURE FOR AUTOMATIC QUALITY CONTROL

4.1 Overview

The complete automated system for quality control, checking the consistency of data and reality is subdivided into the three steps: 1) fully-automatic pre-processing, 2) fully-automatic quality control and 3) interactive post-processing (cf. Fig. 3) . In the following we describe the fully automatic part in detail. It is to be regarded as being a black box for the operator delivering a preliminary quality check for focussing the interactive intervention by the user.

The procedure starts with automatically pre-processing and preparing the GIS data so that it is appropriate as well for the automatic processes as for the interactive analysis by the operator. This pre-processing is performed by the GIS component and compounds e.g. the selection of test area for quality control, establishing the link between object geometry and thematic attributes and supplying an appropriate interface to the knowledge-based photogrammetric component. Due to practical reasons the working units are image tiles of a size of e.g. 2 km × 2 km or interactively selected image areas defined for the quality control. For each tile all types of ATKIS objects and their attributes, that are relevant for quality control are requested from the database and are transferred to the photogrammetric component. Actually these ATKIS objects are exported to interchange formats, that can be read by the knowledge-based photogrammetric component but in future the transfer will be performed by database queries.

The second step, the core of the system for automatic quality control (QCS), which combines the knowledge-based, and the photogrammetric component is running under Linux. It comprises automatic road extraction adapted for the quality check and the comparison of its result to the original data. The quality description is delivered to the GIS component, the ArcInfo 8 system, where the operator performs the well directed pre-editing of those parts of the scene description, which could not be reliably analysed by the automatic process.

4.2 Core Quality Control System QCS

4.2.1 Automatic Road Extraction

International research has produced many different algorithms for road extraction (e.g. Förstner et al. 1999, Baltsavias et al. 2001) each of them being suitable for well-defined extraction tasks. Our concept for checking the quality of roads in general is designed to use different algorithms whereas the selection of the algorithm is performed by the knowledge-based component. We actually apply software developed by C. Wiedemann (cf. Wiedemann et al. 1998a, Wiedemann 2001) at the Chair for Photogrammetry and Remote Sensing at the Technical University Munich. We adapted it to our specific tasks especially by exploiting the GIS scene description and embedded it into the knowledge-based framework for steering and deriving quality statements. The underlying road model partly can be steered by parameters, which are automatically defined or even adapted to the image content by the knowledge-based component.

4.2.2 The use of pre-knowledge for quality control

GIS data in general can provide a valuable source of additional knowledge (cf. Vosselman 1996) and can be used to stabilize the image interpretation tasks. Examples are given in e.g. Quint and Sties (1995) or Wallace et al. (2001). In contrast to cartographic feature extraction solely based on imagery the starting position for quality control of existing data is different, as an initial scene description already is available. In this case algorithms for object extraction can benefit from the information contained in the GIS. This however requires a close and well-defined interaction between image analysis and GIS.

The knowledge we use for road extraction and for evaluation of the differences between extracted roads and objects in the database can be distinguished into object-specific properties and context-specific properties:

- Object-specific properties e.g. are the road type (highway, single/multi track, road, path), road widths or road constitution (asphalt, concrete). These kinds of properties usually are partially represented in the underlying road model of each algorithm and thus characterize the application domain of the algorithm.
- Context-specific properties can be subdivided into global and local context dependencies (cf. Baumgartner et al. 1997).
    - The global context, e.g. the environment through which a linear feature passes or is contained in influences the appearance of the road in the images e.g. by probabilities for having disturbances like shadow, fragmentation or low contrast. The expected appearance of the road is also partially represented in the underlying road model of each algorithm, but in many cases it can be adapted to the scene by parameter steering. Actually we use three types of context regions given by the GIS for extraction and evaluation: rural, forestry and urban. The appropriate parameters are defined by empirical studies.
    - The local context, e.g. the local neighbourhood relations between different objects also influences the appearance of the road in the images by interrelationships like occlusion and shadow, connectivity and parallelism conditions (e.g. buildings cast shadows on roads and buildings in general are connected to roads). It is difficult to model the influence of local context and therefore is not considered in our approach.

The main idea of our procedure is to exploit the initial scene description in the geodatabase to guide and constrain the road extraction by a knowledge-based procedure in the following way:

- by definition of region of interest
- by selection of the appropriate algorithm
- by parameter control of the road extraction
- by parameter control for evaluating the results

4.2.3 Verification of Road Data and Acquisition of Changes

For quality control we distinguish two different functionalities, namely verification and acquisition of changes. The partitioning mainly is motivated by the different amount of knowledge, which can be exploited:

- *Verification:* The verification focuses on those objects, which are described in the database. It is able to check the positional and the thematic accuracy as well as the completeness subelement commission, but not the completeness subelement omission (cf. Table 1). Beneath general context information object specific knowledge defined by the object instances in the database is used during the road extraction.
- *Acquisition of changes*: The acquisition of changes cannot use any specific knowledge defined by the object instances as it aims at checking the completeness subelement omission (cf. Table 1) that is registering additional objects, which are not contained in the database. The knowledge, which is used, is about the scene, about the global context and in general about the objects of interest. The acquisition of changes is executed subsequently to the verification to introduce verified ATKIS objects as reliable pre-information.

Verification: The verification is performed object by object by comparing the existing road data with roads extracted from the images. The geometric and thematic description of each object is transferred to constraints for defining regions of interest, the appropriate algorithm and its parameters used for automatic road extraction for verification of the respective ATKIS object. The differences between extracted roads and the original data are analysed and evaluated. The evaluation result leads to simplified quality measures described by different quality classes. For road elements which could not be initially verified the reason for rejecting the data is analysed by refined verification in a feed-back-loop following the hypotheses and verify paradigm. The generation of new hypotheses is performed by analysing the specific geometric and radiometric situation given by the raster data.

Acquisition of changes: After having verified the existing data the acquisition of changes can be performed. This task is even more difficult as it can be compared to object extraction from scratch, where no constraints are given by the GIS. The only pre-knowledge that can be introduced in this case is given on the one hand by the verified road data which can be used as reliable road parts during the road network generation. On the other hand the context regions can be used for steering the extraction as well as for self-diagnosis of the extraction result. The reliability of the extraction result especially depends on the extraction context and on the underlying low-level extraction used for road network generation. A self-diagnosis is used to derive a traffic-light-solution describing the quality of the data by a qualitative description as it is done during the verification of the existing data.

4.2.4 Quality evaluation

After the road extraction either for verification or for change acquisition, the extraction result has to be compared to the existing geodata to derive a quality measure. The quality measure has to distinguish road verification and acquisition. The quality description is simplified to a so-called traffic-light solution indicating three types of quality attributes: verified, rejected, and ambiguous. Details and a refinement of the decision may be analysed by a feedback loop to denote if the error is caused by wrong attributes or wrong geometry.

- *Quality description in road verification:* For verification we check if and how good an extracted road matches the corresponding ATKIS object. If no road matches the ATKIS object the object is denoted rejected. Otherwise the RMS is derived to further classify into verified and ambiguous objects.
- *Quality description in acquisition of changes:* In acquisition of changes we actually compare the extracted roads to the verified roads using the evaluation scheme proposed by (Wiedemann et al. 1998b). We especially are interested in new objects which are not contained in the verified dataset and which are denoted as changes and are delivered to the operator. As there still are extracted many false road elements a further classification of these changes is required, e.g. using the internal accuracy of the extracted roads. Thus the user interaction can further be reduced to very probable changes.

Fig. 4 shows an example of the automatically derived quality measure and the underlying road extraction used for road verification.

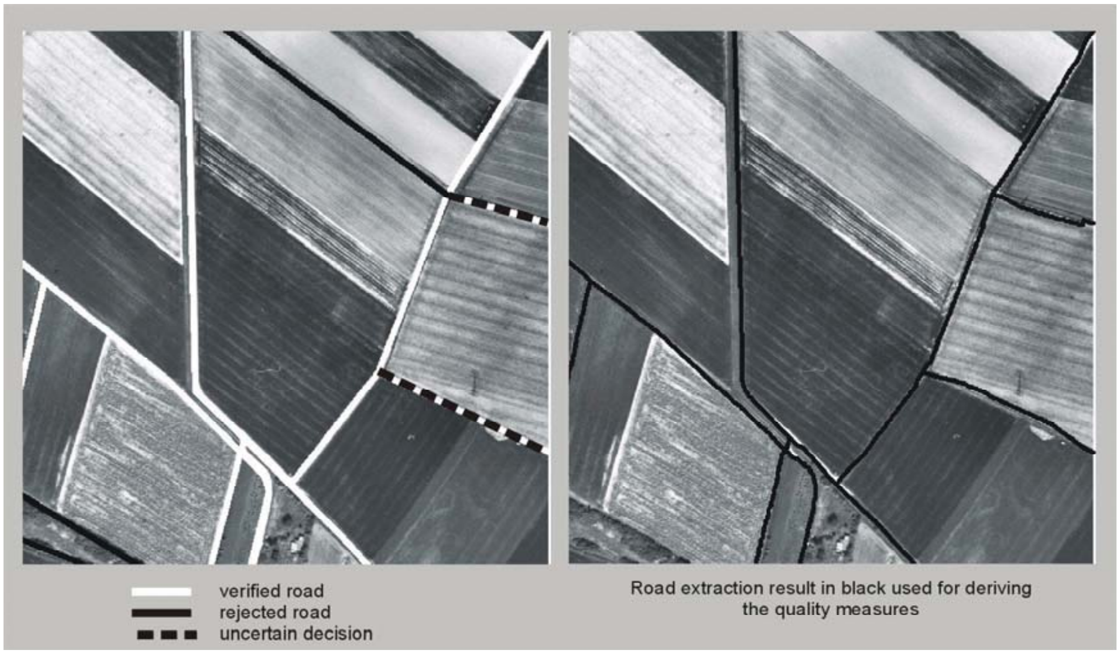


Figure 4: The result of the fully automatic quality control of the verification step as it is transferred to the interactive quality control by an operator.

5. INTERACTIVE QUALITY CONTROL

5.1 Tasks of the Operator

An efficient interlocking of the interactive steps of the operator and the automatic verification procedure as described in Section 4 is essential to guarantee an optimal workflow and a significant increase in productivity. Therefore the operator needs a graphical user interface (GUI) that allows fast and simple access to

- the image data,
- the relevant data from the ATKIS DLMBasis,
- the automatic verification procedure from Section 4,
- the tools to convert the ATKIS data to formats that can be read by the automatic verification procedure,
- the results of the automatic verification procedure,
- the tools for the final editing of the results.

After the automatic verification procedure has run as a batch process in the background, the results are imported and visualized together with the orthoimages on the screen. The quality measures are delivered as attributes of each inspected ATKIS object. They are used for an appropriate visualization to guide the operator to those objects that require his intervention. Of most importance are situations where the knowledge-based system indicates an uncertain decision. In these cases the final decision is made by the operator who classifies it as verified or not verified. All ATKIS objects that could not be verified by the automatic verification procedure have to be checked by the operator, too. Here the decision of the system has to be corrected, if necessary. To ensure that all objects that are classified as uncertain or not verified are processed, they can be stored in a queue that has to be worked of.

5.2 Editing Mode and Record of Quality Description

The editing functionality of the GUI allows the operator to correct or complete the results of the automatic procedure. All ATKIS objects for which any incorrectness has been detected are copied to a special file where all errors are recorded. From a list of items the operator can select the type of the detected error, e.g. wrong attribute road width, geometric deviation exceeds tolerance. Missing objects and copies of objects with large geometry deviations can be edited in this record file. There is no functionality to assign topological relations with other ATKIS objects or to build up a complete topology. This it not necessary since the results are reported to the responsible authority. For this purpose it is sufficient to provide the information needed to locate and identify the errors.

6. CONCLUSIONS

We presented the concept of quality management of the Authoritative Topographic-Cartographic Information System ATKIS® in Germany as it is proposed at the BKG. Within the production flow the vector data actually are automatically checked for conformity with the underlying ATKIS data model in an operational way. Checking the consistency of the data with reality is much more time consuming. To solve this task in an economical way we presented a prototype of a knowledge-based photogrammetric cartographic system, which we developed to speed up the whole production workflow in quality control. This system is designed to increase the efficiency of the updating process by combining automatic procedures with user interaction in a GIS environment. First results show the range of the concept. However the system has to be tested on a large amount of data sets to rate its performance. The quality check on completeness in the sense of commission has to be refined to increase the reliability of the automatically derived result.

REFERENCES :

AdV, 2000, *Progress Report 2000*. http://www.adv-online.de/english/index.htm#top, http://www.adv-online.de/intern/g_stelle/dokument/adv2000_e.zip

Baltsavias, E, Grün, A., van Gool, L., 2001, *Automatic Extraction of Man-Made Objects from Aerial and Space Images(III),* A.A. Balkema Publishers, Lisse / Abingdon / Exton(PA) / Tokio.

Baumgartner A., Eckstein W., Meyer, H. Heipke C., Ebner H., 1997: *Context-supported road extraction*. In Grün, A., Baltsavias, E.P., Henricsson, O., (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag Basel, pp. 299-308.

Endrullis, M., 2000*, Bundesweite Geodatenbereitstellung durch das Bundesamt für Kartographie und Geodäsie (BKG),* In R. Bill and F. Schmidt, eds., ATKIS − Stand und Fortführung, Beiträge zum 51. DVW-Seminar am 25. und 26. September 2000 an der Universität Rostock, Schriftenreihe des DVW, Band 39, Verlag Konard Wittwer, Stuttgart pp. 39-52.

Förstner, W., Liedtke, C.-E. and Bückner, J. eds., 1999*, Semantic Modeling for the Acquisition of Topographic Images and Maps / SMATI 99,* http://www.ipb.uni-bonn.de/SM/smati99.html.

Gülch, E., 2000: *Digital Systems for automatd cartographic feature extraction,* In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, B2, pp. 241-256.

Joos, G., 2000, Zur Qualität von objektstrukturierten Geodaten. In Schriftenreihe des Studienganges Geodäsie und Geoinformation der Universität der Bundeswehr München, Nr. 66.

Lang, F., Förstner, W. 1996: *3D-city modelling with a one-eye stereo system*, In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, B4, pp. 261-266.

Liedtke C.-E., Bückner J., Pahl M., Stahlhut O., 2001: *Knowledge based system for the interpretation of complex scenes*, In E. Baltsavias, A. Gruen, L.,van Gool (eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), A.A.Balkema Publishers, Lisse / Abingdon / Exton(PA) / Tokio, pp. 3-12.

Quint, F., Sties, M., 1995: *Map-based semantic modeling for the extraction of objects from aerial images*, In: Grün, A., Kübler, O., Agouris, P. (eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, Basel, pp. 307-316.

Straub, B.-M., Wiedemann, C., 2000: *Towards the update of geodata by automatic object extraction*, In Serpico S.B. (ed.), SPIE Vol. 4170, Image and Signal Processing for Remote Sensing VI. 304-315.

Vosselman, G., 1996: Uncertainty in GIS supported road extraction, In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, B3, pp. 909-917.

Walter, V., 1999, "Automated GIS data collection and update". In D. Fritsch and R.H. Spiller, eds., *Photogrammetric Week `99,* Wichmann, Heidelberg, pp. 267-280.

Wallace, S.J., Hatcher, M.J., Priestnall, G., Morton, R.D., 2001: *Research into a framework for automatic linear feature identification and extraction*, In E. Baltsavias, A. Gruen, L. van Gool (eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), A.A.Balkema Publishers, Lisse/Abingdon/Exton(PA)/Tokio, pp. 381-390.

Wiedemann, C., 2001: *Extraktion von Straßennetzen aus optischen Satellitenbildern*, Dissertation Lehrstuhl für Photogrammetrie und Fernerkundung, TU München.

Wiedemann, C., Heipke, C., Mayer, H., Hinz, S., 1998a: *Automatic Extraction and Evaluation of Road Networks from MOMS-2P Imagery*, In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 3/1, pp. 285-291.

Wiedemann, C., Heipke, C., Mayer, H., Jamet, O., 1998b: *Empirical evaluation of automatically extracted road axes*, In: Empirical Evaluation Methods in Computer Vision, Ed. Bowyer, K.J., Phillips, P.J. IEEE Computer Society Press, pp. 172-187.

Spatial Data Quality: From Metadata To Quality Indicators and Contextual End-User Manual

Rodolphe Devillers (1), Marc Gervais (2), Yvan Bédard (3) and Robert Jeansoulin (4)

(1) (2) (3) CRG, Pavillon Louis-Jacques Casault, Université Laval,
Québec (QC) G1K 7P4, Canada
rodolphe.devillers@scg.ulaval.ca (1), marc.gervais@scg.ulaval.ca (2), yvan.bedard@scg.ulaval.ca (3)

(4) LSIS-CMI, Université de Provence
39 r. Joliot Curie, 13453 Marseille Cedex 13, France
robert.jeansoulin@cmi.univ-mrs.fr

Abstract

The context within which geospatial data are used has changed significantly during the past ten years. Users have now easier access to geospatial data but typically have less knowledge in the geographical information domain, so have limited knowledge of the risk related to the use of geospatial data. This sometimes leads to faulty decision-making that may have significant consequences. In order to reduce these risks, geospatial data producers provide metadata to help users to assess the fitness for use of the data they are using within the context of their application. However, experience shows that these metadata have several limitations and do not reach their information goal for this new group of non-expert users. In addition, geospatial data are becoming a mass product that has to follow legal requirements related to this class of products. Metadata, as currently defined, do not reach these obligations, especially concerning the requirements for easily understood information about product specifications and potential risks of misuse. This paper describe an approach that aims to reduce these risks of misuse by comparing data producers specifications and data users needs and providing indicators describing data quality to users. The system, named Multidimensional User Manual (MUM), allows the management of geospatial data quality and the communication of the quality information using indicators that can be analysed at different levels of detail.

Keywords: Geospatial data Quality Indicators, GIS Risk, interactive User-Manual, Spatial OLAP, Law.

1. INTRODUCTION

The development of information and communication technologies such as the Internet has significantly modified the context within which geospatial datasets are used. We now face an increase of geospatial datasets being exchanged between people or organisations. Geospatial data can now be easily downloaded on the Internet (e.g. Geospatial data catalogues). It is also possible to download free GIS software that is easy to use for non-expert users (Agumya and Hunter, 1997, Elshaw Thrall and Thrall, 1999, Goodchild, 1995). In addition, many organisations are building new data infrastructures such as Data Warehouses in order to regroup their different data in a common system and share them at different levels of the organisation, facilitating then information diffusion. Decision processes based on geospatial data are also increasing (Hunter, 1999). Geospatial datasets are now used in a wide range of applications such as public health, geology, forestry, transportation, urban planning, etc. (Longley et al., 1999). They are also used today by people at different levels of the organisation (i.e. operational, tactic and strategic levels) whereas it used to be used mainly by experts in geographical information (Longley et al., 1999). Consequently there are new categories of GIS users who are experts in their field of application but not in the geographic information domain.

Such a situation leads to a higher risk of misuse of the GIS data, potentially leading to faulty decision-making that may have significant legal, social and economical impacts. Many cases of Geospatial data misuse are mentioned in the literature and the media, causing an increasing number of legal contentious (e.g. (Beard, 1989, Goodchild and Kemp, 1990, Monmonier, 1994, Curry, 1998)). Most of Geospatial data users do no understand the main geographical information concepts and many of them are not aware of the uncertainty that digital data may contain (Morrison, 1995). They then often integrate several datasets having similar appearance without being aware of the potential risks that may arise from these combinations (Curry, 1998).

The objective of this paper is to present the Multidimensional User Manual (MUM) project that aims at providing to the users of geographical information an automatic contextual manual that helps them evaluate the risk involved when using one or several geospatial datasets for a given region. We thus aim at reducing the number of potential cases of geospatial data misuse.

We will first discuss about some problems arising from data quality terminology and its impact on systems dealing with data quality. We will continue with a discussion about limitations of present metadata in term of communication between data producers and users. We will then see how well these metadata are compliant to legal requirements, using Quebec province legislation as an example. We will then focus on different aspects of geospatial data quality management and communication by describing the general architecture of the MUM in order to identify the different modules of the system. We will present a way to manage geospatial data quality in a multidimensional database and introduce the concept of quality indicators defined from a user perspective. Finally, data quality visualisation issues will be briefly described in the context of the MUM.

2. COMMUNICATIONNAL AND LEGAL LIMITATIONS OF PRESENT METADATA

2.1 Spatial Data Quality issues

Data quality is a very active domain in geographic information research and has a growing interest because of the increase of data exchange (Goodchild, 1995, Veregin, 1999). There is now considerable agreement on the definition of quality in the literature, quality being defined as "fitness for use" (Veregin, 1999). Quality is defined by ISO 8402 as the "totality of characteristics of a product that bear on its ability to satisfy stated or implied needs". This means that to define quality you need both the information on the data being used and on the users needs (e.g. intended use). Although the "fitness of use" definition is frequently referred to, it is surprisingly rare that user needs are actually taken into account. There has only been a relatively small amount of research on measurement of fitness of use (Bédard and Vallière, 1995, Agumya and Hunter, 1997, De Bruin et al., 2001), but there is a strong need for systems that can implement data quality as fitness for use. However, there is an ambiguity in the definition in the term "quality". It is noticeable that data producers often define a product of quality as being consistent with specifications (ex: difference between a database and the nominal ground) while data users define it as meeting or exceeding their expectations (Kahn and Strong, 1998). It is frequent to see products having a numeric value as quality. However, it does not make sense to allow a unique value on a certain scale of quality for a product, as quality may only be measured within the intended use context that may be different for each users, and even among different for the same user.

Another problem is the difficulty in properly communicating the information about data quality. Although several researchers have explored the best way data quality may be visualised (Beard and Mackaness, 1993, Buttenfield and Beard, 1991, Buttenfield, 1993, Leitner and Buttenfield, 2000), the only way actually used to communicate data quality parameters is using metadata. However, the currently produced metadata still have strong limitations in term of communication media for non-expert users as well as for expert users.

2.2 Present spatial metadata limitations

In order to help users to assess the fitness for use between the data and their needs, data producers often provide metadata describing different aspects of the datasets (Guptill, 1999). Metadata are usually stored in text files using “home-made” formats that may be based on national standards (e.g. FGDC). However, data providers are moving to international standards such as ISO and OpenGIS which include both conceptual and implementation specifications. However, these metadata are still often stored in files that are independent from their related data, then, if data are being modified, changes are not always propagated to their associated metadata. As metadata are static in nature, they are not very useful for dynamic operations when using a GIS. In addition, for reasons of cost, time and complexity, metadata are often related solely to the dataset and rarely to individual objects (or even to geometric primitives). As quality is often heterogeneous within a single dataset, it is important to describe it, at least, at the object level as allowed by the most recent standards and high-quality datasets.

Metadata are also often technical descriptions dedicated to experts or professionals, using a terminology that is hermetic for non-expert users (ex: “Polygon topology was verified with the Arc/Info "BUILD" command” – extract from SMMS for GeoMedia quality metadata sample example). Consequently, geospatial metadata are often unused by the users (Harvey, 1998, Timpf et al., 1996). Thus, quality description is inadequate for most users and does not help them to decide if a potentially useful dataset should be acquired and used (Frank, 1998).

2.3 Spatial Metadata and legal issues

In many countries, applicable legislation allows one or more civil liability regimes with whom the objective is to govern interrelations between citizens and to penalize some reprehensible behaviours, should the occasion arise. There are many similarities between the legal systems in different states, especially those having a legislation originating from the Napoleon code (province of Quebec in Canada, France, Belgium, etc.). The analysis of the juridical context related to the MUM project focuses on the legislation actually applied in the province of Quebec in Canada, especially by rules included in the Quebec Civil code (Q.C.c.). Even if the conclusions may not be *fully* applicable in all countries supporting a similar regime, the data producers’ juridical obligations are usually built in these countries on identical concepts.

The two main liability regimes in Quebec are civil *contractual* liability (art. 1458 Q.C.c.) and civil *tort* liability (art. 1457 Q.C.c.). The first one results from a contravention to a *to do* obligation or *not to do* obligation, *temporary* and getting its source in a juridical *act* (contract). The second one results from a violation of a *not to do* obligation, *continuous* and *legal* and getting its source in a juridical *fact* (Baudoin and Deslauriers, 1998, Baudoin and Jobin, 1998). It involves, normally, two parties that are not bound by a convention.

With regard to tort liability, the following constitute a fault for producers or distributors:

- Failure to give warnings and cautions that are clear, complete, and up to date;
- Failure to inform the user adequately about the product risks and dangers;
- Failure to give the means to take precautions against it (directions for use) and on the particularities making it inappropriate to the expected use (ex: road map designed for pedestrian tourism navigation and inappropriate for car navigation).

To these different fault categories, we can also add poor product conception. (Ex: map road designed for car navigation but without considering one-way streets).

Different means of exemption exist for producers and distributors. However, any of these can be admissible with regard to corporeal or moral damages (art. 1474 Q.C.c.). As for material damages, exemption will be effective if absence of security default is demonstrated, that the prejudice origins either from a major force case (event unforeseeable and irresistible, art. 1470 Q.C.c.), or from victim negligent behaviour itself, or finally, from an unknown default considering the science development at the moment of the product commercialisation.

The development of the mass market of consumer goods has greatly influenced the juridical context and had given birth in many places to a distinct liability regime much more severe (more particularly the « *Loi sur la protection du consommateur* (L.P.C.), c. P-40.1 » in Quebec, and the « *Directive 85/374 du Conseil du 25 juillet 1985* relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux dans certains pays d'Europe ») with which the main goal was to frame the relations between physical and moral persons having "consumer" status and trades people. In Quebec, a consumer is defined as *a physical person, except for a trade people who is procuring a good or a service with a commercial purpose* (art. 1 (e), L.P.Q, free translation). The number of court decisions, related to this type of trade has exploded in the past twenty years. This new liability regime are distinguishing itself from the others notably by the following characteristics:

- Additional obligations about formal procedures;
- Vice precedence presumption, i.e. that product default is presumed to be present at the transaction moment;
- Impossibility to plead ignorance or reasonable diligence or absence of risks knowledge (irrefutable presumption, risk relied to the technological uncertainty is burden to the trade people);
- Increased pressure to take the information disclosure initiative;
- Punitive or exemplary damages in addition to usual compensatory damages (art. 272 L.P.C.);
- Interpretation on consumer's behalf (art.17 L.P.C.).

The information industry (which includes geographic information) is changing in the direction of massively increased information production and circulation, loss of privileged contact between producer and consumer, and increasing number of non-experts users, as discussed above. These factors contribute considerably to the increasing risk of bad decisions, bad interpretations, and failing applications. In this context, if it is as fact that to the same causes follow the same effects, contentious increasing may statistically increase in this direction (Montero, 1998).

A Geographic information producer or distributor sees their liability involved if, firstly, they make a fault in the Quebec civil code sense, secondly, this one cause a certain prejudice and, thirdly, if there is a causal relation between fault and prejudice. It is notably at the level of proof burden and the fault concept that the impact of the "*Loi sur la protection du consommateur"* is much more noticeable. Twenty years ago, a consumer taking proceedings against producer need to prove (which was often complex and outside his field of expertise), a product conception default or of the prejudice author's negligent behaviour. Then the legislator, anxious of consumer protection and contractual balance maintenance, reversed the proof burden and requires now that the producer and distributor themselves prove absence of default, the consumer having only to declare that product or service was containing a *default*.

Liability regime interpretation, with regard to informational product, allow us to assimilate to the fault concept the following situations where producer and/or distributor neglects to put in place technical and organisational mechanisms in a way to prevent damages, to reveal the part of uncertainty included into the data, to formulate indications on the value of provided information, and to control the software performance being used for interface between user and informational product (Montero, 1998). For this last obligation, we can expect that the producer or distributor has to do some compatibility tests between geographic information system available on the market and the database in order to inform the user which one is recommended.

Considering the different liability regime mentioned above and the possibility that they can be brought into operation following the commercialisation of geographical information, database producers must ask themselves if the information furnished in a transaction are consistent with the juridical *obligations* which fall on them. With respect to the previous discussion, it is plausible to believe that the current forms of metadata do not allow the producer to respond *adequately* and *totally* to these requirements especially with regard to producers' and distributors' information obligations. In fact, metadata:

1. Rarely contain warnings and cautions with regard to the expected use, and considering the language used, are hardly admissible in front of a civil court with regard to the requirement of being clear and complete;
2. Do not constitute directions for use, which inform user on the product risks and dangers and the means to take precautions against it, even in some cases, when sufficient information allowing one to infer it is present, the interpretation burden falls on the user (Frank, 1998);
3. Can be hardly considered as a technical and organisational mechanism in a way to prevent damages, especially since they are often incomprehensible to non-experts users. (Ex: Mercator Transverse cartographic projection, reference datum NAD 83, etc.) ;
4. Are much too numerous (particularly when they are associated to each entity) to allow efficient consultation and to infer in which measure the expected application will generate *reliable* results, this situation can be considered as much a management and user-friendliness constraint as a juridical constraint.

The user manual conception, the ultimate goal of the MUM project, will take into consideration these obligations of Quebec legislation described in this section.

3. SPATIAL DATA QUALITY MANAGEMENT AND COMMUNICATION

In order to provide to users information about geospatial data quality in an easily understood format, we need to define database structures that may store this information. We suggest an implementation using (i) a single multidimensional database containing both raw data (metadata and other source of information about data quality) being used before comparing information about data quality and users' needs and (ii) another multidimensional database containing the values of comparisons between the aggregated values and the users needs (indicators values) that will be displayed to the users.

3.1 Multidimensional User Manual General Architecture

In order to solve these different problems, we aim at elaborating a system that evaluates data quality according to the "fitness for use" definition. The system would compare data specifications as provided by data producers, with the needs, as expressed by users. It would also take into account the level of risk a user is willing to take for his project. In order to be able to compare these descriptions, the system would formalise them into a product ontology and a user ontology. By comparing these two ontologies (describing the properties of objects, relations, etc.), it would be possible to quantify and qualify the proximity between the two (i.e. the overall quality). The information produced by this comparison would highlight some possible risks that may occur when manipulating the data in some context. For example, a dataset may represent houses. According to data producers, houses may be all constructions that have more than 150 m^2 whereas data users may need all constructions of any size. These risks can then be communicated to users using an interface coupled or integrated with the GIS.

3.2 Spatial Data Quality indicators hierarchy

Information about Geospatial data quality has to be provided to data users in order to help them during their decision processes. Cognitive sciences and decision processes theory teach us that humans often base their decisions on indicators. A physician will for instance look at different indicators when treating an illness (ex: pulse, tongue, temperature) before deciding on the medical treatment to apply. Decision support systems also often use indicators in order to provide high-level aggregated values to strategic decision-makers. In this context, indicators may be defined as "a way of seeing the big picture by looking at a small piece of it" (Jackson community Council, 1999). We need then to define indicators that may inform users more efficiently than present metadata descriptions.

The MUM system organises the indicators following a multidimensional hierarchy (see Figure 1) that may be managed using an On-Line Analytical Processing (OLAP) system (Berson and Smith, 1997, Codd, 1993). This category of tools allows users to navigate at different levels of aggregation within a multidimensional database without having an information overload (only a few pieces of information are displayed at each level). Indicators' values may be displayed using different representations depending on the type of value (street light, yes/no, percentage, speed meter, etc.). Figure 1 uses only a street light display for indicators values, the green light meaning that data is fit for the intended use (low risk), the red one meaning that there may be high risks on the associated indicator and the yellow one being somewhere in between. A user may then be aware of the overall data quality by viewing a single indicator "SD Quality" (for Spatial Data Quality). If he wants to have further detail on this indicator he may use the OLAP function "Drill down" in order to display the indicators of which this high level indicator is composed (in this case, Confidence, Data quality and Risk seriousness). He may then wonder why the Data Quality indicator has a red value and drill down again on this indicator to display the next level of detail, etc.

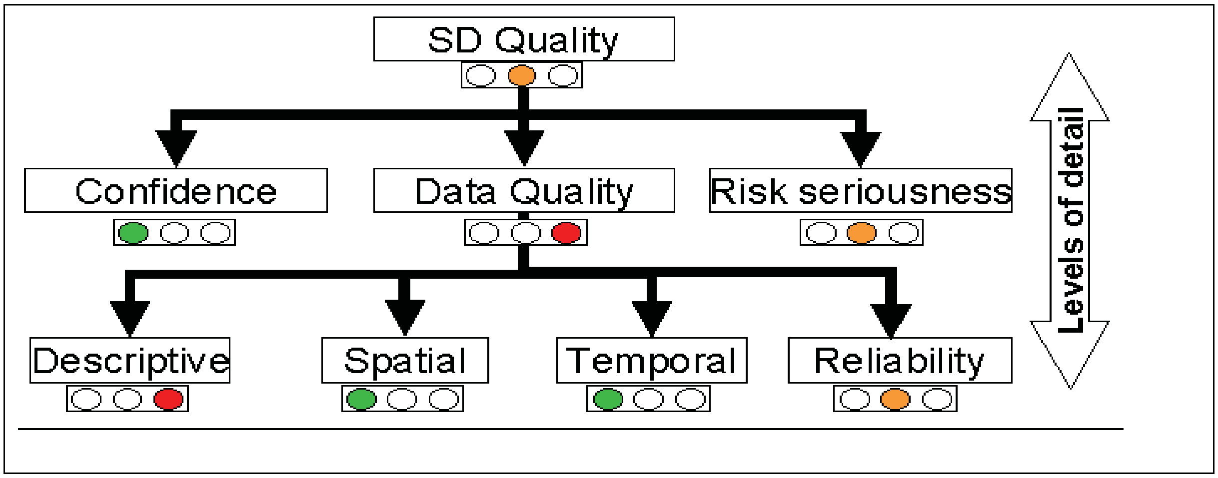


Figure 1. Example of Geospatial Data Quality indicators hierarchy.

3.3 Spatial Data Quality dimensions

Information about Geospatial Data Quality may be organised using different levels of aggregation within the database. In order to manage these data into a multidimensional database, we first need to define the different dimensions the data may follow. Three reference dimensions are used in this structure (geometry, semantic and temporal) in addition of dimensions for each quality indicator. Reference dimensions allow the association of quality values with its level (e.g. Indicator A associated to an instance of the class object lake). Unique combination of reference and quality dimensions allows to have a measure for the quality.

3.3.1 Geometric dimension

Quality metadata may be attached to elements at different levels of detail on the geometric dimension. The levels are:

- Geometric primitives (ex: line segments composing a polygonal lake);
- Complete geometric shape of an object (ex: complete polygon of Lake Placid);
- All geometric shapes of a same object class (ex: all polygons representing lakes);
- Dataset or all geometric shapes of all object classes (ex: topographic map of a region including lakes, rivers, streets and houses).

3.3.2 Semantic dimension

Quality metadata may be attached to elements at different levels of detail on the semantic dimension. The levels are:

- Value (ex: commercial);
- Domain (ex: commercial, industrial, residential);
- Attribute (ex: building type);
- Object Class (ex: building);
- Semantic (ex: semantic of buildings, lakes, rivers and streets).

3.3.3 Temporal dimension

Quality metadata may be attached to elements at different levels of detail on the temporal dimension. These levels may depend on the hierarchy defined by the user, such as:

- Temporal primitives;
- Complete temporality of an attribute (evolution) or an object (existence);
- Complete temporality of an object (evolution thru all attributes + its existence);
- All temporalities of a same object class or of a same attribute type of a same object class;
- Dataset or all temporalities of all object classes.

3.3.4 Quality indicator dimension

Quality indicator hierarchy is organised in term of dimensions in the multidimensional database. The different levels are not generic and may change depending on the system conception.

3.4 Information about quality aggregation

Whereas data producers often documented datasets using a short description at the dataset level, many data producers want now to document their datasets at a more detailed level, such as the entity level. However, a complete documentation of the dataset following, for instance, the ISO standard for geospatial metadata, may require to fill more than 400 fields for each object. Then, the volume of metadata that has to be managed by the system would be much more important than the volume of data. It would be in this case impossible to communicate clearly to users a big picture of the overall quality. There is then a need to provide users an aggregated view of the overall dataset quality. We are suggesting in our approach the use of contextual quality indicators that are an aggregation of several information about quality.

Quantitative data aggregation may be done using some basic formulas such as minimum, maximum or average, or more sophisticated methods, depending on the context (ex: aggregation of horizontal and vertical accuracy to create a "Spatial accuracy" indicator). However, if these methods may work for quantitative data integration, it is not possible to apply them to qualitative data aggregation. These may also be adapted for qualitative data aggregation by other data fusion techniques may be more appropriate; for example, certain logical formalisms may allow, using rules, the definition of more complex aggregation techniques.

3.5 Spatial Data Quality visualisation

Geospatial data quality may be communicated to users in different ways. Quality indicators can be visualized using different representations (see section 4.2) within the GIS interface, such as within a "dashboard" that user may consult when needed. For example, figure 2 presents a Spatial OLAP system that integrates the quality information dashboard. This dashboard includes indicators selected by the user. Indicators' values change when the user adds or removes data in the application. The yellow light on the bottom right of the screen, should always be visible in the GIS environment, representing an aggregation of all the indicators' values displayed on the dashboard.

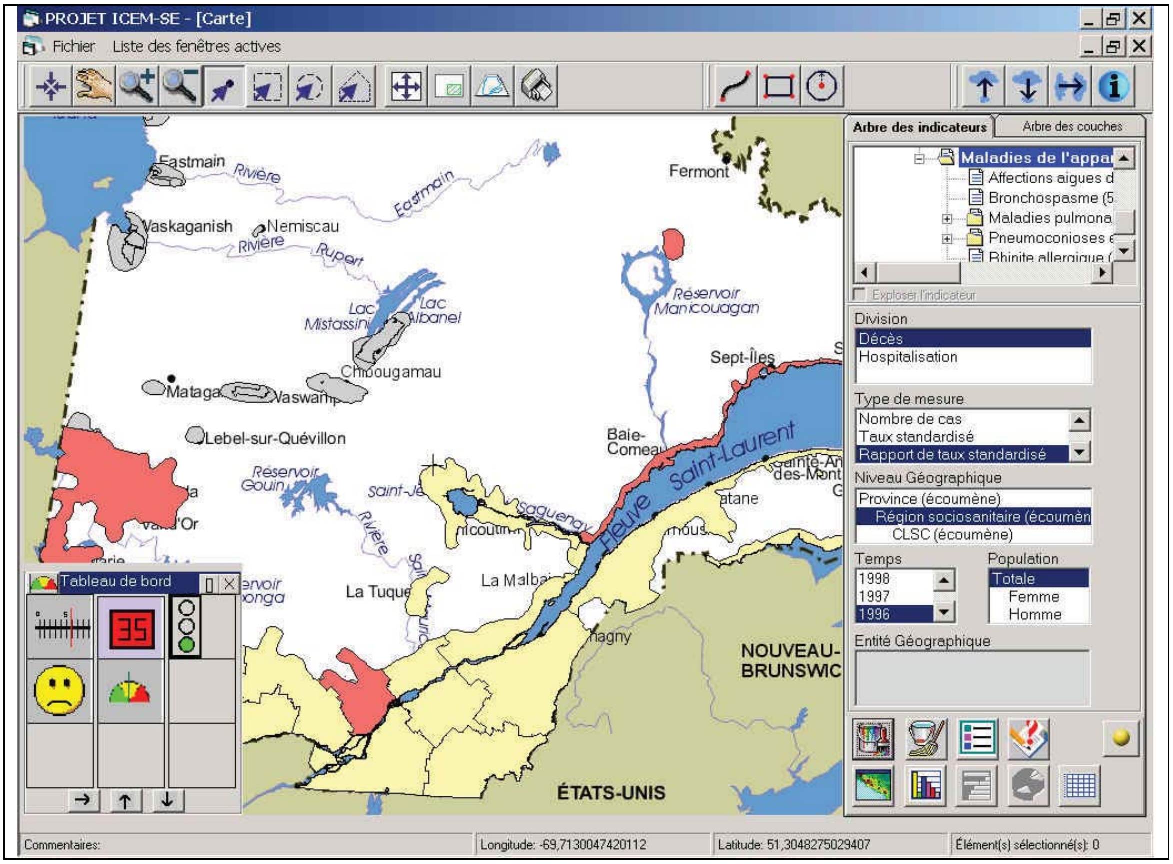


Figure 2. Visualisation of geospatial data quality using a dashboard (bottom left). Each symbol in the dashboard represent an indicator selected by the user.

Another way to represent quality indicator values is to display them directly on maps using Spatial OLAP systems (Rivest et al., 2001). This category of system allows the management and the visualisation of geometric entities at different levels of detail. It may then be possible to directly visualise quality indicators on the associated geometric entities and then to navigate at different levels of aggregation using SOLAP operators (e.g. Drill down). Figure 3 provide an example of Geospatial Data Quality visualisation using a Spatial OLAP system. Users can display quality indicators either with a streetlight representation, or by directly associating the quality indicators to the individual objects, at different levels of detail.

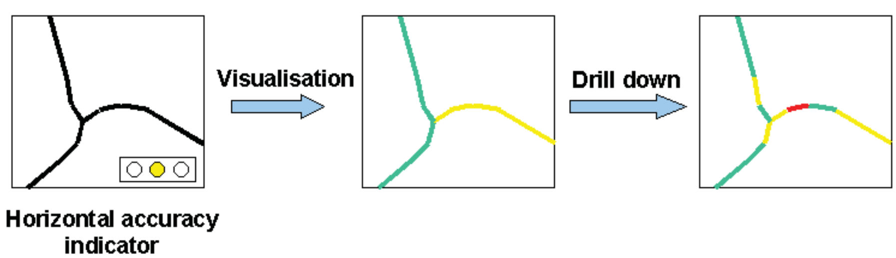


Figure 3. Visualisation of geospatial data quality using indicators and a Spatial OLAP system.

As quality may vary spatially, indicators values have to be updated dynamically when the user change the area visualised.

4. CONCLUSIONS AND REFERENCES

This paper presented a method that allows the management and the communication of spatial data quality to users in order to reduce the risk of misuse. This system, the Multidimensional User Manual (MUM), would respond both to the problem of poor communication with the end users, and to the legal requirements for data producers seem certain to have to follow. We presented the advantages of communicating data quality using indicators, based on the aggregation of different information about data quality, which would be easier to understand for end-users than present metadata. Metadata are fundamental to this approach, since they represent a large part of the information available describing data quality; however we will need more formally defined metadata, using more quantitative data or enumerated values for instance, in order to be able to automatically produce the quality indicators. We described how the quality information and these indicators could be stored at different levels of detail within a multidimensional database and retrieved using an OLAP system. We briefly introduced the challenges in term of data aggregation that will be necessary for the definition of the indicators' values. We then presented how this information about data quality may be communicated to the end-users, whether they are experts in the geographic information domain or not, using both the indicators display at different levels of detail and a Spatial OLAP visualisation system. Each aspect of this User Manual system will be explored in detail in different research projects.

ACKNOWLEDGEMENTS :

The work was funded by the Canadian GEOIDE network, partly by the IST/FET program of the European Community, through the REV!GIS project, the Centre for Research in Geomatics (CRG) and Laval University. Thanks are due to Nic Wilson for the reviewing of the manuscript.

REFERENCES :

Agumya, A. and Hunter, G. J., 1997, "Determining fitness for use of geographic information". *ITC Journal,* 2, pp.109-113.

Baudoin, J.-L. and Deslauriers, P., 1998, *La Responsabilité Civile,* Quebec, Canada.

Baudoin, J.-L. and Jobin, P.-G., 1998, *Les Obligations,* , Quebec, Canada.

Beard, K., 1989, "Use error: the neglected error component". In *AUTO-CARTO 9,* Baltimore, Maryland, pp. 808-817.

Beard, K. and Mackaness, W., 1993, "Visual Access to Data Quality in Geographic Information Systems". *Cartographica,* 30, pp. 37-45.

Bédard, Y. and Vallière, D., 1995, "Qualité des données à référence spatiale dans un contexte gouvernemental". Université Laval, Québec, pp. 53.

Berson, A. and Smith, S. J., 1997, *Data Warehousing, Data Mining and OLAP (Data Warehousing / Data Management),* New York.

Buttenfield, B. and Beard, K., 1991, "Visualizing the quality of spatial information". In *AUTO-CARTO 10*, pp. 423-427.

Buttenfield, B. P., 1993, "Representing Data Quality". *Cartographica,* 30, pp. 1-7.

Codd, E. F., 1993, "Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate". E. F. Codd and Associates, pp. 24 p.

Curry, M. R., 1998, *Digital places: Living with Geographic Information Technologies,* Routeledge, London & New York.

De Bruin, S., Bregt, A. and Van de Ven, M., 2001, "Assessing fitness for use: the expected value of spatial data sets". *International Journal of Geographical Information Systems,* 15, pp. 457-471.

Elshaw Thrall, S. and Thrall, G. I., 1999, "Desktop GIS software". In *Geographical Information Systems*, Vol. Vol. 1 - Principles and Technical Issues (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, New-York, pp. 331-345.

Frank, A., 1998, "Metamodels for Data Quality Description". In *Data Quality in Geographic Information - From Error to Uncertainty* (Eds, Goodchild, M. and Jeansoulin, R.) Editions Hermes, , pp. 192.

Goodchild, M. F., 1995, "Sharing Imperfect Data". In *Sharing Geographic Information* (Eds, Onsrud, H. J. and Rushton, G.) Rutgers University Press, New Brunswick, NJ, pp. 413-425.

Goodchild, M. F. and Kemp, K. K., 1990, "NCGIA Core Curriculum in GIS". National Center for Geographic Information and Analysis, University of California, Santa Barbara CA.

Guptill, S. C., 1999, "Metadata and data catalogues". In *Geographical Information Systems*, Vol. Management Issues and Applications (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, Inc., pp. 677-692.

Harvey, F., 1998, "Quality Needs More Than Standards". In *Data Quality in Geographic Information - From Error to Uncertainty*(Eds, Goodchild, M. and Jeansoulin, R.) Editions Hermes, pp. 192.

Hunter, G. J., 1999, "Managing uncertainty in GIS". In *Geographical Information Systems*, Vol. Management Issues and Applications (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, Inc., pp. 633-641.

Kahn, B. K. and Strong, D. M., 1998, "Product and Service Performance Model for Information Quality: An Update". In *Conference on Information Quality,* Cambridge, MA: Massachusetts Institute of Technology.

Leitner, M. and Buttenfield, B., 2000, "Guidelines for the Display of Attribute Certainty". *Cartography and Geographic Information Science,* 27, pp. 3-14.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (Eds.), 1999, *Geographical Information Systems,* John Wiley & Sons, New York.

Monmonier, M., 1994, "A Case Study in the Misuse of GIS: Siting a Low-Level Radioactive Waste Disposal Facility in New-York Sate". In *Conference on Law and Information Policy for Spatial Databases*Tempe (AZ) USA.

Montero, E., 1998, *La responsabilité civile du fait des bases de données,* Belgium.

Morrison, J. L., 1995, In *Elements of spatial data quality* (Ed, inc., E. S.) New York, pp. 202.

Rivest, S., Bédard, Y. and Marchand, P., 2001, "Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP)". *Geomatica,* 55, pp. 539-555.

Timpf, S., Raubal, M. and Werner, K., 1996, "Experiences with Metadata". In *Symposium on Spatial Data Handling, SDH'96, Advances in GIS Research II*, Vol. 2 (Eds, Kraak, M.-J. and Molenaar, M.) IGU, Delft, The Netherlands, pp. 12B.31 - 12B.43.

Veregin, H., 1999, "Data quality parameters". In *Geographical Information Systems*, Vol. Principles and Technical Issues (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, Inc., pp. 177-189.

Web Based Services For Data Quality Evaluation – A User-Oriented Approach

Anders Östman

Division of Geographic Information Technology
Luleå University of Technology
SE-971 87 LULEÅ
SWEDEN

Key words: Spatial data quality, spatial data usability, quality assurance routines, open systems

Abstract

During the last decade, much emphasis have been put on standardising the components of spatial data quality. As a result, standardised quality measures related to geometric accuracy, thematic accuracy, completeness and consistency have been defined. According by a survey performed by the ICA, most data producers at the national level use these measures when specifying the data quality. However, the same study also report that the users of the data don't request such specifications to a high degree. The reason for this has not been investigated, but one possible explanation is that these measures are difficult to interpret or use and that there is a lack of software and procedures in using these quality measures.

The quality assurance routines used today, are usually suited for data producers, who produce large volumes of data in a repetitive manner. In such cases, much effort can be spent on data quality assurance routines with preserved or decreased production costs. But most data producers don't produce the same amount of information. As a consequence, other approaches for establishing data quality assurance routines have to be used.

Luleå University of Technology have addressed this question in its research during the last years. The approach is to build Web-based services that a user can use for quality evaluation and quality assurance. The services are based on open systems, as defined by the specifications from the Open GIS Consortium (OGC). Currently, the services being implemented are used for error propagation studies using Monte-Carlo simulation. In the (near?) future, other services will also be implemented. The services are currently not robust enough to be used in real production. As a consequence, they are not yet publicly available.

Outside the GI field, the term "usability" is often used in user-centred contexts. Here quality is only one, but often important, dimension of usability. Other important dimensions are for instance the accuracy of achieved goals, the efficiency of the procedures, freedom from discomfort etc. There are several different definitions of usability, also within the ISO family of standards. It is however not the purpose of this presentation to analyse all these definitions in detail. But the conclusion is anyway that

many other fields outside the GI arena consider the term "data usability" is more user centred and that data quality is an essential part of data usability.

The Association of GI Laboratories in Europe (AGILE), is an organisation consisting of more than 50 members from more than 20 European countries. The members in AGILE are mainly universities with laboratories within the GI sector. A working group on spatial data usability is currently being established. In November 2001, a workshop was held in Wageningen, where scientists discussed scientific issues and action plans. If OEEPE and ISPRS consider establishing international activities within the field of spatial data quality, a co-operation with the AGILE working group is encouraged.

Participants to the Workshop On Spatial Data Quality Management

NO	Name & Surname	Email, Tel & Fax	Country	Function
1	Dursun BAK Major-General	dbak@hgk.mil.tr	Turkey	Host & Opening Address Speaker
2	M.Orhan ALTAN Prof.Dr.	oaltan@itu.edu.tr Tel: ++90 212 285 38 10 Fax: ++90 212 285 65 87	Turkey	Host & Organizer & Session #1 Chairman & Workshop Reporter
3	Hayati TASTAN Dr-Ing.LtC.	htastan@hgk.mil.tr Tel :++90 312 5952185 Fax : ++90 312 3201495	Turkey	Organizer & Speaker & Breakout Session#3 Leader & Workshop Reporter
4	Andrew FRANK Prof.Dr.	frank@geoinfo.tuwien.ac.at	Austria	Keynote Address Speaker
5	Costas ARMENAKIS Prof.Dr.	armenaki@NRCan.gc.ca	Canada	Session#2 Chairman
6	Andrew SMITH Prof.Dr.	absmith@ordsvy.gov.uk Tel:++44 (0) 23 8079 2378 Fax: ++44 (0)23 8079 2078	UK	Speaker & Breakout Session#1 Leader
7	Oystein ANDERSEN Prof.Dr.	oystein.andersen@ikf.nlh.no Tel: ++47 64948843 Fax: ++47 64948856	Norway	Speaker & Session#3 Chairman
8	Jorgen GIVERSEN Prof.Dr.	jgg@kms.dk Tel:++45 35875707 Fax: ++45 35875067	Denmark	Speaker &
9	Andreas BUSH Dr.-Ing.	Busch@ifag.de Tel :++ 49 69 6333 312 Fax : ++ 49 69 6333 441	Germany	Speaker & Breakout Session#1 Leader
10	Rudolphe DEVILLERS Dr.-Ing.	rodolphe.devillers@scg.ulaval.ca Tel:++418 656-2131 #3203 Fax:++418 656-3607	Canada	Speaker & Breakout Session#2 Leader
11	Anders ÖSTMAN Prof.Dr.	anos@sb.luth.se Tel: ++46 920 91975	Sweden	Speaker & Breakout Session#2 Leader
12	Cevat ULKEKUL Major-General (Retired)		Turkey	Participant
13	Mustafa ONDER Dr.Col.	monder@hgk.mil.tr	Turkey	Participant
14	Oktay AKSU Eng.LtC.	oaksu@hgk.mil.tr	Turkey	Participant
15	Jean Pierre BEEKMAN		Belgium	Participant
16	Nihat KANDAL Surveying Eng., PhD.	nkandal@tapu.gov.tr	Turkey	Participant
17	Orhan ERCAN Surveying Eng., PhD.	oercan@tapu.gov.tr	Turkey	Participant
18	S. SABUNCUOGLU Civil Eng.	ubm@ubm.com.tr	Turkey	Participant
19	Murat Anıl ARISOY Civil Eng.		Turkey	Participant
20	Yaşar HASEYÜPOGLU Surveying Eng.	hacieyup@ubm.com.tr	Turkey	Participant
21	Hande DEMİREL Surveying Eng., PhD.		Turkey	Participant
22	Zaide DURAN Surveying Eng., PhD.	hdemirel@itu.edu.tr	Turkey	Participant
23	Hayati KOYUNCU Dr.-Ing.	hkoyuncu@sieurasia.com	Turkey	Participant
24	Serdar CORMAN		Turkey	Participant
25	Ahmet KARABURUN Surveying Eng., MSc.	akaraburun@islem.com.tr	Turkey	Participant