Unit #11

Data Protection

MIS5214

Agenda

- In the News – Section 001
- Data protection by design
- System Security Plan
  - Cloud computing specifications
  - Security control inheritance

Data protection by design and default…

Data protection capabilities must work from beginning to end of data processing to enable protection of individuals' personal data by default

Art. 25 GDPR

Data protection by design and by default

(1) Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

(2) The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

(3) An approved certification mechanism pursuant to Article 42 may be used as an element to demonstrate compliance with the requirements set out in paragraphs 1 and 2 of this Article.

Danezis, G. et al. (2014) "Privacy and Data Protection by Design", European Union Agency for Network and Information Security (ENISA)

D' Acquisto, G. et al. (2015) "Privacy by design in big data", European Union Agency for Network and Information Security (ENISA)

Key General Data Protection Regulation (GDPR) requirements:

1. **Collection** of personal data is **fully avoided or minimized** at the earliest stage of processing
2. Data subjects give **specific, informed and explicit consent** to the processing of their data
3. Data subjects have **right to access, review and rectify** their personal data
4. Data subjects have the **right to withdraw given consent** with effect for the future and
   - Block access
   - Constrain processing and use
   - Erase their personal data
5. Personal **data obtained for one purpose must not be processed for other purposes** not compatible with the original purpose

Achieving “Privacy by Design” is difficult

Privacy is a complex, multifaceted and contextual notion

Not the primary requirement of an information system

May come into conflict with other requirements

> “…privacy and data protection features are… ignored by traditional engineering approaches when implementing desired functionality.
>
> - *This ignorance is caused by limitations of awareness and understanding of developers and data controllers as well as lacking tools to realize privacy by design”*

Danezis, G. et al. (2014) “Privacy and Data Protection by Design”, European Union Agency for Network and Information Security (ENISA)

Privacy and Data Protection by Design

"Although the concept has found its way into legislation as the… European General Data Protection Regulation, **its concrete implementation remains un-clear at the present moment**"

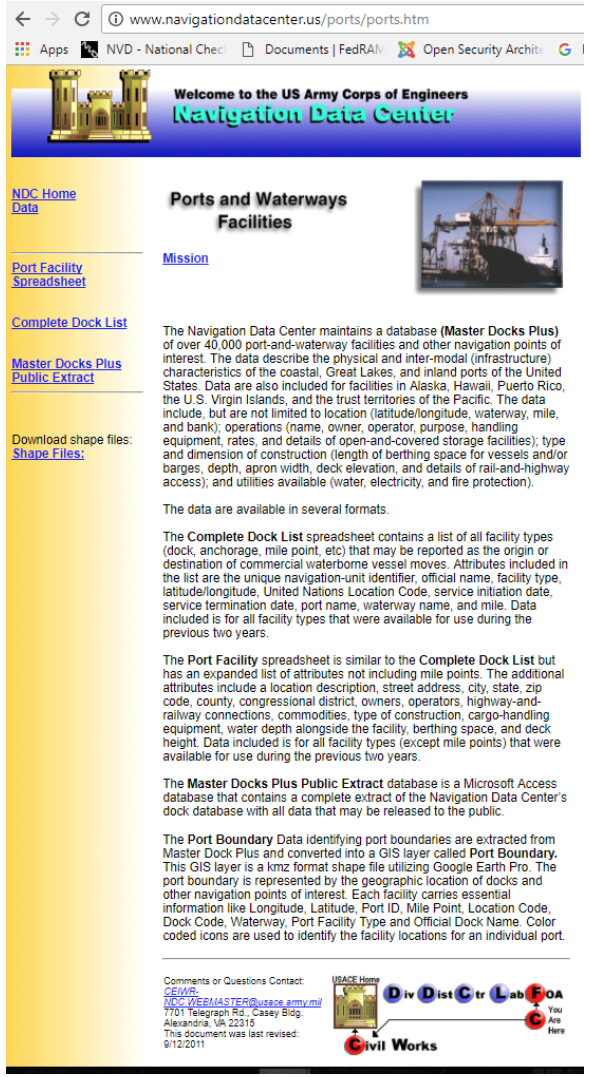
Danezis, G. et al. (2014) "Privacy and Data Protection by Design", European Union Agency for Network and Information Security (ENISA)

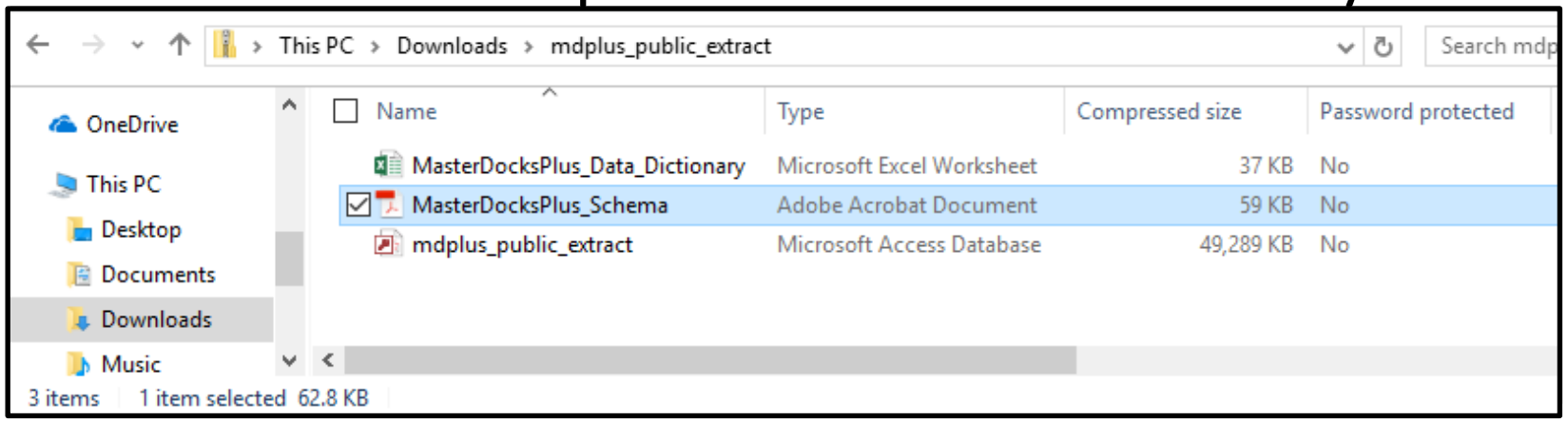
To start: Where can we look in descriptions of an information system for data subjects' personal information?

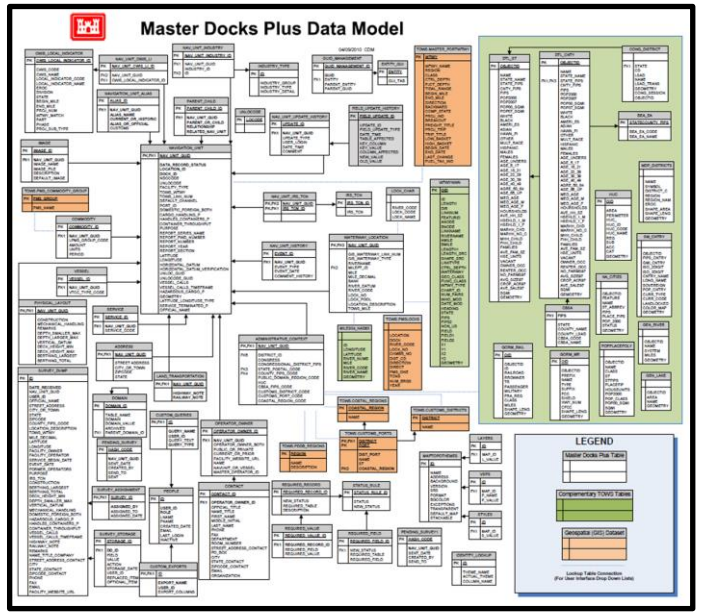
Descriptions of the following system components are expected:

- **Infrastructure.** The physical structures, IT, and other hardware
  - For example: facilities, computers, equipment, mobile devices, and telecommunications networks
- **Software.** The application programs and IT system software that supports application programs
  - For example: operating systems, middleware, and utilities
- **People.** The personnel involved in the governance, operation, and use of a system
  - For example: developers, operators, entity users, vendor personnel, and managers
- **Procedures.** The automated and manual procedures.
  - For example: System descriptions and plans, data flow diagrams, user guides and technical documentation (data models and data dictionaries)
- **Data.** Data files, database tables, transactional data streams, data processed by the system, and system outputs

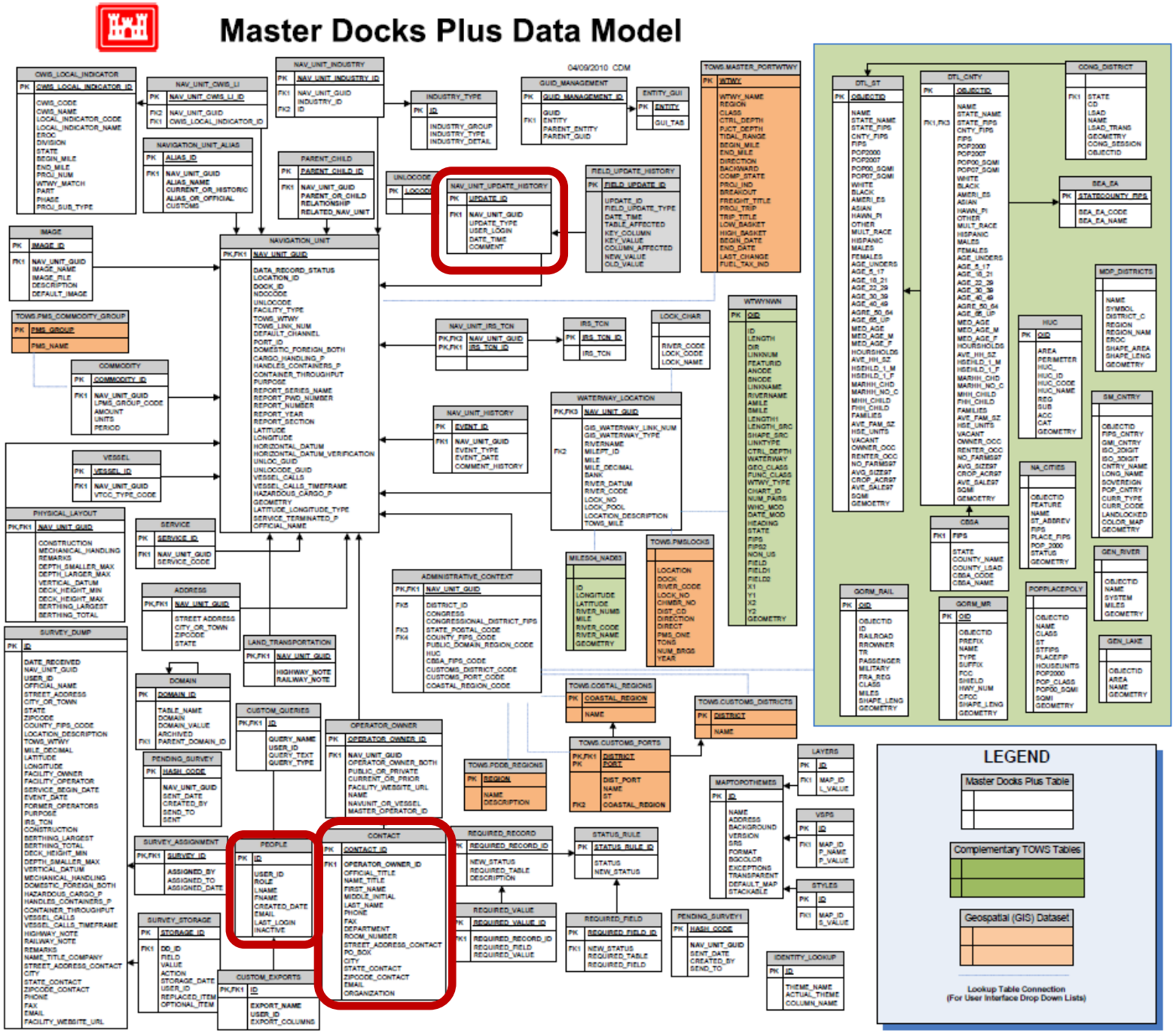
Example – Looking for documentation of data subjects' personal information in an enterprise information system







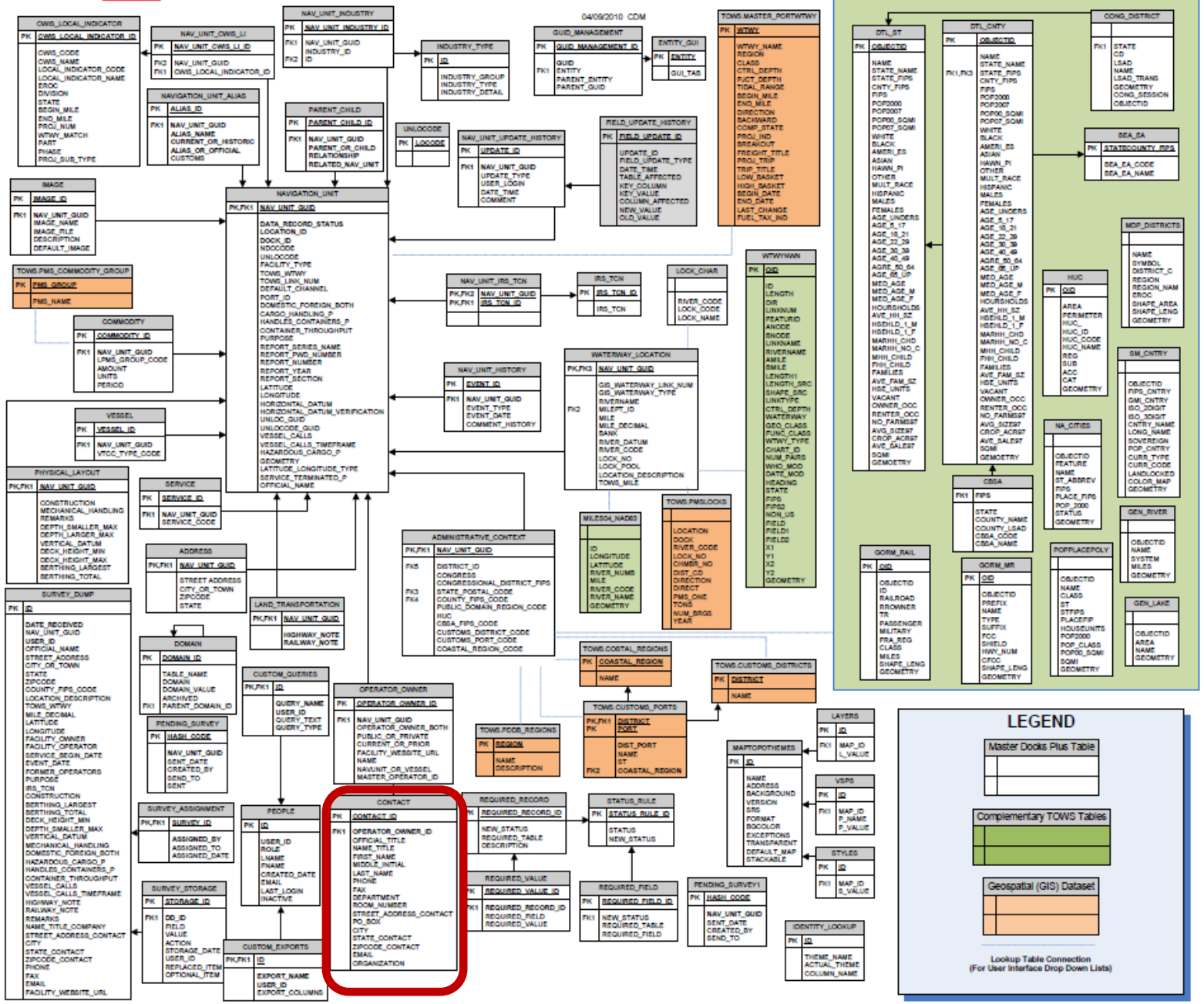
How to document where data subjects' personal information is stored and how it is used within a database ?



Legend

Personal information stored here

Master Docks Plus Data Model



OPERATOR_OWNER	
PK	**OPERATOR_OWNER_ID**
FK1	**NAV_UNIT_GUID**
	OPERATOR_OWNER_BOTH
	PUBLIC_OR_PRIVATE
	CURRENT_OR_PRIOR
	FACILITY_WEBSITE_URL
	NAME
	NAVUNIT_OR_VESSEL
	MASTER_OPERATOR_ID

CONTACT	
PK	**CONTACT_ID**
FK1	**OPERATOR_OWNER_ID**
	OFFICIAL_TITLE
	NAME_TITLE
	FIRST_NAME
	MIDDLE_INITIAL
	LAST_NAME
	PHONE
	FAX
	DEPARTMENT
	ROOM_NUMBER
	STREET_ADDRESS_CONTACT
	PO_BOX
	CITY
	STATE_CONTACT
	ZIPCODE_CONTACT
	EMAIL
	ORGANIZATION

How can we document where data subjects' personal information is stored and how it is used within a database ?

Typical Information System Database Data Dictionary

MasterDocksPlus_Data_Dictionary [Read-Only] - Excel

Home Insert Page Layout Formulas Data Review View ACROBAT Power Pivot Tell me what you want to do...

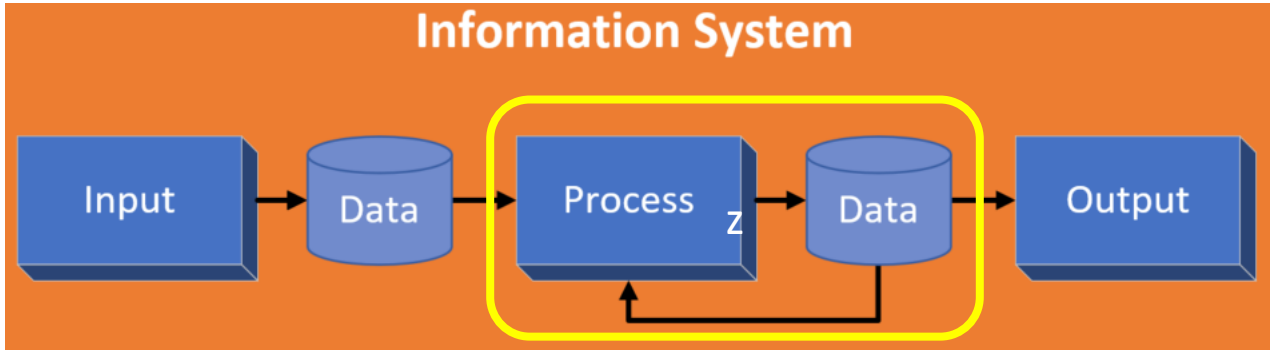
A	B	C	D	E	F	G	H	I	J	K
MD+ Field Name	**MD+ Field Type**	**MD+ Field Size**	**Suggested Field Size**	**Primary Key**	**Foreign Key**	**Notes**	**Domain Values**	**Constraints**	**Filemaker Migration Field**	**TOWS Migration Field**
Table Name: **Contact**										
Contact info for an owner or operator of a navigaton unit. Each owner or operator may have multiple contact records. This data was migrated from Filemaker										
Contact_ID	Number	38	12,0	Y		Unique identifier for contact records		Not Null	none	none
Operator_Owner_ID	Number	38	12,0		Y	Identifies the associated operator_owner record.	Operator_Owner_ID from mdpclient.operator_owner	Not Null	none	none
City	Character	100	100						city_mail	
Department	Character	150	150						department	
Email	Character	150	150						email_facility	
Fax	Character	50	50						fax	
First_Name	Character	50	50						first_name	
Last_Name	Character	60	60						last_name	
Middle_Initial	Character	30	1						mi	
Name_Title	Character	40	40						mr_or_mrs	
Official_Title	Character	100	100						title	
Phone	Character	50	50						phone	
PO_Box	Character	50	50						po_box_no	
Room_Number	Character	50	50						room_no	
State_Contact	Character	2	2				State_Abbr from mdpgis.mdp_states		state_for_mail	
Street_Address_Contact	Character	100	100						street_only	
Zipcode_Contact	Character	31	10						zip	
Organization	Character	150	150			The organization that this contact belongs to.			organization	

Some challenging data protection requirements may be solved with techniques presented here…

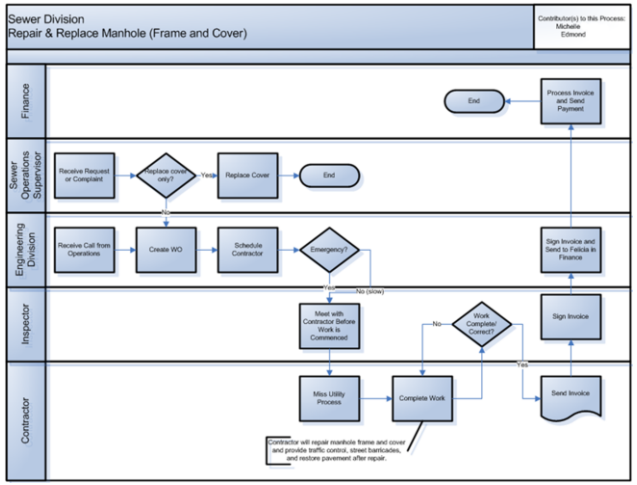
1. **Collection** of personal data is **fully avoided or minimized** at the earliest stage of processing
2. Data subjects give **specific, informed and explicit consent** to the processing of their data
3. Data subjects have **right to access, review and rectify** their personal data
4. Data subjects have the **right to withdraw given consent** with effect for the future and
   - Block access
   - Constrain processing and use
   - Erase their personal data
5. Personal **data obtained for one purpose must not be processed for other purposes** not compatible with the original purpose

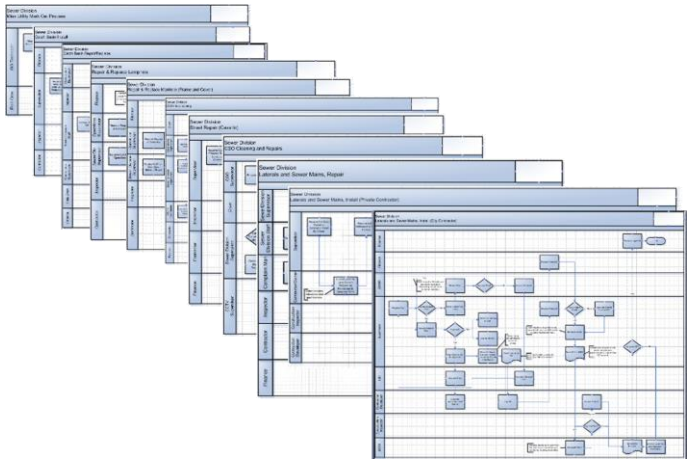
As a practical matter…

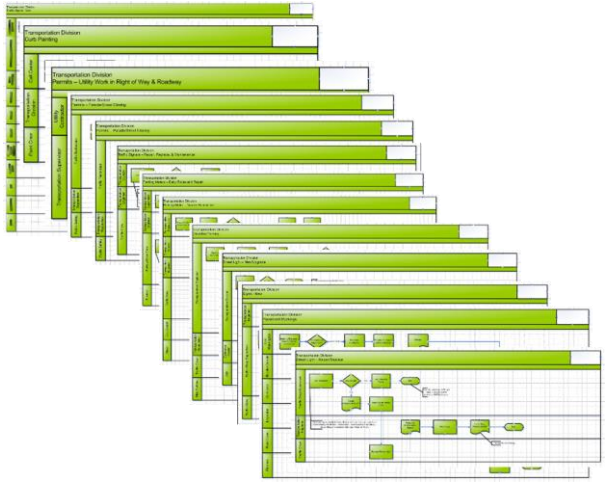
Data within information systems are often stored and organized as datasets within files and/or databases…



Regardless of application, there is reliance on data processing workflows to produce and use information

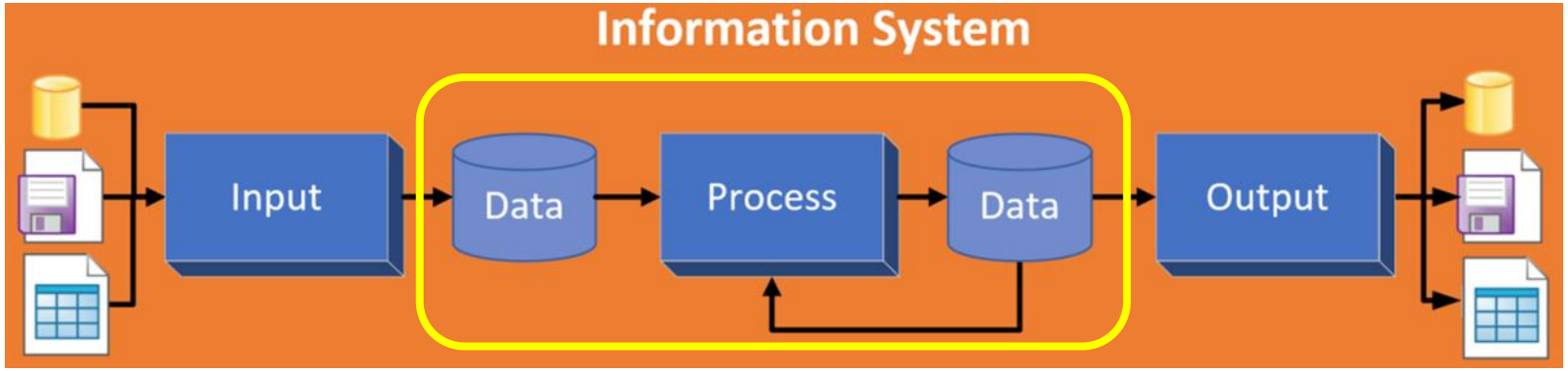




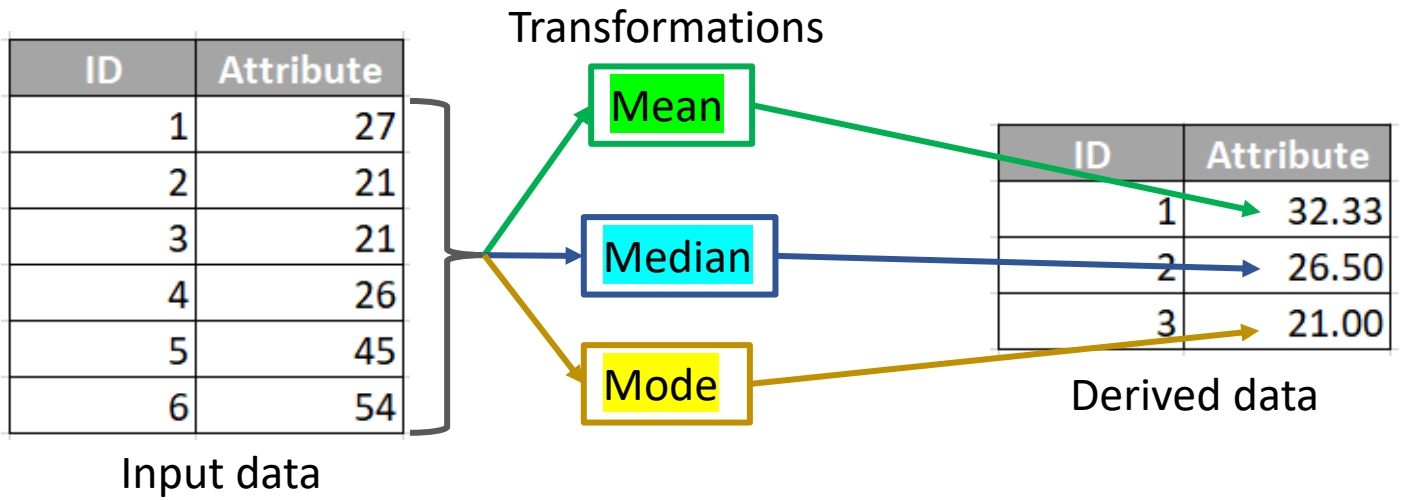


Data processing often transforms existing data into new data, which is a double-edged sword…

- *The resulting database may have more information than the older version*



- *The **meaning** of the new information, however, **is exogenous and not found in the data itself***



ID	Attribute
1	27
2	21
3	21
4	26
5	45
6	54

ID	Attribute
1	32.33
2	26.50
3	21.00

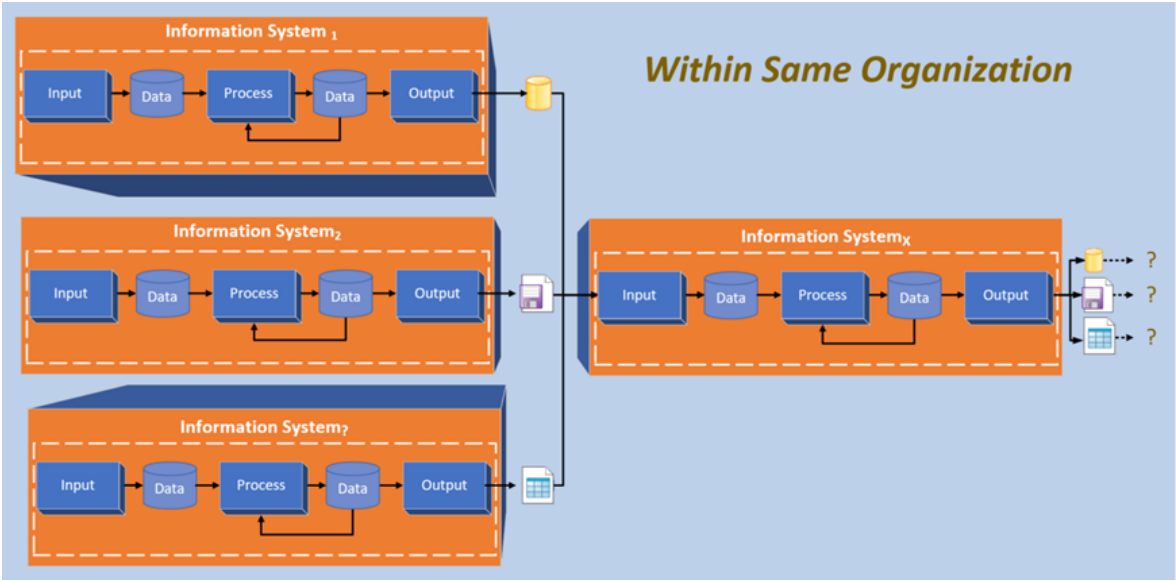
Evaluating & judging data's "fitness for use"

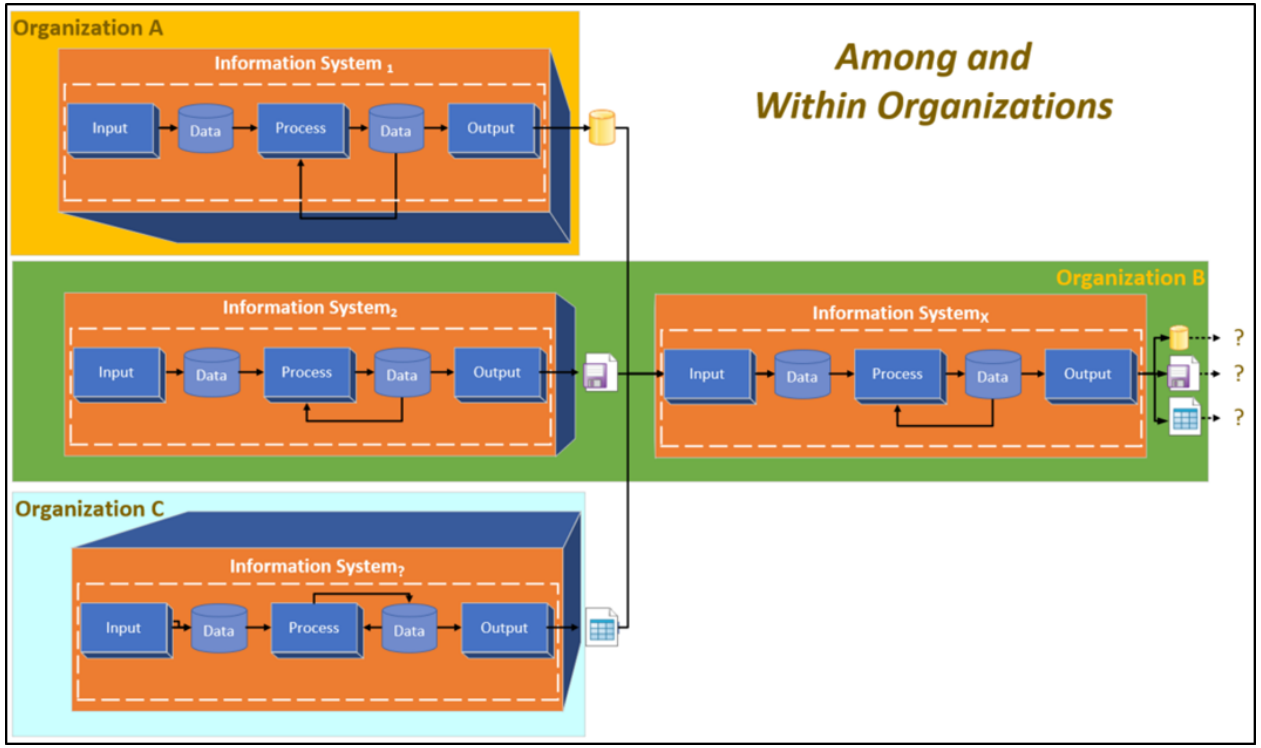
- **Is not the responsibility of the producer**
- **Is the responsibility of the user …and IT Auditor**

Data produced for one purpose is often used to serve other purposes

Data producers should provide information about data that permit informed determinations of fitness for use

Datasets are often exchanged without information needed to determine their fitness for use…





Provenance

Provenance traces back to 1294 in Old French as a derivative of the Latin *provenire*

- *To come from, to be due to, be the result of*

In the art domain, provenance entails an artifact's complete ownership history

The Bridge at Villeneuve-la-Garenne
1872
by Alfred Sisley British

Traditional Provenance

Durand-Ruel, Paris, August 23, 1872 [1];
Catholina Lambert, New Jersey;
Lambert sale, American Art Association, Plaza Hotel, New York, NY,
February 21, 1916 until February 24, 1916, no. 67;
Durand-Ruel, Paris, until at least 1930;
purchased by Simon Bauer, Paris, by June 1936 [2];
anonymous sale, Parke-Bernet Galleries, Inc., February 25, 1970, no. 19 [3];
Sam Salz, Inc., New York, NY;
purchased by Museum, May 1971.

Notes:
[1] bought from the artist.
[2] Listed and illustrated in "List of Property Removed from France during the War 1939-1945" (no. 7114, as belonging to Simon Bauer).
[3] "Highly Important Impressionist, Post-Impressionist & Modern Paintings and Drawings", illustrated.

Standardizing Museum Provenance — David Newbury *(@workergnome)*

Newbury, D. (2017) "Standardizing Museum Provenance for the Twenty-First Century", from talk given at the Yale Center for British Art

There is an established research process for obtaining an artifact's trusted provenance

- *This information is highly valued, particularly to authenticate real versus fraudulent works*

"Provenance" is now increasingly used in a broad range of fields with various degrees of conflation of two closely related but distinct concepts of *trust* and *metadata*

Tullis, J.A. et al., 2016, "Geoprocessing, Workflows, and Provenance", in Remote Sensing Handbook: Remotely Sensed Data Characterization, Classification, and Accuracies, edited by P. Thenkabail, Vol. 1., pp. 401-422, Boca Raton, FL: CRC Press.

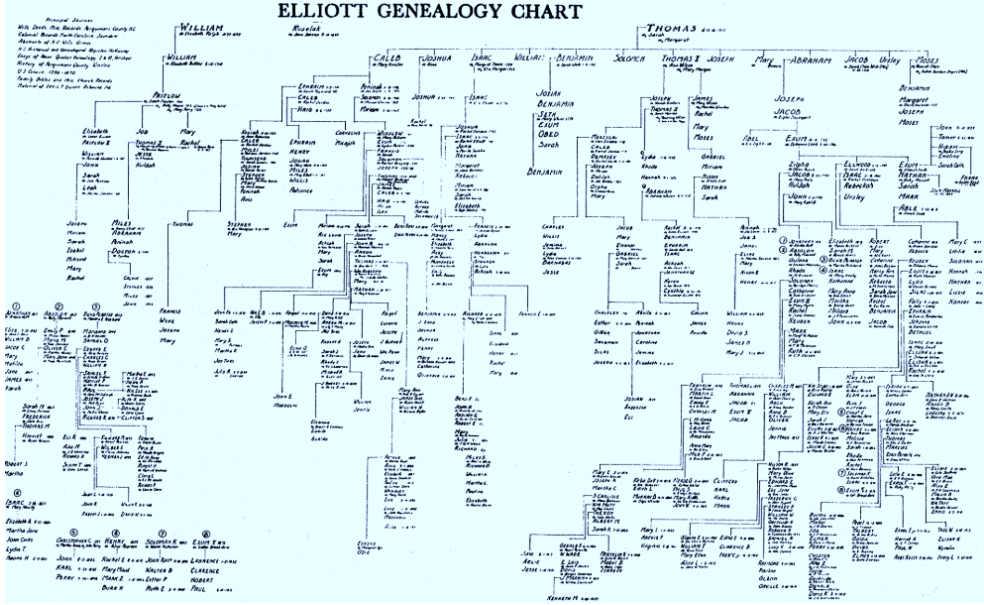
Provenance

W3C Provenance Incubator Group's definition of provenance (in a web resource context):

- Provenance is a record that describes entities and processes involved in producing and delivering or influencing a resource
- Provenance provides a critical foundation for assessing authenticity, enabling trust, and allowing reproducibility
- Provenance assertions are contextual metadata that can become important records with their own provenance

https://www.w3.org/TR/prov-primer/

Provenance and data lineage



"Data provenance" and "data lineage" is used here interchangeably, overlooking subtle differences in their meanings

- Data provenance suggests process history
- Data lineage implies a kind of genealogy or data pedigree record relative to both
  1. Sources of data
  2. Processing applied to the sources to produce an information product

This presentation explores how data lineage metadata can aid understanding and establish trust of data

Early metadata standards for documenting lineage of data produced with Geographic Information Systems

FGDC-STD-001-1998

NSDI

National Spatial Data Infrastructure

Content Standard for Digital Geospatial Metadata

Metadata Ad Hoc Working Group
Federal Geographic Data Committee

Federal Geographic Data Committee
Department of Agriculture ● Department of Commerce ● Department of Defense ● Department of Energy
Department of Housing and Urban Development ● Department of the Interior ● Department of State
Department of Transportation ● Environmental Protection Agency
Federal Emergency Management Agency ● Library of Congress
National Aeronautics and Space Administration ● National Archives and Records Administration
Tennessee Valley Authority

EUROPEAN STANDARD
NORME EUROPÉENNE
EUROPÄISCHE NORM

EN ISO 19115-1

April 2014

ICS 35.240.70

Supersedes EN ISO 19115:2005

English Version

Geographic information —
Metadata —
Part 1: Fundamentals
(ISO 19115-1:2014)

Information géographique —
Métadonnées —
Partie 1: Principes de base
(ISO 19115-1:2014)

Geoinformation —
Metadaten —
Teil 1: Grundsätze
(ISO 19115-1:2014)

This European Standard was approved by CEN on 22 February 2014.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.

cen

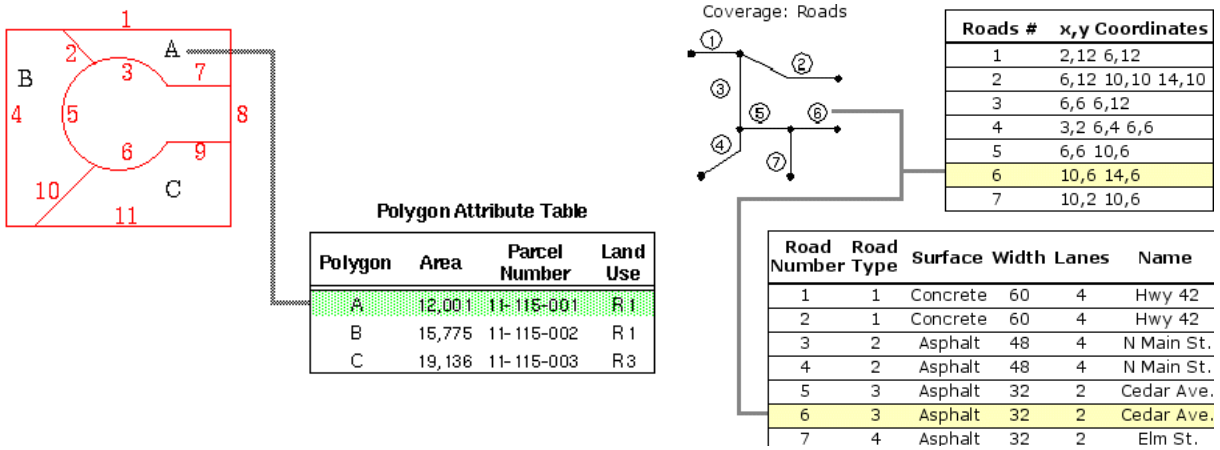
EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

CEN-CENELEC Management Centre: Avenue Marnix 17, B-1000 Brussels

Ref. No. EN ISO 19115-1:2014 E

Geographic Information System (GIS)

- Provides similar data import, query, manipulation, analysis (e.g. statistics), reformat, display/visualization, output and report capabilities as other information systems
- Also organize their data in
  - Data base management systems
  - File systems

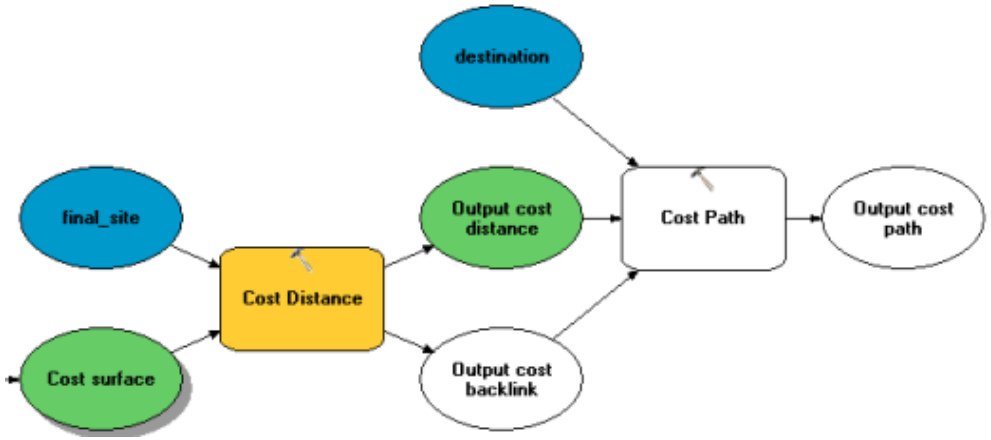


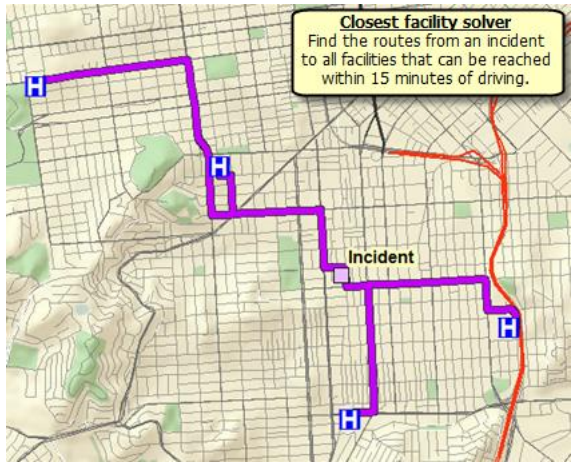
Roads #	x,y Coordinates
1	2,12 6,12
2	6,12 10,10 14,10
3	6,6 6,12
4	3,2 6,4 6,6
5	6,6 10,6
6	10,6 14,6
7	10,2 10,6

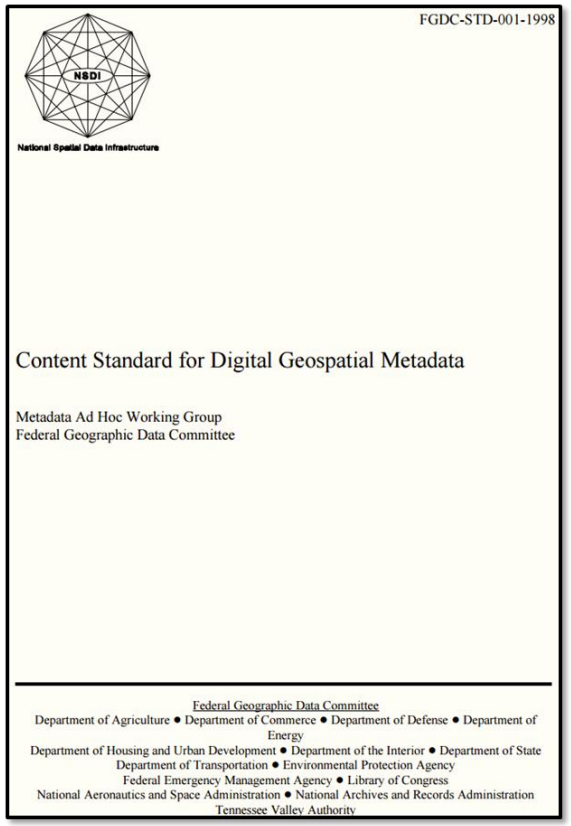
Polygon	Area	Parcel Number	Land Use
A	12,001	11-115-001	R1
B	15,775	11-115-002	R1
C	19,136	11-115-003	R3

Road Number	Road Type	Surface	Width	Lanes	Name
1	1	Concrete	60	4	Hwy 42
2	1	Concrete	60	4	Hwy 42
3	2	Asphalt	48	4	N Main St.
4	2	Asphalt	48	4	N Main St.
5	3	Asphalt	32	2	Cedar Ave.
6	3	Asphalt	32	2	Cedar Ave.
7	4	Asphalt	32	2	Elm St.

- With the addition of spatial analysis and cartographic mapping capabilities







FGDC-STD-001-1998

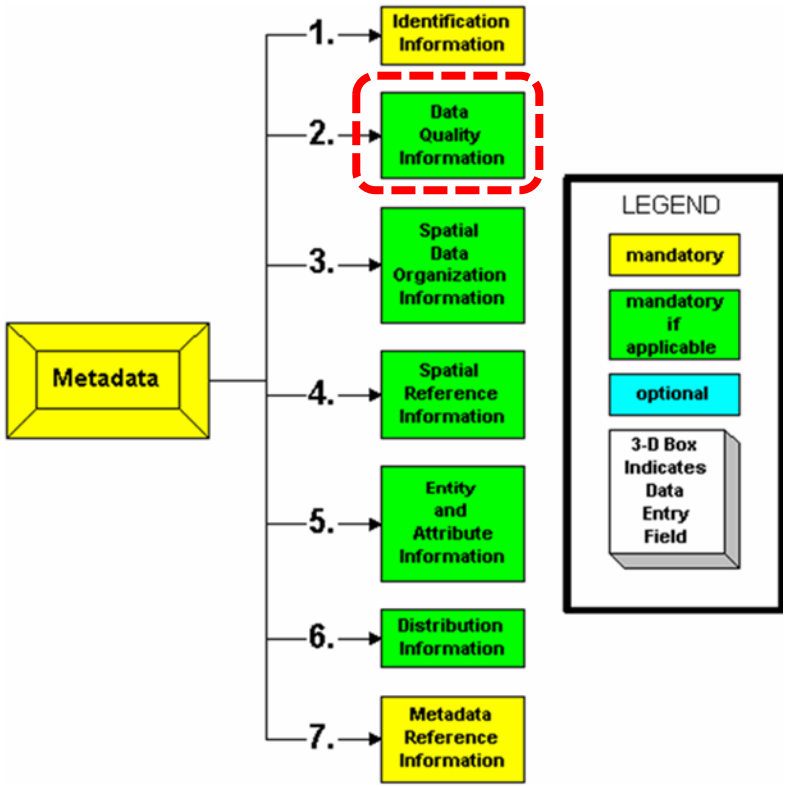
NSDI

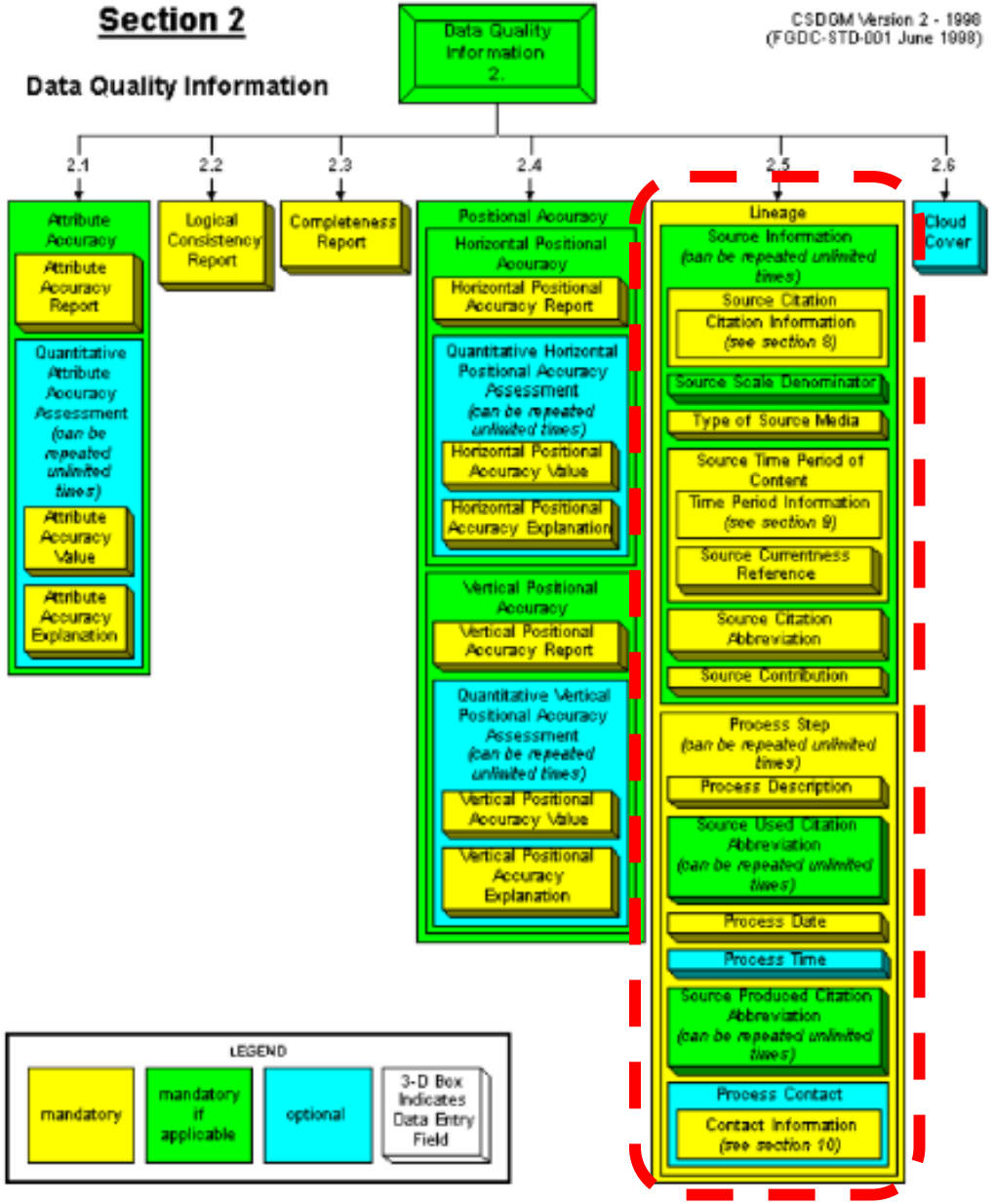
National Spatial Data Infrastructure

Content Standard for Digital Geospatial Metadata

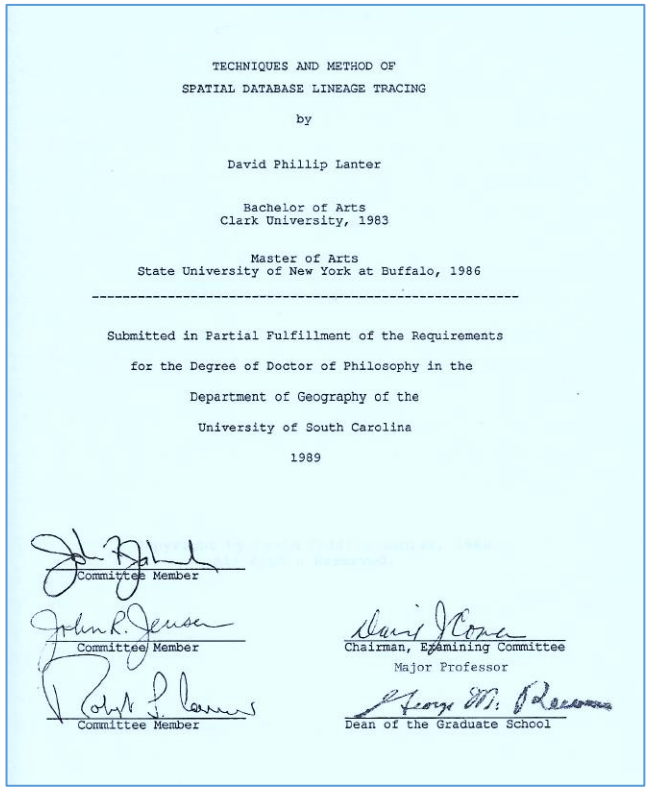
Metadata Ad Hoc Working Group
Federal Geographic Data Committee

Federal Geographic Data Committee
Department of Agriculture • Department of Commerce • Department of Defense • Department of Energy
Department of Housing and Urban Development • Department of the Interior • Department of State
Department of Transportation • Environmental Protection Agency
Federal Emergency Management Agency • Library of Congress
National Aeronautics and Space Administration • National Archives and Records Administration
Tennessee Valley Authority





1st automated capability for tracking the lineage of data throughout their processing in information systems



TECHNIQUES AND METHOD OF
SPATIAL DATABASE LINEAGE TRACING

by

David Phillip Lanter

Bachelor of Arts
Clark University, 1983

Master of Arts
State University of New York at Buffalo, 1986

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in the

Department of Geography of the

University of South Carolina

1989

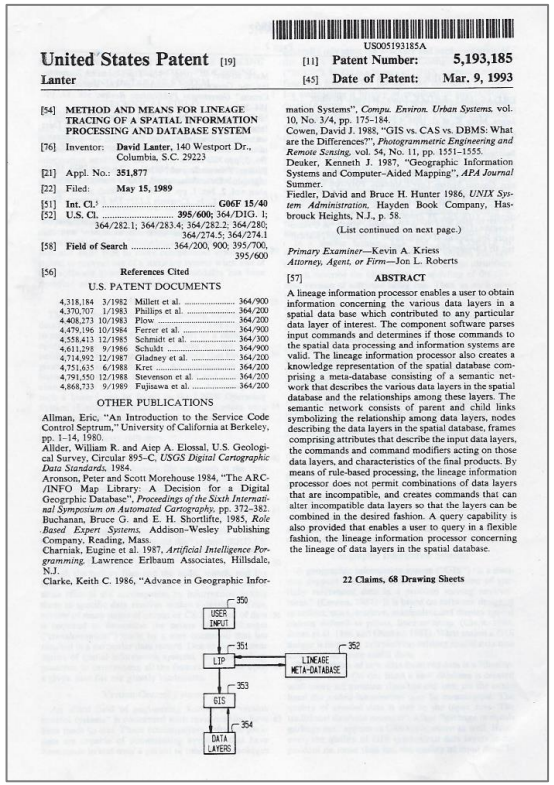
Committee Member

Committee Member

Chairman, Examining Committee
Major Professor

Committee Member

Dean of the Graduate School



US005193185A

United States Patent [19]
Lanter

[11] **Patent Number: 5,193,185**
[45] **Date of Patent: Mar. 9, 1993**

[54] **METHOD AND MEANS FOR LINEAGE TRACING OF A SPATIAL INFORMATION PROCESSING AND DATABASE SYSTEM**

[76] Inventor: **David Lanter**, 140 Westport Dr., Columbia, S.C. 29223

[21] Appl. No.: **351,877**

[22] Filed: **May 15, 1989**

[51] **Int. Cl.**[5] **G06F 15/40**

[52] **U.S. Cl.** **395/600**; 364/DIG. 1; 364/282.1; 364/283.4; 364/282.2; 364/280; 364/274.5; 364/274.1

[58] **Field of Search** 364/200, 900; 395/700, 395/600

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,318,184	3/1982	Millett et al.	364/900
4,370,707	1/1983	Philips et al.	364/200
4,408,273	10/1983	Plow	364/200
4,479,196	10/1984	Ferrer et al.	364/900
4,558,413	12/1985	Schmidt et al.	364/300
4,611,298	9/1986	Schuldt	364/900
4,714,992	12/1987	Gladney et al.	364/200
4,751,635	6/1988	Kret	364/200
4,791,550	12/1988	Stevenson et al.	364/200
4,868,733	9/1989	Fujisawa et al.	364/200

OTHER PUBLICATIONS

Allman, Eric, "An Introduction to the Service Code Control Septrum," University of California at Berkeley, pp. 1–14, 1980.
Allder, William R. and Aiep A. Elossal, U.S. Geological Survey, Circular 895–C, *USGS Digital Cartographic Data Standards*, 1984.
Aronson, Peter and Scott Morehouse 1984, "The ARC-/INFO Map Library: A Decision for a Digital Geogrphic Database", *Proceedings of the Sixth International Symposium on Automated Cartography*, pp. 372–382.
Buchanan, Bruce G. and E. H. Shortliffe, 1985, *Role Based Expert Systems*. Addison-Wesley Publishing Company, Reading, Mass.
Charniak, Eugine et al. 1987, *Artificial Intelligence Porgramming*. Lawrence Erlbaum Associates, Hillsdale, N.J.
Clarke, Keith C. 1986, "Advance in Geographic Information Systems", *Compu. Environ. Urban Systems*, vol. 10, No. 3/4, pp. 175–184.
Cowen, David J. 1988, "GIS vs. CAS vs. DBMS: What are the Differences?", *Photogrammetric Engineering and Remote Sensing*, vol. 54, No. 11, pp. 1551–1555.
Deuker, Kenneth J. 1987, "Geographic Information Systems and Computer–Aided Mapping", *APA Journal* Summer.
Fiedler, David and Bruce H. Hunter 1986, *UNIX System Administration*, Hayden Book Company, Hasbrouck Heights, N.J., p. 58.

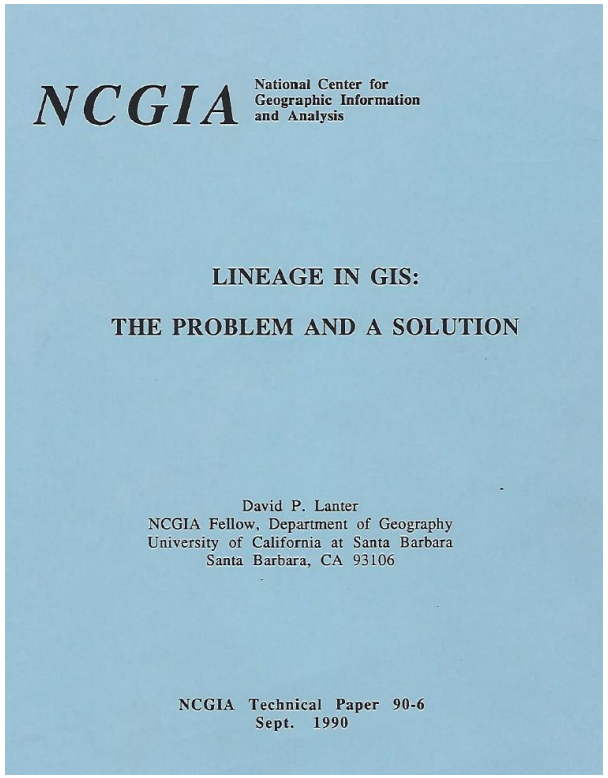
(List continued on next page.)

Primary Examiner—Kevin A. Kriess
Attorney, Agent, or Firm—Jon L. Roberts

[57] **ABSTRACT**

A lineage information processor enables a user to obtain information concerning the various data layers in a spatial data base which contributed to any particular data layer of interest. The component software parses input commands and determines if those commands to the spatial data processing and information systems are valid. The lineage information processor also creates a knowledge representation of the spatial database comprising a meta-database consisting of a semantic network that describes the various data layers in the spatial database and the relationships among these layers. The semantic network consists of parent and child links symbolizing the relationship among data layers, nodes describing the data layers in the spatial database, frames comprising attributes that describe the input data layers, the commands and command modifiers acting on those data layers, and characteristics of the final products. By means of rule-based processing, the lineage information processor does not permit combinations of data layers that are incompatible, and creates commands that can alter incompatible data layers so that the layers can be combined in the desired fashion. A query capability is also provided that enables a user to query in a flexible fashion, the lineage information processor concerning the lineage of data layers in the spatial database.

22 Claims, 68 Drawing Sheets



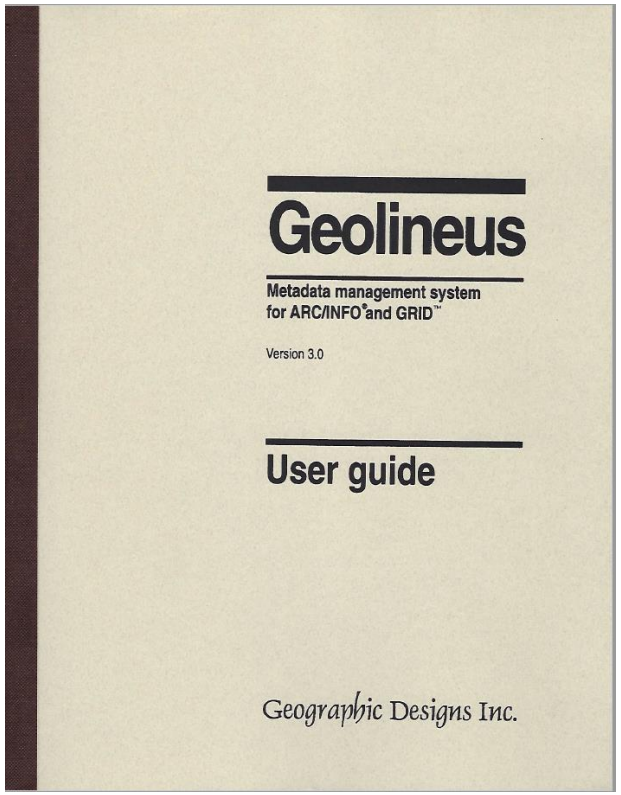
NCGIA National Center for Geographic Information and Analysis

LINEAGE IN GIS:

THE PROBLEM AND A SOLUTION

David P. Lanter
NCGIA Fellow, Department of Geography
University of California at Santa Barbara
Santa Barbara, CA 93106

NCGIA Technical Paper 90-6
Sept. 1990



Geolineus

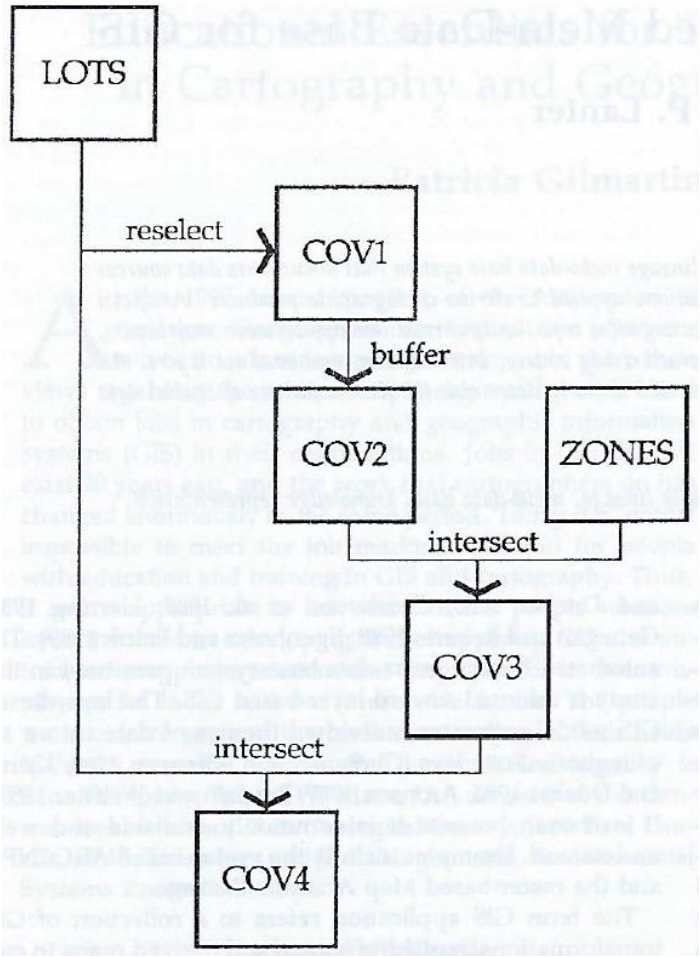
Metadata management system for ARC/INFO® and GRID™

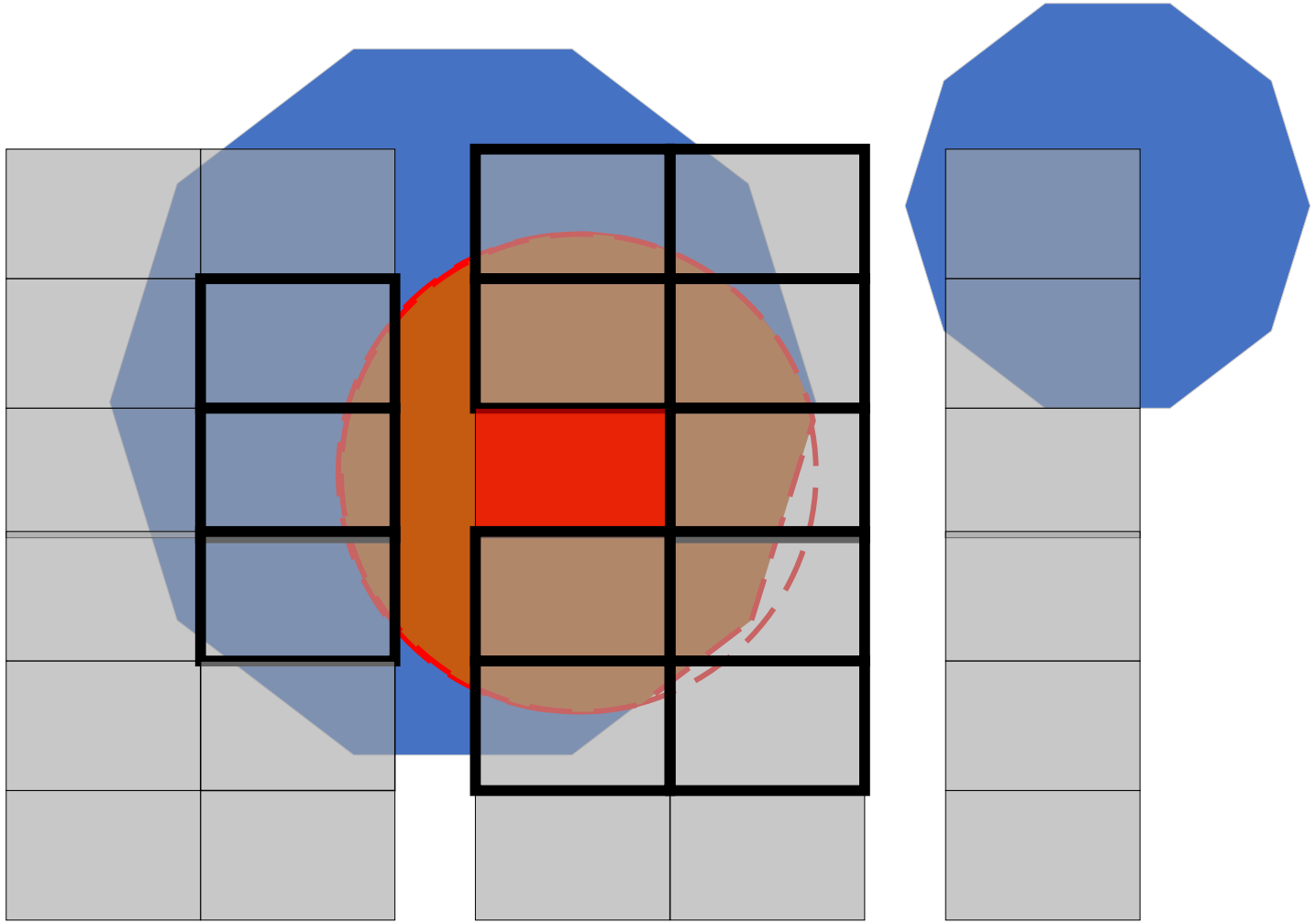
Version 3.0

User guide

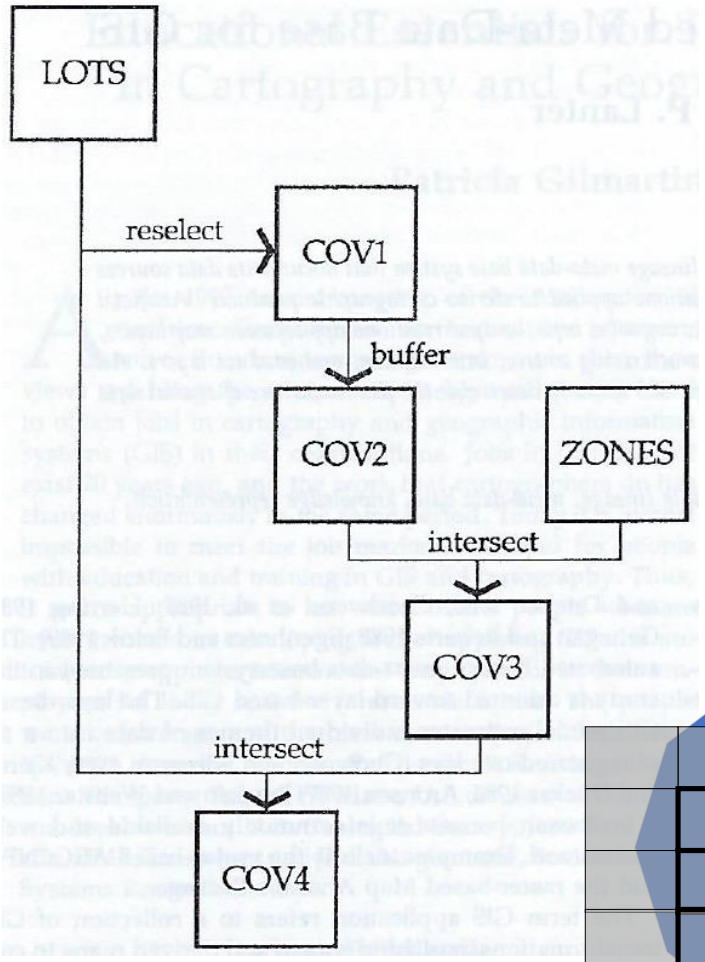
Geographic Designs Inc.

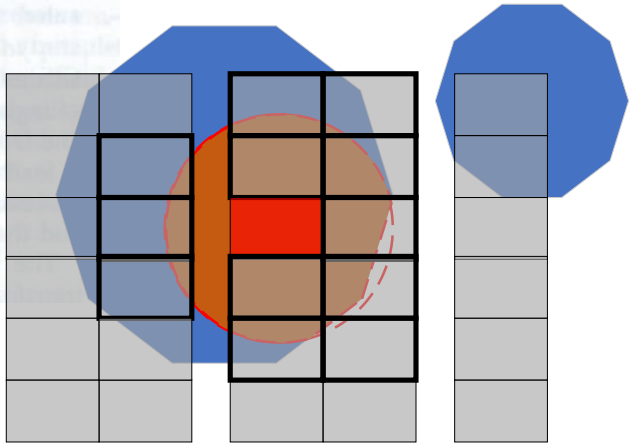
Information processing steps in the head of the user (PhD student) as he worked on a class assignment and transformed the LOTS and ZONES datasets to derive COV4…





Information processing steps in the head of the GIS user as he transformed the LOTS and ZONES datasets to derive COV4…



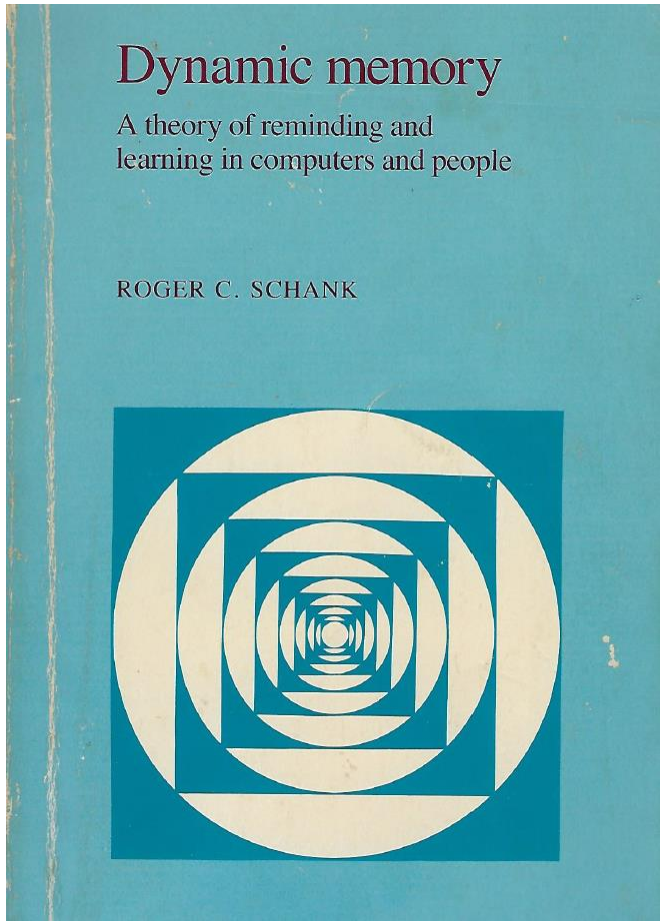


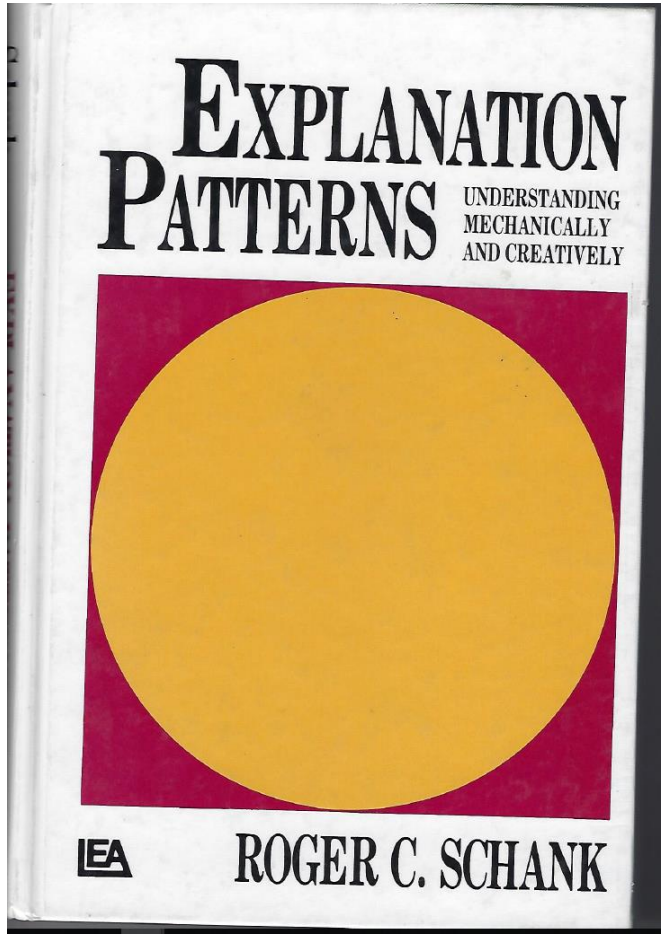
Datasets presented by the operating system after data processing concluded…

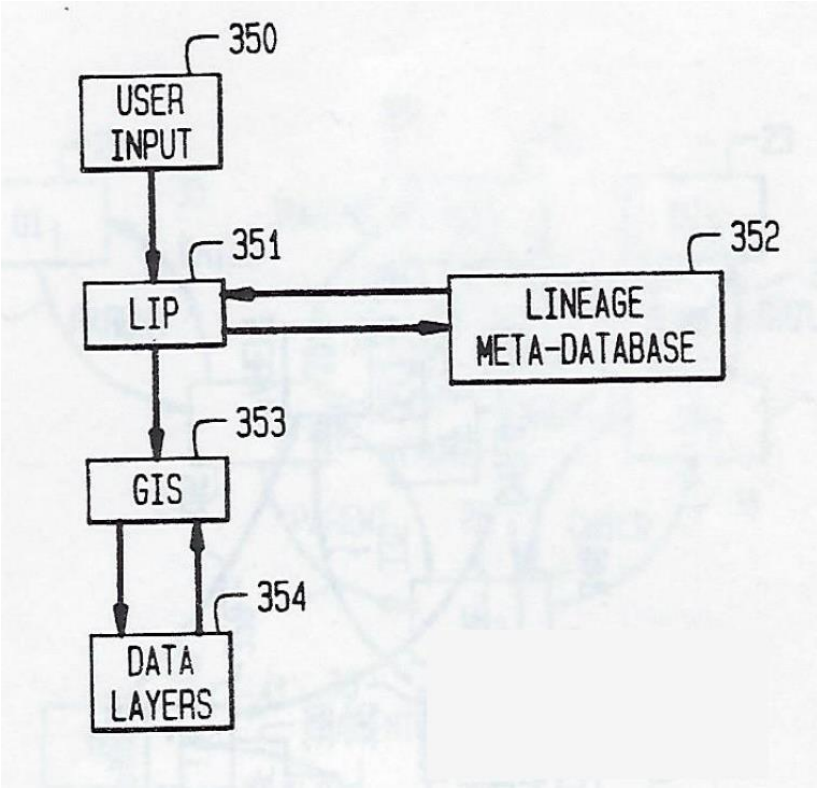
Datasets organized as files in folders

.	<DIR>	5-05-89	10:26a
..	<DIR>	5-05-89	10:26a
COV1	<DIR>	5-24-89	11:35p
LOTS	<DIR>	5-05-89	10:26a
INFO	<DIR>	5-05-89	10:26a
ZONES	<DIR>	5-05-89	10:27a
OUTPUT	<DIR>	5-05-89	10:27a
ONELOT	<DIR>	5-06-89	11:52a
DAV1	<DIR>	5-31-89	1:35p
FINAL	<DIR>	5-06-89	12:27p
COV3	<DIR>	5-24-89	11:46p
COV4	<DIR>	5-24-89	11:51p
BUF	<DIR>	5-06-89	12:21p
COV2	<DIR>	5-24-89	11:42p
DAV3	<DIR>	5-31-89	1:45p
DAV4	<DIR>	5-31-89	1:49p
DAV2	<DIR>	5-31-89	1:42p

The PhD student wondered: "How can I program the computer to help me remember what I knew about the data I was processing when I was processing it?"

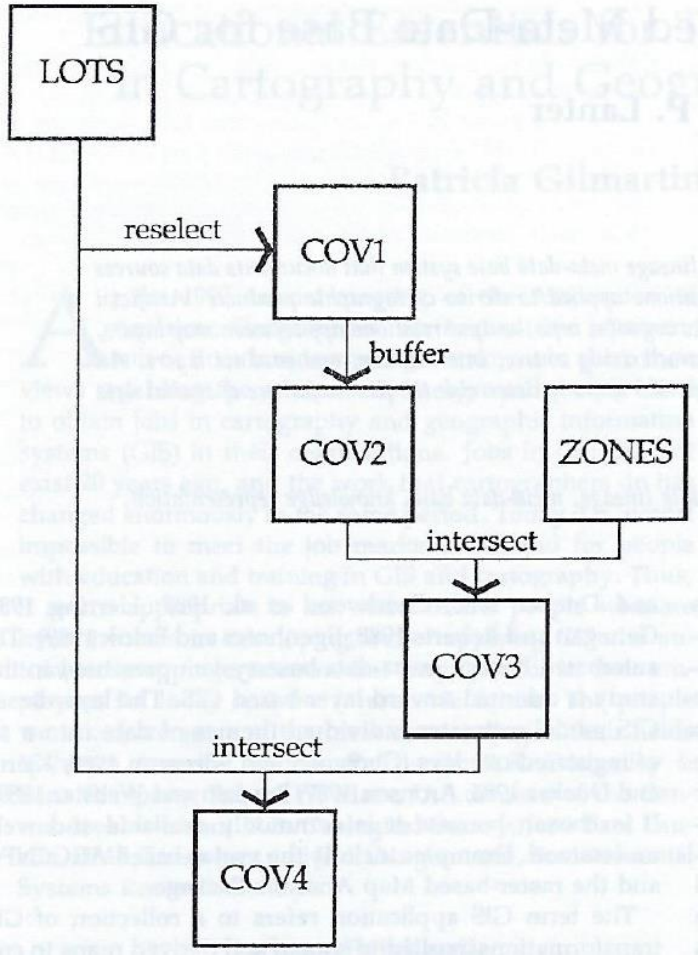


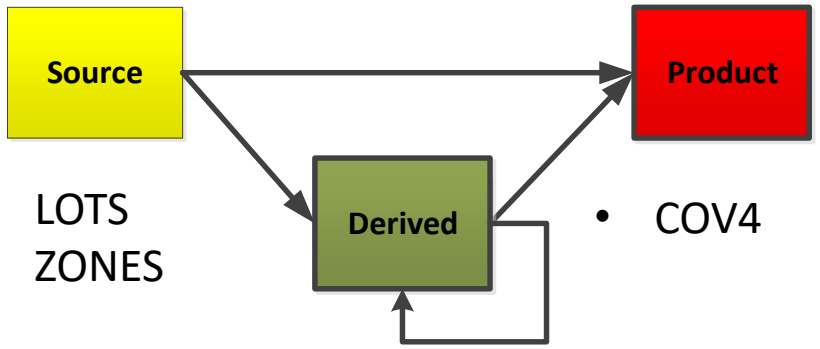




LIP = Lineage Information Processor

He wondered: “How do we understand differences among datasets created during processing applications?”

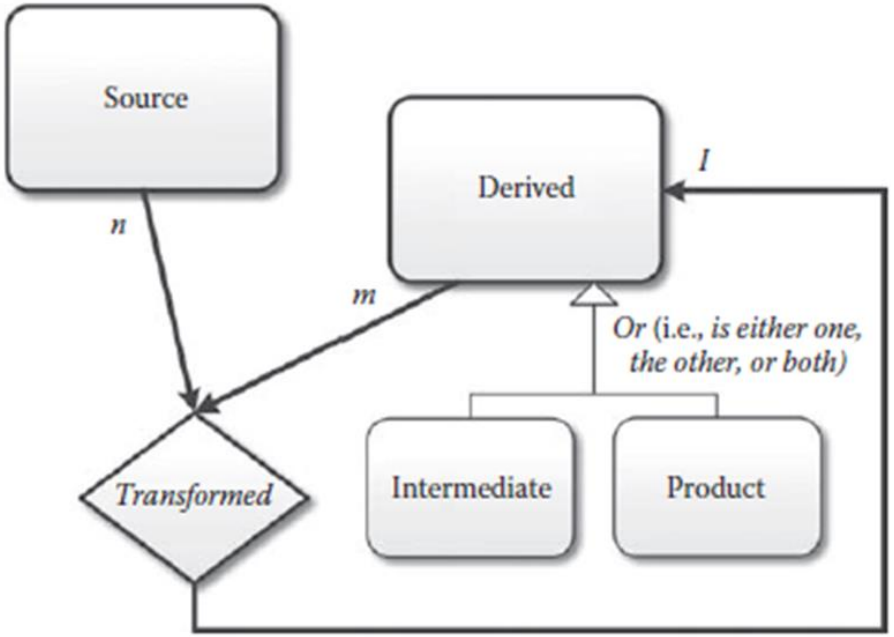




- LOTS
- ZONES

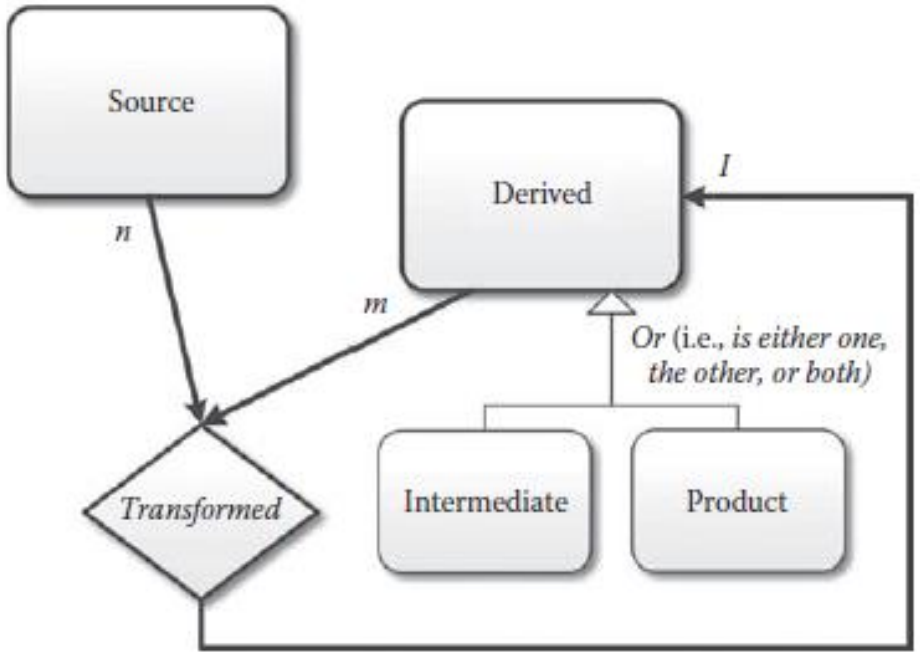
- COV4

- COV1
- COV2
- COV3



Data lineage vocabulary helped him understand & communicate how data is processed in an information system

and also aids thinking about how to meet privacy by design requirements

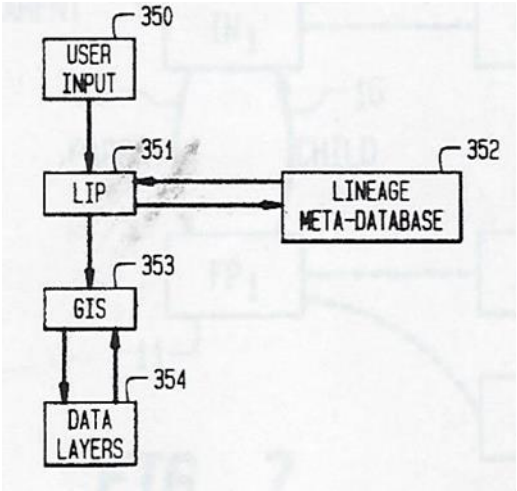


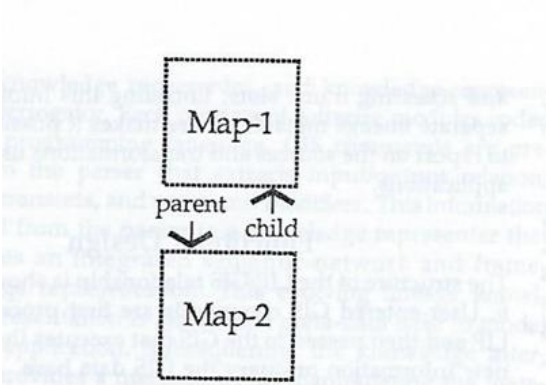
Source datasets *may contain personal data*

Derived datasets inherit this personal data from their input

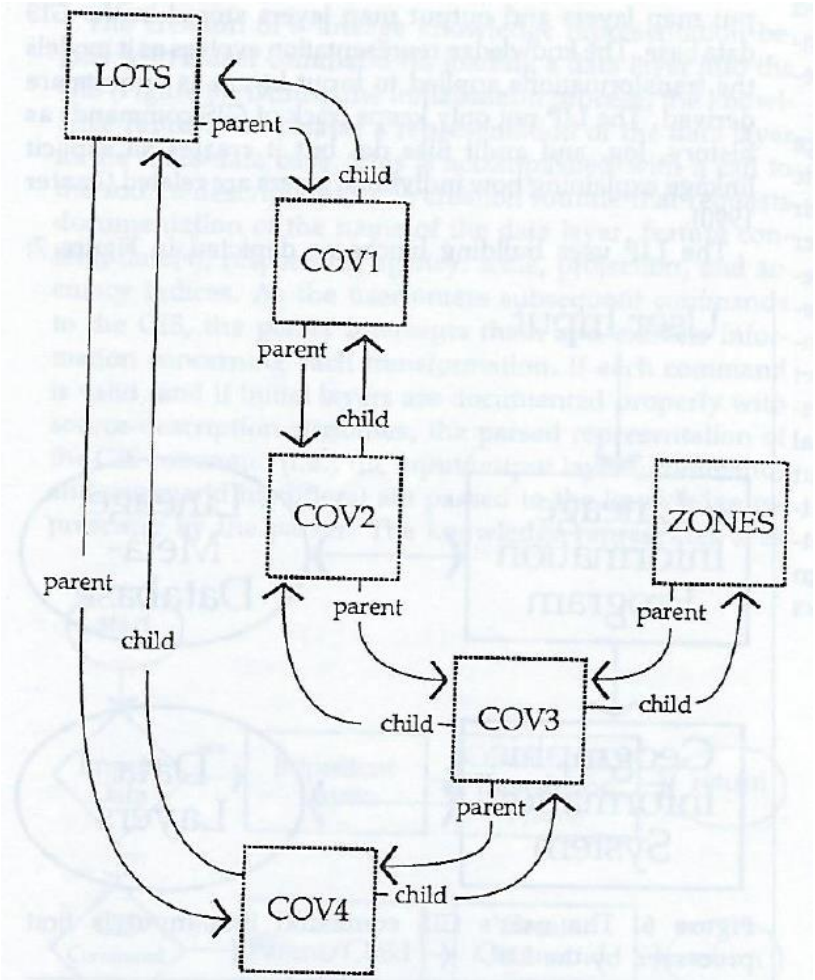
- *Using transformations such as:*
  - *Relational database joins and relates*
  - *Queries, arithmetic, statistical, spatial processing…*

Adding semantic “parent” & “child” metadata links enables deductions about relationships among input & output datasets…





Semantic parent and child links between layers.



Input datasets provided with parent links pointing to output datasets can answer the question: ***Who am I the parent of?***

Output datasets’ child links connect them back to their input datasets can answer the question: ***Who am I the child of?***

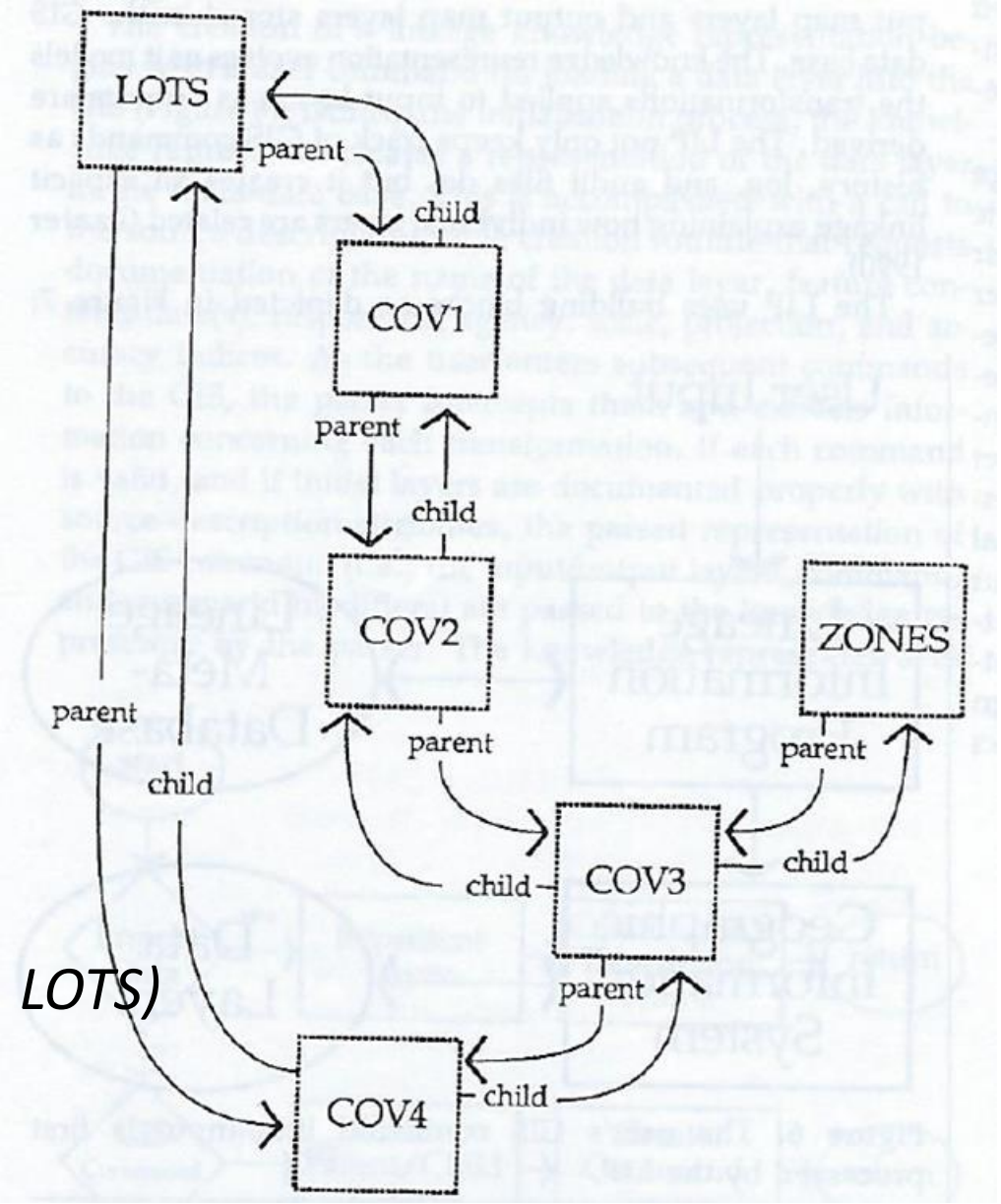
Descendants function traces parent links to identify all datasets derived from a source or other derived input dataset used within the application.

```
(defun decendents (map)
   (cond ((null map) nil)
         ((null (car (get map 'parent)))
            (print (append (list map)
                           (is a product map layer) (terpri))))
      (t
         (cond((null (cdr (get map 'parent)))
                 (decendents (car (get map 'parent))))
              (t (decendents (car (get map 'parent)))
                 (decendents (cadr (get map 'parent))))))))
```

Descendants ("LOTS") = (COV1, COV2, COV3, COV4)

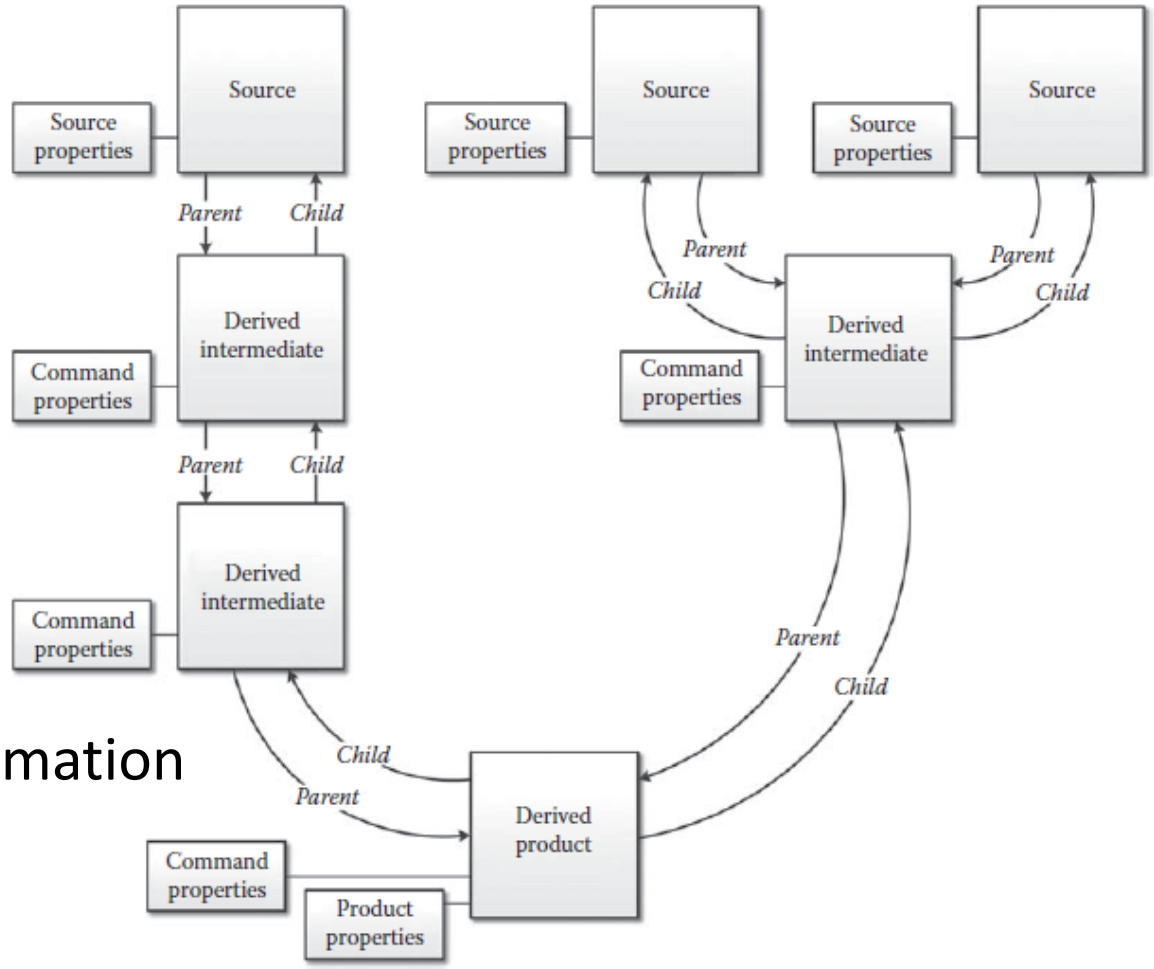
Ancestors function traces child links to identify input datasets used to create a derived dataset

Ancestors ("COV4") = (LOTS, COV3, ZONES, COV2, COV1, LOTS)



Source properties can include:

- Originating organization
- Data content (i.e. entity and attribute definitions)
- Timeliness (e.g. when collected, when acquired,…)
- Accuracy
- Confidentiality & privacy categorization of attributes



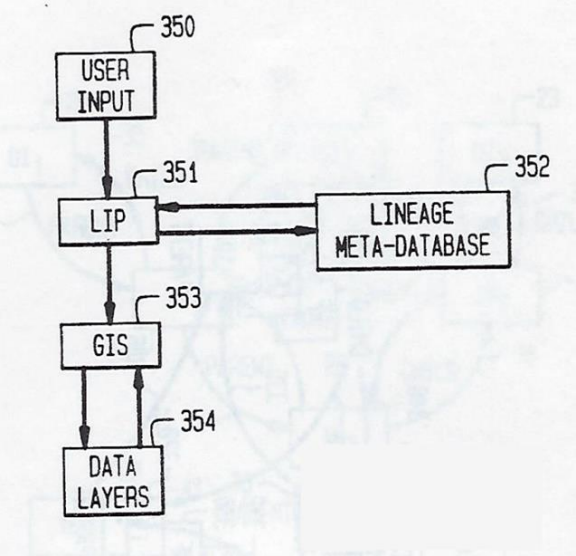
Command properties include details of the transformation

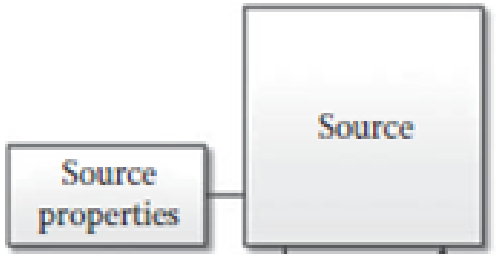
Product properties include the product's

- Intended goal
- Users
- When published
- Responsible manager,…

Meet Geo_lineus
source metadata input

```
(geo_lineus)I am Geo_lineus
Please give me information or ask questions: import cover landuse
landuse

What is the source name? landuse-landcover

Containing what cartographic features? hydrography urban
agriculture wetland

What is the source date? 3/12/75

What is the source agency? USGS

What is the source scale? 1/24000

What is the source projection? UTM

What is the source accuracy? +-80 meters

Thank You!
```



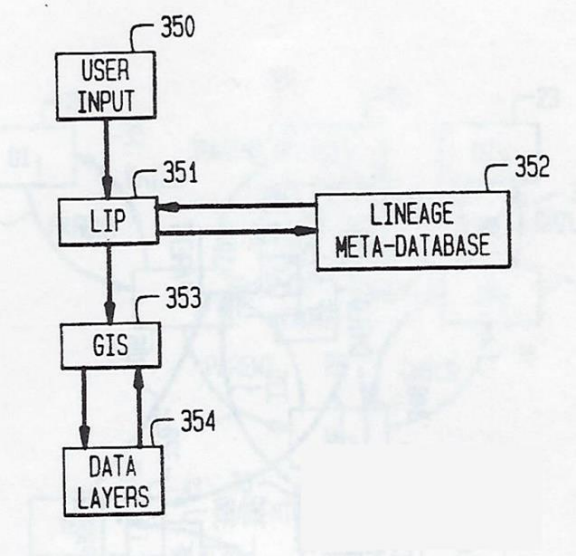


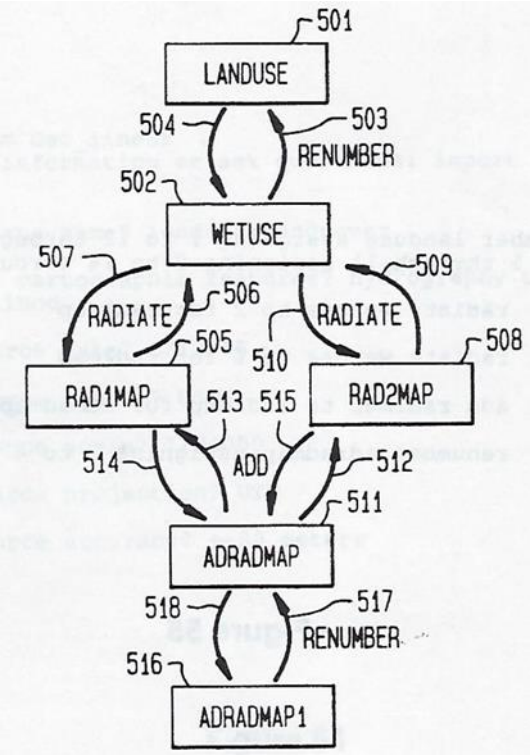
SOURCE DESCRIPTION FRAME

SOURCE:	Digital line graph
FEATURES:	Hydrography
S_DATE:	4/7/83
AGENCY:	USGS
SCALE:	1:100,000
PROJECTION:	Mercator
ACCURACY:	+-10 meters Horiz

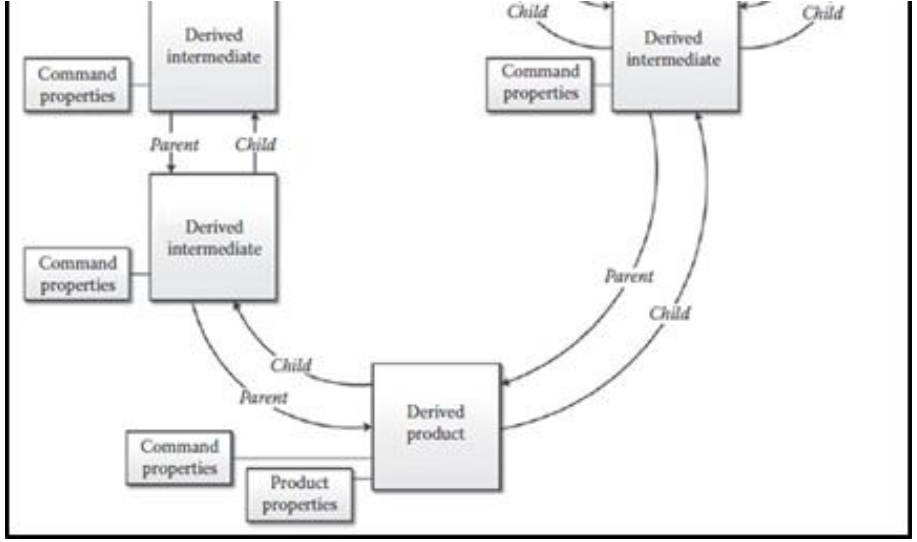
Command metadata input…

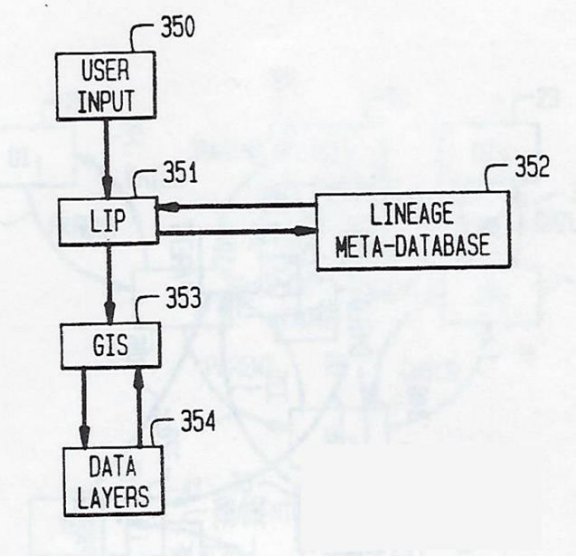
```
(geo_lineus)

(I AM GEO_LINEUS)

(PLEASE GIVE ME INFORMATION OR ASK QUESTIONS) (renumber landuse
assigning 1 to 2 through 13 assigning 0 to 1 through 11 assigning
0 to 14 through 18 for wetuse)

(I UNDERSTAND) (radiate wetuse to 2 for rad1map)

(I UNDERSTAND) (radiate wetuse to 6 for rad2map)

(I UNDERSTAND) (add rad1map to rad2map for adradmap)
```



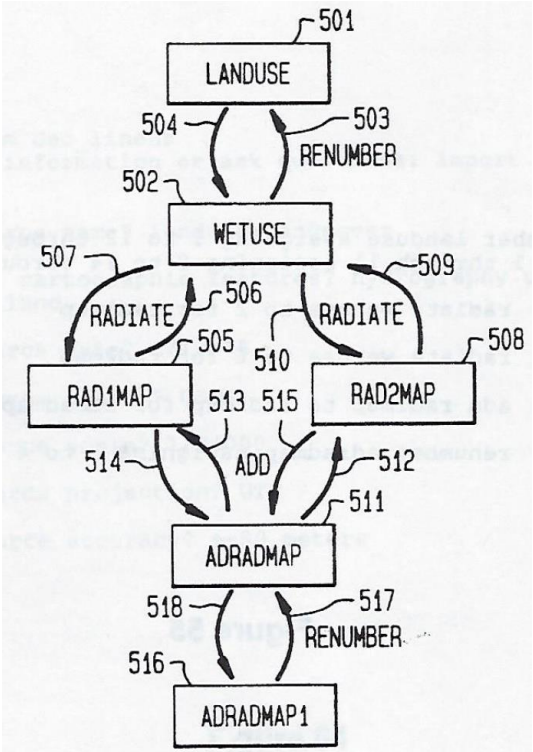


Inputting product metadata…

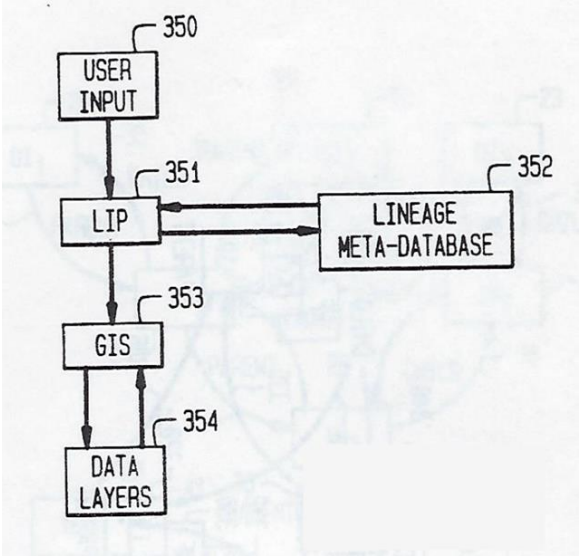




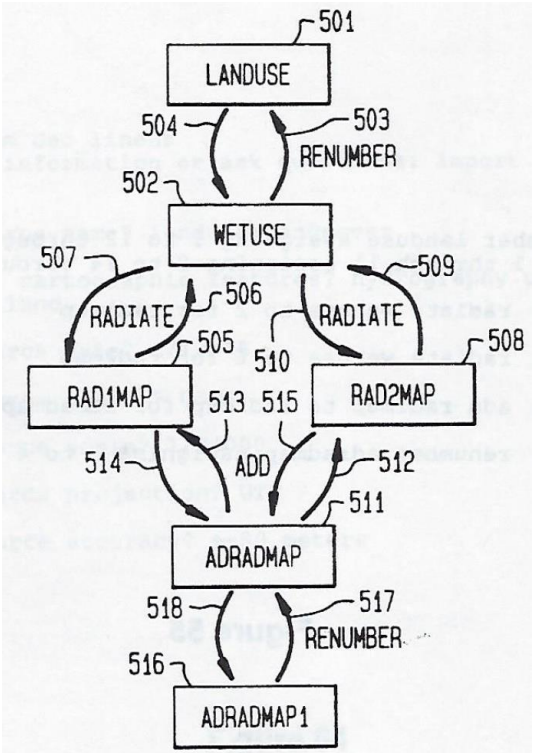

```
export cover adradmap1 eco_zones

What is the product's name? eco_zones

What is the product's use? Environmental protection of wetlands

Who are the product's users? Dept of Health and Environ.
Conservation

Who is responsible for the product? Diego Essinger

What is the product's release date? 3/5/89

Thank You!
```



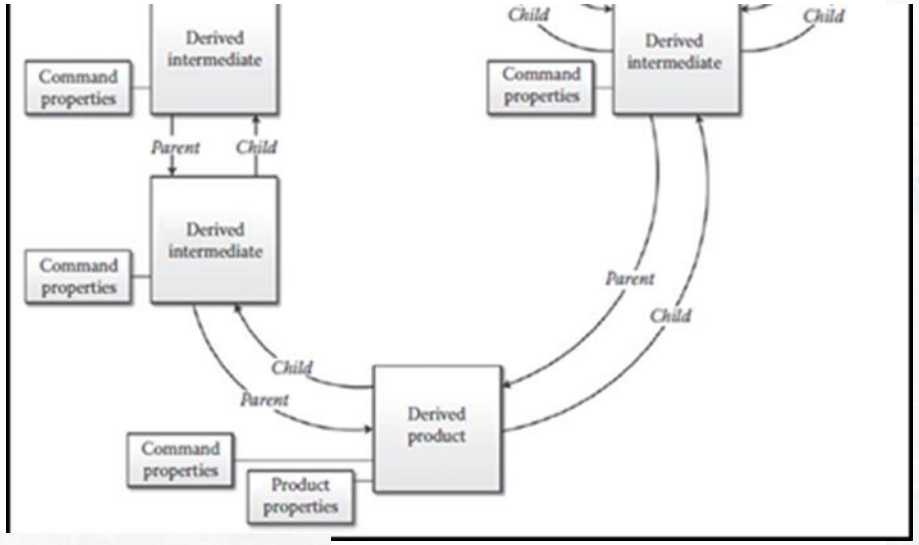
Querying data lineage metadata…

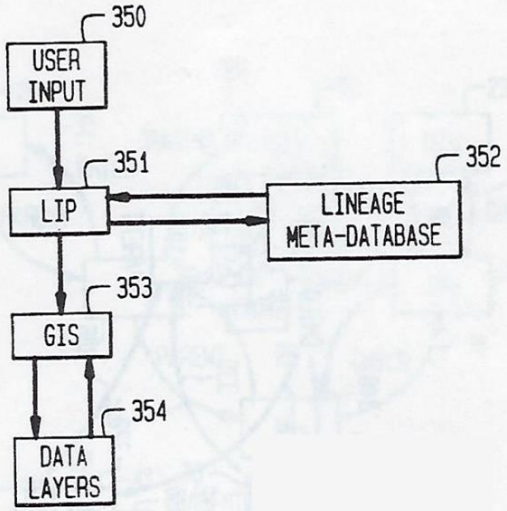



```
Is landuse a parent of adradmap

(YES INDEED LANDUSE IS A PARENT OF ADRADMAP)
```



Querying metadata…





What is the lineage of adradmap1

(INPUT TO ADRADMAP1 IS ADRADMAP COMMAND IS RENUMBER)

(INPUT TO ADRAPMAP IS RAD2MAP RAD1MAP COMMAND IS ADD)

(INPUT TO RAD2MAP IS WETUSE COMMAND IS RADIATE)

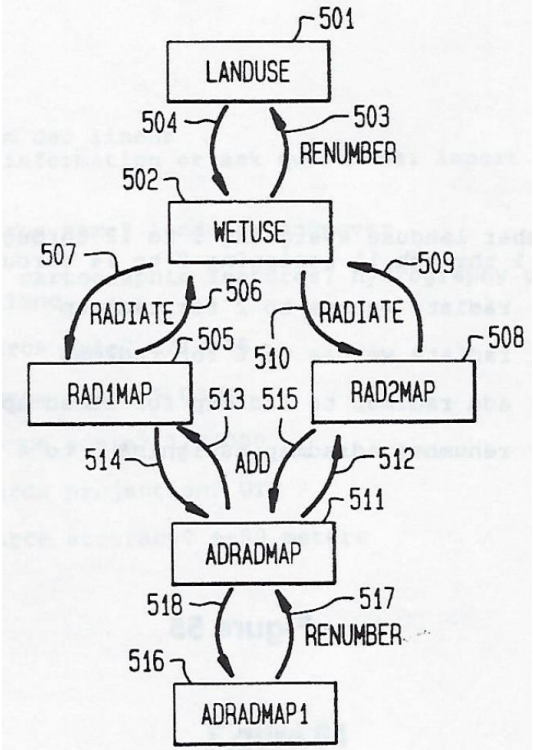
(INPUT TO WETUSE IS LANDUSE COMMAND IS RENUMBER)

(LANDUSE IS AN ORIGINAL MAP LAYER)

(INPUT TO RAD1MAP IS WETUSE COMMAND IS RADIATE)

(INPUT TO WETUSE IS LANDUSE COMMAND IS RENUMBER)

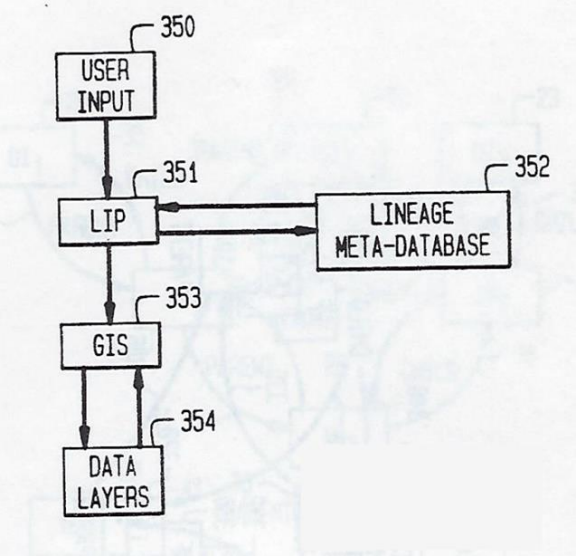
(LANDUSE IS AN ORIGINAL MAP LAYER)



Querying metadata…

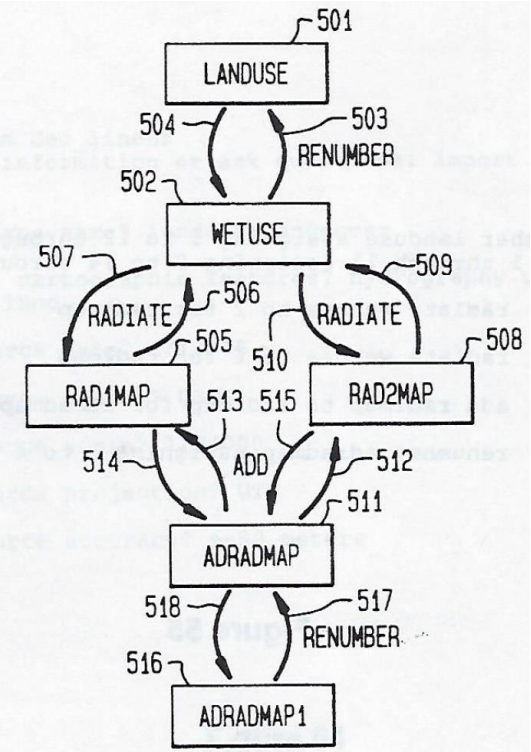
What are the final products of landuse

(ADRADMAP1 IS A PRODUCT MAP LAYER)



Why is rad2map a parent of adradmap1

(BECAUSE RAD2MAP IS A PARENT OF ADRADMAP AND ADRADMAP IS A PARENT OF ADRADMAP1)



In-class demonstration…

South Carolina Department of Commerce

…a manager of decision support who was visiting the class from the South Carolina State Economic Development Board exclaimed…

"I really need <u>that</u> to understand the data and maps my staff are producing with our GIS!"

This spurred the student to wondering:
Is this a solution to a significant problem?

One authority stated that geographic information system users:

"…are not generally explicitly aware of the source or accuracy of their data….

There may be several levels of abstraction and generalization between the cartographic product and the …data that was originally used to produce it.

Once a map product is produced, none of the associated data used to produce it remains available.

That is, the product is divorced from the quality, reliability and timeliness of the source material…. Only by knowing the source and derivation methods can we begin to establish data accuracy."

McKeown, D.M. Jr. 1987, "The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems", IEEE Transactions of Geoscience and Remote Sensing, vol. GE-25, no. 3, pp. 330-347

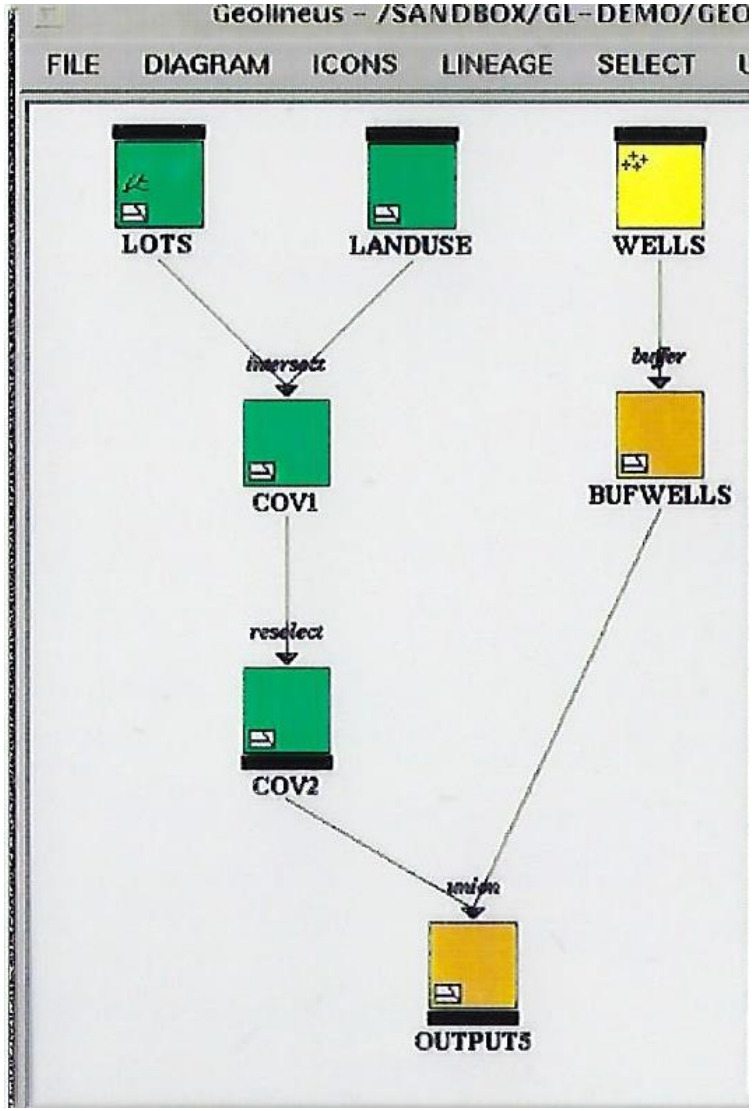
Is this a solution to a significant problem?

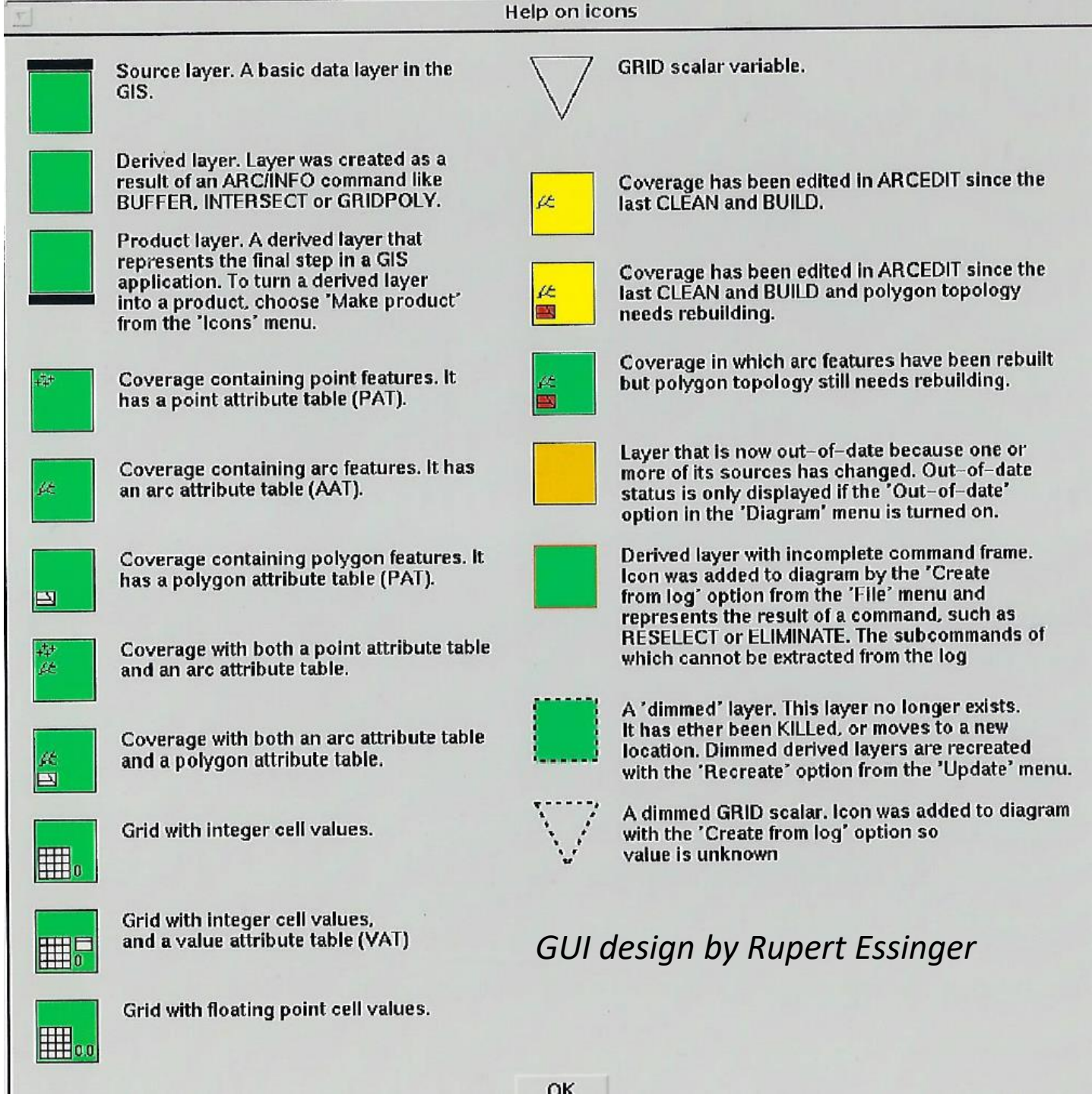
Another pair of authorities indicated:

"Of six interrelated spatial data quality components: lineage, positional accuracy, attribute accuracy, logical completeness and currency, lineage is the only one that is 'not testable' in the course of spatial data processing.

Vonderohe, A.P. and N.R. Chrisman 1985, "Tests to Establish the Quality of Digital Cartographic Data: Some Examples from the Dane County Land Records Project", Proceedings of the Seventh International Symposium on Computer-Assisted Cartography, American Congress on Surveying and Mapping; Falls Church, VA, pp. 552-559

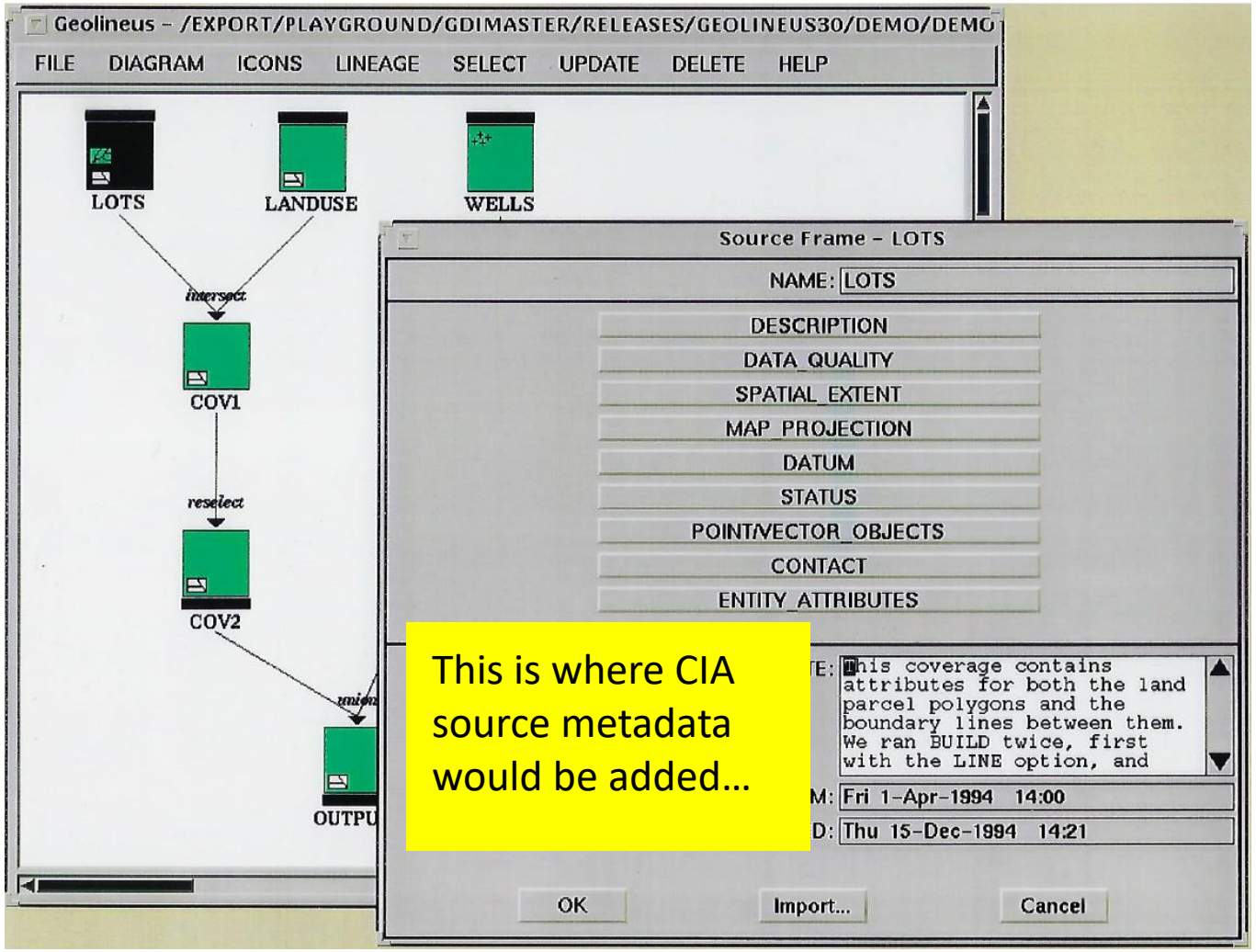
Adding a graphical user interface…



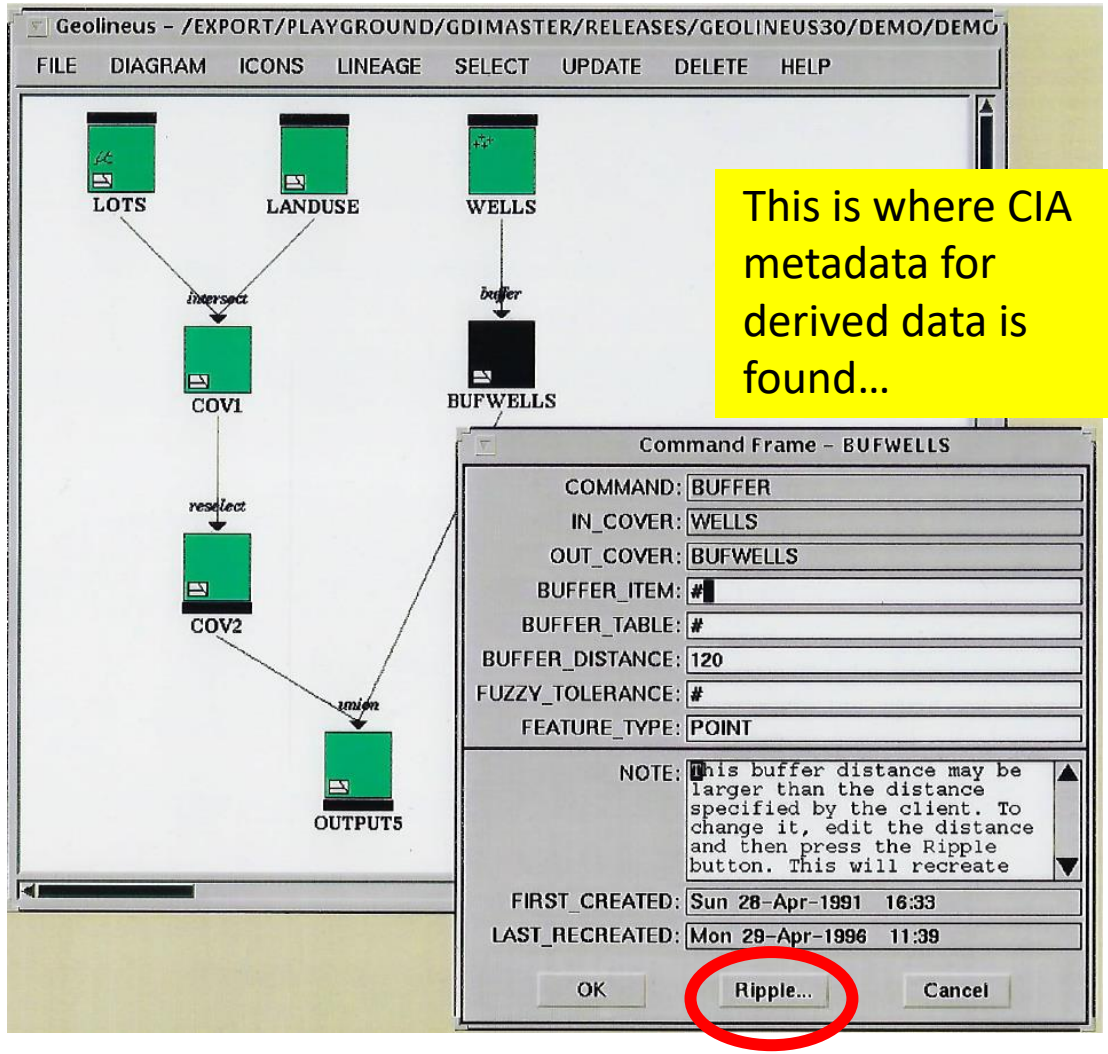


GUI design by Rupert Essinger

Working with source and command metadata

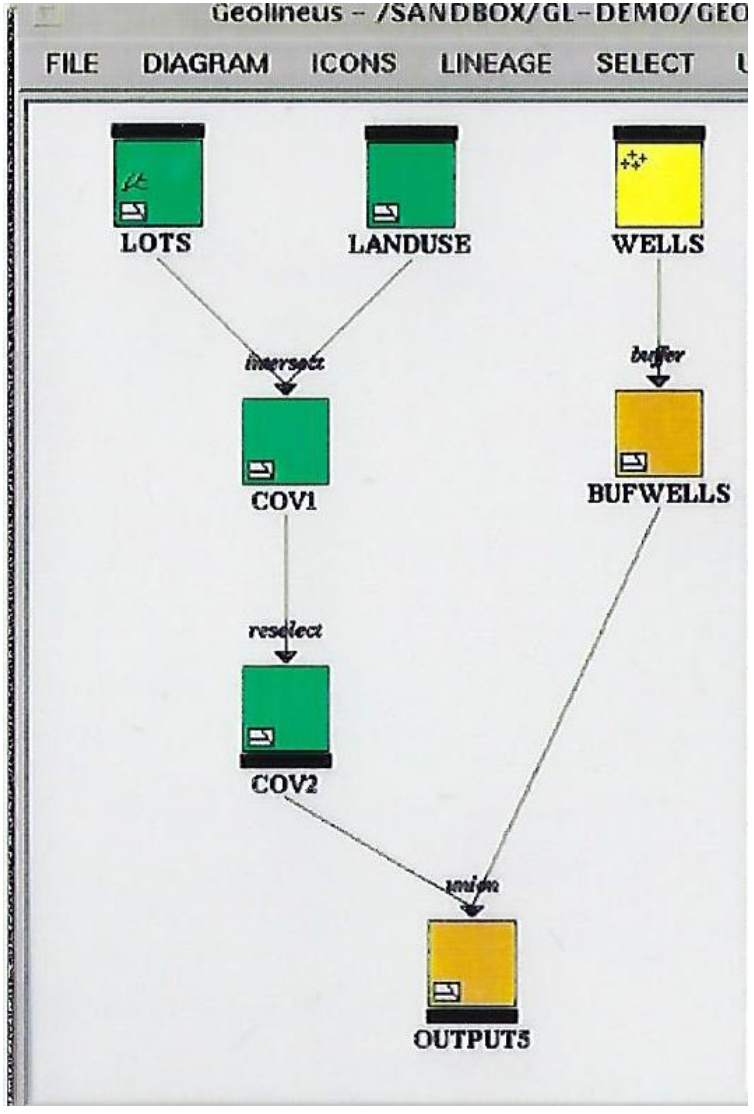


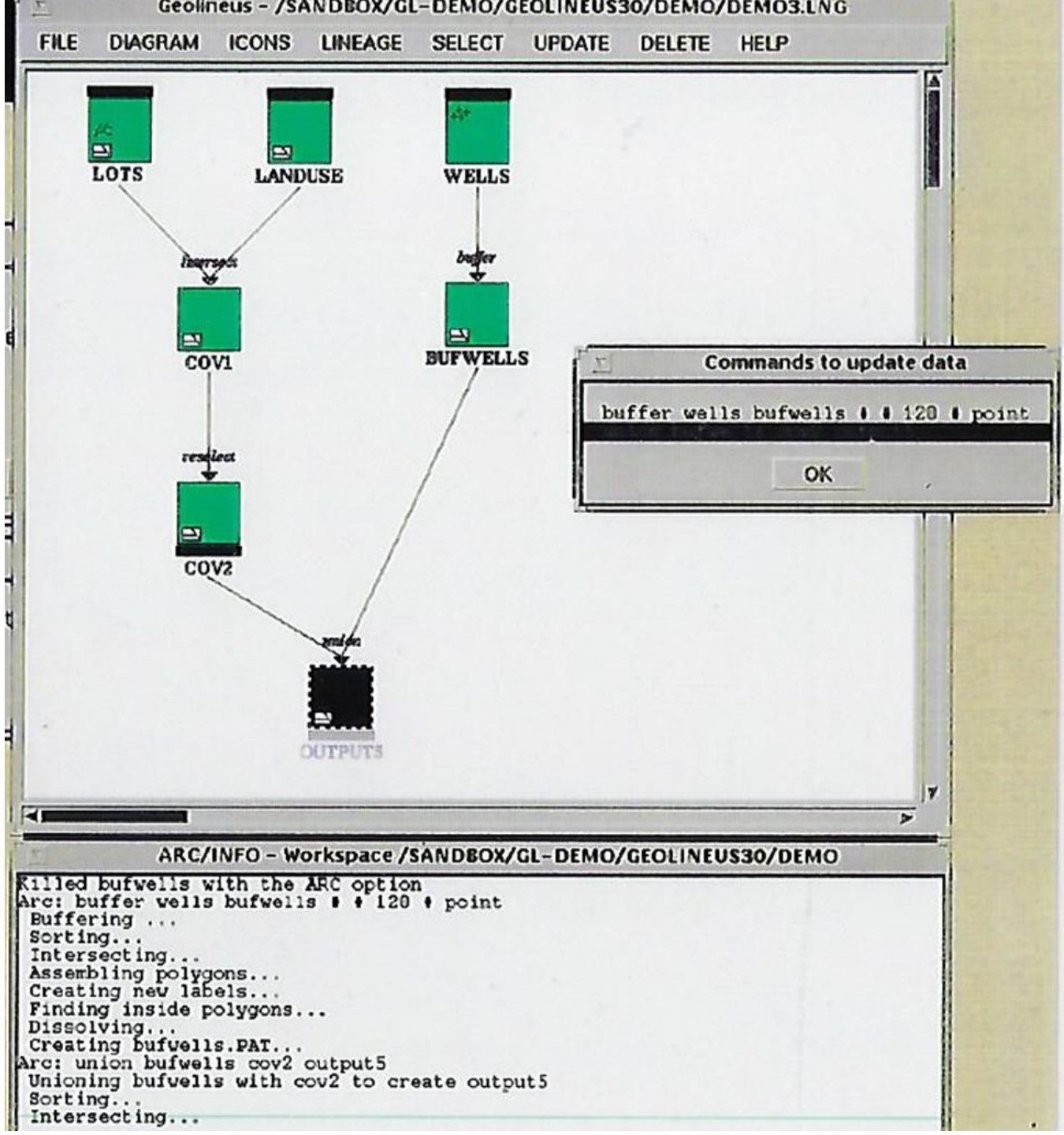
This is where CIA source metadata would be added…



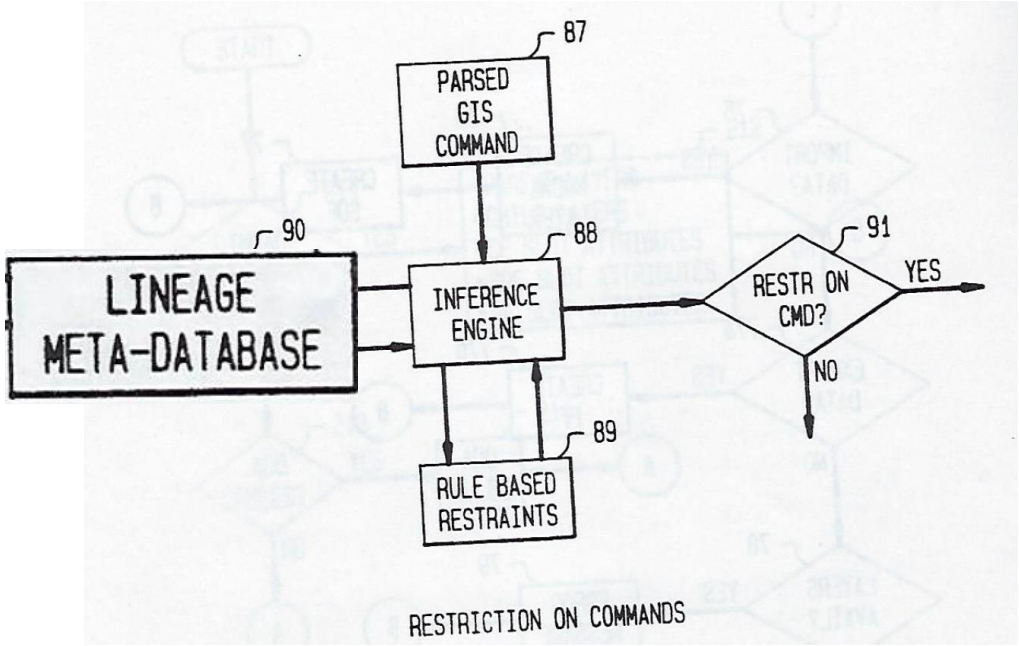
This is where CIA metadata for derived data is found…

Update propagation…

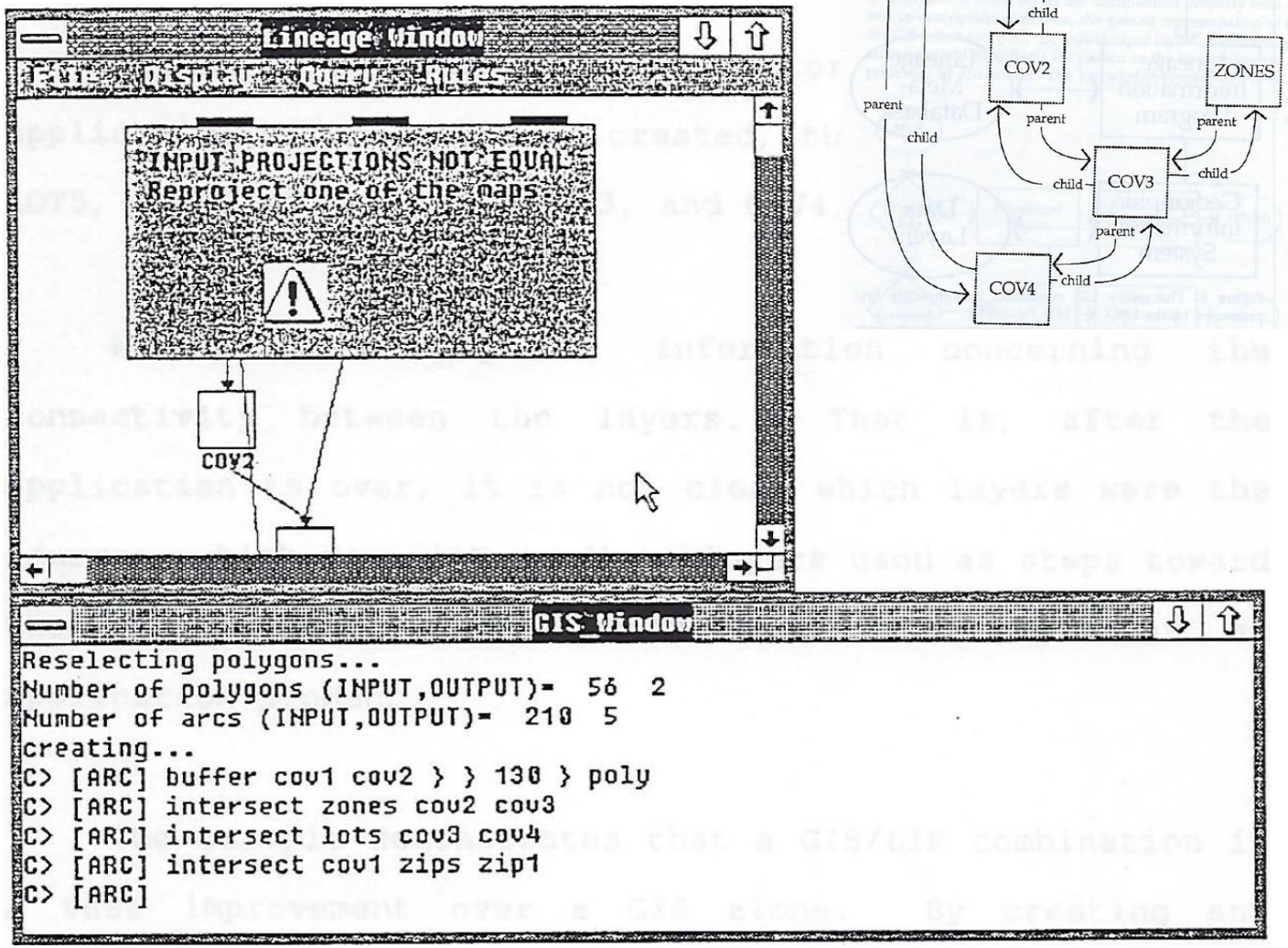


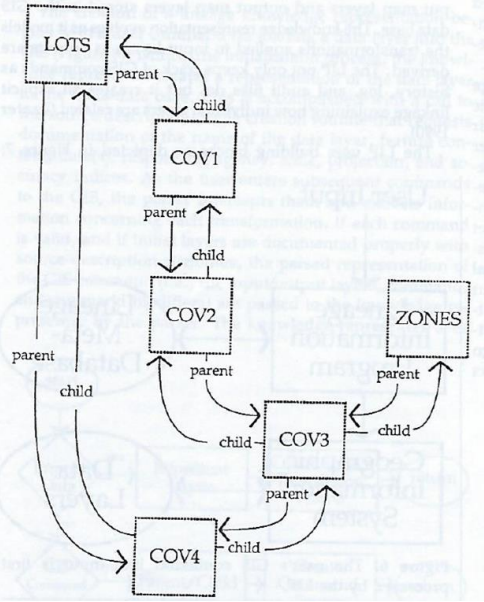


Source metadata-based integrity constraints



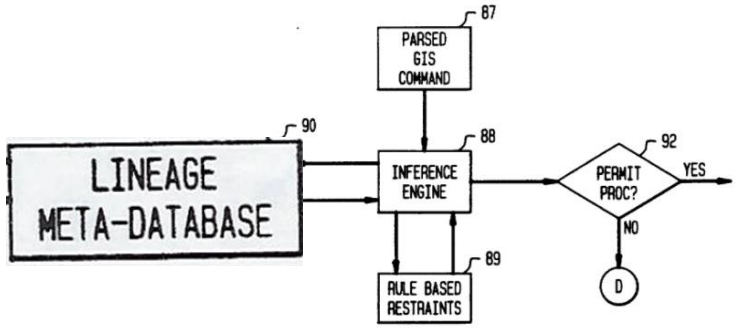

```
(setq intersect_rules

    '((rule intersect1
          (if (not (equal (scale inmap)
                          (scale intersectmap))))
          (then ("INPUT SCALES NOT EQUAL" )))

      (rule intersect2
          (if (not (equal (projection inmap)
                          (projection intersectmap))))
          (then ("INPUT PROJECTIONS NOT EQUAL")
                ("Reproject one of the maps.")) )))
```

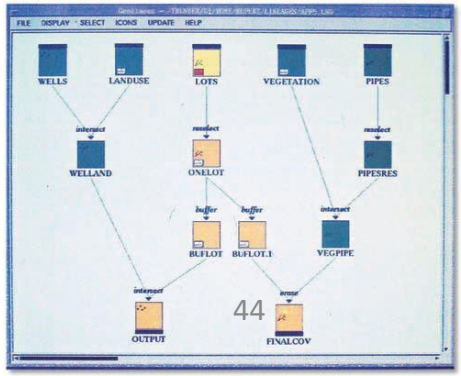




Data lineage metadata can help information systems meet key data privacy by design requirements, including:

- Enabling data subjects access, review and rectify their personal data
- Enabling data subjects to withdraw given consent with effect for the future by:
  a. Blocking access to their personal data
  b. Constraining processing and usage of their personal data
  c. Erasing their personal data
- Blocking and restricting personal data obtained for one purpose from being processed for other purposes not compatible with the original purpose





…it also enables data quality modeling

A Research Paradigm for Propagating Error in Layer-Based GIS

David P. Lanter and *Howard Veregin*[*]
Department of Geography/NCGIA, University of California, Santa Barbara, CA 93106

ABSTRACT: This paper focuses on the nature of error in spatial databases and the implications of this error for spatial data transformations in GIS applications. It describes an error propagation research paradigm as an information flow linking successively more formal components of error propagation in a GIS context. These components include development of conceptual models of error, creation of formal indices to measure error in spatial databases, implementation of mathematical functions to transform error indices and model the propagation of error as it is processed, and evaluation of the indices to gain insight into the utility of conceptual models used in error measurement and propagation. The paradigm enables researchers to formulate, manipulate, and experiment with components of error propagation to determine their implications for decision making. The applicability of the paradigm is illustrated with a simple GIS application in which error is propagated from sources to final product through a sequence of data transformation functions.

INTRODUCTION

GEOGRAPHIC INFORMATION SYSTEMS PROVIDE USERS WITH convenient and consistent mechanisms for applying automated transformation functions to manipulate and analyze spatial data. These capabilities expand the role and increase the value of spatial databases used in a variety of decision-making contexts. Such systems, however, often lack capabilities for establishing the accuracy and validity of products derived to support decisions. That is, a GIS provides a means of deriving new information without simultaneously providing a mechanism for establishing its reliability. The literature detailing GIS applications shows that there is a lack of concern for error in spatial databases and its propagation through sequences of data transformation functions. In such applications input data quality is often not ascertained, functions are applied to these data without regard for the accuracy of derived products, and these products are presented without an associated estimate of their reliability or an indication of the types of error they may contain.

Such omissions do not imply that errors are of such low magnitude that they can simply be ignored. Rather, they reflect the lack of a standard framework for modeling how error is propagated through sequences of data transformation functions. Paradoxically, an enormous volume of research has been carried out on the question of spatial database accuracy and the errors introduced by various types of data transformation (Goodchild and Gopal, 1989; Veregin, 1989a). Numerous indices have been developed to measure spatial and aspatial dimensions of error in databases, and methods have been proposed for modeling the ways in which data transformation functions modify and introduce error. Much of this research, however, has been carried out in isolation from the broader context of error propagation modeling in a GIS environment. There is a lack of a methodology for specifying the interactions among these various error indices and models of error propagation. That is, there is no accepted paradigm for modeling error propagation that explicitly recognizes the interdependence between basic concepts of spatial database accuracy and formal methods of error propagation in an actual system.

Figure 1 illustrates an informational flow linking successively more formal components of error propagation modeling and is presented as a possible error propagation research paradigm. The conceptual model of error reflects notions of what error signifies in a particular context. This ontological issue is of fundamental importance because error in spatial databases is inherently multi-dimensional. The utility of different dimensions of error is a function of context defined by the requirements of the uses and the classes of geographical data under consideration. Once determined, significant dimensions of error must be represented numerically as an index or set of indices for error measurement. This permits error propagation to be implemented by an error propagation function. Such functions model how a particular type of error is modified as spatial data are processed by a given data transformation function. Automated error propagation functions can be used to track errors present in source data through specific sequences of data transformation functions to determine the quality of a GIS derived data product.

The sections that follow discuss conceptual models of error for geographic data, indices to measure those errors, and functions to propagate the indices in a GIS application. Error propagation research is facilitated by a computer program for testing error indices and error propagation functions. The program utilizes a meta-data model of a GIS application allowing users to characterize data sources with error indices and implement

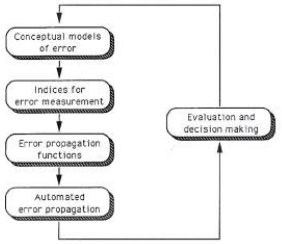


FIG. 1. An error propagation research paradigm.

[*]Presently with the Dept. of Geography, Kent State University, Kent, OH 44242.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 58, No. 6, June 1992, pp. 825–833.

0099-1112/92/5806-825$03.00/0

Pergamon

Comput., Environ. and Urban Systems, Vol. 19, No. 1, pp. 23–36, 1995

0198-9715/95 $9.50 + .00

0198-9715(94)00032-8

DATA-QUALITY ENHANCEMENT TECHNIQUES IN LAYER-BASED GEOGRAPHIC INFORMATION SYSTEMS

Howard Veregin

Department of Geography, Kent State University, Kent, OH

David P. Lanter

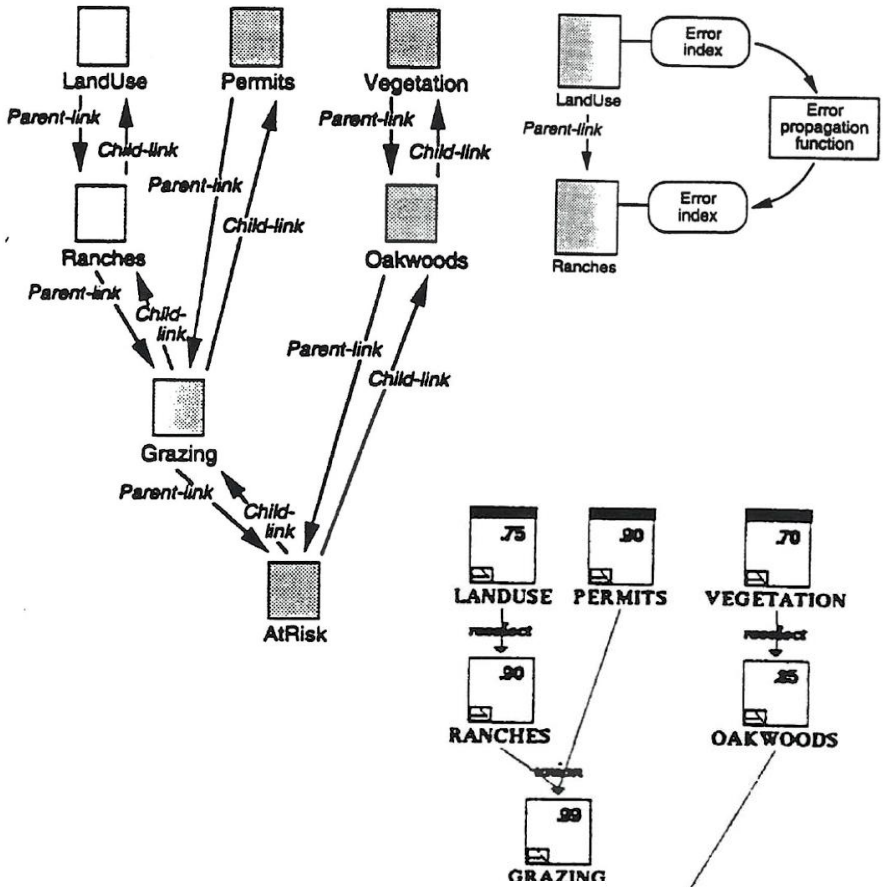
Department of Geography, University of California, Santa Barbara, CA

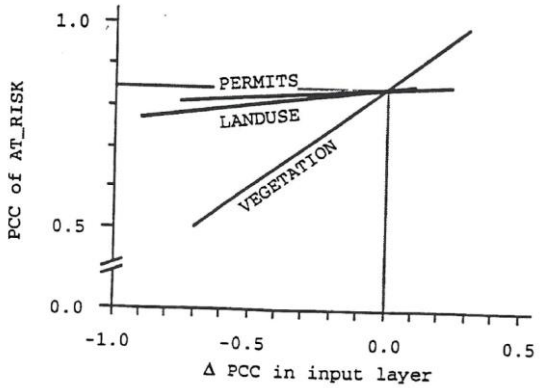
ABSTRACT. *This study deals with the general issue of data quality in geographic information systems (GIS). The specific focus is the propagation of source data errors through GIS data-transformation functions. The objective of the study is to describe quality enhancement (QE) tools that can be used to improve the quality of derived data products. These tools allow users to explore the error characteristics of their databases, devise optimal strategies for improving the accuracy of derived data, and enhance the reliability of information used for decision-making purposes.*

INTRODUCTION

The need for data-quality assessment techniques and quality-assurance procedures in geographic information systems (GIS) no longer needs much by way of formal introduction or justification. In recent years the issue has received a great deal of attention from the GIS community. There has been a significant increase in the number of journal and conference articles related to the issue of GIS data quality, and several books have recently been published on the topic (Goodchild & Gopal, 1989; Hunter, 1991). The importance of data quality is also reflected in the recent adoption of the Spatial Data Transfer Standard (SDTS), which includes a data-quality component, as a Federal Information Processing Standard (FIPS) to serve all segments of the U.S. federal geographical information processing community. Nor is this type of activity restricted to the United States, as evinced by work on data quality being conducted by the International Cartographic Association, the efforts to promote European data standards (such as DIGEST

Requests for reprints should be sent to Dr. H. Veregin, Department of Geography, Kent State University, Kent, OH 44242-0001, e-mail: veregin@humboldt.kent.edu.





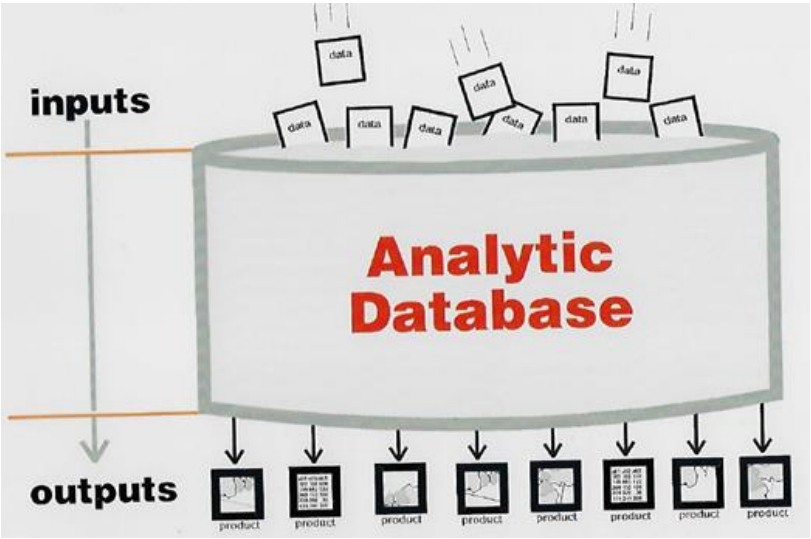
Case Study: First data provenance IT Audit

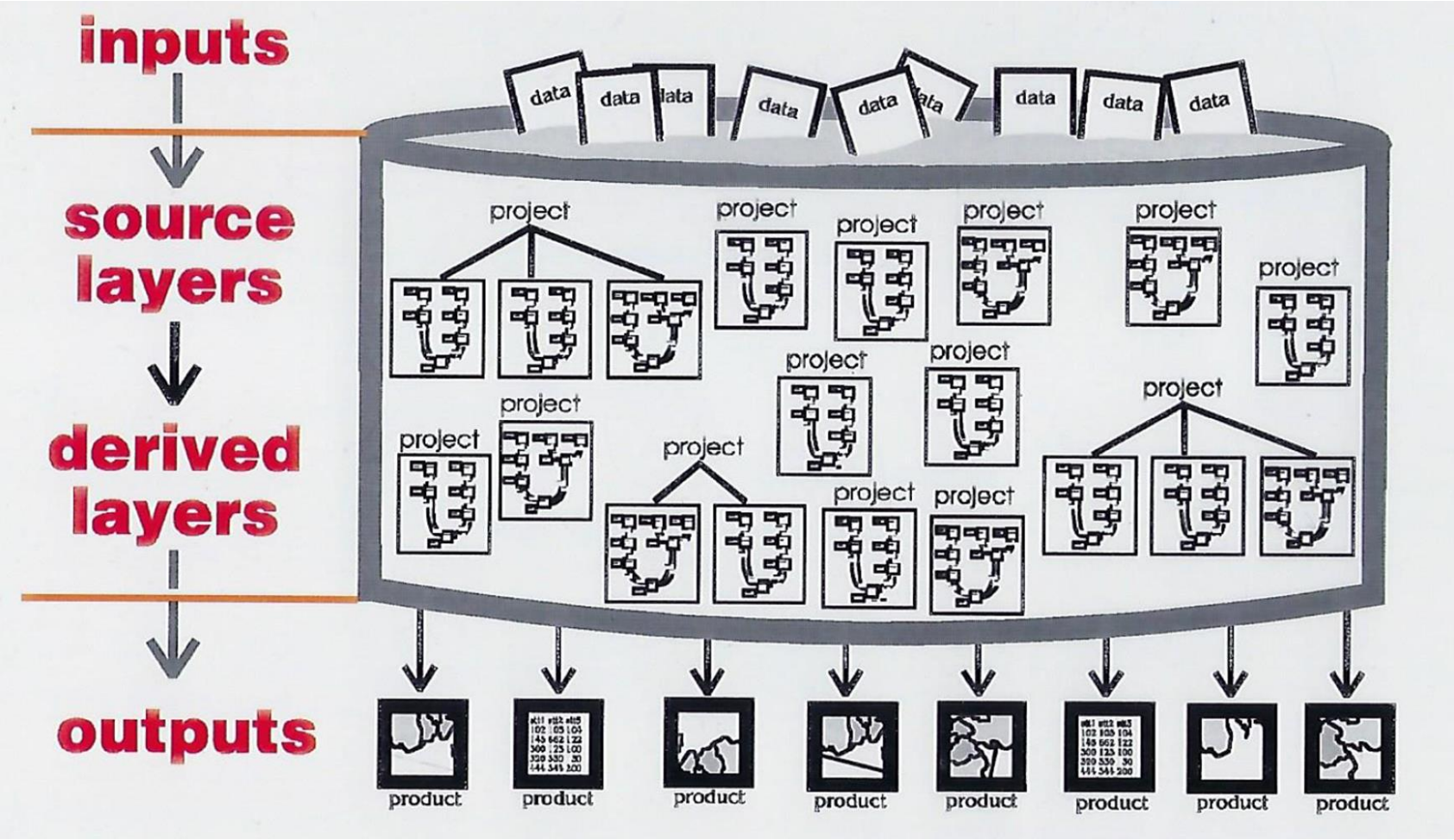
conducted in 1992 for Southern California Edison

Focus of the audit:

1. Document and help management understand the quality of their decision support data
2. Test scientific replicability of data used in decision making

Data provenance audit problem…





Extraction of metadata of data and processing

Geolineus user guide

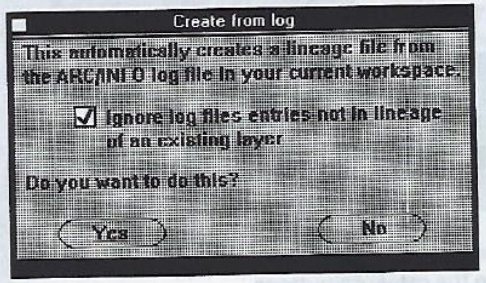
Contents

To install Geolineus see the separate 'Geolineus Release N
Instructions' document.

Creating a new lineage diagram

The Geolineus "Create from log" option in the "File" menu automatically creates a lineage diagram for an ARC/INFO workspace by reading the workspace's ARC/INFO **log file**. The workspace log file is maintained by ARC/INFO and records the commands and their parameters that have been performed on the layers in that workspace. When "Create from log" reads a workspace's log file it looks for ARC/INFO commands that process data (see "Help on commands" from the Geolineus "Help" menu for a list of these commands) and creates a lineage diagram to represent the processing.that has taken place.

1. Make sure you are in the ARC/INFO workspace (page 11) you want to document.
2. Select "Create from log" from the "File" menu. This box pops up ( ).

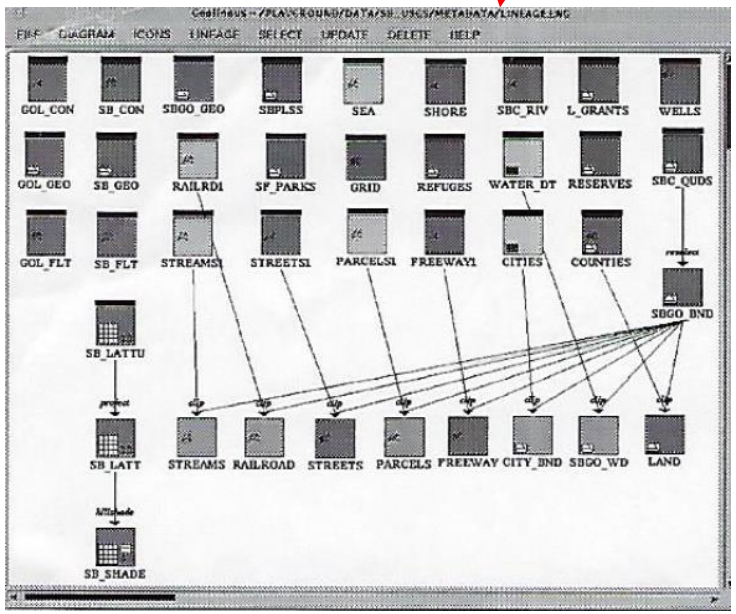


The check option enables you to choose whether or not the diagram that "Create from log" will create will include lineage for layers that no longer exist. Normally, Geolineus will **ignore** any lines in the log file that do not contribute to the lineage of an existing layer. This results in a lineage diagram that documents the **current state of the workspace**.

If you uncheck the option, Geolineus creates a diagram using **all** the lines in the log file, even if they are in the lineage of layers that no longer exist. This results in a diagram showing what has **happened previously** in the workspace **in addition** to its current state. Use this option for example to create a diagram from a log file for which the data is unavailable.

Log Files

198923021442	1	3	OARCPLOT
198923021442	0	10	OBUILD NISLAND POLY
198923021442	0	1	OEXTERNAL NISLAND
198923021503	20	44	OARCPLOT
198923021505	0	3	OPOLYGRID NISLAND
198923021512	2	15	Opolygrid nisland
198923021514	1	24	Ogridpoly nisland.svf nigrid 662795 680175 30 30
198923021516	2	6	Oarcplot
198923021520	2	4	Oarcplot
198923021520	0	2	Oarcplot
198923021520	0	0	Oexternal nisland
198923021520	0	1	Oexternal nigrid
198923021520	0	3	Oarcplot
198923021526	5	71	Oarcedit
198923021530	0	1	ORENAME NIGRID NIG30
198923021533	3	72	OPOLYGRID NISLAND GR10.SVF
198923021536	3	85	OGRIDPOLY GR10.SVF NI10 662795 680175 10 10



Lineage metadata enabled audit of data and processing

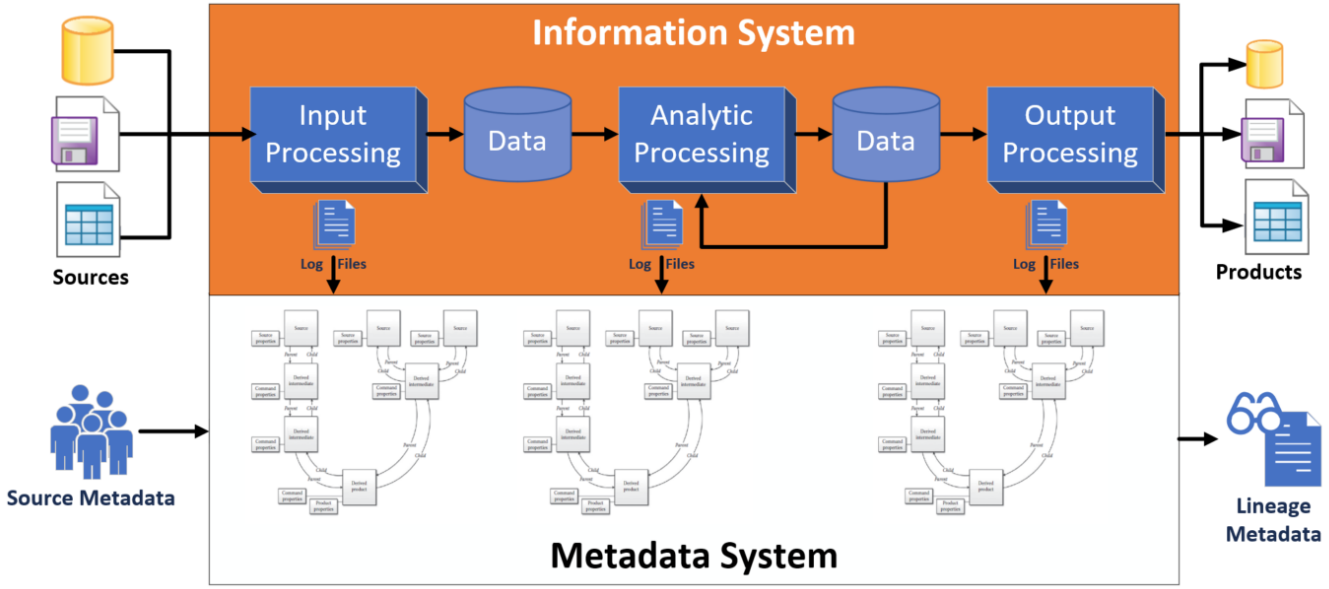
at Southern California Edison

9 visits with SCE's GIS Lab's technical staff in 1992, collected:

1. Descriptions of 14 data processing projects

2. Metadata for data sources that were acquired and imported into the enterprise GIS database for the projects

3. Processing log files for the projects



Lineage metadata enabled audit of data and processing

at Southern California Edison

1. Descriptions of 14 data processing projects

…for 7 corporate divisions were examined:

- Customer Service
- Engineering
- Environmental Research
- Information Services
- Power Generation
- Project Development
- Sewer & Hydrologic Engineering

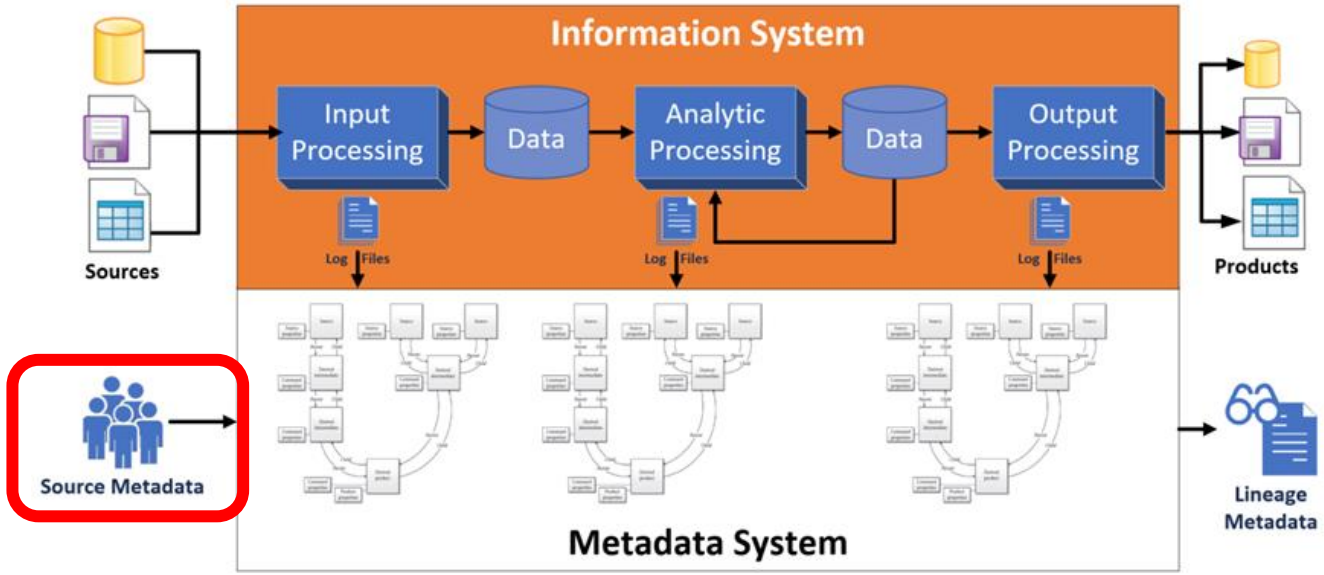
Project	Output	Deliverable
1	1 map	Spatial distribution of SCE substations relative to important features
2	5 maps	SCE's Service Territory and its various features
3	1 map	SCE's Service Territory and various features
4	1 map	Areas in Redlands CA near power lines containing sensitive species
5	1 map	Areas in Victorville CA near transmission lines containing sensitive species
6	1 map	Route of proposed pipeline from Mandalay facility to Ormond Beach facility
7	data file	Locations of historic sites in Redlands CA
8	database	Land use information for species habitat study
9	1 map	Land use, street network, elevation contours in areas around microwave stations
10	Map	Land use and street network reference map of Ormond Beach area
11	21 maps data file	3 maps each for 7 dam/reservoir sites in SCE Territory; Data file of calculated terrain units for use in hydrologic modeling project
12	database	Environmental site suitability models for locating artificial reef to mitigate impact of San Onofre Nuclear Generation Station as requirement of operation permit
13	1 map	SCE Service Territory's relationships between switching and intermediate processing centers
14	2 maps	Congressional boundaries and demographic data

Linage metadata enabled audit of data and processing

at Southern California Edison

2. Identified data acquired from internal and external sources and collected metadata on these data

- Entity types (“features”) and attribute content
- Format
- Area covered
- Geographic scale and spatial resolution
- Location coordinate system
- Spatial projection
- Supplying agency
- Original source organization
- Original publication date
- Production source date
- Responsible staff member
- Statement of data quality

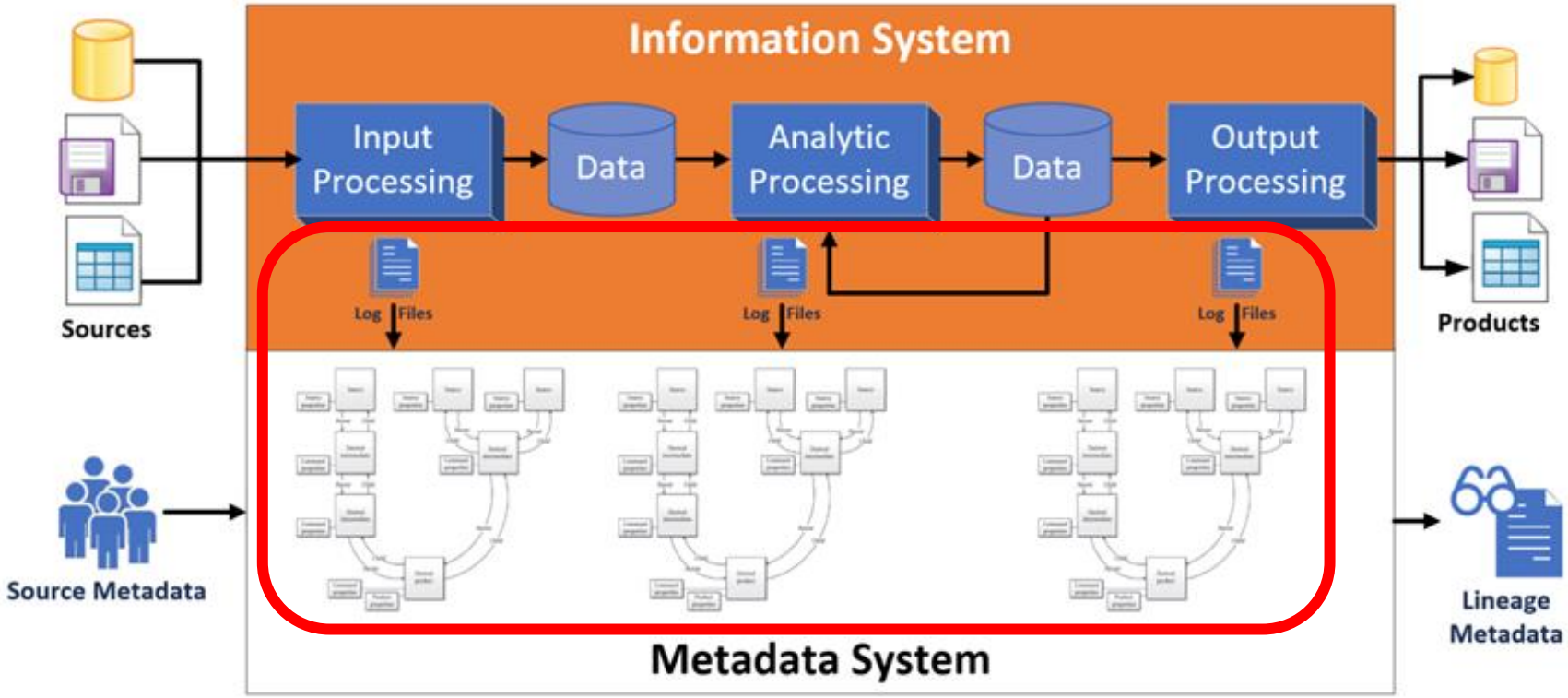


Metadata enabled audit of data and processing

at Southern California Edison

3. Processing log files obtained for each of the 14 projects

Reverse engineered lineage metadata from the log files



Metadata enabled audit of GIS data and processing

at Southern California Edison

GIS Lab analysts identified 54 data files input into the Information System to support their projects, obtained from:

- Internal client department
- Other internal departments
- California state agencies
- Outside consultants

Log processing identified 806 datasets referenced in the log files :

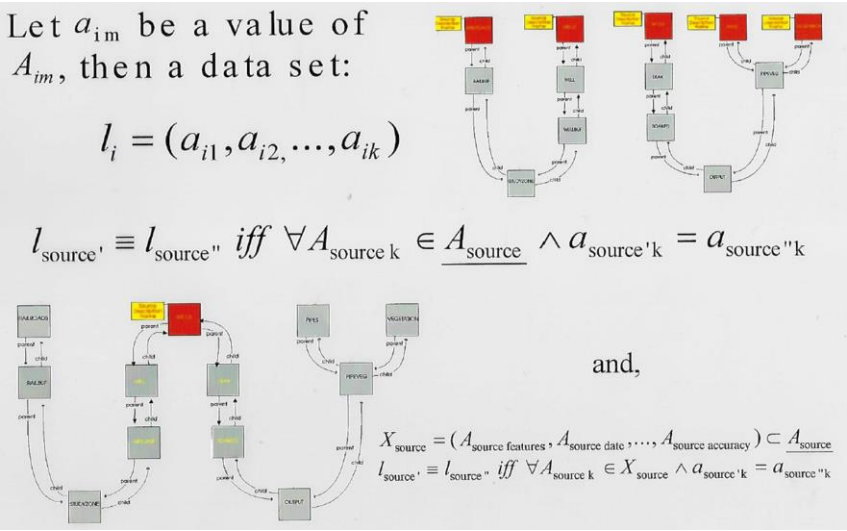
- 487 source datasets (i.e. lacking child links pointing to inputs)
- 319 derived datasets

Metadata enabled audit of GIS data and processing

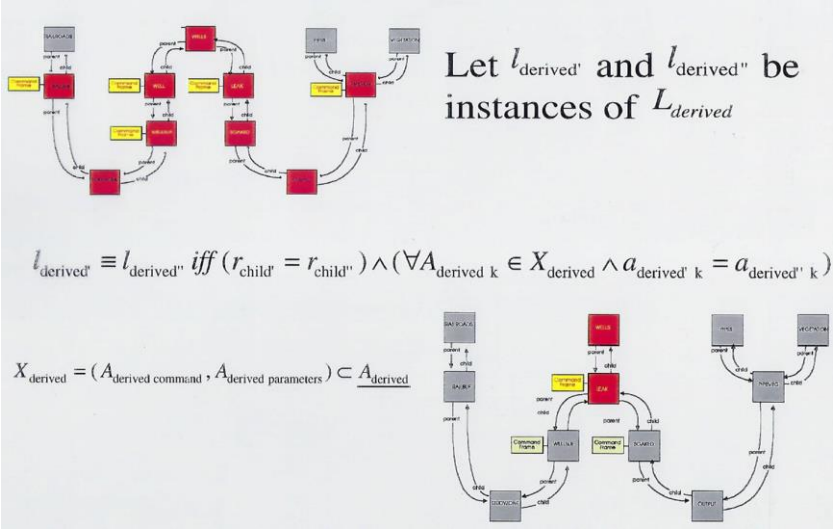
at Southern California Edison

Next step… would have focused on use of metadata analysis to identify **commonalities and differences** in:

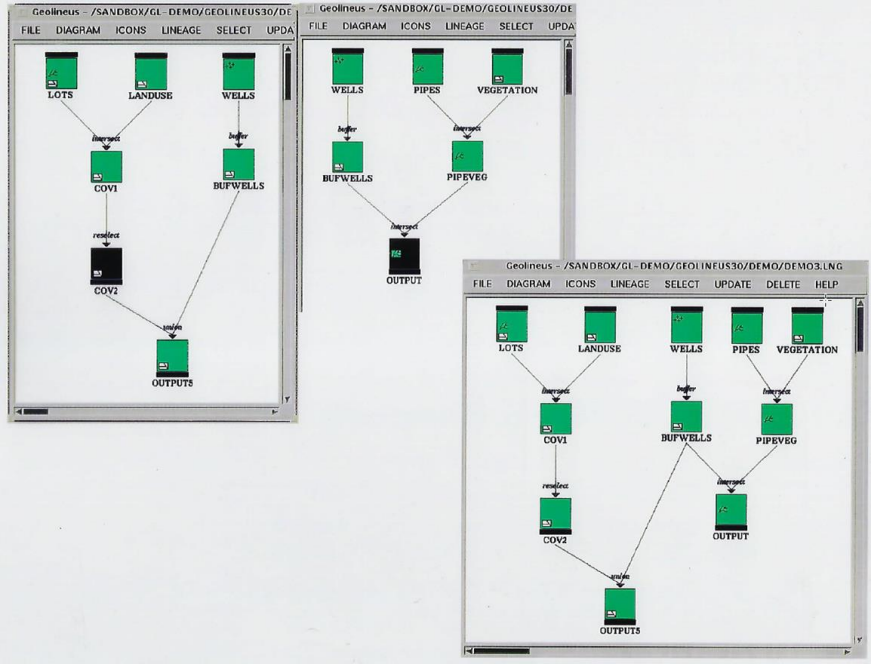
1. Source data usage
2. Analytical processing logic



Source equivalence testing



Derived equivalence testing



Metadata enabled audit of GIS data and processing

at Southern California Edison

…instead we found:

1. Much metadata for documenting the data sources were missing…

   - GIS Lab Technical Staff analysts were unable to remember much about the data they had used in earlier projects

   - Of the 54 data files used as input to the GIS database:
     - 89% were of unknown Spatial Projections
     - 79% were of unknown Original Publication Dates
     - 70% were of unknown Scales and Spatial Resolutions
     - 68% were from unknown Original Source Organizations
     - 43% contained attributes and spatial data assumed “fit for use” but untested

Metadata enabled audit of GIS data and processing

at Southern California Edison

We also found:

2. Lack of naming conventions for identifying primary data source files and source datasets once they were imported into the Information System
   - For example,
     - "TER" used as mnemonic device to name datasets after import:
       - 5 datasets in Project 1: TERBND, TER.MRK, TERMRK1, TERMRK2, and TERMERK3
       - 3 datasets in Project 2: TERRITORY, SCE-TERR, SCE-TERR2
       - Information Analysts could not differentiate them

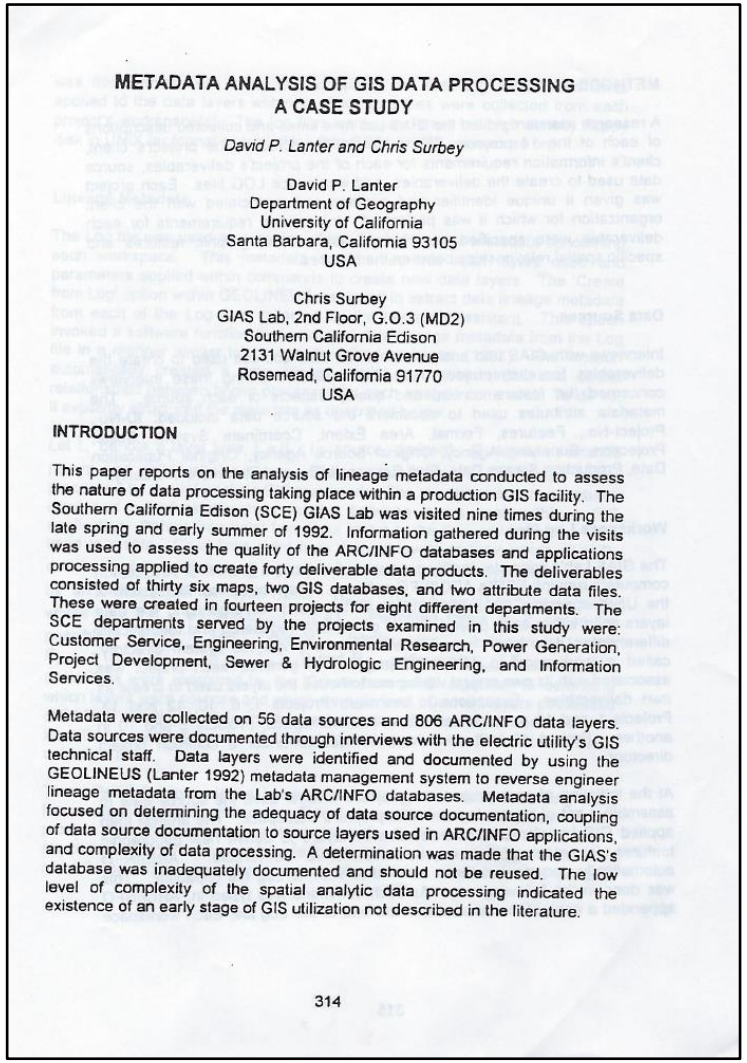
Utility company only had one service territory boundary, there were 8 different versions of it. Without taking the itme to visually inspect and compare the actual data – it was not clear what, if any, significant differences existed among the versions

Metadata enabled audit of GIS data and processing

at Southern California Edison

Recommendation:

- GIS Lab's *"…database was inadequately documented and should not be reused."*



METADATA ANALYSIS OF GIS DATA PROCESSING
A CASE STUDY

David P. Lanter and Chris Surbey

David P. Lanter
Department of Geography
University of California
Santa Barbara, California 93105
USA

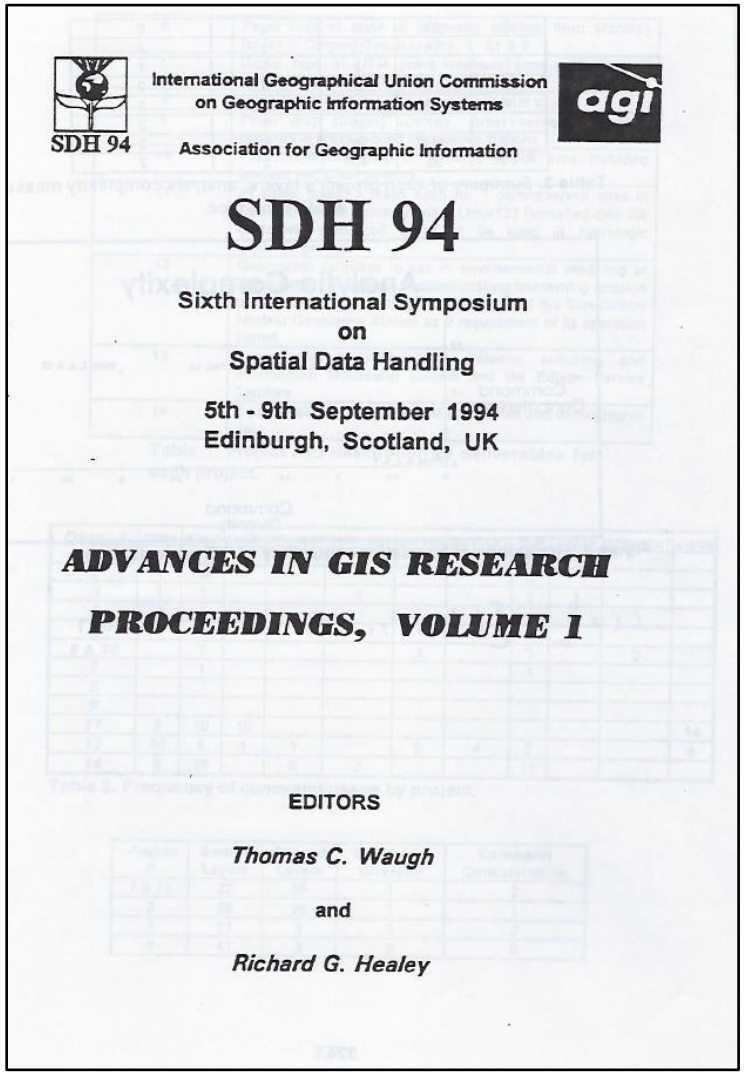
Chris Surbey
GIAS Lab, 2nd Floor, G.O.3 (MD2)
Southern California Edison
2131 Walnut Grove Avenue
Rosemead, California 91770
USA

INTRODUCTION

This paper reports on the analysis of lineage metadata conducted to assess the nature of data processing taking place within a production GIS facility. The Southern California Edison (SCE) GIAS Lab was visited nine times during the late spring and early summer of 1992. Information gathered during the visits was used to assess the quality of the ARC/INFO databases and applications processing applied to create forty deliverable data products. The deliverables consisted of thirty six maps, two GIS databases, and two attribute data files. These were created in fourteen projects for eight different departments. The SCE departments served by the projects examined in this study were: Customer Service, Engineering, Environmental Research, Power Generation, Project Development, Sewer & Hydrologic Engineering, and Information Services.

Metadata were collected on 56 data sources and 806 ARC/INFO data layers. Data sources were documented through interviews with the electric utility's GIS technical staff. Data layers were identified and documented by using the GEOLINEUS (Lanter 1992) metadata management system to reverse engineer lineage metadata from the Lab's ARC/INFO databases. Metadata analysis focused on determining the adequacy of data source documentation, coupling of data source documentation to source layers used in ARC/INFO applications, and complexity of data processing. A determination was made that the GIAS's database was inadequately documented and should not be reused. The low level of complexity of the spatial analytic data processing indicated the existence of an early stage of GIS utilization not described in the literature.

314



SDH 94

International Geographical Union Commission
on Geographic Information Systems

agi

Association for Geographic Information

SDH 94

Sixth International Symposium
on
Spatial Data Handling

5th - 9th September 1994
Edinburgh, Scotland, UK

ADVANCES IN GIS RESEARCH

PROCEEDINGS, VOLUME 1

EDITORS

Thomas C. Waugh

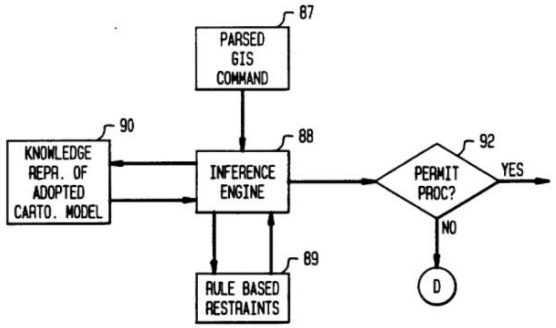
and

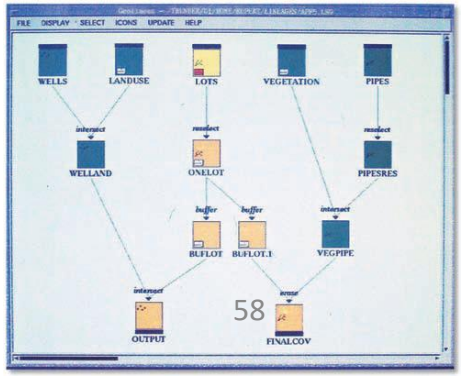
Richard G. Healey

This resulted in a follow-on consulting contract to provide help SCE's data analysts with guidelines and standards for developing decision support data with data provenance documentation

Conclusion: Data lineage metadata can help information systems meet key data privacy by design requirements, including:

- Enabling data subjects access, review and rectify their personal data
- Enabling data subjects to withdraw given consent with effect for the future by:
  a. Blocking access to their personal data
  b. Constraining processing and usage of their personal data
  c. Erasing their personal data
- Blocking and restricting personal data obtained for one purpose from being processed for other purposes not compatible with the original purpose



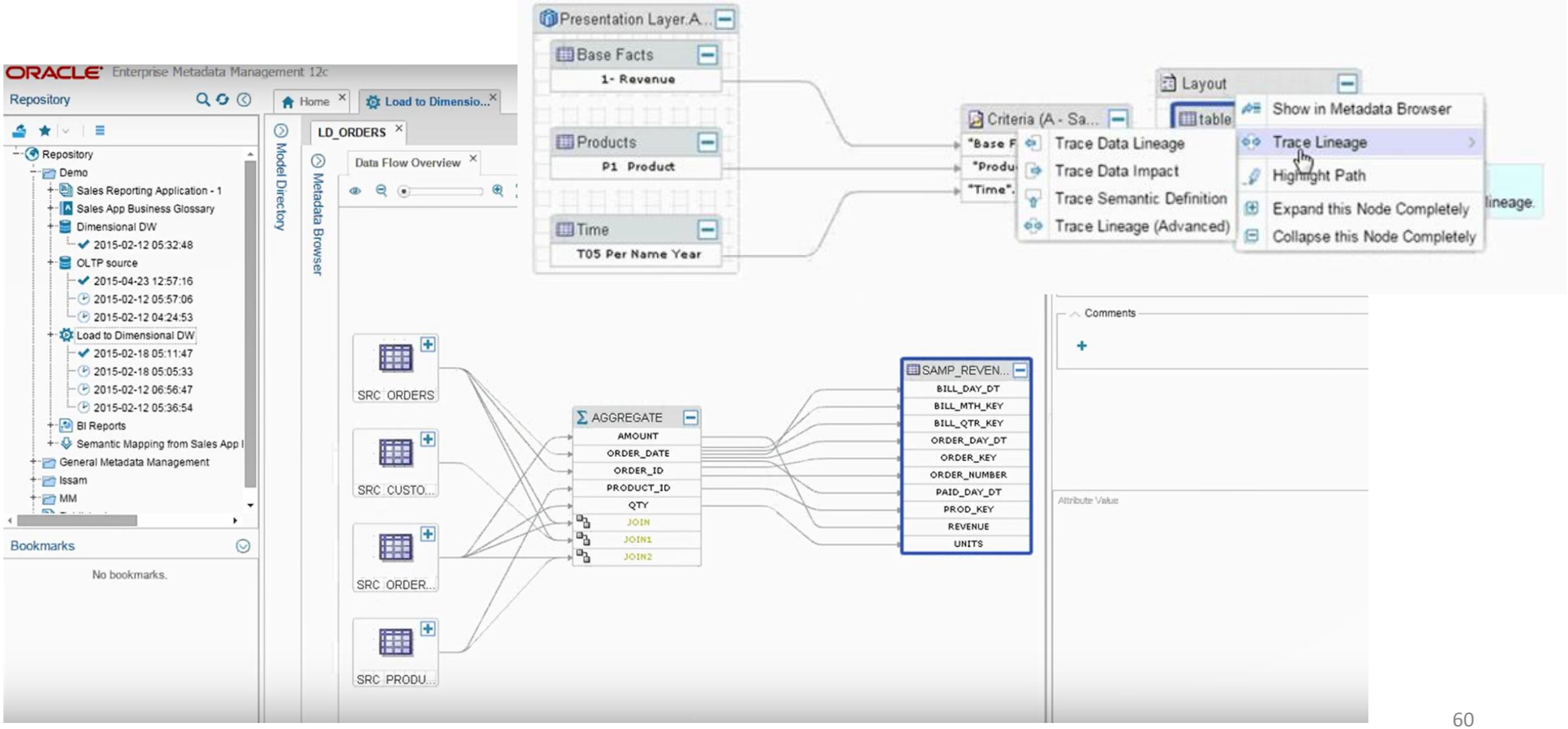


Conclusion:

Data lineage metadata can be used to help information system developers meet key data protection by design requirements:

1. Data subjects have **right to access, review and rectify** their personal data
2. Data subjects have the **right to withdraw given consent** with effect for the future and
   - Block access
   - Constrain processing and use
   - Erase their personal data
3. Personal **data obtained for one purpose must not be processed for other purposes** not compatible with the original purpose

Outlook: Commercial database management systems are beginning to include lineage metadata capabilities for tracking attribute values processed and transformed among relational database tables ...

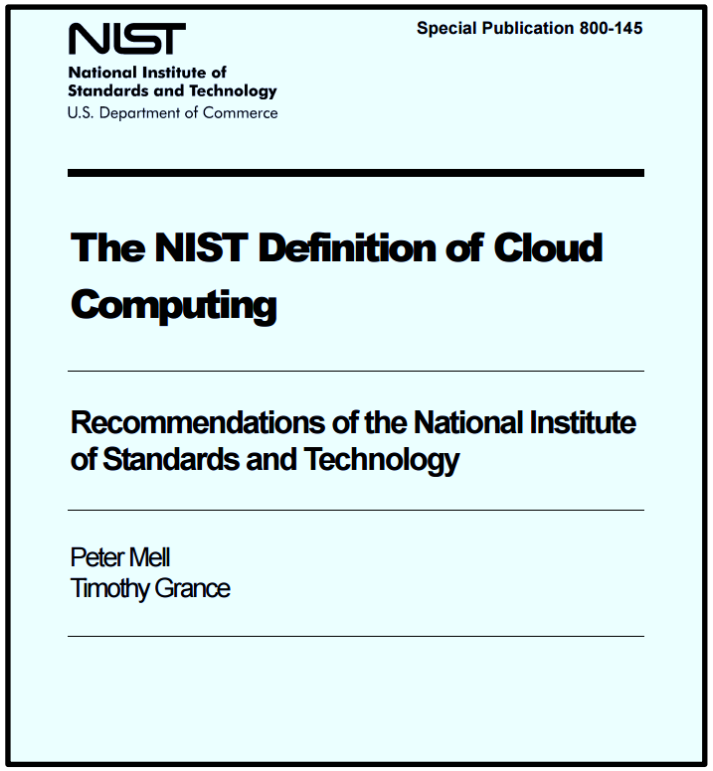


Agenda

- ✓ Data protection by design
- System Security Plan
  - Cloud computing specifications
  - Security control inheritance

Cloud computing

Cloud computing enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction


NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Special Publication 800-145

The NIST Definition of Cloud Computing

Recommendations of the National Institute of Standards and Technology

Peter Mell
Timothy Grance

Essential Characteristics of Cloud Computing

1. **On-demand self-service**
2. **Broad network access**
3. **Resource pooling**
4. **Rapid elasticity**
5. **Measured service**

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Which Service Model(s) of cloud computing is your project's information system providing to your end users?

TABLE OF CONTENTS

Table 3.1 provides a summary of the key attributes of the CSO.

Table 3.1 System Information

System Information	
CSP Name:	<Insert CSP Name> <Insert CSP Abbreviation, as appropriate>
CSO Name:	<Insert CSO Name> <Insert CSO Abbreviation, as appropriate>
FedRAMP Package ID:	<Insert FedRAMP Package ID>
Service Model:	<Choose one: IaaS, PaaS, SaaS, IaaS/PaaS, IaaS/PaaS/SaaS, IaaS/SaaS, PaaS/SaaS, LI-SaaS>
Digital Identity Level (DIL) Determination (SSP Appendix E):	<Choose one: IAL3/FAL3/AAL3, IAL2/FAL2/AAL2, IAL1/FAL1/AAL1>
FIPS PUB 199 Level (SSP Appendix K):	<Choose one: High, Moderate, Low, LI-SaaS>
Fully Operational as of:	<Insert MM/DD/YYYY>
Deployment Model:	<Choose one: Public Cloud, Government-Only Cloud, Hybrid Cloud>
Authorization Path:	<Choose one: Joint Authorization Board Provisional Authorization, Agency Authorization>
General System Description:	<Insert CSO Name> is delivered as [a/an] [insert based on the Service Model above] offering using a multi-tenant [insert based on the Deployment Model above] cloud computing environment. It is available to [Insert scope of customers in accordance with instructions above (for example, the public, federal, state, local, and tribal governments, as well as research institutions, federal contractors, government contractors etc.)].

3 Service Models of Cloud Computing

Infrastructure as a Service (IaaS)

Provides processing, storage, networks, and other fundamental computing resources

Consumer is able to deploy and run arbitrary software, which can include operating systems and applications

- The consumer does not manage or control the underlying cloud infrastructure,
  - but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls)

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

3 Service Models of Cloud Computing

Platform as a Service (PaaS)

Consumer is provided capability to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider

- The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage,
  - but has control over the deployed applications and possibly configuration settings for the application-hosting environment

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

3 Service Models of Cloud Computing

Software as a Service (SaaS)

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure

- Accessible from various client devices through either a thin client interface, such as a web browser or a program interface
- The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Which cloud deployment model is your project's information system based on?

TABLE OF CONTENTS

Table 3.1 provides a summary of the key attributes of the CSO.

Table 3.1 System Information

System Information	
CSP Name:	<Insert CSP Name> <Insert CSP Abbreviation, as appropriate>
CSO Name:	<Insert CSO Name> <Insert CSO Abbreviation, as appropriate>
FedRAMP Package ID:	<Insert FedRAMP Package ID>
Service Model:	<Choose one: IaaS, PaaS, SaaS, IaaS/PaaS, IaaS/PaaS/SaaS, IaaS/SaaS, PaaS/SaaS, LI-SaaS>
Digital Identity Level (DIL) Determination (SSP Appendix E):	<Choose one: IAL3/FAL3/AAL3, IAL2/FAL2/AAL2, IAL1/FAL1/AAL1>
FIPS PUB 199 Level (SSP Appendix K):	<Choose one: High, Moderate, Low, LI-SaaS>
Fully Operational as of:	<Insert MM/DD/YYYY>
Deployment Model:	<Choose one: Public Cloud, Government-Only Cloud, Hybrid Cloud>
Authorization Path:	<Choose one: Joint Authorization Board Provisional Authorization, Agency Authorization>
General System Description:	<Insert CSO Name> is delivered as [a/an] [insert based on the Service Model above] offering using a multi-tenant [insert based on the Deployment Model above] cloud computing environment. It is available to [Insert scope of customers in accordance with instructions above (for example, the public, federal, state, local, and tribal governments, as well as research institutions, federal contractors, government contractors etc.)].

4 Deployment Models of Cloud Computing

Public cloud

The cloud infrastructure is provisioned for open use by the general public

- It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

4 Deployment Models of Cloud Computing

Private cloud

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units)

- It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

4 Deployment Models of Cloud Computing

Community cloud

Provisioned for use by a specific community of consumers from organizations with shared concerns

- It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

4 Deployment Models of Cloud Computing

Hybrid cloud

A composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities

- …but are bound together by standardized or proprietary technology that enables data and application portability

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Agenda

- ✓ Data protection by design
- ✓ Cloud computing specifications
- Security control origination
- Team project SSP progress review and discussion

Security Control Origination

Security control “inheritance” exist when

an information system or application receives protection from security controls developed, implemented, assessed, authorized, and monitored by entities other than those responsible for the system or application

NIST SP 800-53

IA-5 (3)	Control Summary Information
Responsible Role:	
Parameter IA-5(3)-1:	
Parameter IA-5(3)-2:	
Parameter IA-5(3)-3:	
Parameter IA-5(3)-4:	
Implementation Status (check all that apply): ☐ Implemented ☐ Partially implemented ☐ Planned ☐ Alternative implementation ☐ Not applicable	
Control Origination (check all that apply): ☐ Service Provider Corporate ☐ Service Provider System Specific ☐ Service Provider Hybrid (Corporate and System Specific) ☐ Configured by Customer (Customer System Specific) ☐ Provided by Customer (Customer System Specific) ☐ Shared (Service Provider and Customer Responsibility) ☐ Inherited from pre-existing FedRAMP Authorization for Click here to enter text. , Date of Authorization	

Control Origination

Many controls needed to protect organizational information systems are inheritable by other systems, e.g.

- Security awareness training
- Incident response plans
- Physical access to facilities
- Rules of behavior
- Public Key Infrastructure [PKI]
- Authorized secure standard configurations for clients/servers
- Access control systems
- Boundary protection
- Cross-domain solutions

Control Origination

Control Origination (check all that apply):

- ☐ Service Provider Corporate
- ☐ Service Provider System Specific
- ☐ Service Provider Hybrid (Corporate and System Specific)
- ☐ Configured by Customer (Customer System Specific)
- ☐ Provided by Customer (Customer System Specific)
- ☐ Shared (Service Provider and Customer Responsibility)
- ☐ Inherited from pre-existing FedRAMP Authorization for Click here to enter text. , Date of Authorization

- Indicate what sections of the security control are inherited and provide a description of what is inherited
- If a entire control is inherited, it must be clear to the Assessor what is inherited
- The writer does not need to describe how the leveraged service is performing the particular function
  - That detail is found in the SSP of the leveraged system from which the control is inherited

If a published policy is referenced as the basis for an inherited security control, make sure that published document is provided as an attachment, or a supporting artifact with the SSP when submitted for FedRAMP review

https://www.fedramp.gov/weekly-tips-cues-february-15-2017/

Control Origination

IA-5 (3)	Control Summary Information
Responsible Role:	
Parameter IA-5(3)-1:	
Parameter IA-5(3)-2:	
Parameter IA-5(3)-3:	
Parameter IA-5(3)-4:	
Implementation Status (check all that apply): ☐ Implemented ☐ Partially implemented ☐ Planned ☐ Alternative implementation ☐ Not applicable	
Control Origination (check all that apply): ☐ Service Provider Corporate ☐ Service Provider System Specific ☐ Service Provider Hybrid (Corporate and System Specific) ☐ Configured by Customer (Customer System Specific) ☐ Provided by Customer (Customer System Specific) ☐ Shared (Service Provider and Customer Responsibility) ☐ Inherited from pre-existing FedRAMP Authorization for Click here to enter text. , Date of Authorization	

Agenda

- ✓ITACS Mentoring Program
- ✓Catch up… Centralized Remote Access Control Technologies
- ✓Data protection by design
- ✓System Security Plan
  - ✓Cloud computing specifications
  - ✓Security control inheritance