

# MIS 4596

Data Privacy

Class 5

# Agenda

- Online privacy
- Privacy and data protection by design  
*...with data provenance and lineage metadata*

The background of the slide features a blurred, blue-tinted image of a crowd of people, likely at a public event or conference, with their silhouettes and movements softened to create a sense of a busy, populated environment.

# **ONLINE PRIVACY: HOW DID WE GET HERE?**

California Consumer Privacy Act (CCPA, 2018)  
&  
California Privacy Rights Act (CPRA, 2020)



# California voters approve Prop. 24, ushering in new rules for online privacy



---

## CORONAVIRUS AND PANDEMIC >

---

Concordia University coronavirus 'outbreak' attributed to more than 50 'false positives'

---

Are L.A. County's new COVID restrictions really necessary? We talk to the experts

---

Coronavirus infections are higher than ever, COVID-19 deaths are not. Why?

---

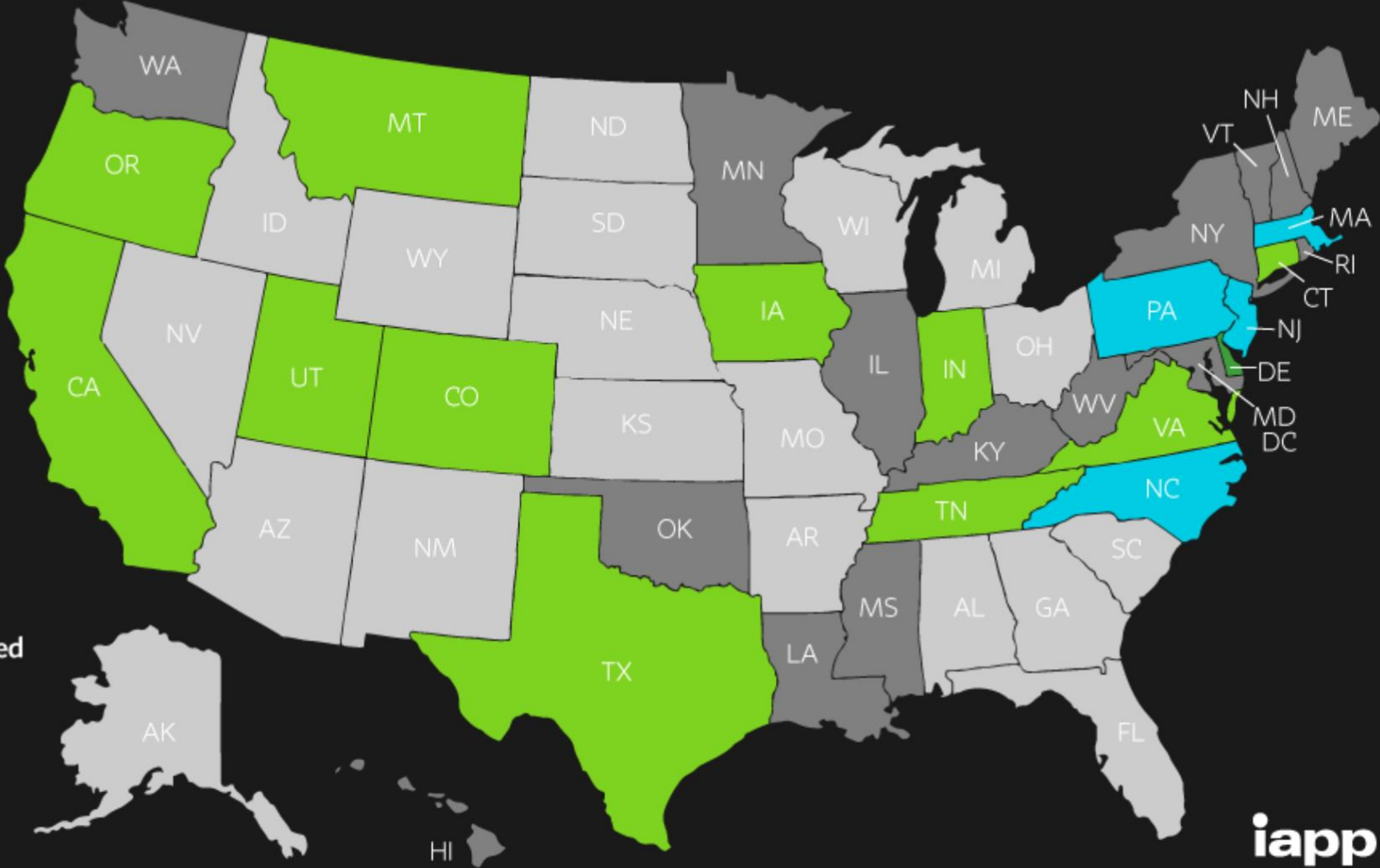
# State Data Privacy Laws in Effect

- **California Consumer Privacy Act (CCPA) – Effective date: 1/1/2020**
- California Privacy Rights Act (CPRA) – Effective date: 1/1/2023
- Colorado Privacy Act (CPA) – Effective date: 1/1/2023
- Virginia Consumer Data Protection Act – Effective date: 1/1/2023

# US State Privacy Legislation Tracker 2023

## STATUTE/BILL IN LEGISLATIVE PROCESS

- Introduced
- In committee
- In cross chamber
- In cross committee
- Passed
- Signed
- Inactive bills
- No comprehensive bills introduced



Last updated: 8/4/2023



[https://iapp.org/media/images/resource\\_center/State\\_Comp\\_Privacy\\_Law\\_Map.png](https://iapp.org/media/images/resource_center/State_Comp_Privacy_Law_Map.png)  
<https://iapp.org/resources/article/us-state-privacy-legislation-tracker/>



# Hilton Hotels fined for credit card data breaches

🕒 1 November 2017



🔗 Share



Top Story

**Ex-Marine  
bar attack**

The bar was  
country mus  
opened fire,

🕒 1 hour ago

**US Supreme  
tribs**

🕒 3 hours ago

**Russia pr  
Democrat**

🕒 2 hours ago

Feature

# Hilton Hotels fined for credit card data breaches

🕒 1 November 2017

f 🗨️ 🐦 ✉️ Share

Hilton's \$700,000 fine for data breach impacting 350,000 customers

Top Sto

**Ex-Marine  
bar attack**

The bar was  
country mus  
opened fire,

🕒 1 hour ag

**US Supre  
ribs**

🕒 3 hours a

**Russia pr  
Democrat**

🕒 2 hours a

Feature

# Hilton Hotels fined for credit card data

# breaches Under the European Union's

🕒 1 November 2017

General Data Protection

Regulation (GDPR) the fine  
would have been 4% of  
Hilton's global revenue  
\$420,000,000

Top Sto

Ex-Marine  
bar attack

The bar was  
country mus  
opened fire,

🕒 1 hour ag

US Supre  
ribs

🕒 3 hours a

Russia pr  
Democrat

🕒 2 hours a

Feature

# GDPR requires data security 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.



## 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

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)

# 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**”

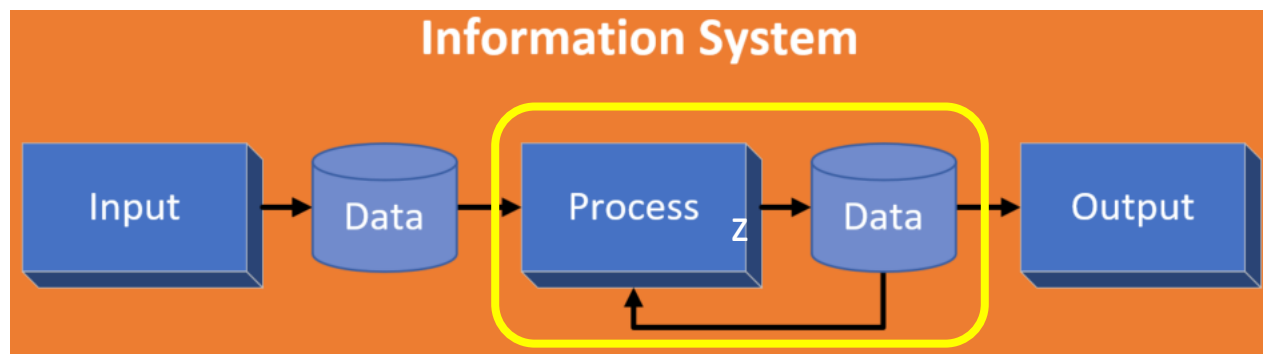
Danezis, G. et al. (2014) “Privacy and Data Protection by Design”,  
European Union Agency for Network and Information Security (ENISA)

# Some challenging data protection requirements may be solved with techniques presented in this webinar...

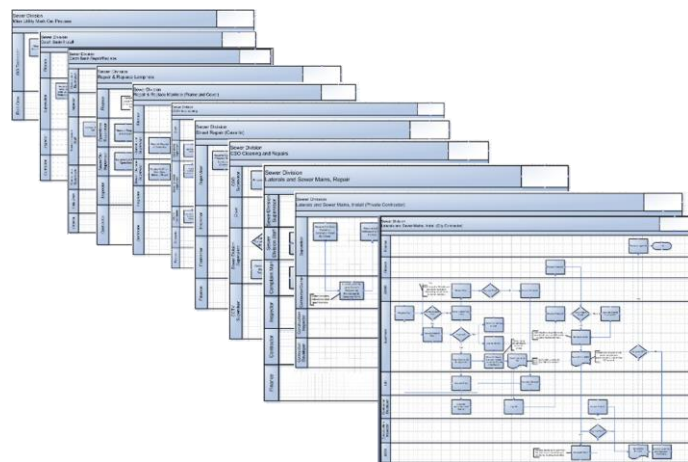
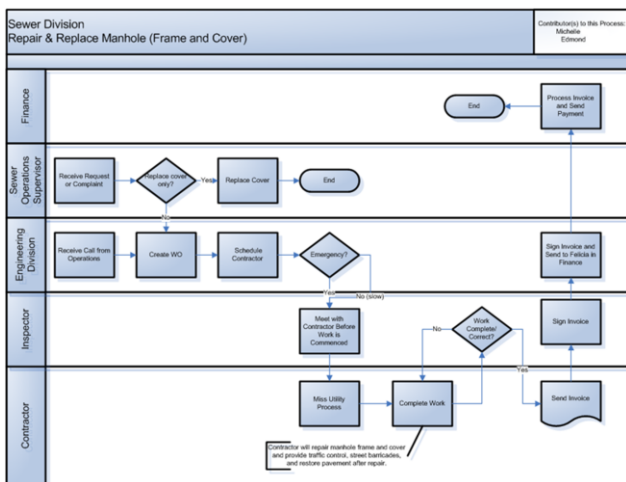
1. **Collection** of personal data is **fully avoided or minimized** at the earliest stage of processing
2. Data subjects give **specific, informed and explicit consent** to the processing of their data
3. Data subjects have **right to access, review and rectify** their personal data
4. Data subjects 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...



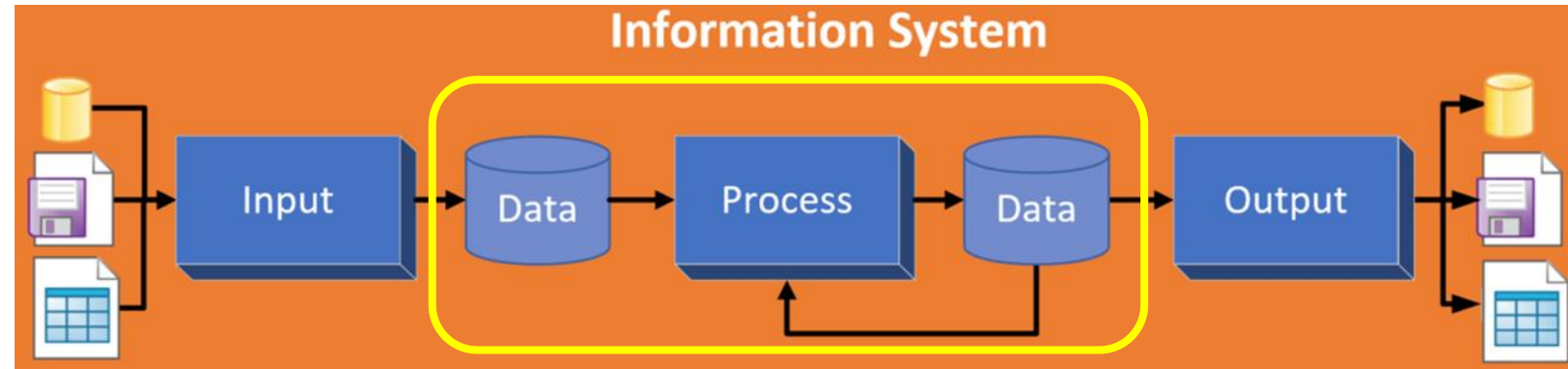
Regardless of application, there is reliance on data processing workflows to produce and use information



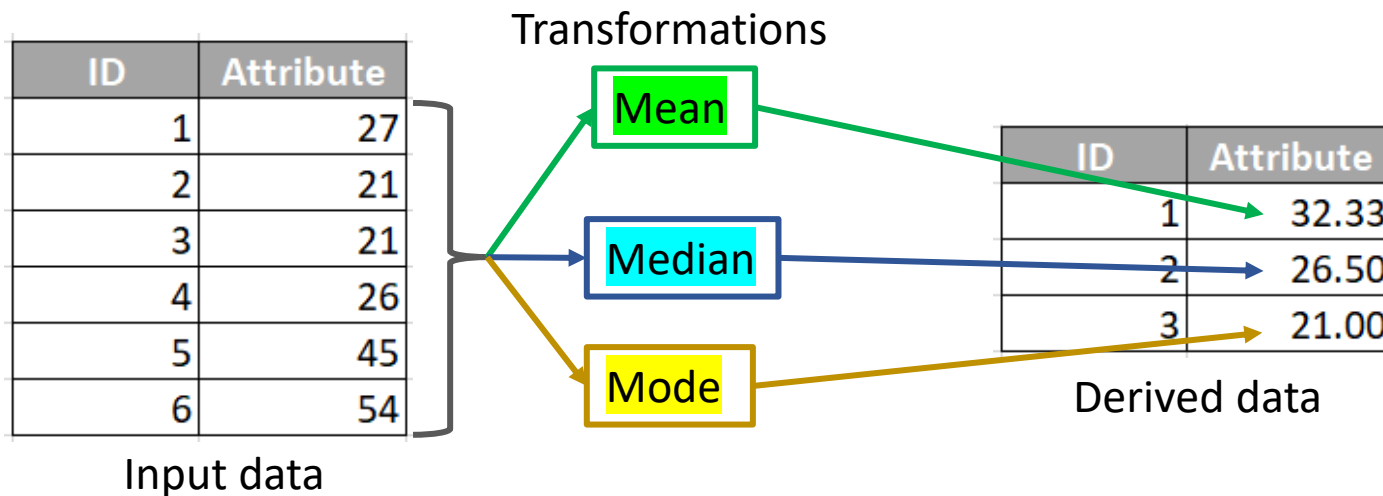


Data processing often transforms existing data into new data, which is a double-edged sword...

➤ *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*



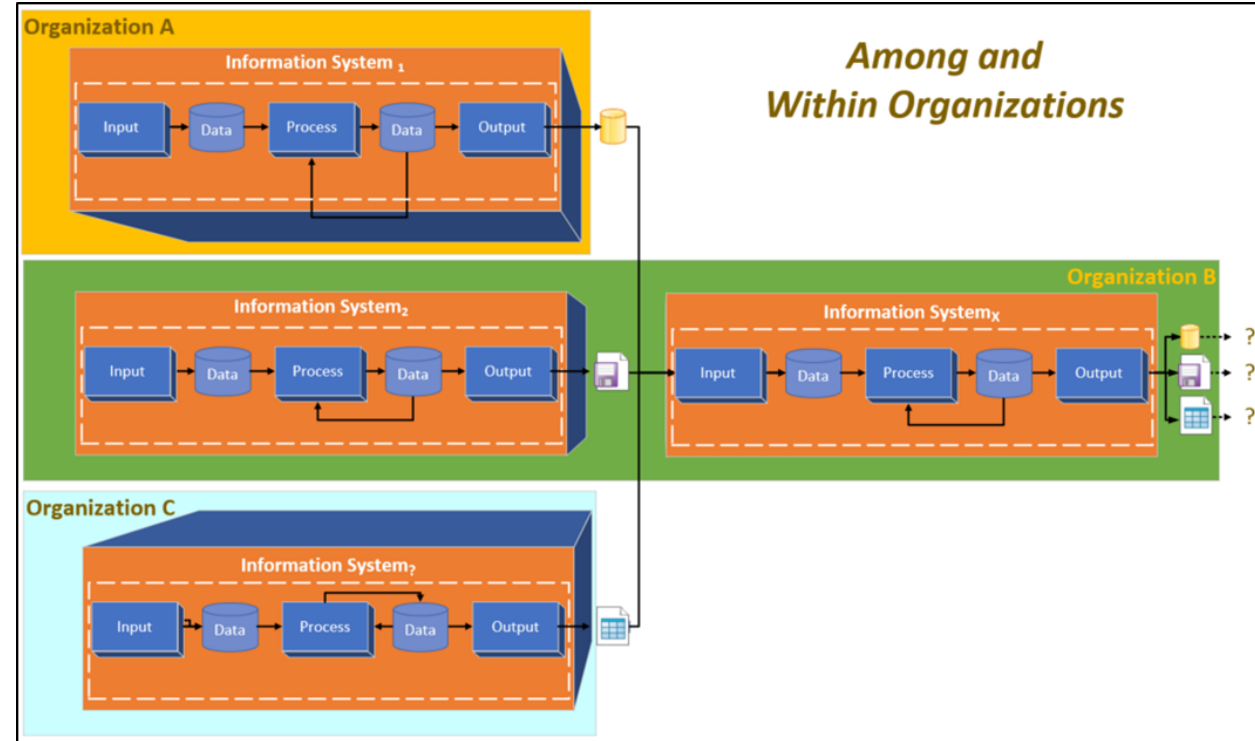
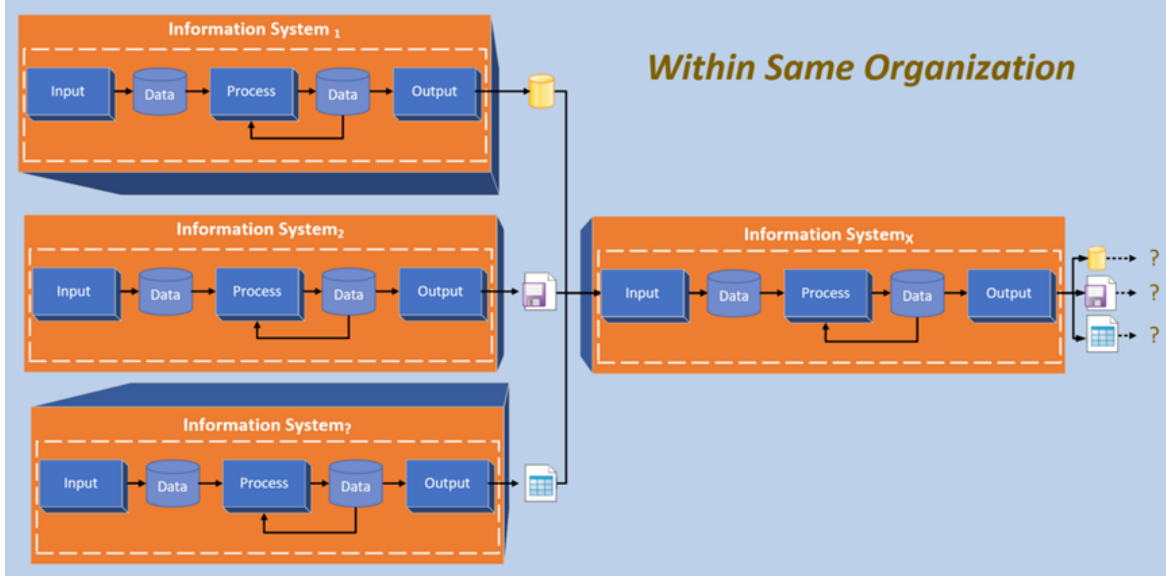
# Evaluating & judging data's "fitness for use"

- **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

## 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.

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

Standardizing Museum Provenance – David Newbury (@workergnome)

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

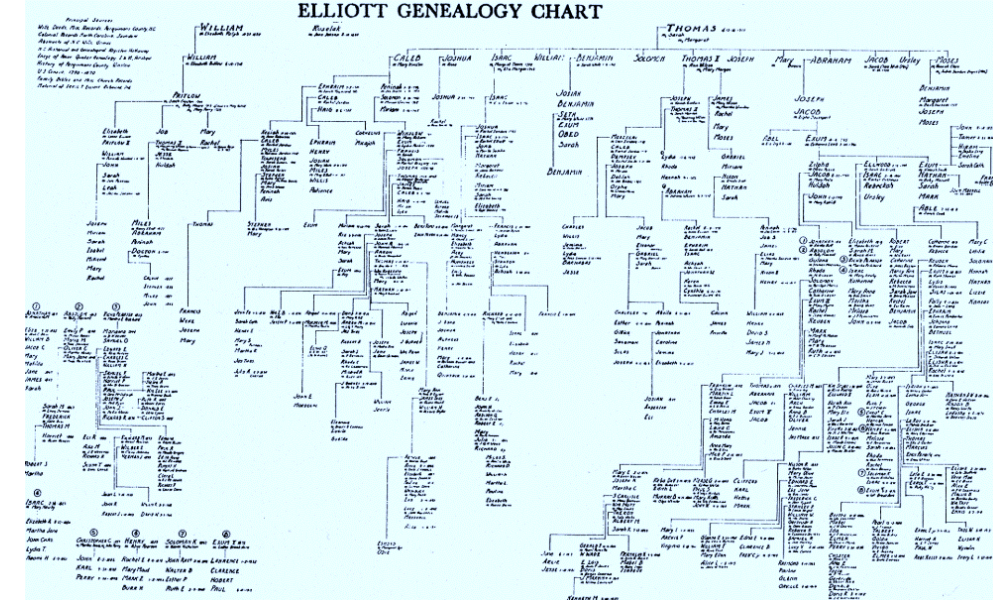
<https://www.w3.org/TR/prov-primer/>

[W3C = World Wide Web Consortium](#)

# 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



***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



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 **EN ISO 19115-1**  
NORME EUROPÉENNE  
EUROPÄISCHE NORM

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.



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

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

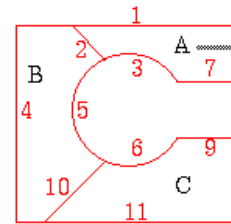
---

© 2014 CEN All rights of exploitation in any form and by any means reserved worldwide for CEN national Members. Ref. No. EN ISO 19115-1:2014 E

# Geographic Information System (GIS)

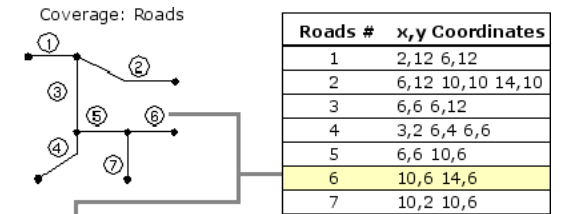
- Provides similar data import, query, manipulation, analysis (e.g. statistics), reformat, display/visualization, output and report capabilities as other information systems

- Also organize their data in
  - Data base management systems
  - File systems



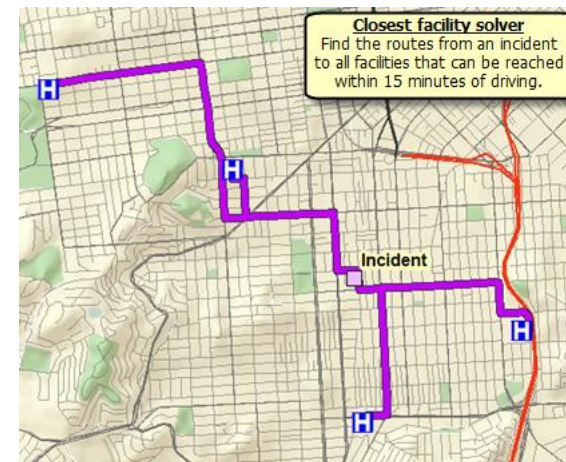
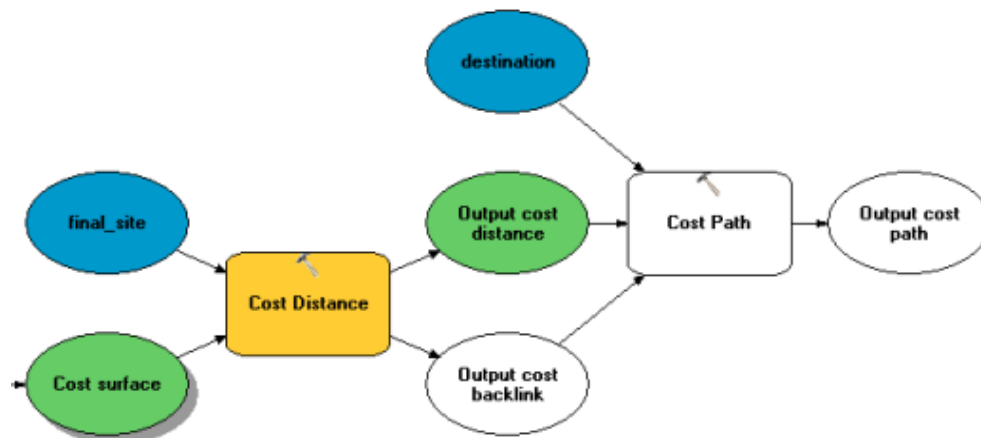
Polygon Attribute Table

Polygon	Area	Parcel Number	Land Use
A	12,001	11-115-001	R 1
B	15,775	11-115-002	R 1
C	19,136	11-115-003	R 3



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







National Spatial Data Infrastructure

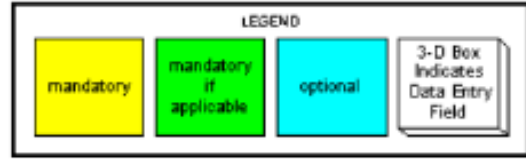
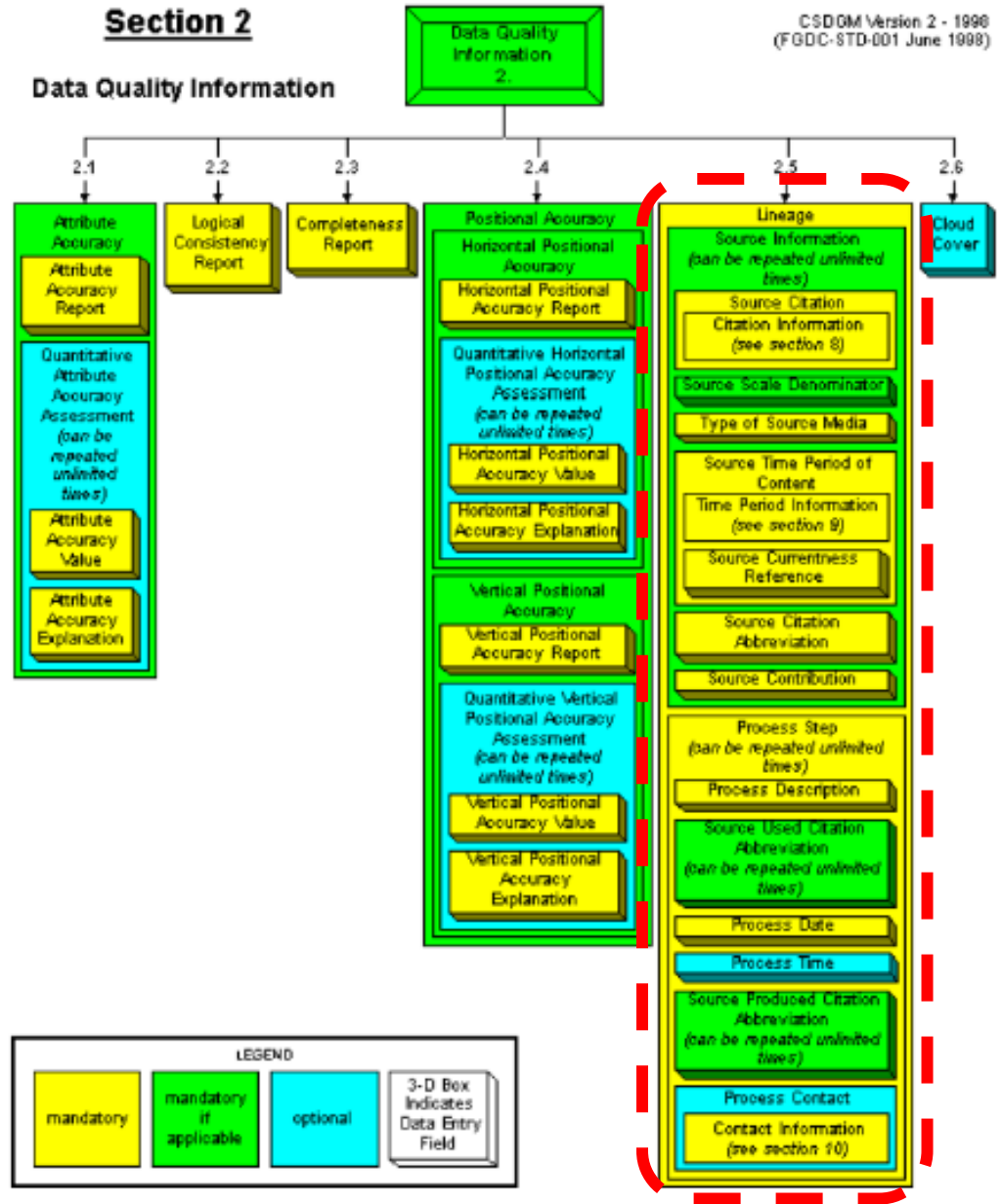
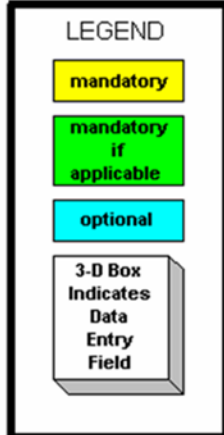
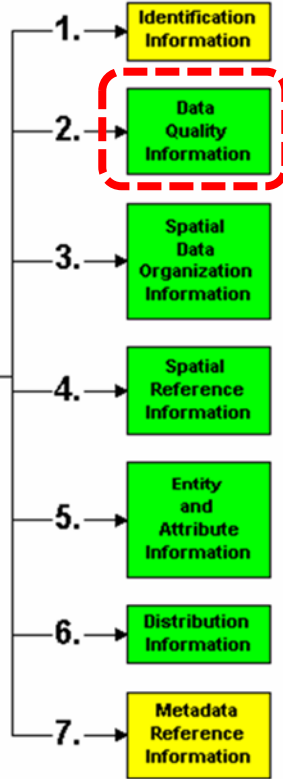
FGDC-STD-001-1998

# 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



# 1<sup>st</sup> application 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


*John B. Jones*  
Committee Member

*John K. Jensen*  
Committee Member

*Robert J. Lawrence*  
Committee Member

*David P. Lanter*  
Chairman, Examining Committee  
Major Professor

*George M. Paciona*  
Dean of the Graduate School



US005193185A

**United States Patent** [19] [11] **Patent Number:** **5,193,185**  
Lanter [45] **Date of Patent:** **Mar. 9, 1993**

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

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

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

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

[51] **Int. Cl.:** 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

**References Cited**

**U.S. PATENT DOCUMENTS**

4,318,184	3/1982	Millet et al.	364/900
4,370,707	1/1983	Phillips et al.	364/200
4,408,273	10/1983	Plew	364/200
4,479,196	10/1984	Ferrer et al.	364/900
4,558,413	12/1985	Schmidt et al.	364/900
4,611,298	9/1986	Schuhli	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	Fujiwara et al.	364/200

**OTHER PUBLICATIONS**

Allman, Eric. "An Introduction to the Service Code Control System." University of California at Berkeley, pp. 1-14, 1980.

Allder, William R. and Aisp 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 Geographic Database", *Proceedings of the Sixth International Symposium on Automated Cartography*, pp. 372-382.

Buchanan, Bruce G. and E. H. Shortliffe, 1985, *Rule-Based Expert Systems*, Addison-Wesley Publishing Company, Reading, Mass.

Charniak, Eugene et al., 1987, *Artificial Intelligence Programming*, 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.

Cohen, David J. 1988, "GIS vs. CAS vs. DBMS: What are the Differences?", *Photogrammetric Engineering and Remote Sensing*, vol. 54, No. 11, pp. 1551-1555.

Deuser, 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. Kriesis  
**Attorney, Agent, or Firm**—Jon L. Roberts

**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

```

graph TD
    250[USER INPUT] --> 251[LTP]
    251 --> 252[LINEAGE META-DATABASE]
    251 --> 253[GIS]
    253 --> 254[DATA LAYERS]
    
```

# 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

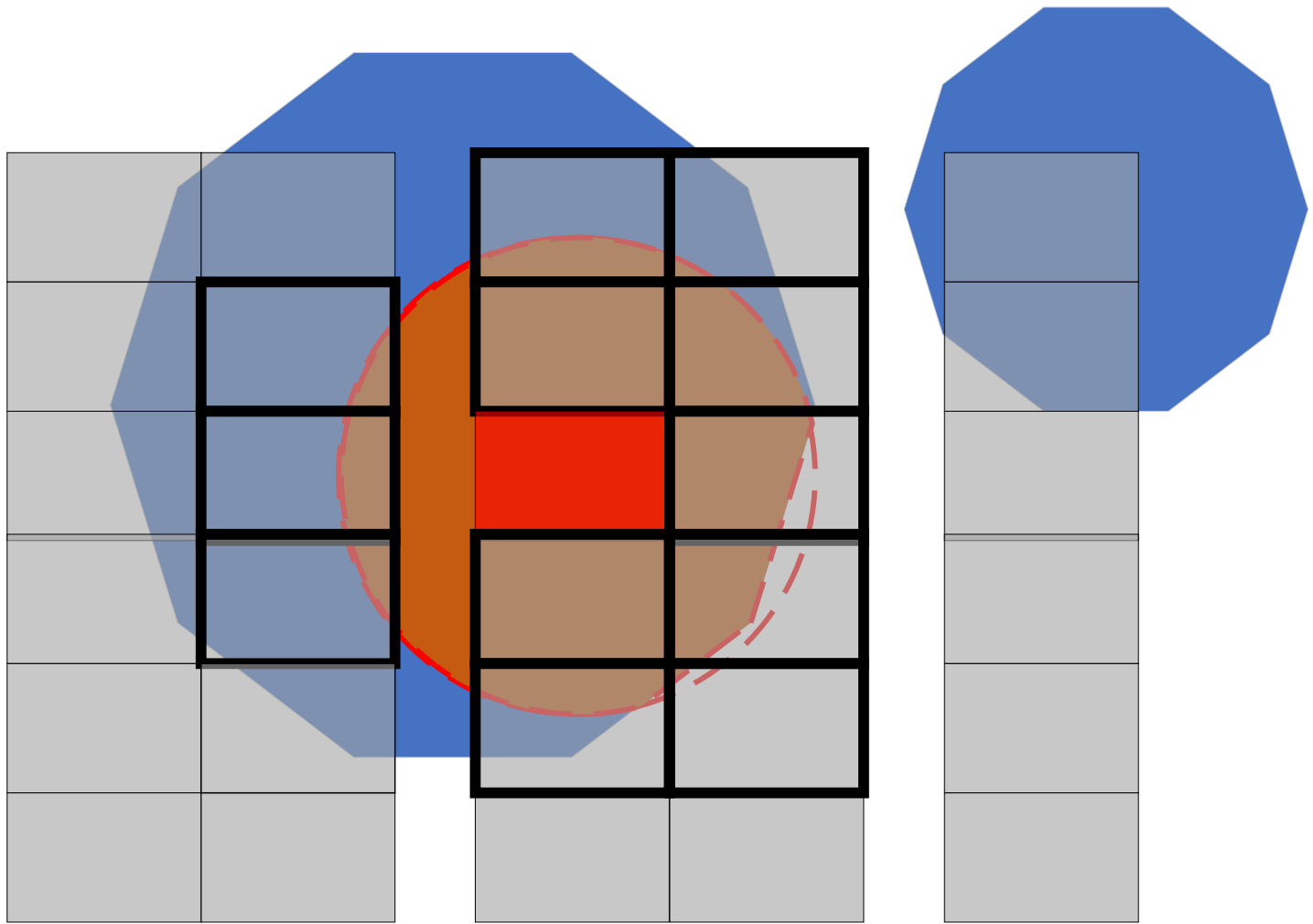
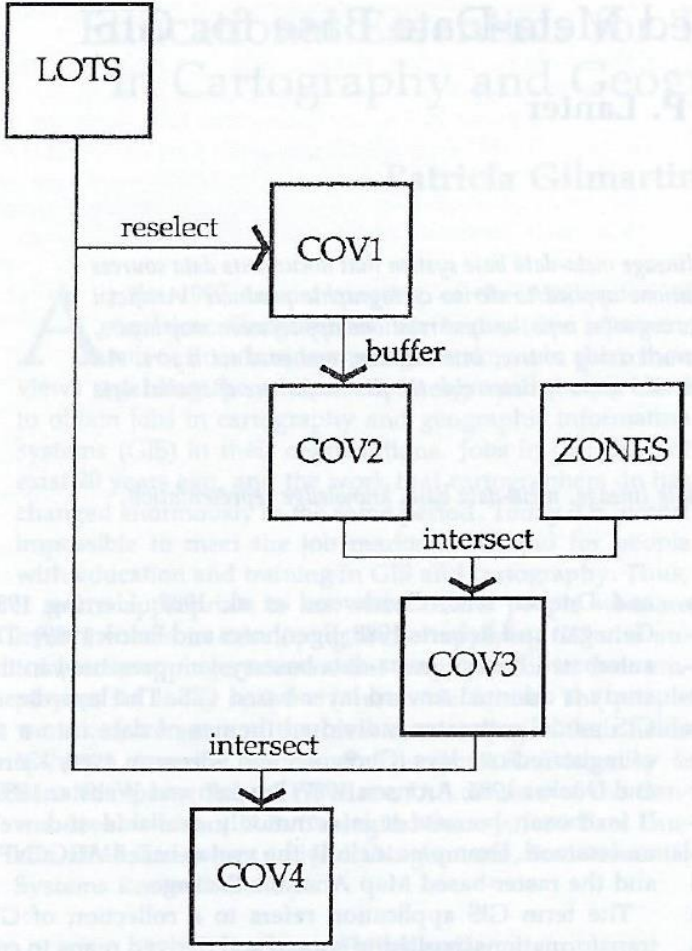
Metadata management system  
for ARC/INFO and GRID™

Version 3.0

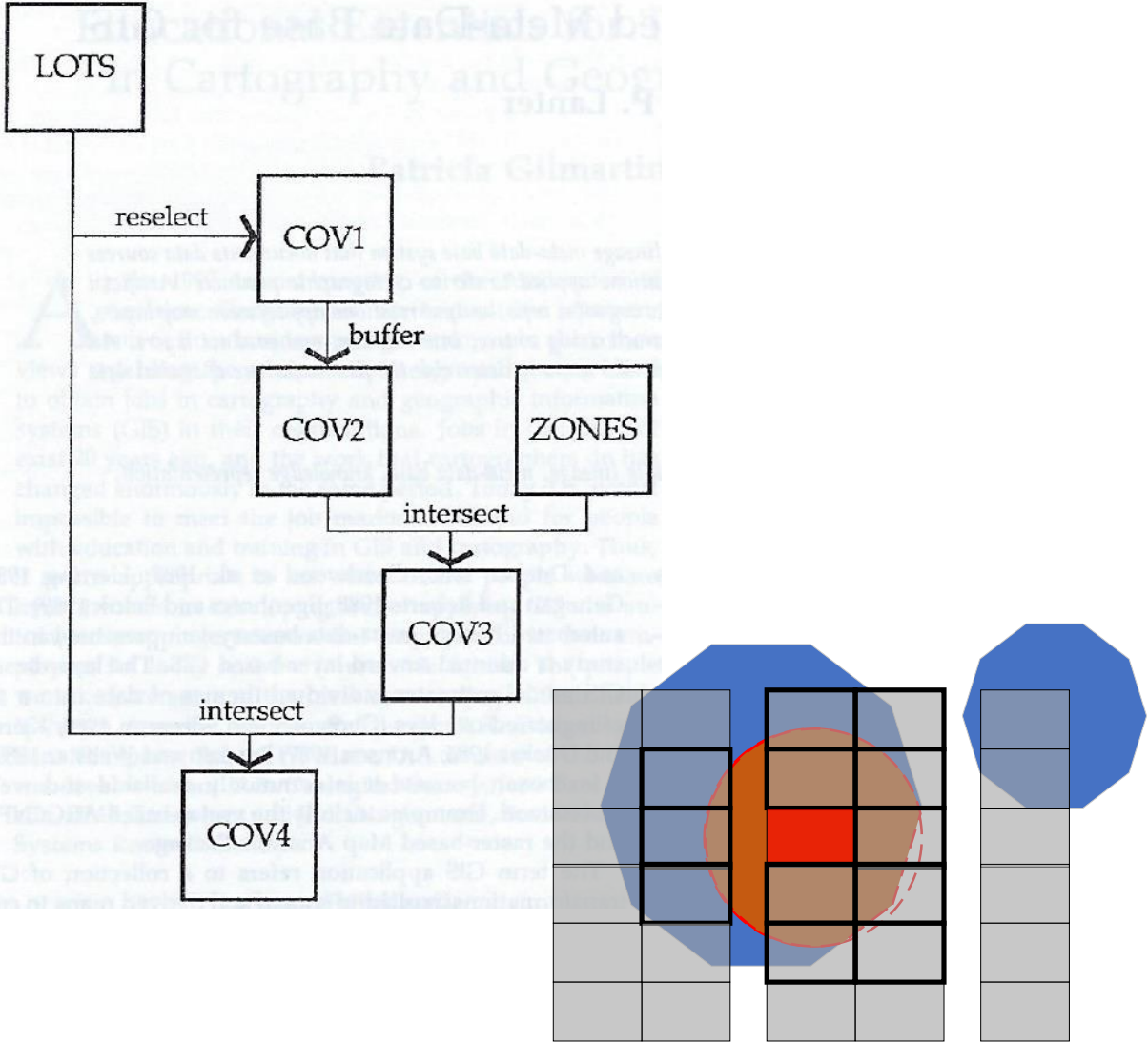
## User guide

Geographic Designs Inc.

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



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



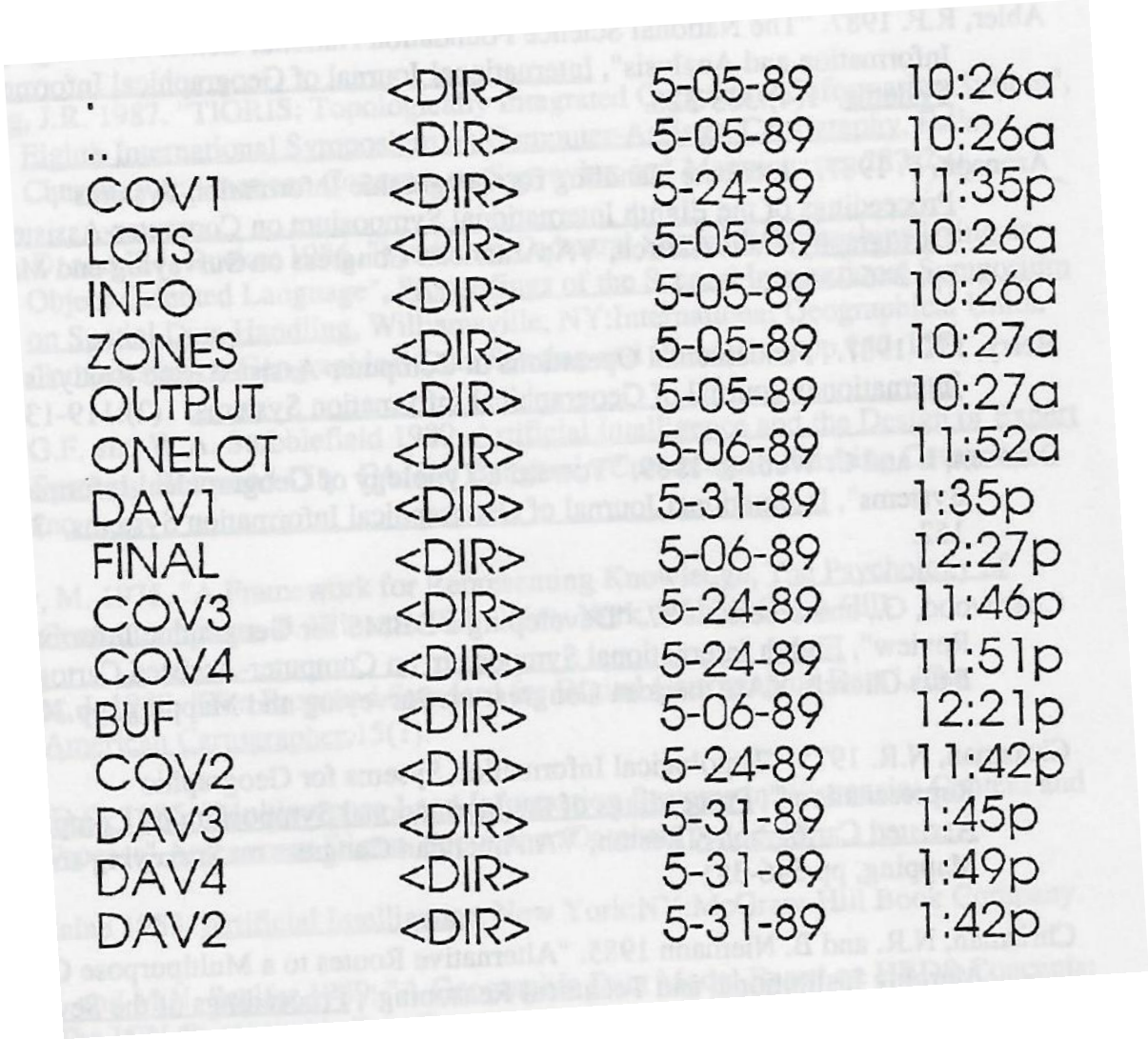
Datasets stored on the computer after data processing analysis was over...

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

# Datasets stored on the computer after data processing analysis was over...

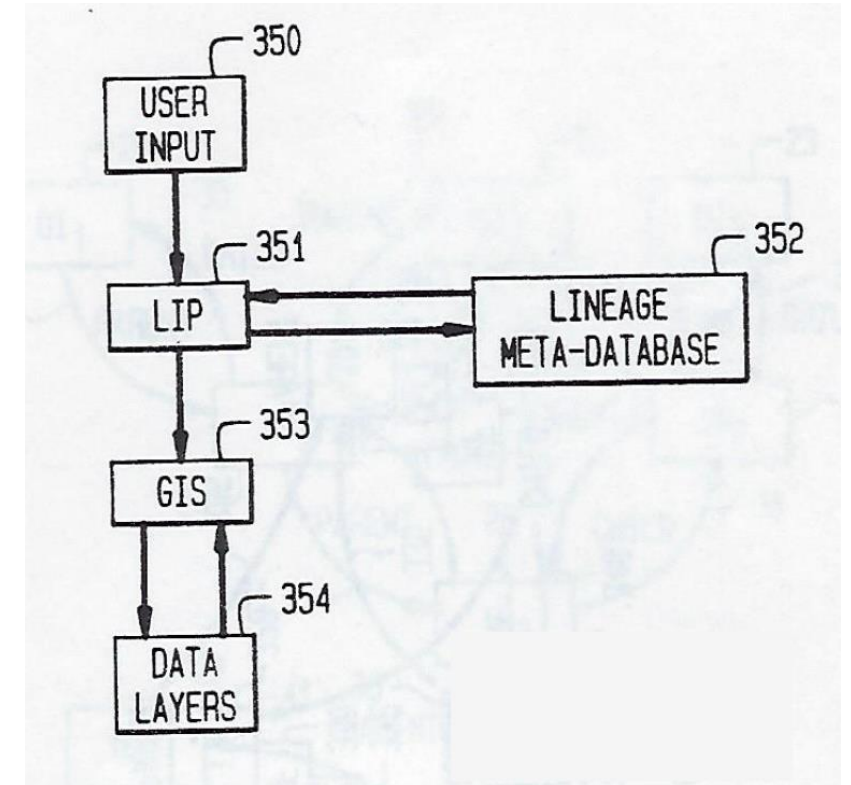
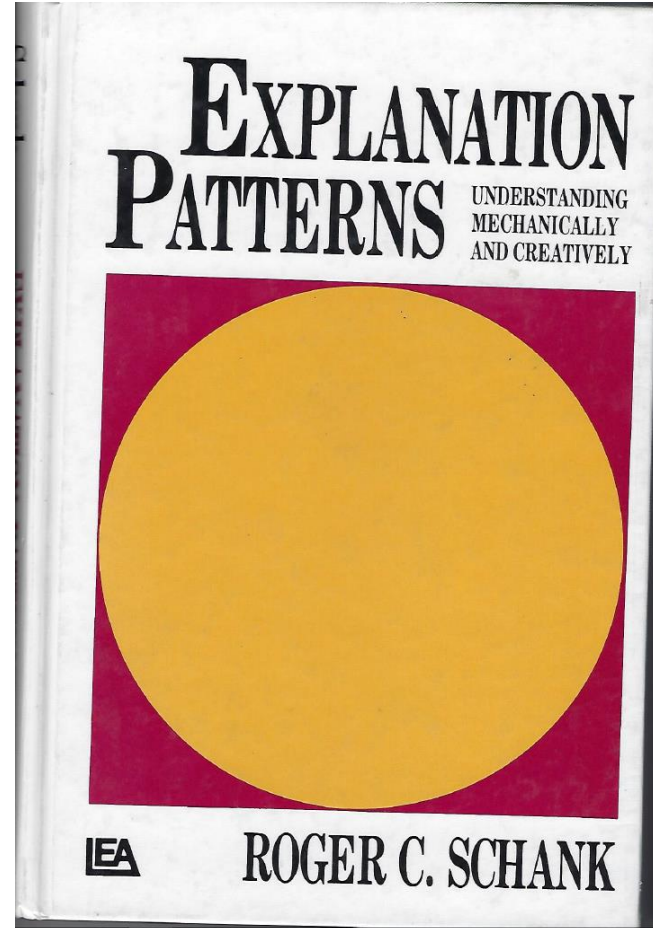
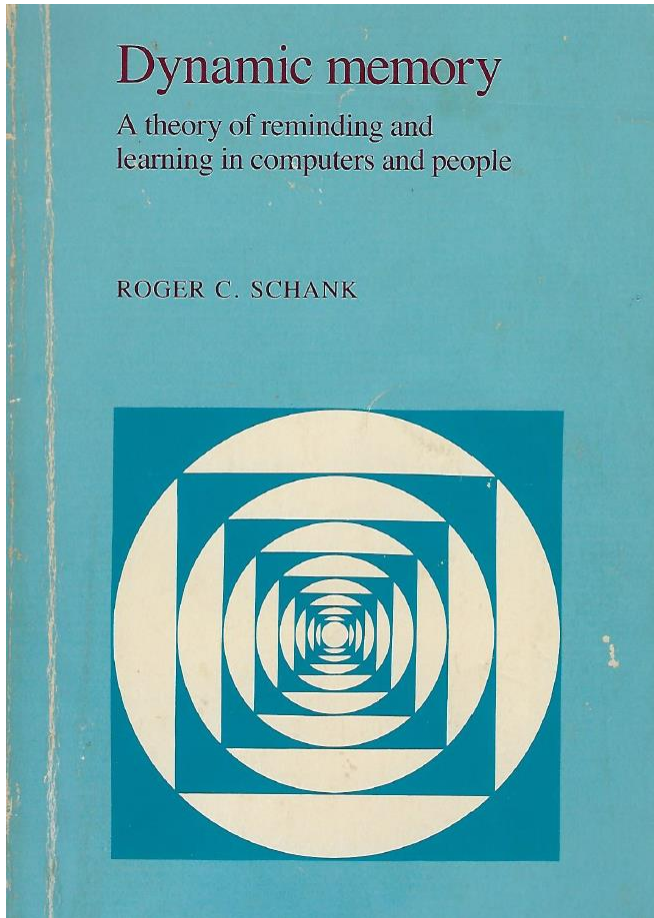
Datasets organized as files in folders



.	<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

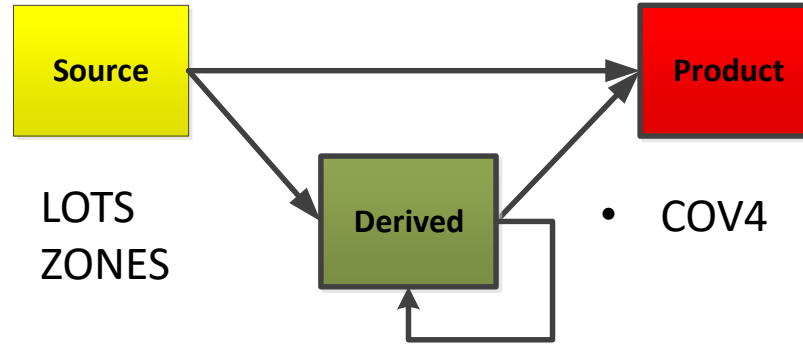
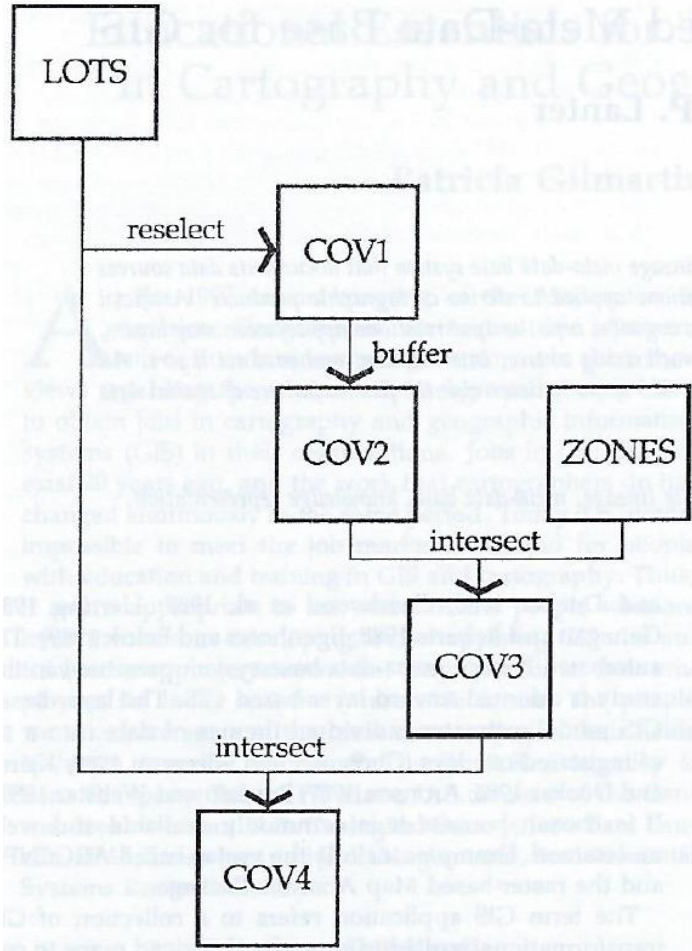
*After the work on the analysis was completed, it was difficult to remember which dataset was goal (i.e. end result)*

# *How can I program the computer to help me remember what I knew about the data I loaded and processed on my computer?*



*LIP = Lineage Information Processor*

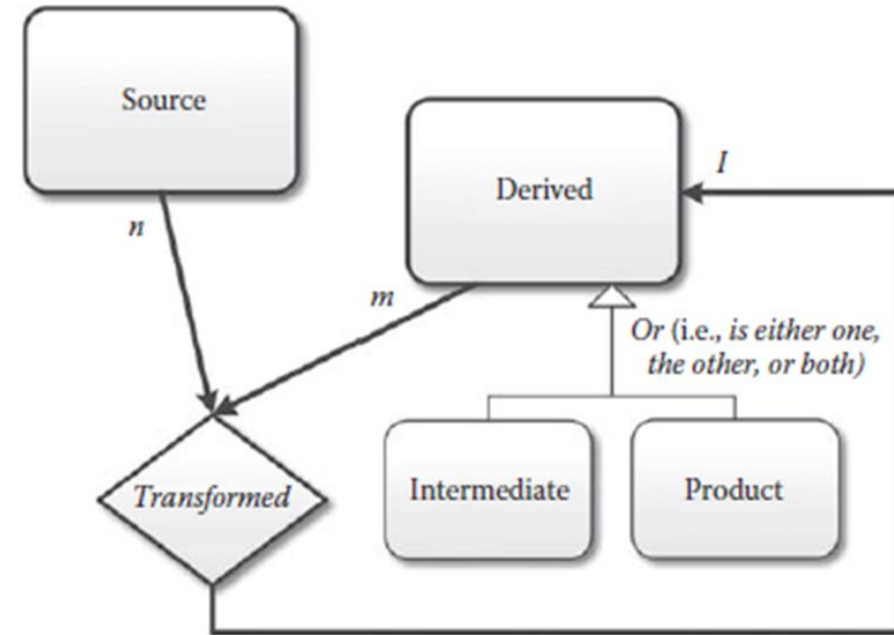
# How do we understand differences among datasets created during processing applications?



- LOTS
- ZONES

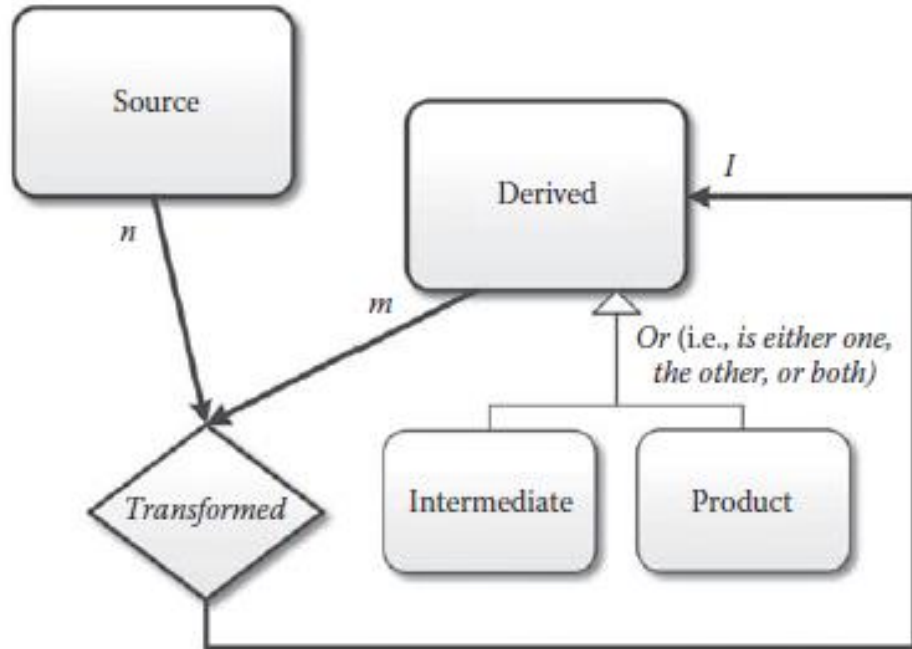
- COV4

- COV1
- COV2
- COV3



Data lineage vocabulary helps communicate how data is processed in an information system

***and can aid thinking about how to meet privacy by design requirements***



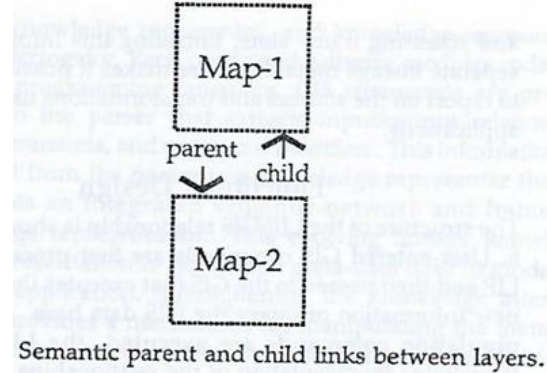
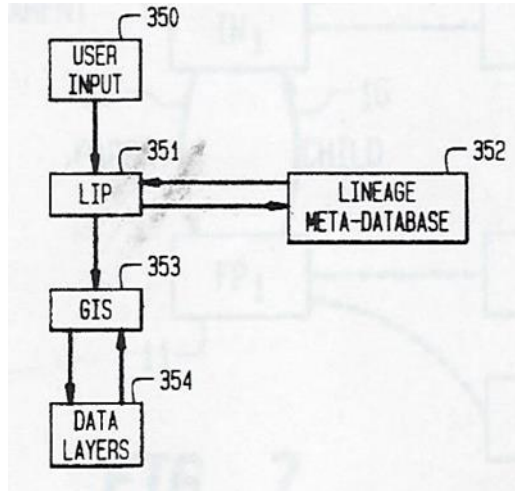
**Source datasets** *may contain personal data*

Derived datasets inherit this personal data from their input

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

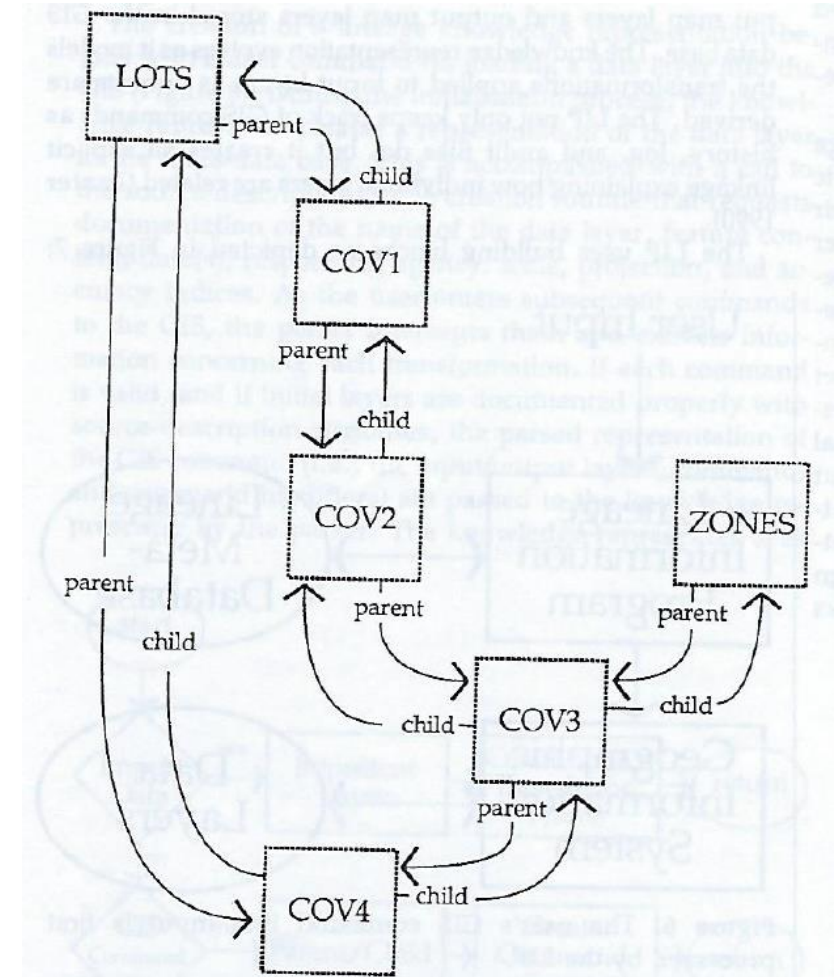


# Semantic “parent” & “child” metadata links added to enable deductions about relationships among input & output datasets...



**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?***



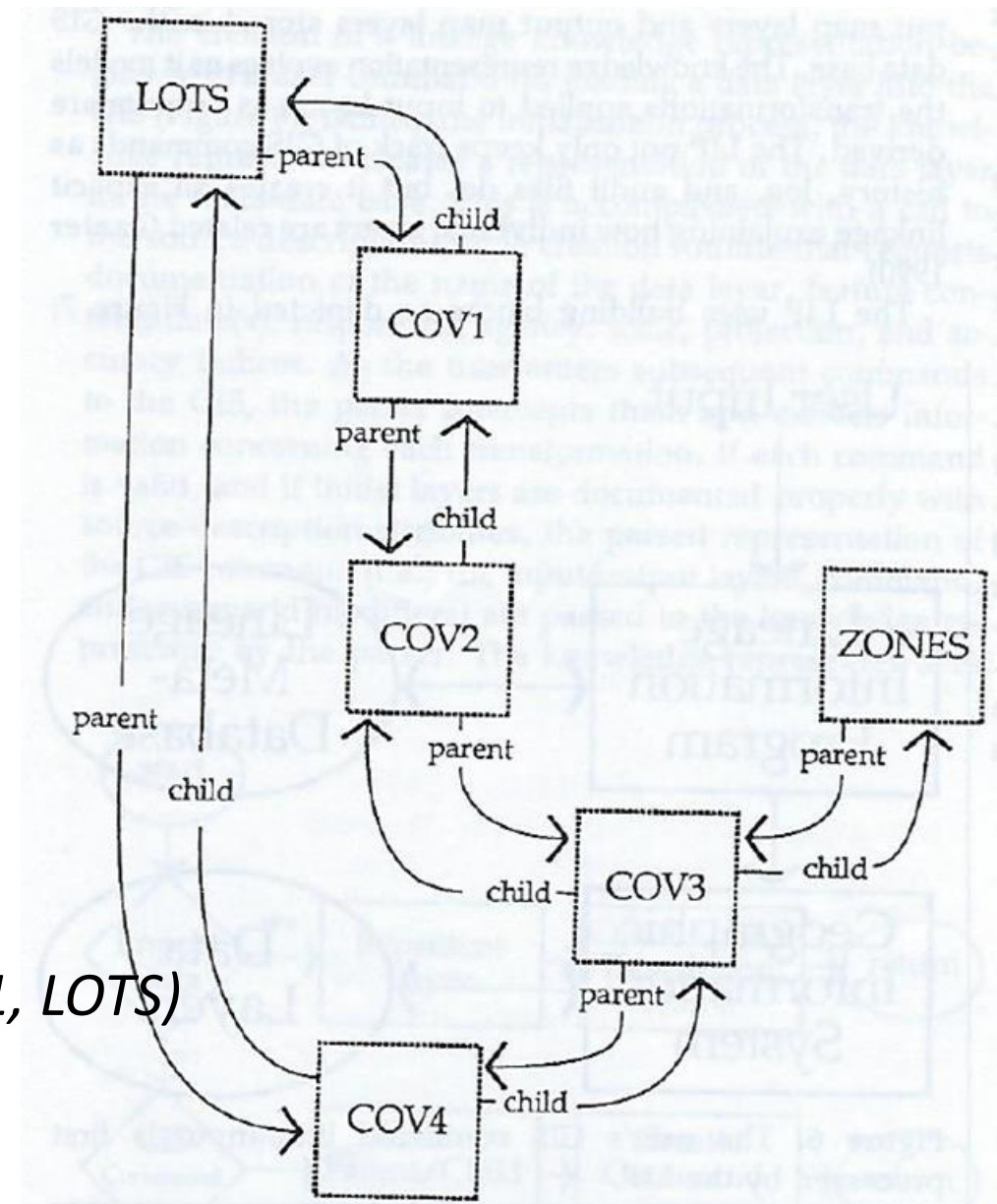
**Descendants** function traces parent links to identify all datasets derived from a source or other derived input dataset used within the application.

```
(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)*



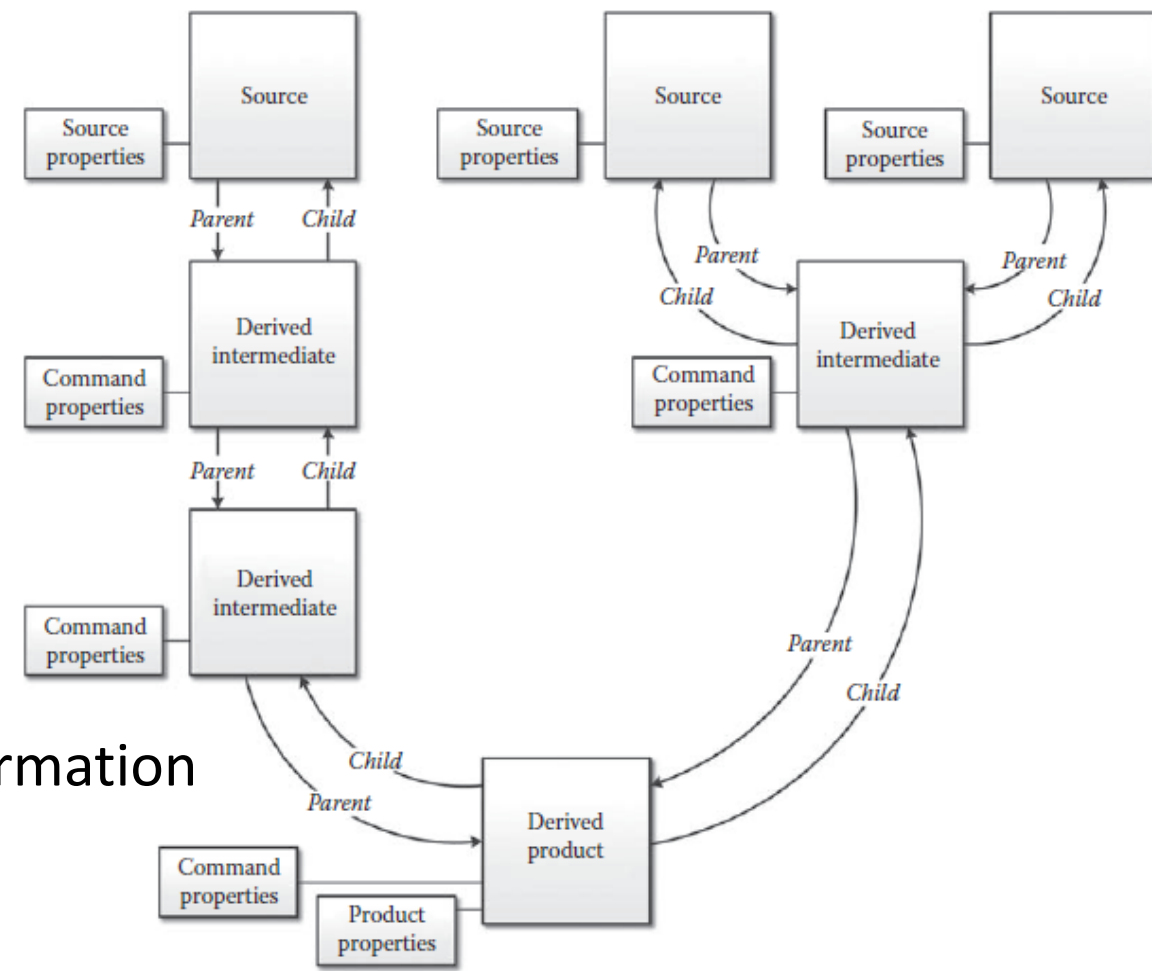
## Source properties can include:

- Originating organization
- Data content (i.e. entity and attribute definitions)
- Timeliness (e.g. when collected, when acquired,...)
- Accuracy
- Confidentiality security categorization of attributes
  - Privacy sensitivity of attributes
- Integrity categorization of attributes...
- Availability categorization...

## 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
```

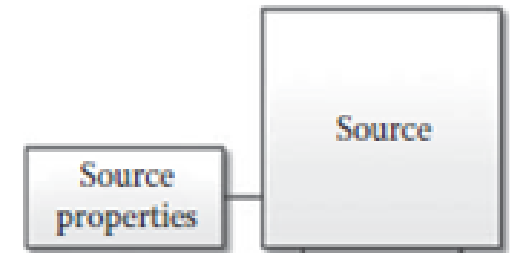
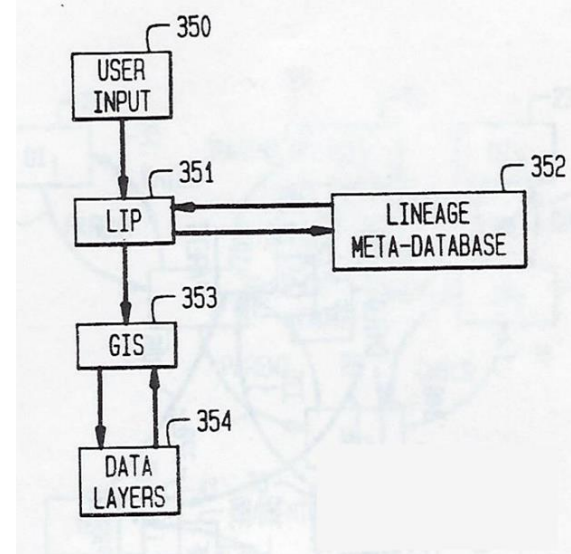
```
What is the source agency? USGS
```

```
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)

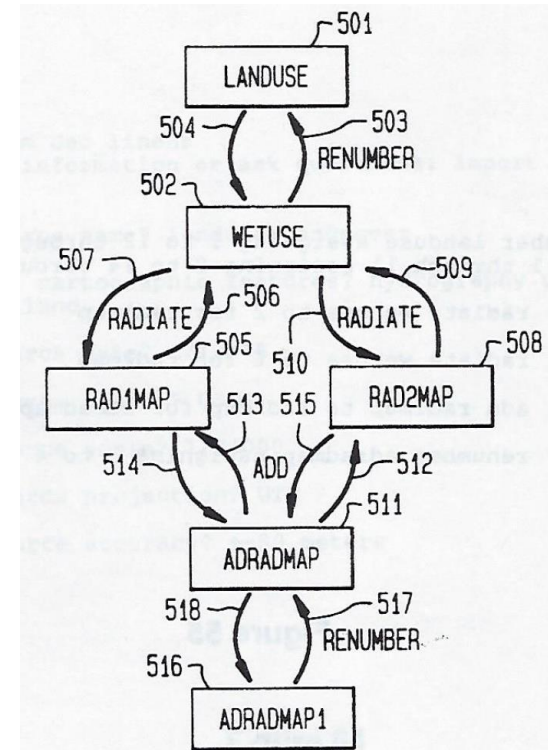
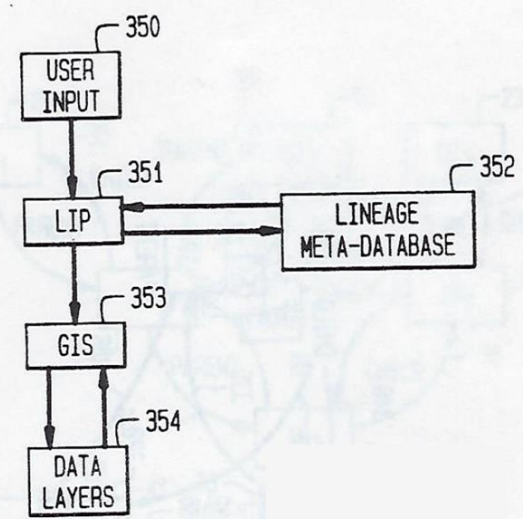
(I AM GEO\_LINEUS)

(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)

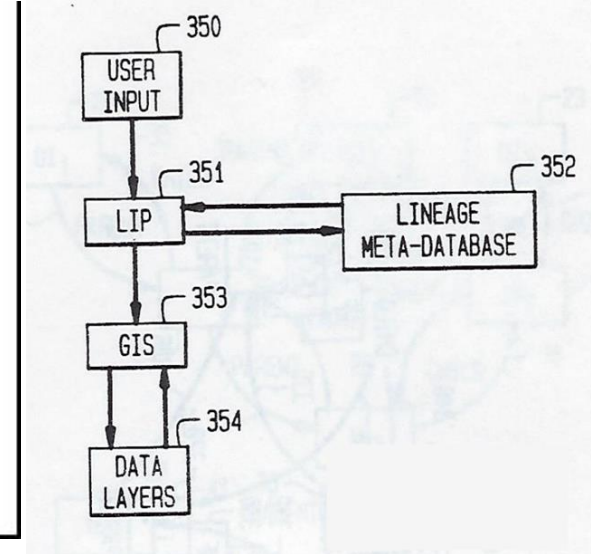
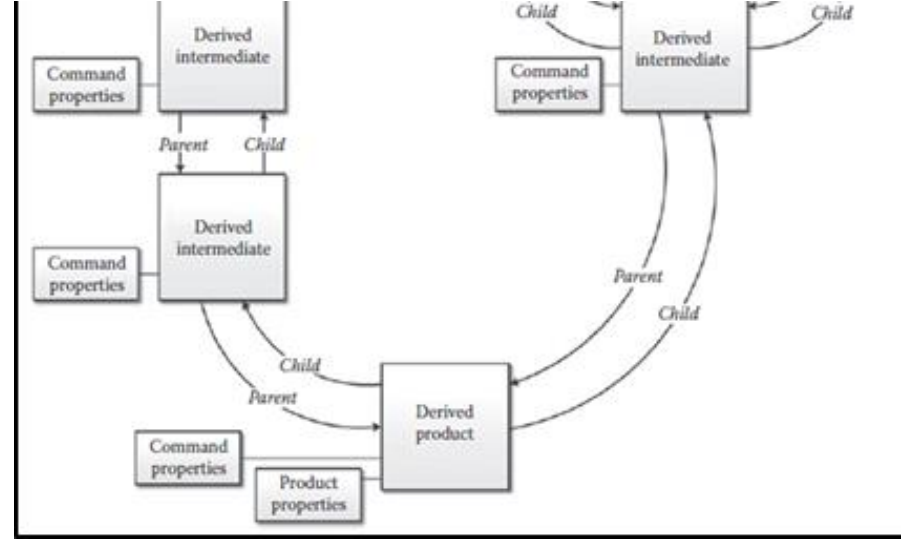
(I UNDERSTAND) (radiate wetuse to 2 for rad1map)

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

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



# Product Metadata input...



```
export cover adradmap1 eco_zones
```

What is the product's name? eco\_zones

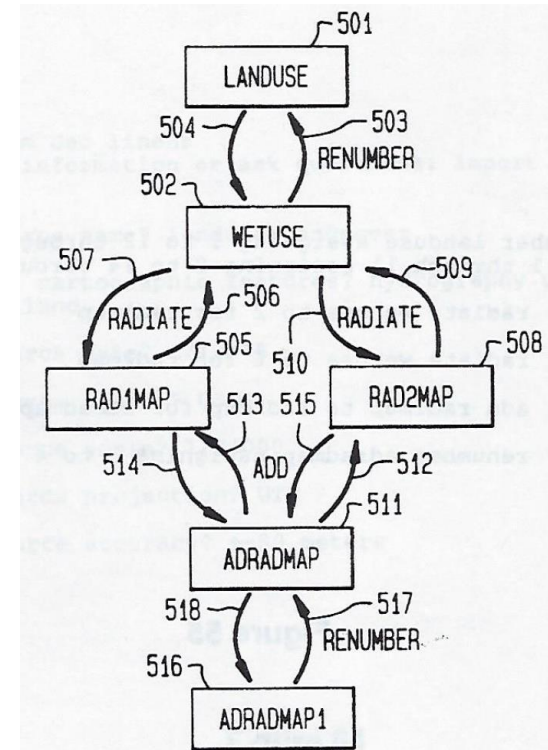
What is the product's use? Environmental protection of wetlands

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

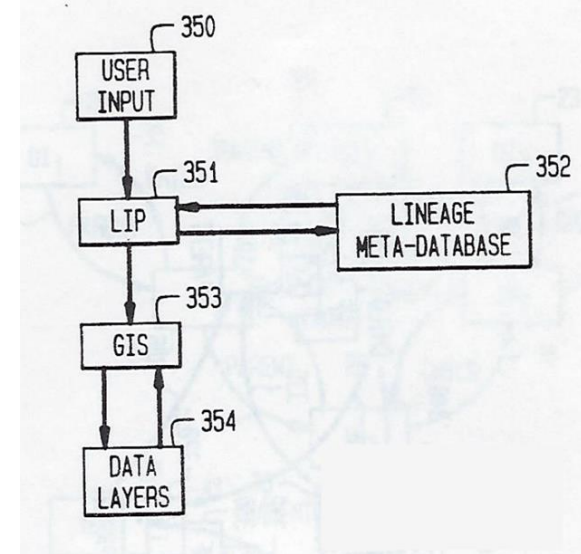
Who is responsible for the product? Diego Essinger

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

Thank You!

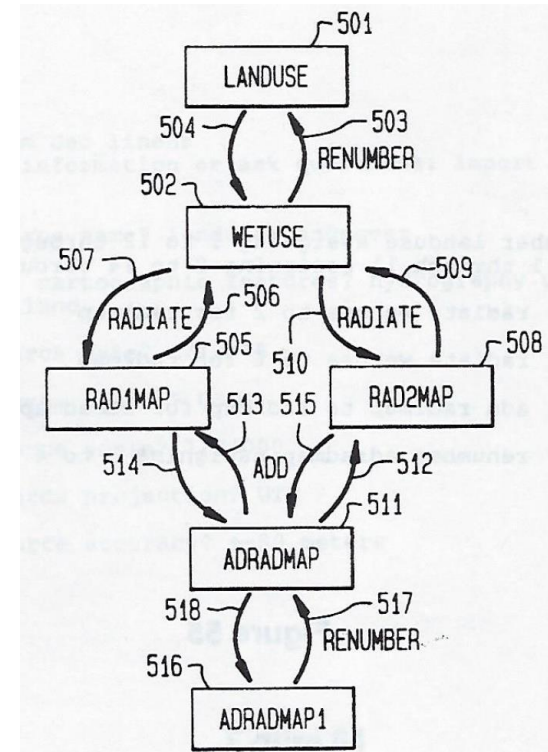


# Querying metadata...

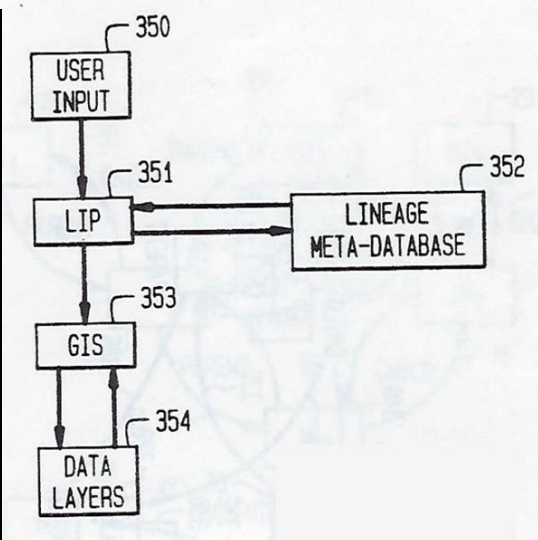
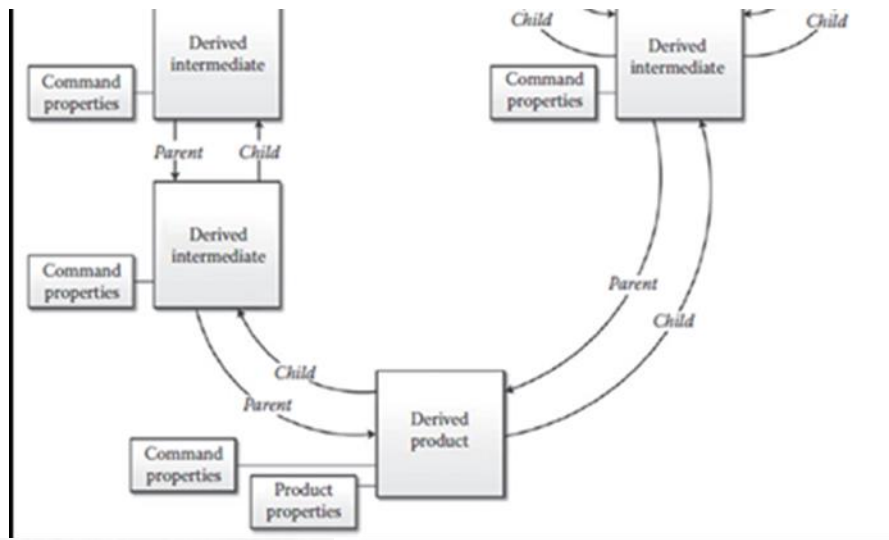


Is landuse a parent of adradmap

(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)

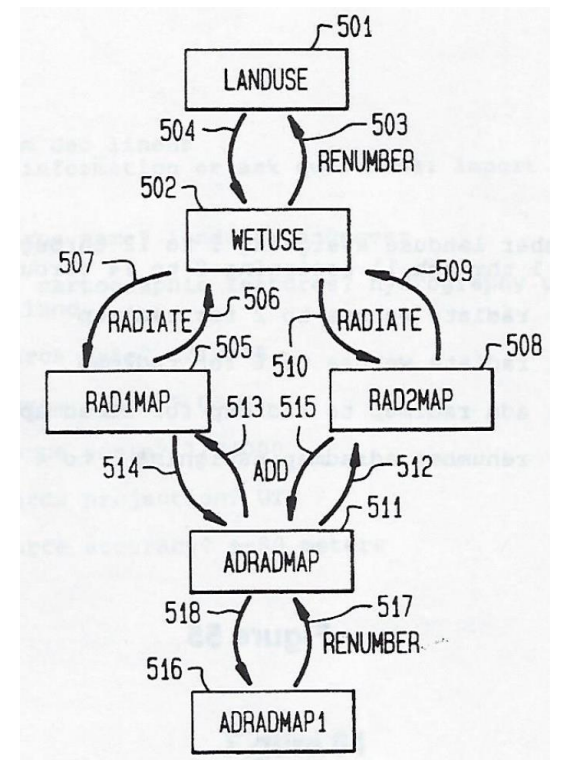
(LANDUSE IS AN ORIGINAL MAP LAYER)

(INPUT TO RAD1MAP IS WETUSE COMMAND IS RADIATE)

(INPUT TO WETUSE IS LANDUSE COMMAND IS RENUMBER)

(LANDUSE IS AN ORIGINAL MAP LAYER)

+

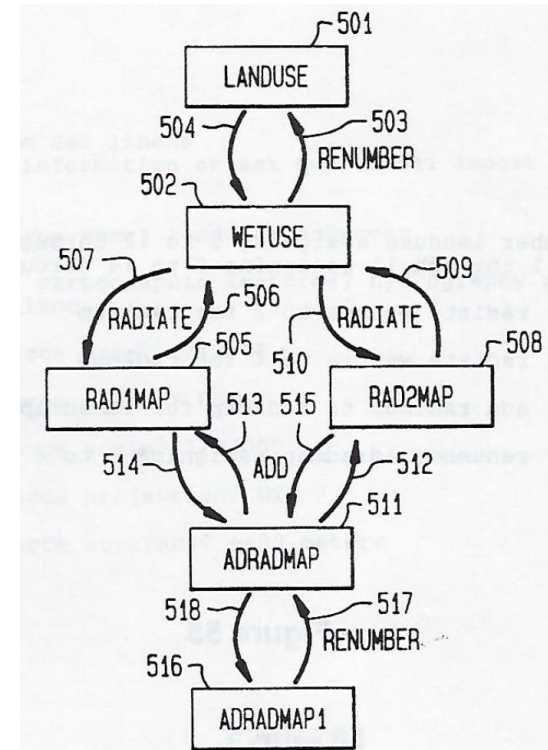
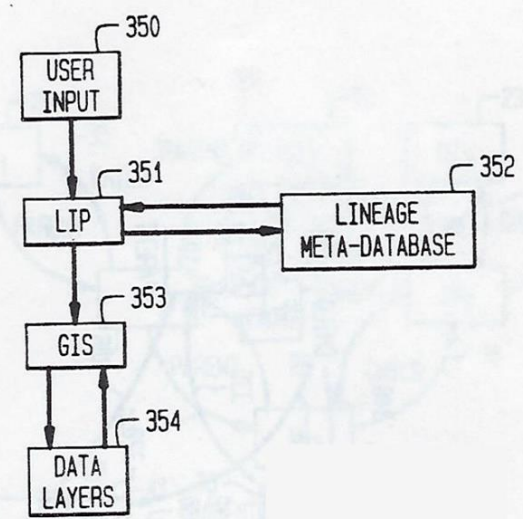




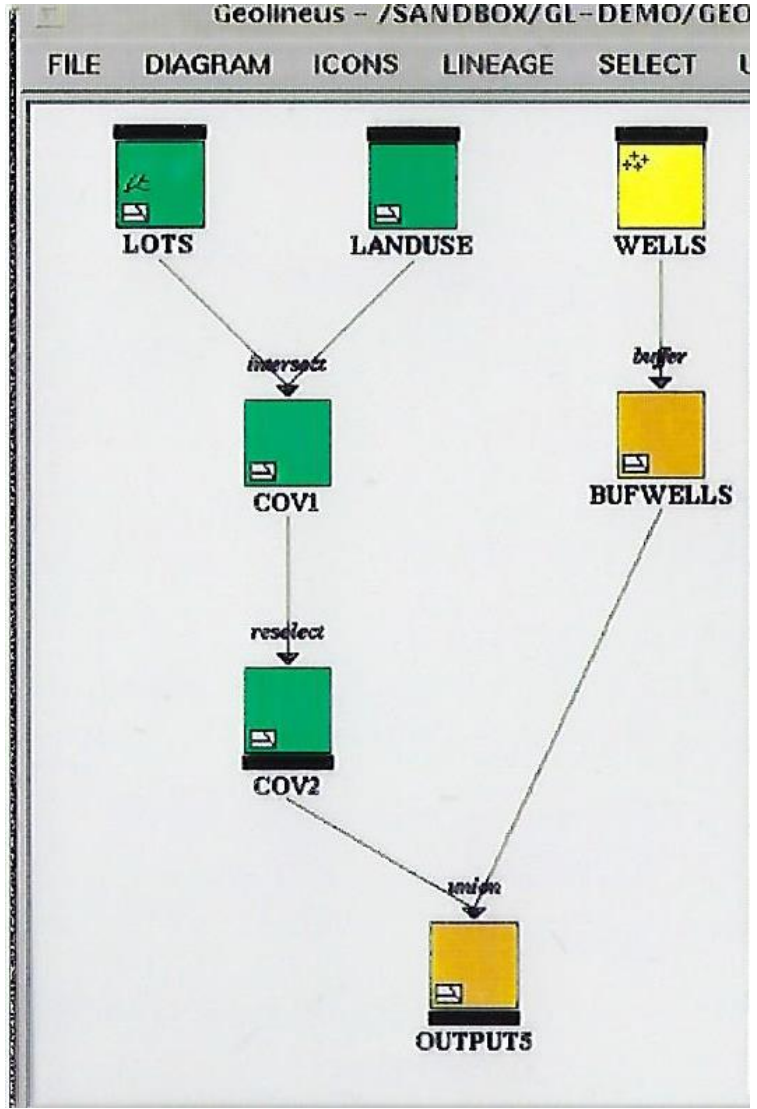
# Querying metadata...

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)



# Adding a graphical user interface...



Help on icons

	Source layer. A basic data layer in the GIS.		GRID scalar variable.
	Derived layer. Layer was created as a result of an ARC/INFO command like BUFFER, INTERSECT or GRIDPOLY.		Coverage has been edited in ARCEDIT since the last CLEAN and BUILD.
	Product layer. A derived layer that represents the final step in a GIS application. To turn a derived layer into a product, choose 'Make product' from the 'Icons' menu.		Coverage has been edited in ARCEDIT since the last CLEAN and BUILD and polygon topology needs rebuilding.
	Coverage containing point features. It has a point attribute table (PAT).		Coverage in which arc features have been rebuilt but polygon topology still needs rebuilding.
	Coverage containing arc features. It has an arc attribute table (AAT).		Layer that is now out-of-date because one or more of its sources has changed. Out-of-date status is only displayed if the 'Out-of-date' option in the 'Diagram' menu is turned on.
	Coverage containing polygon features. It has a polygon attribute table (PAT).		Derived layer with incomplete command frame. Icon was added to diagram by the 'Create from log' option from the 'File' menu and represents the result of a command, such as RESELECT or ELIMINATE. The subcommands of which cannot be extracted from the log
	Coverage with both a point attribute table and an arc attribute table.		A 'dimmed' layer. This layer no longer exists. It has either been KILLED, or moves to a new location. Dimmed derived layers are recreated with the 'Recreate' option from the 'Update' menu.
	Coverage with both an arc attribute table and a polygon attribute table.		A dimmed GRID scalar. Icon was added to diagram with the 'Create from log' option so value is unknown
	Grid with integer cell values.		
	Grid with integer cell values, and a value attribute table (VAT)		
	Grid with floating point cell values.		

GUI design by Rupert Essinger

OK

# Working with source and command metadata

The screenshot shows the Geolineus interface with a workflow diagram on the left. The workflow starts with three source layers: LOTS, LANDUSE, and WELLS. LOTS and LANDUSE are combined via an 'intersect' operation to create COV1. COV1 is then processed by a 'reselect' operation to create COV2. Finally, COV2 and WELLS are combined via a 'union' operation to create OUTPUT. A 'Source Frame - LOTS' dialog box is open, displaying metadata for the LOTS layer. A yellow callout box points to the 'DESCRIPTION' field.

**Source Frame - LOTS**

NAME: LOTS

DESCRIPTION

DATA QUALITY

SPATIAL EXTENT

MAP PROJECTION

DATUM

STATUS

POINT/VECTOR OBJECTS

CONTACT

ENTITY ATTRIBUTES

DESCRIPTION: This coverage contains attributes for both the land parcel polygons and the boundary lines between them. We ran BUILD twice, first with the LINE option, and

DATE: Fri 1-Apr-1994 14:00

DATE: Thu 15-Dec-1994 14:21

Buttons: OK, Import..., Cancel

**This is where CIA source metadata would be added...**

The screenshot shows the Geolineus interface with a workflow diagram on the left. The workflow starts with three source layers: LOTS, LANDUSE, and WELLS. LOTS and LANDUSE are combined via an 'intersect' operation to create COV1. COV1 is then processed by a 'reselect' operation to create COV2. WELLS is processed by a 'buffer' operation to create BUFWELLS. Finally, COV2 and BUFWELLS are combined via a 'union' operation to create OUTPUTS. A 'Command Frame - BUFWELLS' dialog box is open, displaying metadata for the BUFWELLS layer. A yellow callout box points to the 'COMMAND' field, and a red circle highlights the 'Ripple...' button.

**Command Frame - BUFWELLS**

COMMAND: BUFFER

IN\_COVER: WELLS

OUT\_COVER: BUFWELLS

BUFFER\_ITEM: #

BUFFER\_TABLE: #

BUFFER\_DISTANCE: 120

FUZZY\_TOLERANCE: #

FEATURE\_TYPE: POINT

NOTE: This buffer distance may be larger than the distance specified by the client. To change it, edit the distance and then press the Ripple button. This will recreate

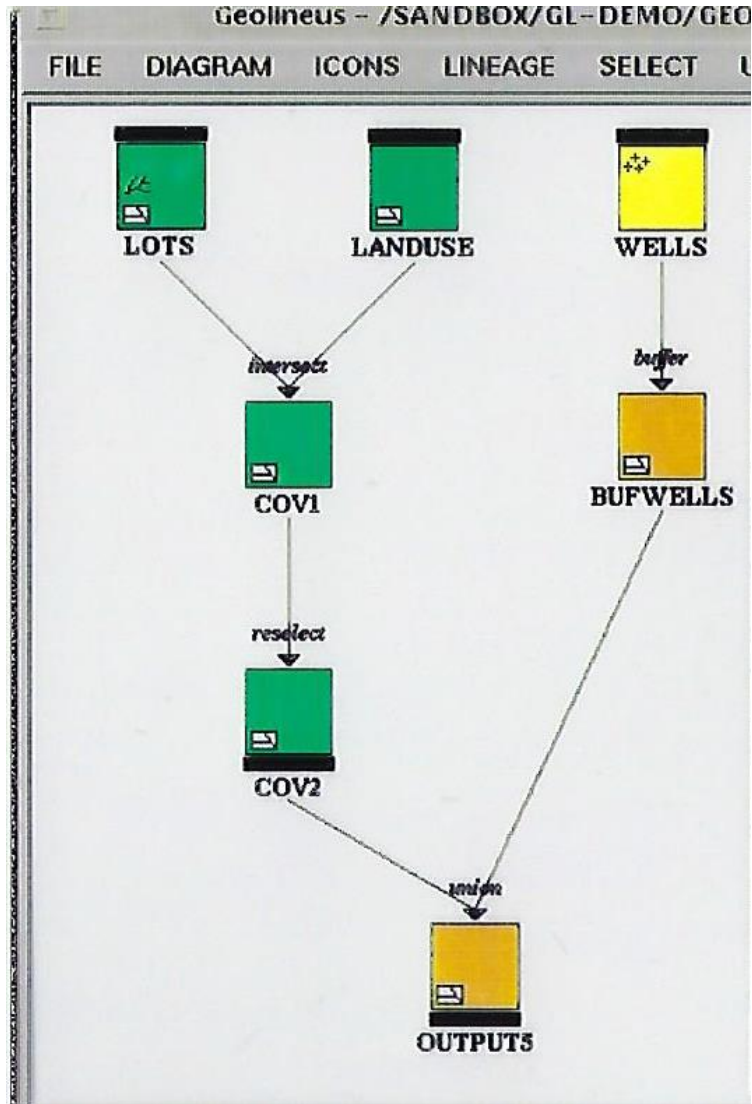
FIRST\_CREATED: Sun 28-Apr-1991 16:33

LAST\_RECREATED: Mon 29-Apr-1996 11:39

Buttons: OK, **Ripple...**, Cancel

**This is where CIA metadata for derived data could be added...**

# Update propagation...



Geolineus - /SANDBOX/GL-DEMO/GEOLINEUS30/DEMO/DEMO3.LNG

FILE DIAGRAM ICONS LINEAGE SELECT UPDATE DELETE HELP

LOTS LANDUSE WELLS

intersect

COV1

buffer

BUFWELLS

resselect

COV2

union

OUTPUTS

Commands to update data

buffer wells bufwells # # 120 # point

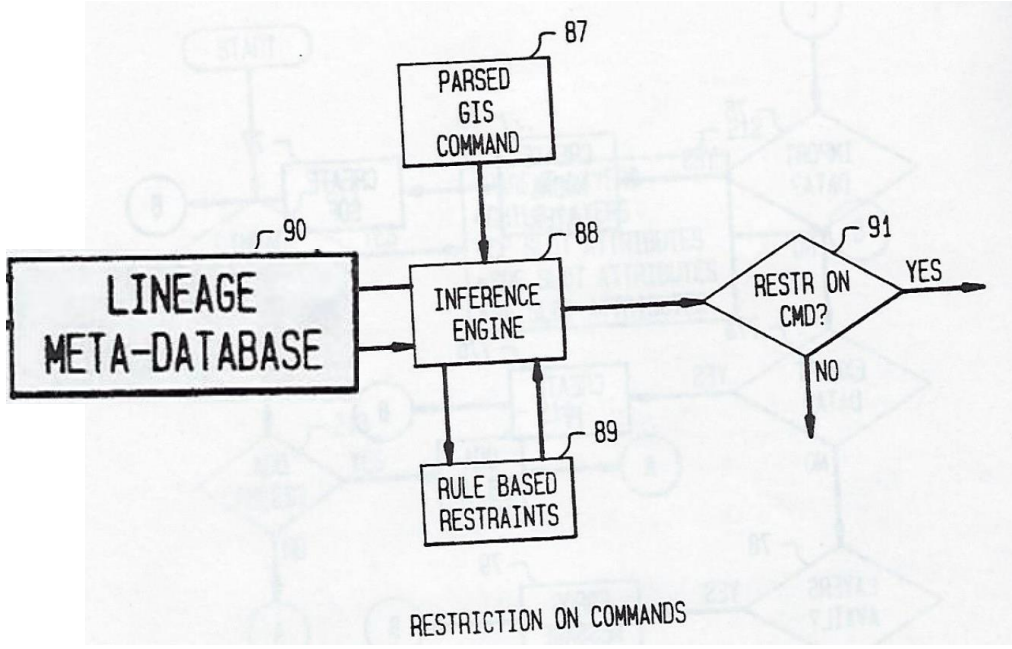
OK

ARC/INFO - Workspace /SANDBOX/GL-DEMO/GEOLINEUS30/DEMO

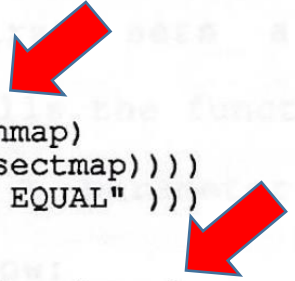
```
Killed bufwells with the ARC option
Arc: buffer wells bufwells # # 120 # point
Buffering ...
Sorting...
Intersecting...
Assembling polygons...
Creating new labels...
Finding inside polygons...
Dissolving...
Creating bufwells.PAT...
Arc: union bufwells cov2 output5
Unioning bufwells with cov2 to create output5
Sorting...
Intersecting...
```

Detailed description: This screenshot shows the Geolineus workspace with the same data flow as the diagram. A dialog box titled 'Commands to update data' is open, displaying the command 'buffer wells bufwells # # 120 # point' and an 'OK' button. Below the workspace, the ARC/INFO command window shows the execution of these commands, including 'Killed bufwells with the ARC option', 'Arc: buffer wells bufwells # # 120 # point', 'Buffering ...', 'Sorting...', 'Intersecting...', 'Assembling polygons...', 'Creating new labels...', 'Finding inside polygons...', 'Dissolving...', 'Creating bufwells.PAT...', 'Arc: union bufwells cov2 output5', 'Unioning bufwells with cov2 to create output5', 'Sorting...', and 'Intersecting...'.

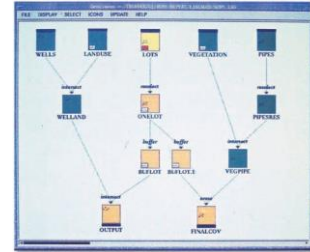
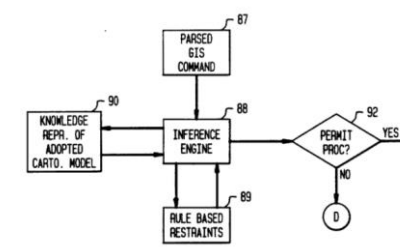
# Data source metadata based integrity constraint



```
(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.")) )))
```



# Conclusion:



Data lineage metadata can help information system developers meet key data protection by design 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**

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

The screenshot displays the Oracle Enterprise Metadata Management 12c interface. On the left, the 'Repository' pane shows a tree structure with 'Sales Reporting Application - 1' and 'Dimensional DW'. The main workspace shows a 'Data Flow Overview' for 'LD\_ORDERS'. It illustrates a data lineage path: source tables (SRC ORDERS, SRC CUSTO..., SRC ORDER..., SRC PRODU...) feed into an 'AGGREGATE' table with columns like AMOUNT, ORDER\_DATE, ORDER\_ID, PRODUCT\_ID, QTY, and JOIN. This aggregate table then feeds into a 'Presentation Layer.A...' table named '1- Revenue' with columns 'Base Facts', 'Products', and 'Time'. A context menu is open over the '1- Revenue' table, with 'Trace Lineage' selected. The menu options include 'Trace Data Lineage', 'Trace Data Impact', 'Trace Semantic Definition', 'Trace Lineage (Advanced)', 'Show in Metadata Browser', 'Trace Lineage', 'Highlight Path', 'Expand this Node Completely', and 'Collapse this Node Completely'. An inset image in the bottom right corner shows the cover of an Oracle White Paper titled 'Addressing GDPR Compliance Using Oracle Data Integration and Data Governance Solutions'.

Addressing GDPR Compliance Using Oracle Data Integration and Data Governance Solutions

ORACLE WHITE PAPER | DECEMBER 2017

# Agenda

- ✓ Online privacy
- ✓ Privacy and data protection by design
- ✓ *...with data provenance and lineage metadata*