A Comprehensive Integration and Framework Strategy for Earth Observation Data

H. Roice Nelson, Jr.
Geophysical Development Corporation
11 March 2005

Introduction

Earth Observations are being redefined through digital remote sensing and inexpensive sensors which provide semi-automated means to acquire, capture, store, process, interpret, document, and retrieve data and information. There are more data available today than ever before. Data access, usage, and distribution issues revolve around the variety of formats and the lack of simple indices so relevant data can be easily accessed.

The initiative outlined in this memo requires endorsement and participation by government, private industry, and academic entities to implement national indices for earth observation data which will enable identification and retrieval of relevant legacy data, automatic indexing of new data, easy identification of data gaps and where new data collection needs to focus, and improved communication between peers, starting with easy identification of those with common data interests via the proposed indices.

Background

Johann Gutenberg changed the world with the printing press in 1450. However, the printing press would never have made the change without the simple enhancement we know as the index. For many years after the first printing press, books started to stack up and people could not find relevant information in them. It wasn't until the index was invented that books found widespread usage. Similarly satellite earth observations have been made for decades, and yet this data is not having the impact it could have because people who could use the data can not easily get access to it. They can not find their way through systems and bureaucracies to identify the data they need to solve their problem.

The U.S. National Spatial Data Infrastructure has put in place a legal, policy, and institutional framework for distributing spatial data. GEOSS, GEO, IEOS, and IW-GEO are organizations charged with making the earth observation data available to the U.S. tax paying public who provided the funds to collect this data. However, like a library without a cataloging system full of books without indexes, there is no one-click way for people to access earth observation data. Commercial initiatives like Microsoft's terraserver (http://terraserver.microsoft.com) are providing access to aerial photographs and topography maps. This initiative only begins to touch data from about earth's linked and integrated geological, chemical, biological, ecological, and climatological systems, and only accesses a small fraction of the data available through IW-GEO, et al.

As identified in the IW-GEO Strategic Plan for the U.S. Integrated Earth Observation System, these factors create a vital need for "a comprehensive and integrated data management and communications strategy to effectively integrate the wide variety of Earth observations across disciplines, institutions, and temporal and spatial scales." The international hydrocarbon exploration and production industry, where the author has over 30 years of experience as an innovator and problem solver, struggles with similar data management issues - specifically because the key component in discovering new reserves for the foreseeable future will be through reinterpretation of existing data. The key to this reinterpretation is the integration of all data, from whatever source, to create the best possible subsurface earth model in order to minimize economic risk and maximize the probability of discovering new reserves. As with IW-GEO Earth Observation data, the volume and scope of these data are vast and multidisciplinary.

In both cases, the integration of available and anticipated data requires common indices for spatial location (Figure 1), temporal occurrence, activity or use or process, and data attributes. This type of indexing, which has been developed and is currently in limited use, allows users to immediately access and incorporate any on-line or indexed data into a project using a standard browser point and click interface. These technologies are available and provide a technical foundation for the initiative proposed herein.

Long Range Thinking – Near Term Action

The indexing of Earth Observation Data is the fundamental prerequisite for sound long range policy decisions, management decisions, and personal decisions to meet societal, scientific, and economic imperatives.

What is required is a U.S. government backed and funded initiative which encourages and enables the application and widespread use of these indexing concepts within government groups involved with IW-GEO, including:

- DOC/NIST, DOC/NOAA,
- DOD, DOE,
- DHHS/NIEHS, DSH/FEMA,
- DOI/USGS,
- DOS,
- DOT,
- EPA,
- NASA,
- NSF,
- Tennessee Valley Authority,
- USAID, USDA, and
- the 60 + countries in GEO.

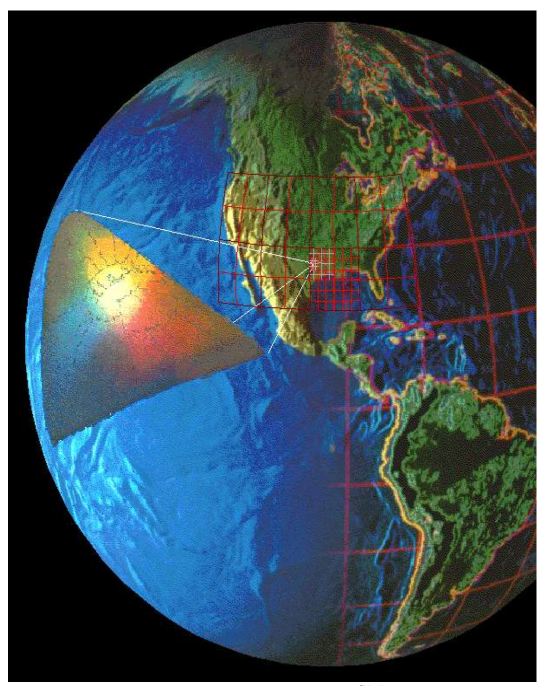


Figure 1. The Infinite Grid[SM] illustrated with various grids overlaid on globe.

Such an indexing project is strategic for a coherent and successful U.S. Strategic Plan for an Integrated Earth Observation System. Combining the oft quoted premise of the information age that "information is more important than data" with a modification of a famous oil field quote from Wallace Pratt that "solutions are first found in the mind," implies that ready access to such accurate indices (information) becomes of critical importance to successful earth observation policies and programs.

Knowing who knows about what processes and data, when critical events happened or are anticipated, where there are exploitable trends, as well as how others have met similar challenges and opportunities will enable creation of the required measurement and monitoring feedback loops necessary to:

- support mineral and hydrocarbon exploration,
- monitoring deadly disease outbreaks,
- optimize food production,
- monitor ecological disasters,
- optimize water usage,
- make quality air quality forecasts,
- improve drought prediction,
- predict storm surges,
- minimize the impact of coastal storms and oil spills,
- reduce pollution,
- optimize fishing and recreation,
- etc.

I recommend IW-GEO take immediate action to reach a consensus on the format, content, and extension rules of the four required indices. Then set policy requiring data collected by supporting organizations be indexed spatially, temporally, by process, and by data type. This allows XML (eXtended hypertext Markup Language) tagged retrieval of relevant data by end-users, where computer translation makes the XML tags invisible.

Figures 2 and 3 illustrate the information which can be shared given common spatial indexing. These Infinite Grid^{SM} maps show a topography map overlaid with 1^{o} longitude by 30' latitude grids. These grid cells have been colored red if there is a "recently" active volcano or colored white if there are know gas hydrate deposits. Figure 2 shows these distributions for the entire world, and Figure 3 shows these distributions for the United States and Mesoamerica. In Infinite Grid^{SM} speak, these are referred to as IG4 maps.

An IG1 map is the entire globe. An IG2 map has 45^{o} by 22.5^{o} grids overlaid. An IG3 map has 5^{o} x 2.5^{o} grids overlaid. And an IG4 map has 1^{o} by 30'grids overlaid on the background map information. To illustrate how the Infinite Grid^{SM} index name works, downtown Houston, Texas is indexed as 26.83.55.55. The IG2 reference is 26, which is the 2nd second IG2 cell in positive X and the 6th IG2 cell in positive Y. The IG3 reference is 83, which is the 8th positive X and the 3rd positive Y cell inside IG2 cell 26. Similarly the IG4 reference is 55 for the 5th positive X and Y cell inside IG3 cell 26.83. So with a single number (26835555), downtown Houston is indexed at 7'30" longitude by 3'45" latitude resolution. This index allows automatic XML referencing for this area.

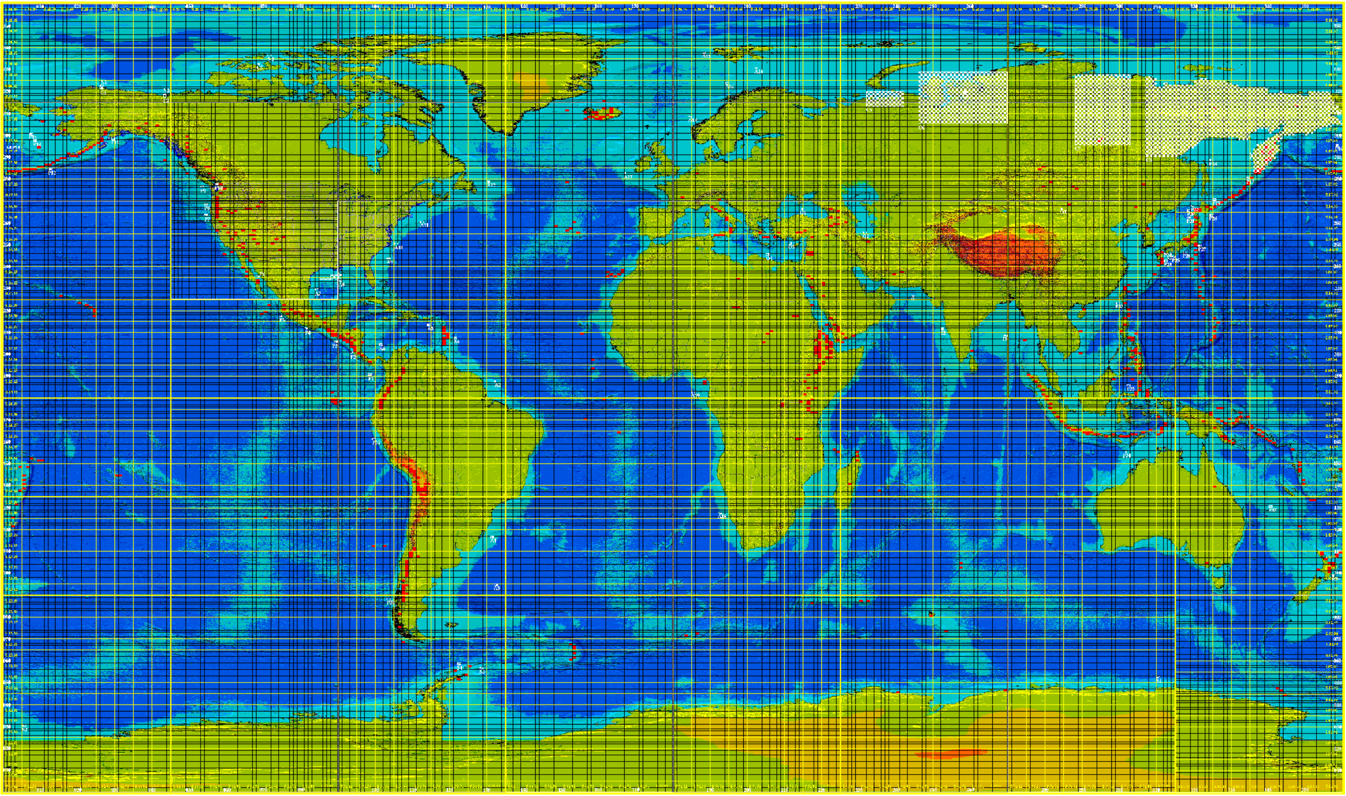


Figure 2. Infinite Grid[SM] global map of distribution of volcanoes (red) and gas hydrates (white).

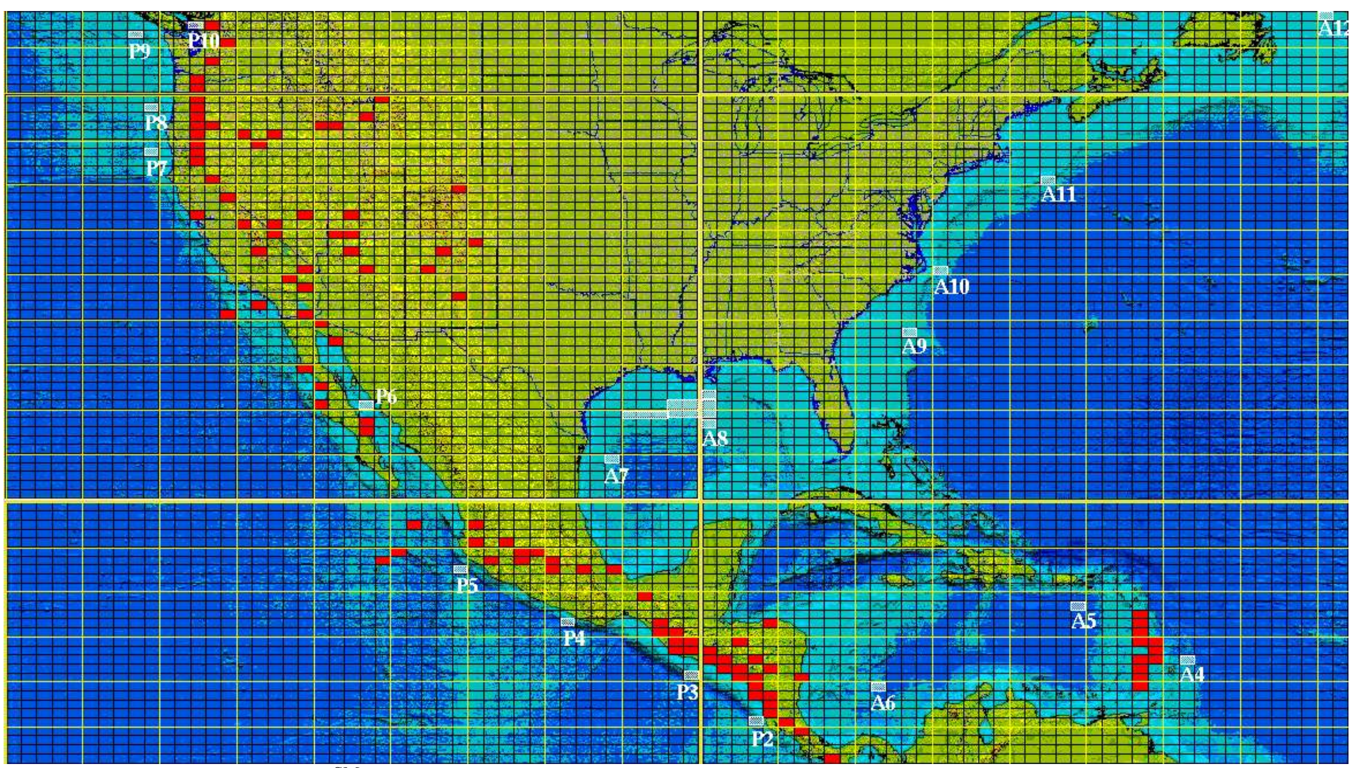


Figure 3. Infinite Grid[SM] U.S. map of distribution of volcanoes (red) and gas hydrates (white).

Similarly time can be indexed using a linear version of the Infinite Grid^{SM} index named the Timedex^{SM}. Figure 4 summarizes the indices used for various time intervals.

Billion Years	Million Years	Million Years	Million Years	Thousand Years	Thousand Years	Thousand Years	Thousand Years	Years
16	2,000	256	32	4,000	512	64	8	1,000
14	1,750	224	28	3,500	448	56	7	875
12	1,500	192	24	3,000	384	48	6	750
10	1,250	160	20	2,500	320	40	5	625
8	1,000	128	16	2,000	256	32	4	500
6	750	96	12	1,500	192	24	3	375
4	500	64	8	1,000	128	16	2	250
2	250	32	4	500	64	8	1	125
A11 to A18	G61 to G68	G51 to G58	G41 to G48	G31 to G38	G21 to G28	G11 to G18	H41 to H48	H31 to H38

Years	Years	Months	Days	Days	Hours	Minutes	Minutes	Seconds
128	16	24	96	12.0	36.0	320	40	320
112	14	21	84	10.5	32.5	280	35	280
96	12	18	72	9.0	27.0	240	30	240
80	10	15	60	7.5	22.5	200	25	200
64	8	12	48	6.0	18.0	160	20	160
48	6	9	36	4.5	13.5	120	15	120
32	4	6	24	3.0	9.0	80	10	80
16	2	3	12	1.5	4.5	40	5	40
H21 to H28	H11 to H18	P71 to P78	P61 to P68	P51 to P58	P41 to P48	P31 to P38	P21 to P28	P11 to P18

Figure 4. The Timedex^{SM}, an interval time index that works across Astrophysical (A), Geologic (G), Historical (H), and Project (P) times.

For activity indexing the U.S. Air Force developed an easy to use process language, named IDEF (loosely meaning the Information Definition Exchange Format), which provides one computer word indexing of any process from an IDEF-0 Activity Model. By standardizing vocabularies across agencies with a common activity model language, it becomes possible for everyone in earth observation enterprises to converse with anyone else in the enterprise with a better level of comprehension. The automatic translation from one level to another creates a continuous improvement methodology that allows a true enterprise wide ecology of mind. IDEF has fourteen defined levels:

- IDEF-0 Function / Activity Modeling
- IDEF-1 Information / Data Modeling
- IDEF-2 Simulation Modeling
- IDEF-1X Data Modeling
- IDEF-3 Process Description Capture
- IDEF-4 Object Oriented Design
- IDEF-5 Ontology Description Capture
- IDEF-6 Design Rationale Capture
- IDEF-7 Information System Audit Method
- IDEF-8 User Interface Modeling
- IDEF-9 Scenario Driven Information System Design Specification
- IDEF-10 Implementation Architecture Modeling
- IDEF-11 Information Artifact Modeling
- IDEF-12 Organization Modeling
- IDEF-13 Three Schema Mapping Design
- IDEF-14 Network Design

Data Type indexing is a function of content. The data type index we developed for the oil industry, which could be built upon to include all Earth Observation Data, includes:

1. Format of Data
    a. Text
    b. Numbers
    c. String (Vector)
    d. Table
    e. Image
    f. Photograph
    g. Animation
    h. Movie
    i. Model
    j. Simulation
2. Type of Display
    a. Outcrop
    b. Well Log Cross-Section
    c. Well Core Data
    d. Seismic Sections
    e. Culture Map
    f. Topographic Map
    g. Bathymetry Map
    h. Surface Geology Map
    i. Satellite Map
    j. Integrated 3-D Immersive Reality Model of Outcrop, Well Log, Seismic, and Other Data
3. Episodic Depositional Analysis
    a. Basin Tectonic Setting
    b. Structural Style within Basin
    c. Geologic Cycle / Sequence Duration
4. Stratal Pattern Cycles (1-10 million years)
    a. Termination Description
    b. Eustatic Boundaries
    c. Tectonic Controlled Boundaries
    d. Boundary Description
    e. Intervals: Stratigraphy Cycles Associated with Depositional Systems (1-10 million years)
    f. Reflection Configurations or Stratal Patterns
    g. External Form of a Sequence
5. Depositional Systems (rocks deposited in a particular environment)
    a. Permeability
    b. Porosity
    c. Petrography (range of values)
    d. Color
    e. Rock Systems
    f. Lithofacies
    g. Grain Size
    h. Primary Sedimentary Structures Contemporaneous with Deposition
    i. Sedimentary Structure
    j. Paleo-Bathymetry and Paleo-Topography
    k. Facies (Sub-Environment)
6. Exploration Categories, Plays, and Prospects
    a. Source Rocks (Type of Organic Material)
    b. Seal Rocks (Pore Pressure, Type Seal, Thickness of Seal)
    c. Reservoir Rocks (Porosity, Permeability)
    d. Structural Traps
    e. Stratigraphic Traps Not Adjacent to Unconformities
    f. Diagenetic
    g. Traps Adjacent to Unconformities (Below Unconformity)
    h. Traps Adjacent to Unconformities (Above Unconformity)
    i. Combination Traps (Undifferentiated)

Once IW-GEO has initiated the indexing project and set policy regarding indexing of Earth Observation Data for participating governmental agencies, the next step is to encourage adoption by user communities. This includes presentations and dialog with professional societies and commercial entities with a vested interest.

Professional Societies are logical enablers. In the geological and oil and gas exploration and production community the key societies include AAPG (American Association of Exploration Geophysicists), AGI (American Geological Institute), AGU (American Geophysical Union), OTC (Offshore Technology Conference), SEG (Society of Exploration Geophysicists), SEPM (Society for Sedimentary Geology), SIPES (Society of Independent Professional Earth Scientists), SPE (Society of Petroleum Engineers), their local affiliates, their international counterparts, and their foundations.

Commercial entities with a vested interest include any company which sells data or services that include earth observation in agriculture, energy, finance, insurance, transportation, real estate, retail and wholesale trade, manufacturing, etc. Given definition of common indices, there will all kinds of creative ways of interacting with integrated earth observation data. For instance, there is a company that sells CD's of telephone directories named Select Phone. These CD's have every public land phone number in the United States on them and are updated quarterly. In addition to the phone numbers and addresses, Select Phone keeps track of four levels of SIC (Standard Industrial Codes) for each business, and their CD's include the longitude and latitude within about 40 feet for each phone number. Figure 5 shows a map of the density of SIC 15 businesses at an IG5 level (7'30" by 3'45" grid cells). SIC-15 consists of building construction, general contractors, and operative builders. Besides showing where to place a plywood distribution center within 15 miles of the majority of potential clients, imagine the power of being able to overlay this information on earth observation data.

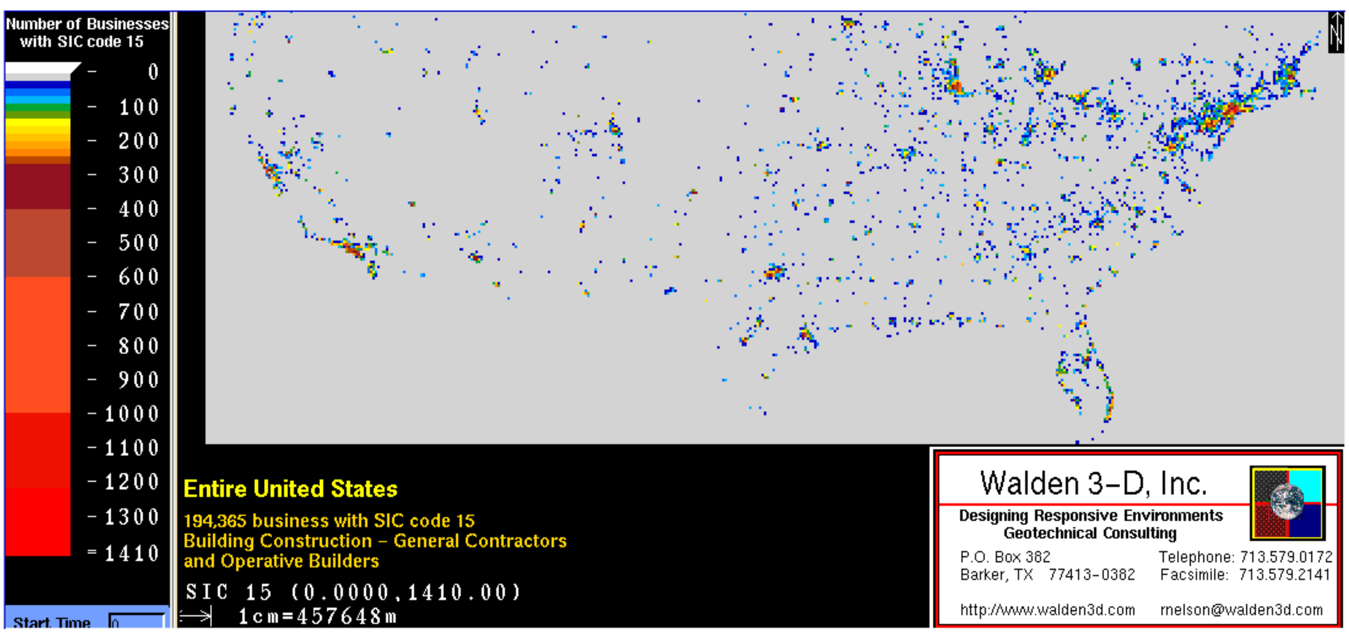


Figure 5. Map of the United States showing the density of SIC-15 businesses.

SIC codes inherently contain a relationship to earth systems (farming, mining, investment, insurance, urban planning, consumer interest, manufacturing, etc.). Indexing the spatial and temporal distribution of these activities and then modeling resource flow - from extractions to manufacturing to end use – enables optimal use of earth observation data. Tying activity indices to resource flow and displaying time indexed snapshots of the spatial distribution of data types creates animated visualizations of human activity. This visualization can be directly compared to weather (climate over time) and to natural disasters like an earthquakes, tsunami, a drought, a deadly disease outbreak, or the relationship of harmful algal blooms to industries which are literally upstream.

Distributed earth observation data needs to be correlated with population, business, and resource flow. Interpolation between the census framework - using commercial data like phone directories and traffic counts - creates an integrated growth function. It is key to spatially and temporally index human activity data, including experience, interest, and regulatory responsibility. Controls will need to be put in place to discourage misuse of these data criminally or politically.

Grid-to-grid operations of Infinite Grid^{SM} indexed earth observation data will enable spread-sheet analysis and distribution. Figures 6, 7, 8, and 9 shows an example of mapping rainfall magnitude and flood damaged residences against a census tract map and table using Excel. A limitation which is quickly reached using this approach is the 256 column limitation in Excel. Raster based mapping and indexing is a subtle but significant improvement over traditional vector GIS mapping. Vector GIS systems are valuable as background maps for smooth scale changes, and it is not feasible to have a simple single computer word spatial index for a line boundary in the same way the Infinite Grid^{SM} allows XML tags to be tied to specific grid cell areas. Because the cells form a matrix, it is feasible to develop matrix math models of grid indexed earth observation data which predict movement across time. Time-lapse measurements can be inverted to derive these algorithms. This will allow the measurement and monitoring of urban expansion boundaries, wetland disappearance and growth, etc.

Atmospheric concentration of CO_2 and other atmospheric constituents, as well as climate variables such as precipitation, land use, and land cover, which undergo regional and local changes, have significant environmental and human impacts. Using a specific example from the IW-GEO Strategic Plan, earth observation models are scientists' primary tools for: (1) integrating observations into a comprehensive analysis of the climate system; (2) forecasting of the climate system on multiple time/space scales; and (3) simulating the impact of a particular gap or enhancement of the comprehensive GEOSS. Simple indexing of new types of data, for instance (1) putting CO_2 and aerosol sensors on commercial airplanes and tying the shelf-space contribution to tax relief or (2) new "Dust" sensors, can improve models and save significant unnecessary end-user costs. Spatial and temporal indices allow mapping of the spatial extent of various regulatory agency responsibilities, as well as the distribution of sensors. Integration with DEM (Digital Elevation Maps) and GIS databases, with GPS and models will enable new creativity and innovation to data mine and developed guided analytics for web based analysis and distribution of earth observation data via map images.

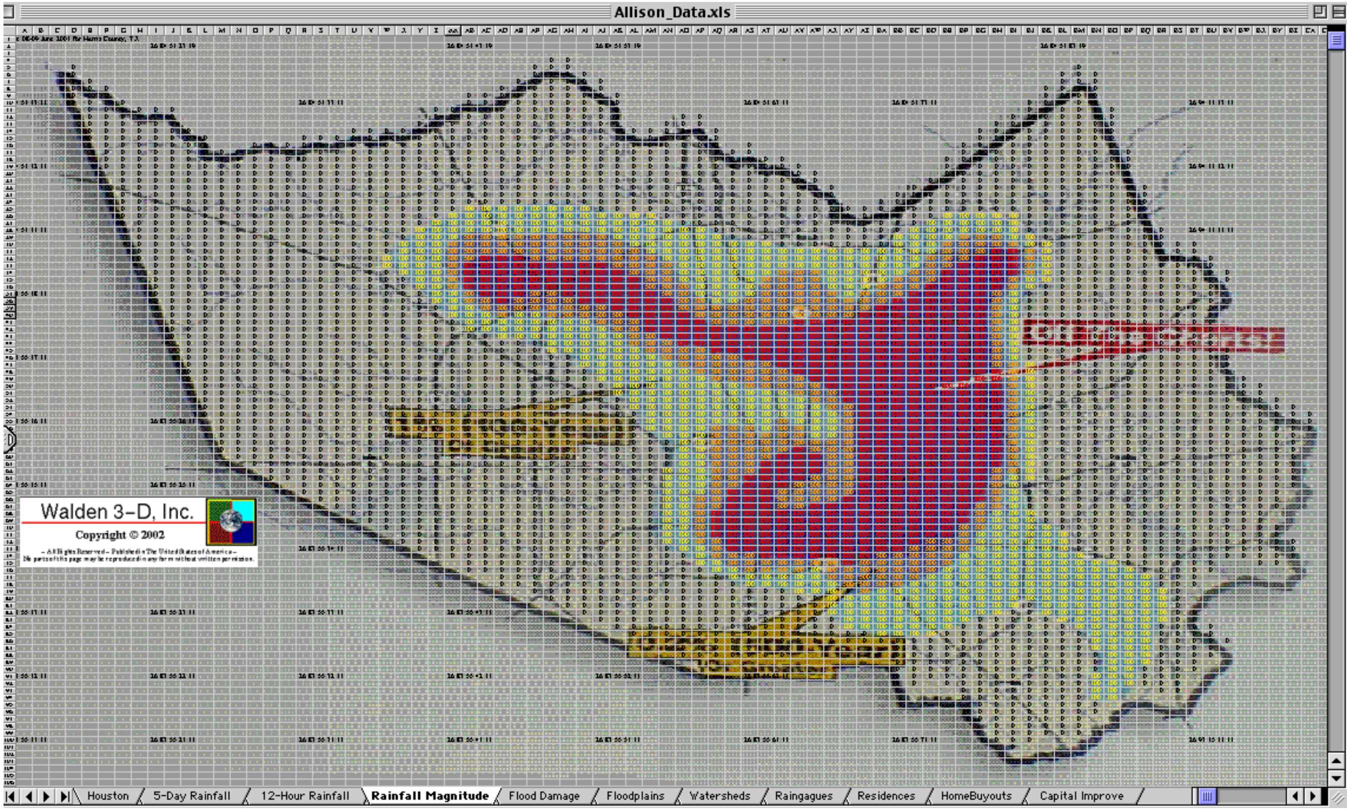


Figure 6. Rainfall Magnitude for Allison in Harris County, Texas mapped with an Excel spread-sheet.

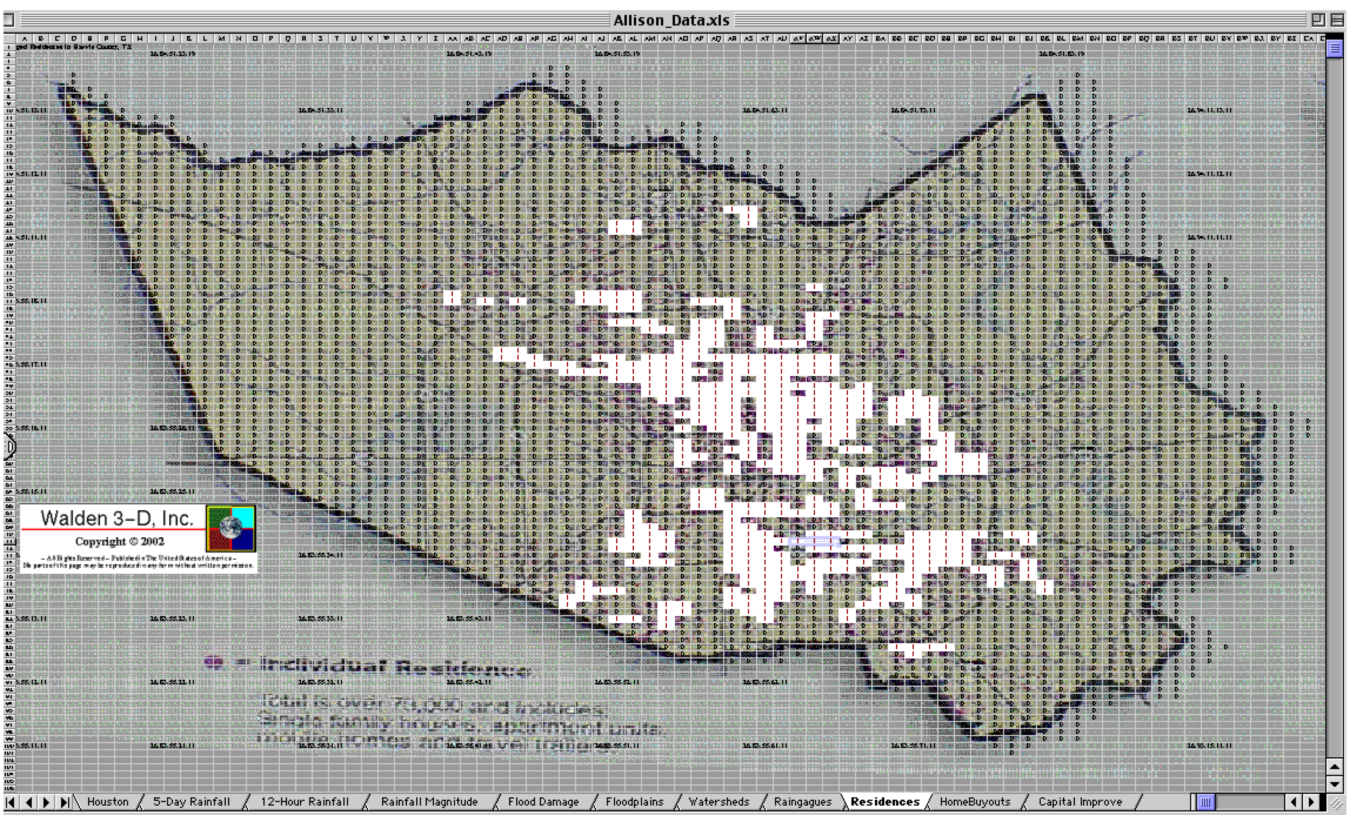


Figure 7. Location of damaged residences from Tropical Storm Allison in Harris County, Texas.

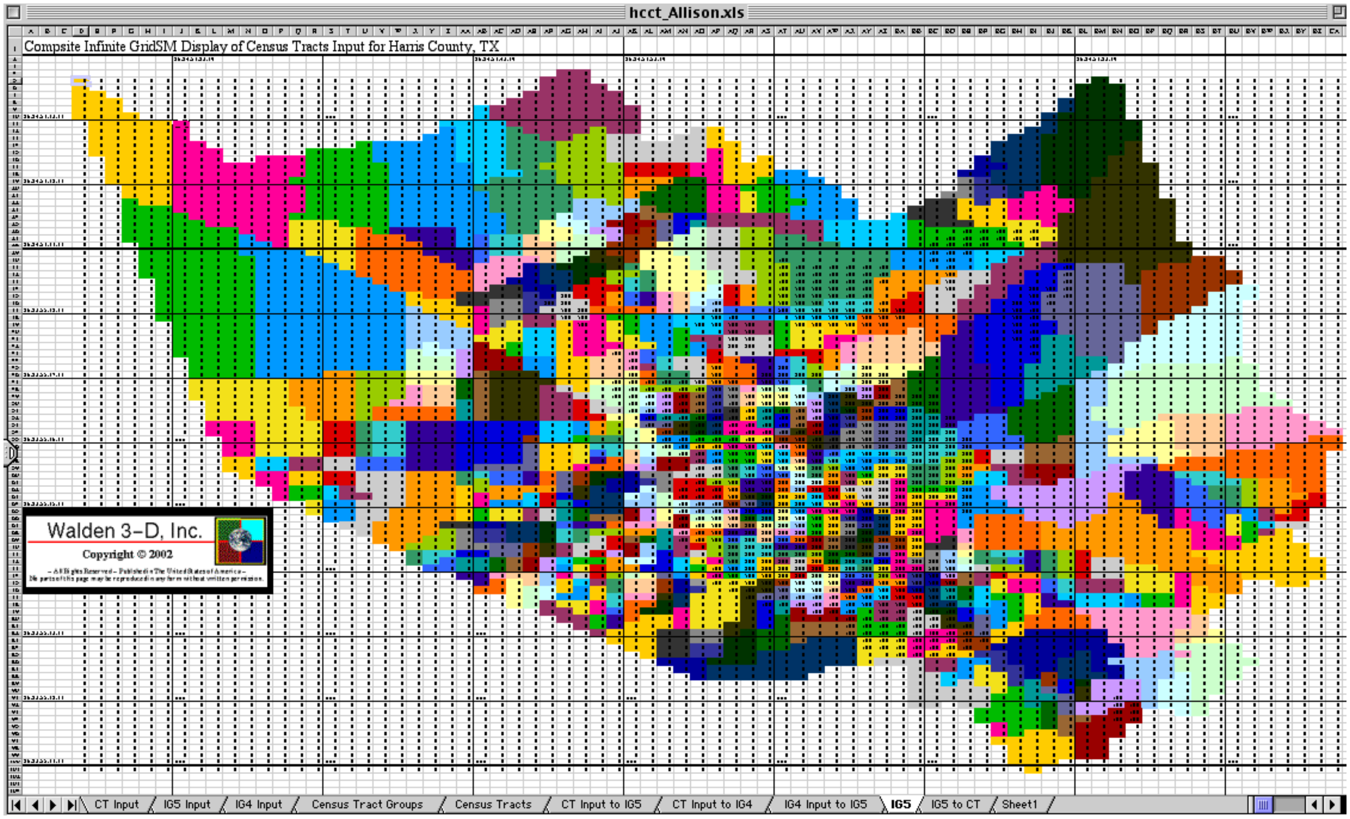


Figure 8. Rainfall magnitude overlaid on a map of census tract boundaries in Harris County, TX.

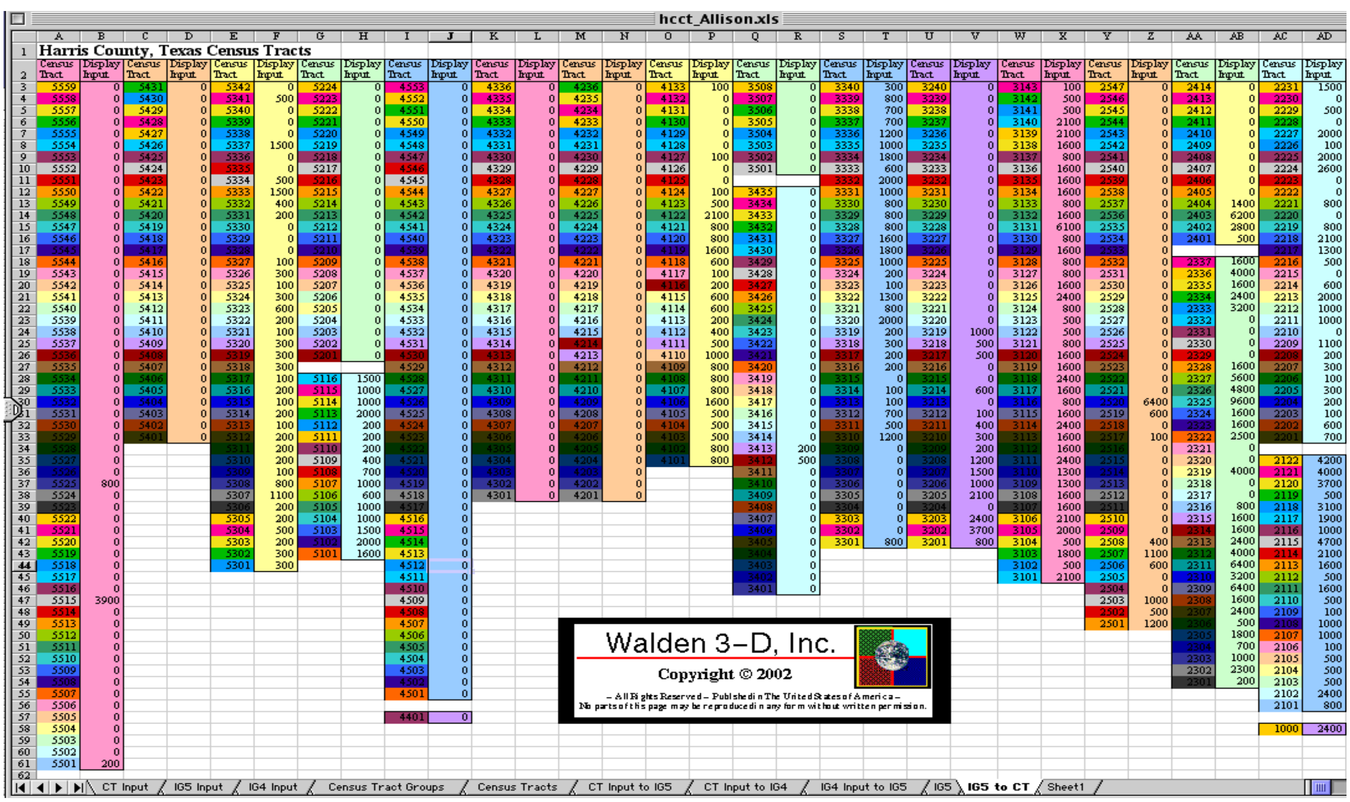


Figure 9. Automatic translation of rainfall magnitude to a census tract table for Harris County, TX.

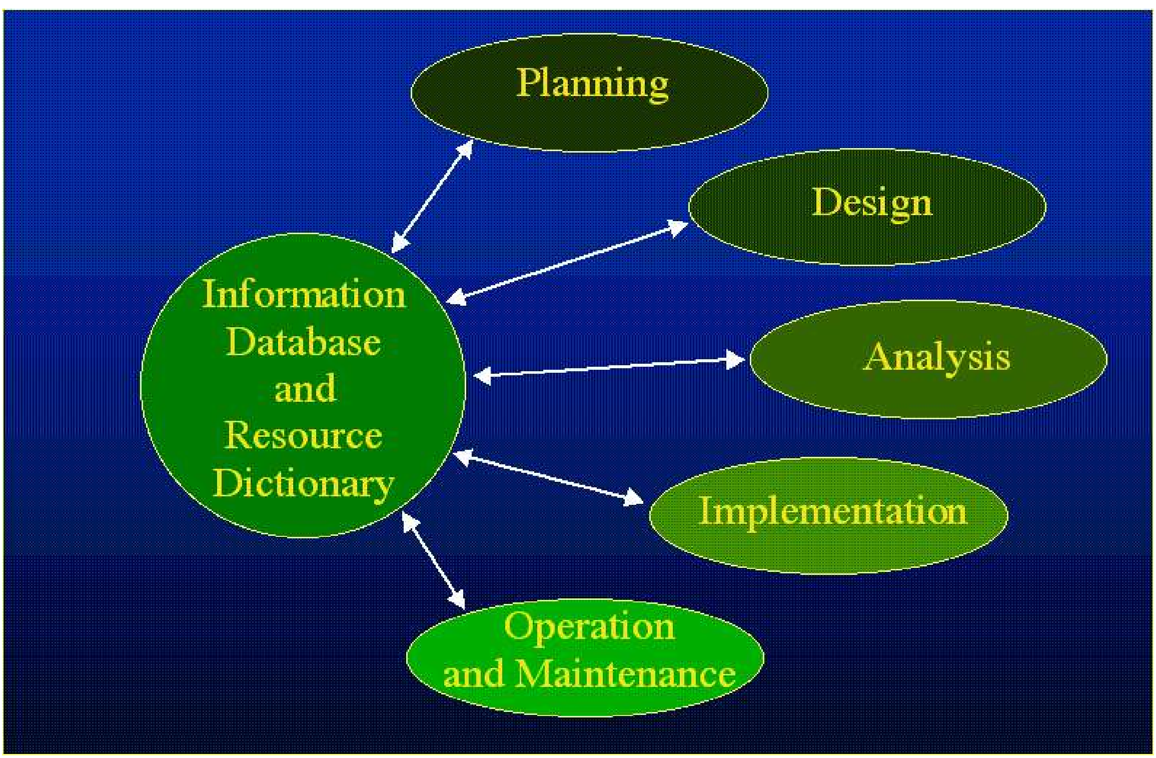
Conclusions

Like the exploration and production industry, government initiatives like NASA and other providers of earth observation data have a long history of being slow to adopt new approaches to doing day-to-day work. However, once a case is made, through demonstration or logic, there tends to be rapid and widespread acceptance.

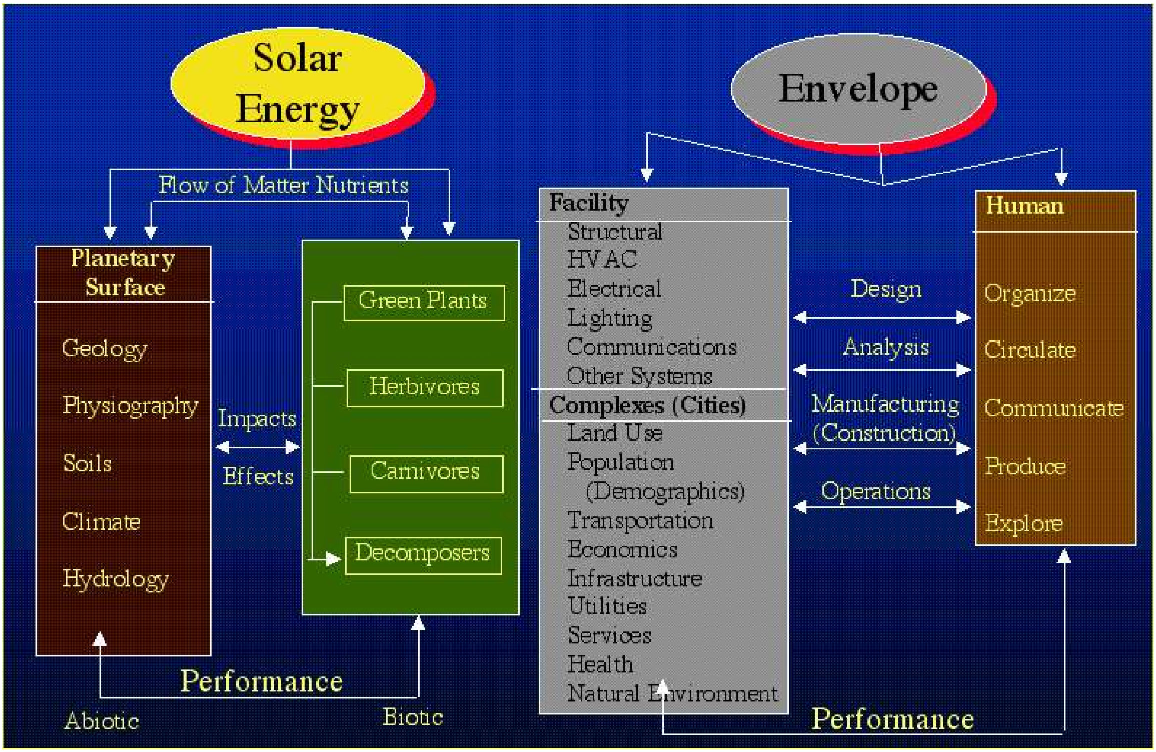
Integrated interpretation of data from the ultimate integrated system humanity relies on, namely Earth, requires interactive data access that needs data context, which requires data indexing. Government initiatives like IW-GEO have the opportunity to set policy, like usage of the indexing proposed herein, which is beyond the scope of one off implementation by professional societies and commercial entities.

This project requires the long range vision and the short-term action of those who want to ensure our secure, continuous and affordable long-term access to and widespread and optimal use of earth observation data.

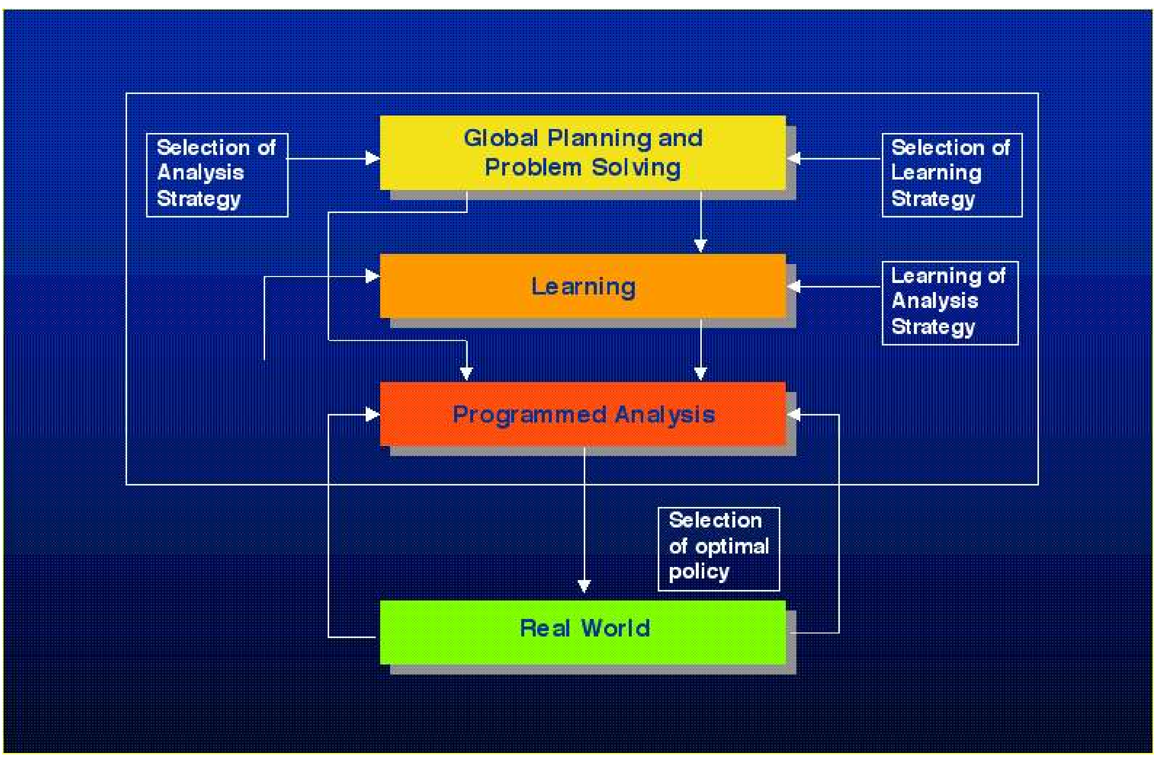
Appendix – Concept supporting slides from a 1990 Presentation by the late Bill Bavinger, who invented and named the Infinite Grid℠:



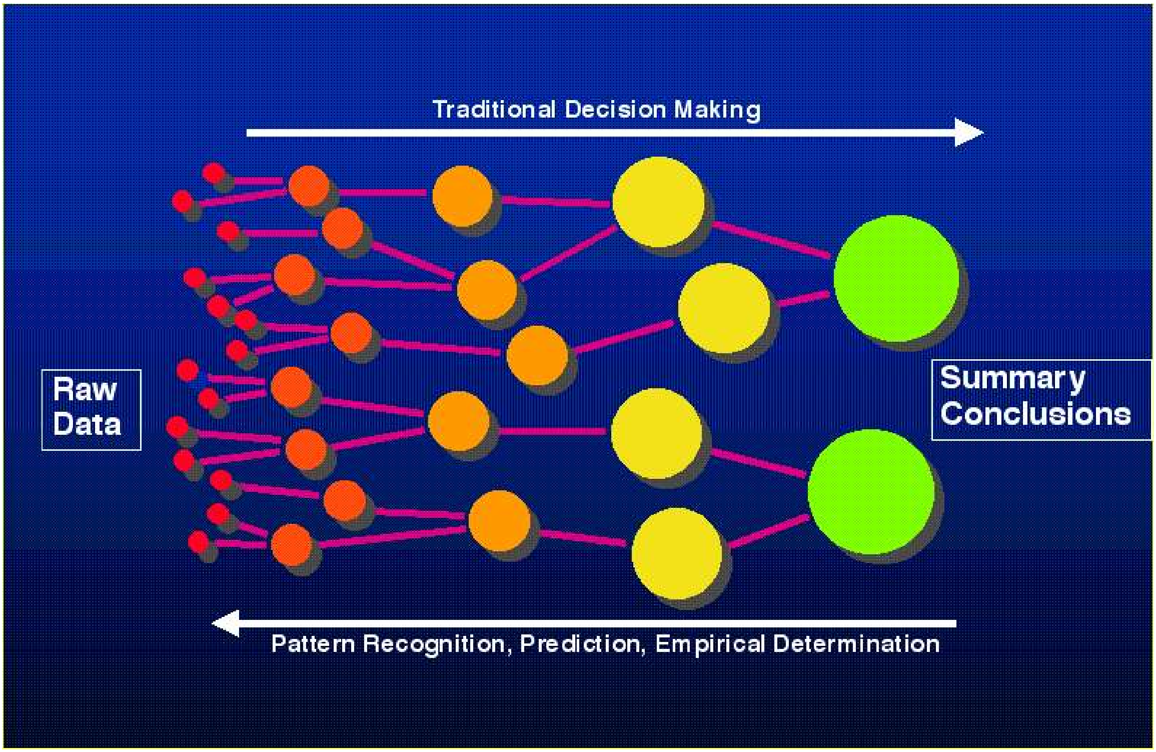
Slide 1. Information Designing.



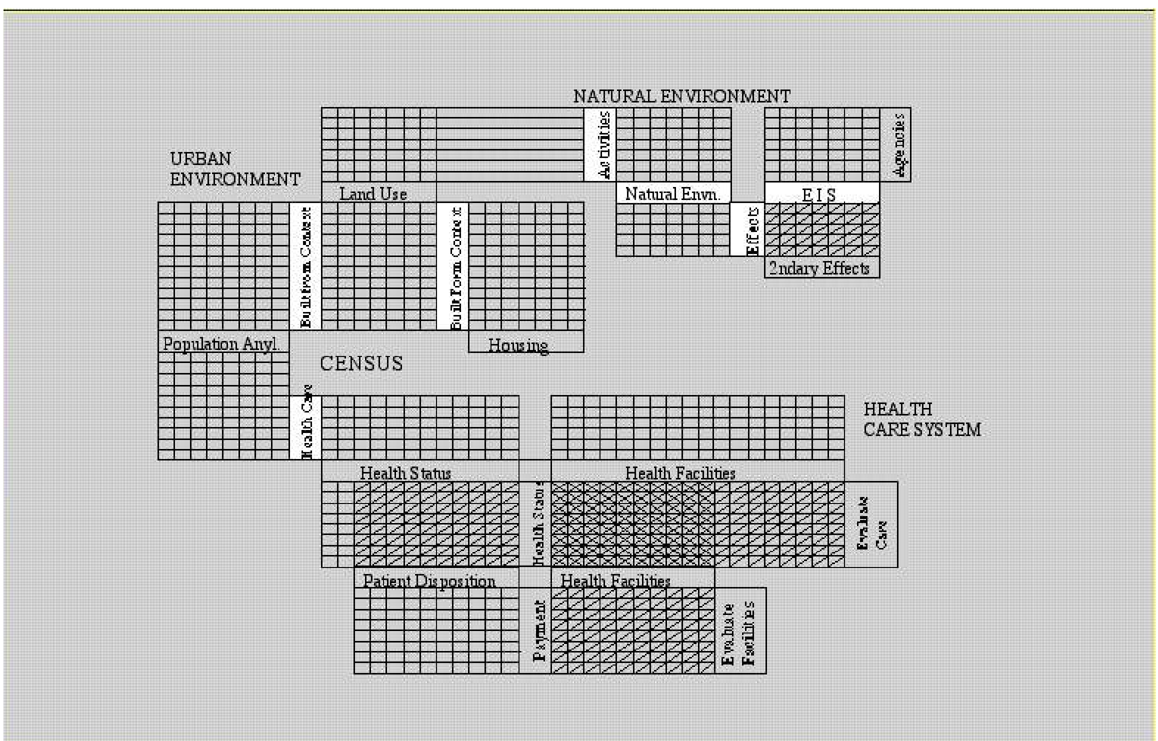
Slide 2. Application to a Project.



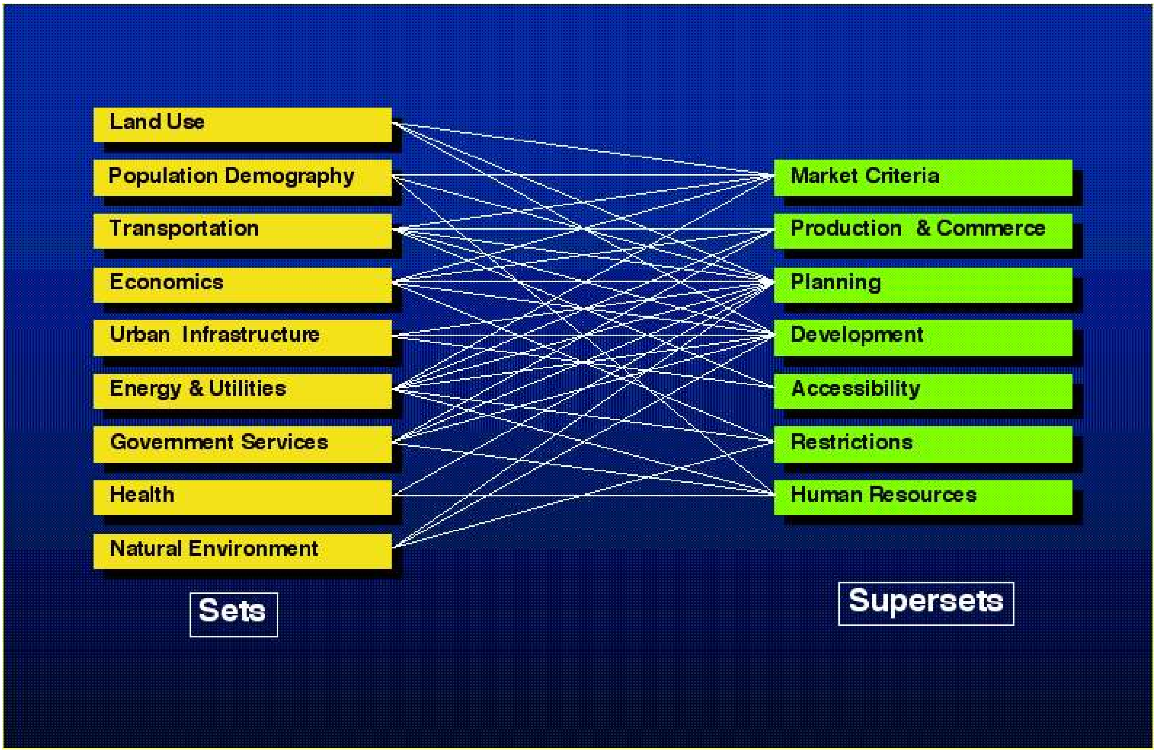
Slide 3. Analysis System.



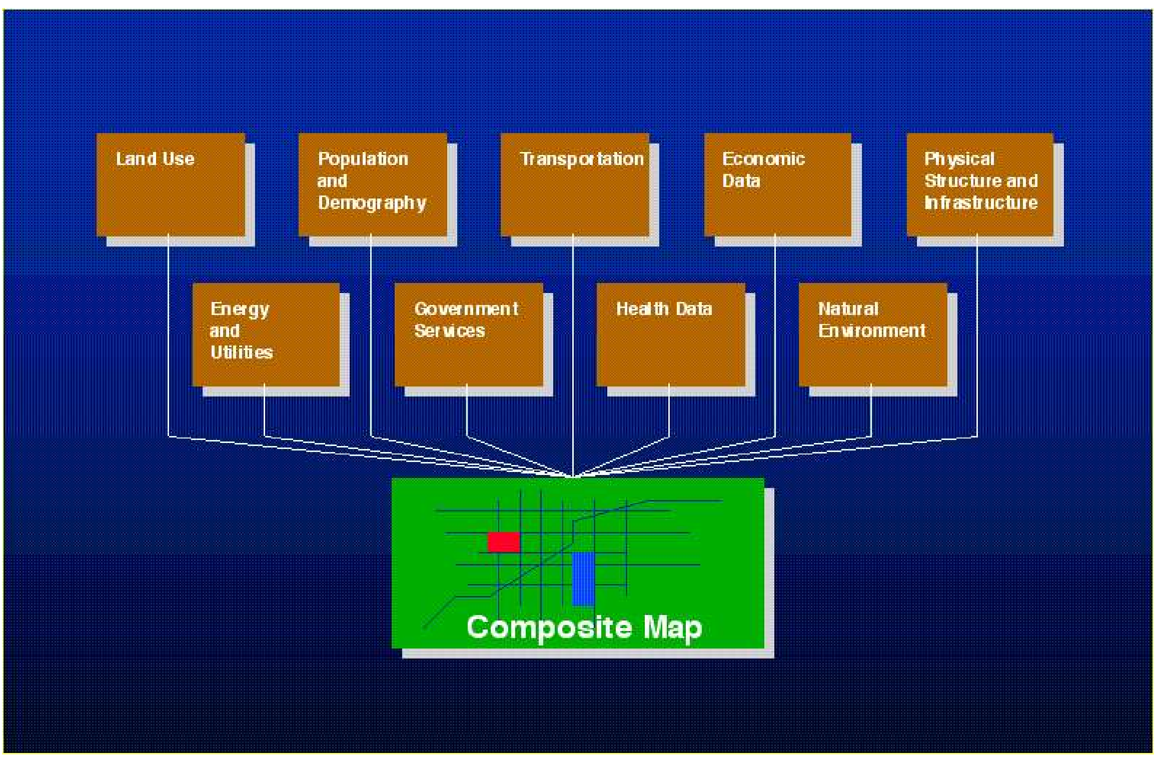
Side 4. Decision Making.



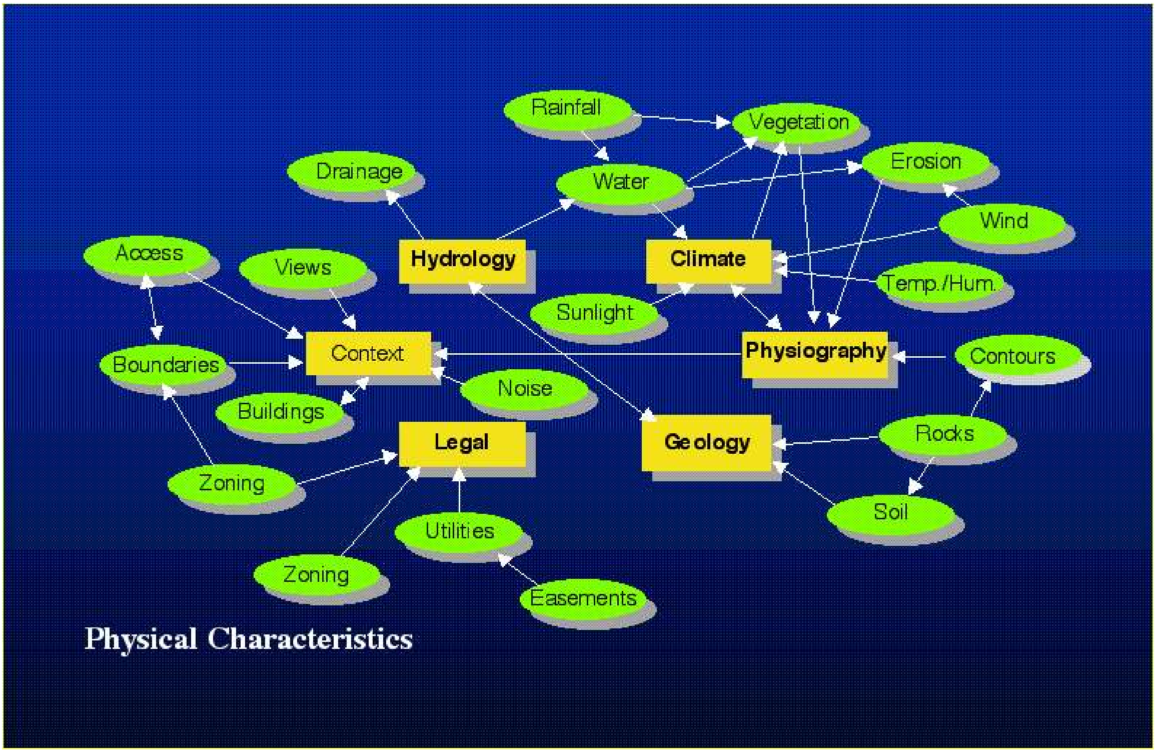
Slide 5. Interlocking Matrices.



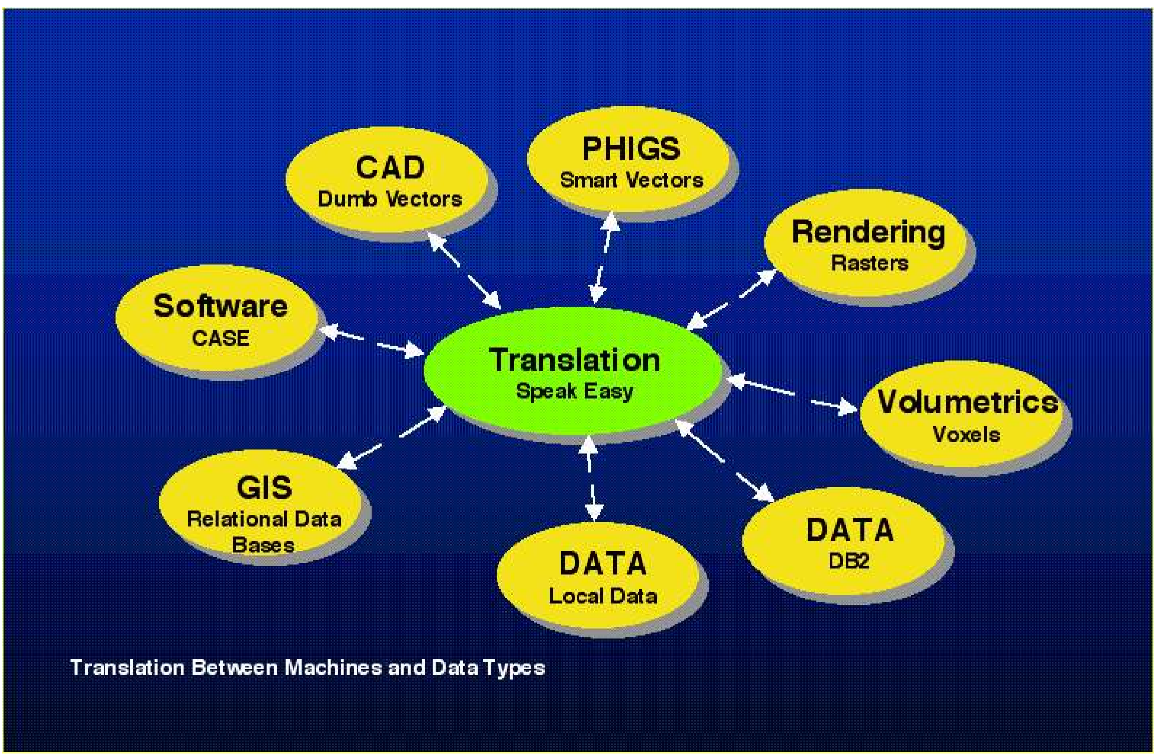
Slide 6. Urban Classification System.



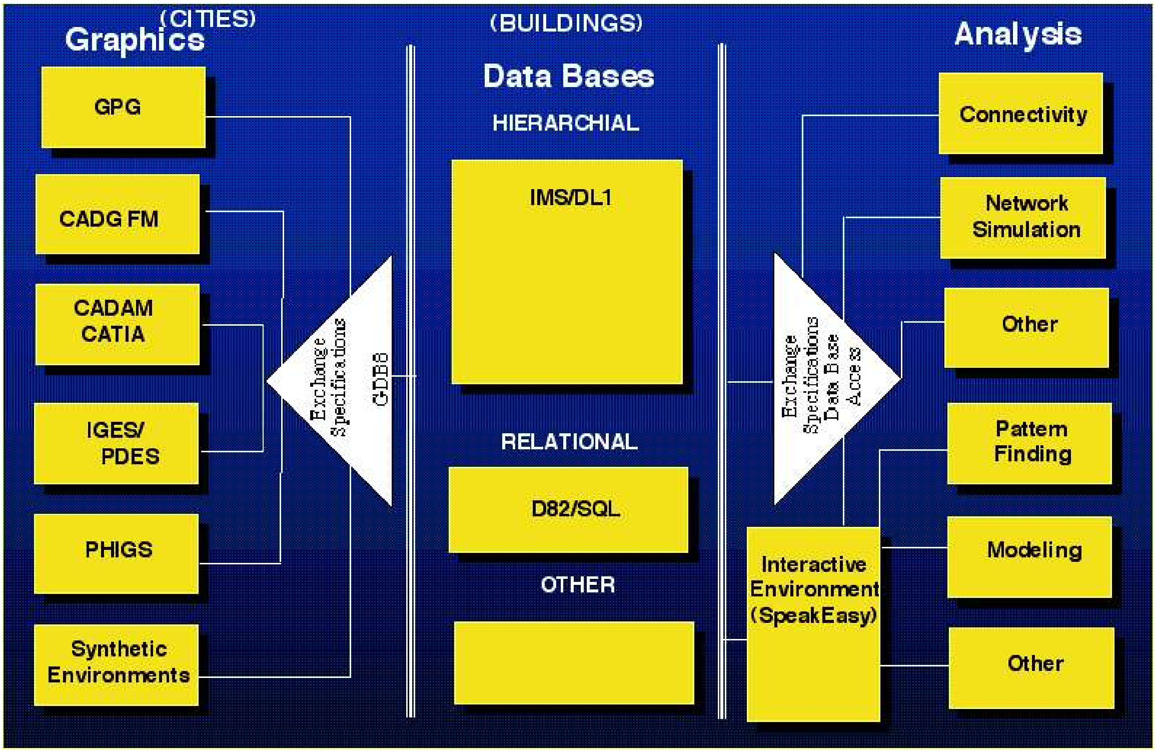
Slide 7. Mapped Urban Information Processing.



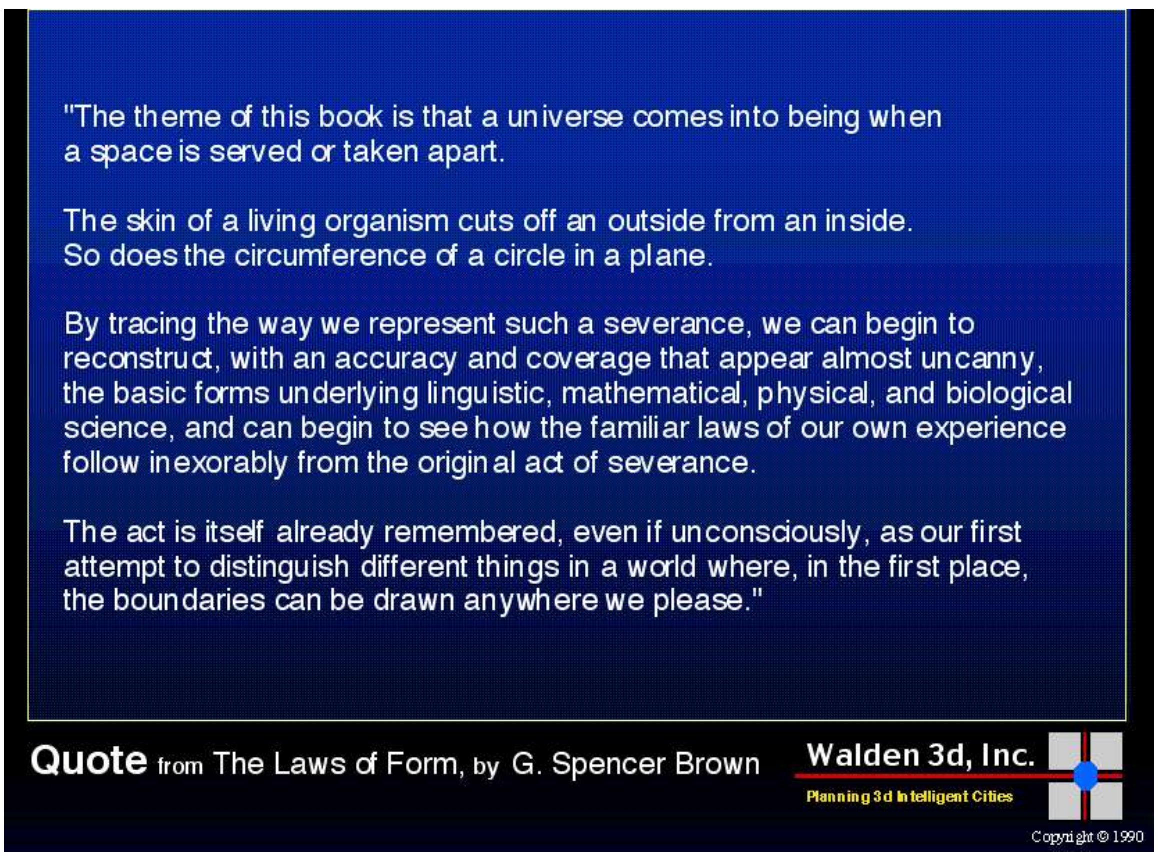
Slide 8. Site Hierarchy.



Slide 9. Information Processing by Translation.



Slide 10. Software Systems Integration.



Slide 11. Envelopes.