Professional Development Short Course On:

Exploring Data: Accessing, Understanding and Visualizing Data To Gain Insight.

Instructor:

Ted Meyer
Dr. Brand Fortner

ATI Course Schedule: **http://www.ATIcourses.com/schedule.htm**

ATI's Exploring Data: Visualization course **http://www.aticourses.com/data_presentation_and_visualization.htm**

Applied Technology Institute (ATI)

Stay Current In Your Field • Broaden Your Knowledge • Increase Productivity

349 Berkshire Drive • Riva, Maryland 21140
888-501-2100 • 410-956-8805
Website: www.ATIcourses.com • Email: ATI@ATIcourses.com

www.ATIcourses.com

Boost Your Skills with On-Site Courses Tailored to Your Needs

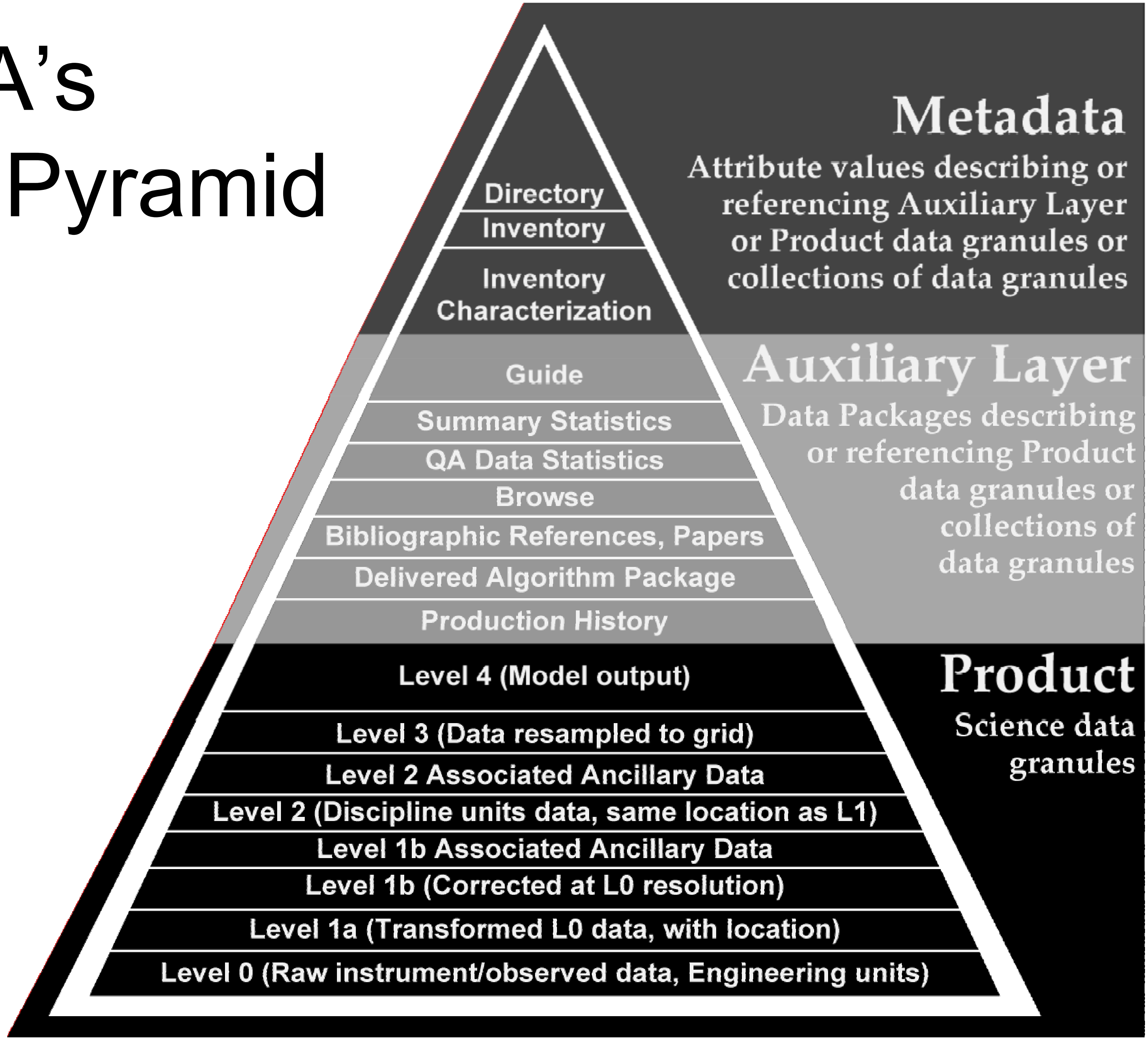
349 Berkshire Drive
Riva, Maryland 21140
Telephone 1-888-501-2100 / (410) 965-8805
Fax (410) 956-5785
Email: ATI@ATIcourses.com

The Applied Technology Institute specializes in training programs for technical professionals. Our courses keep you current in the state-of-the-art technology that is essential to keep your company on the cutting edge in today's highly competitive marketplace. Since 1984, ATI has earned the trust of training departments nationwide, and has presented on-site training at the major Navy, Air Force and NASA centers, and for a large number of contractors. Our training increases effectiveness and productivity. Learn from the proven best.

For a Free On-Site Quote Visit Us At: http://www.ATIcourses.com/free_onsite_quote.asp

For Our Current Public Course Schedule Go To: http://www.ATIcourses.com/schedule.htm

NASA's Data Pyramid



Universe of Data

- Data Elements
- Data Structures (Objects)
- Data Collections
- Data Systems

- Metadata

When Dealing with Data

- How are the numbers stored?
- How is the data Organized
- What is the dimensionality of the data?
- Is the data on a grid?
- What is the best way to analyze the data?

Data Characteristics

- Numeric, symbolic (or mix)
- Scalar, vector, or complex structure
- Various units
- Discrete or continuous
- Spatial, quantity, category, temporal, relational, structural
- Accurate or approximate
- Dense or sparce
- Ordered or non-ordered
- Disjoint or overlapping
- Binary, enumerated, multilevel
- Independent or dependent
- Multidimensional
- Single or multiple sets
- May have similarity or distance metric
- May have intuitive graphical representation (e.g. temperature with color)
- Has semantics which may be crucial in graphical consideration

Numbers in Computers

- Quantitative: Numeric vs. Non-numeric data
  - Categorical Data: Finite set
  - Text
- Number Types
  - Binary
    - Bytes, Integers, Floating point
    - Fixed precision
    - Not readable by humans
    - Storage and processing efficient
  - ASCII
    - Text, Characters
    - Variable precision
    - Human readable
    - Storage and processing inefficient

Evaluating Number Types

- Storage and processing efficiency
- Data range required
- Numeric precision required
- Calculation issues
- Portability

Bytes

- 8 bits represents 2^8 (256) distinct values
- Unsigned and Signed
  - Twos-complement
- Representation
  - Hexadecimal, Octal, Decimal, Binary

Integers

- Short and long integers
- Signed and unsigned
- Fixed point numbers
  - Scale and Offset

    number = scale x (value + offset)

Floating Point Numbers

- Single precision (4-byte)
  - 9.10956 x 10^{-28}
  - sign exponent mantissa
  - 0 -28 910956 decimal
  - 0 -1011010 10010000001011000110001 binary
  - IEEE Standard 754
- Double precision (8-byte)
- Zero, NaN, INF, Complex numbers, Extended

ASCII Text Numbers

- ASCII Characters
- Numbers
  - Exponential notation
  - Delimiters - space, comma, tab
  - Line separators
  - Position formats
- Unicode (16-bit characters)

Storage and Processing Efficiency

- Bytes - efficient use of disk space and CPU cycles
- Integers - efficient use of disk space and CPU cycles, especially if no FPU
- Floating point - less disk efficient, needs FPU to be processing efficient
- ASCII Text - disk and processing hog, no direct access unless position formatted, processing requires translation

Data Range

- Bytes - 0 to 255 unsigned or -128 to 127 signed, easy to exceed range
- Integers - depends on size, -32768 to 32767 for 2-byte integers, easy to exceed range
- Floating point - very large, but user needs to know when to use double precision
- ASCII Text - limited only by capabilities of reading software (most software is limited to integer or floating point ranges)

Numeric Precision

- Bytes - always one
- Integers - precision is always one for integers, 1/scale for fixed point
- Floating point - can vary depending upon a variety of numerical factors, but the maximum is about 7 and 15 decimal digits for single and double precision numbers
- ASCII Text - limited only by capabilities of reading software (most software is limited to integer or floating point precision)

Calculation Issues

- Bytes - dangerous because of likely overflow
- Integers - dangerous because of likely overflow
- Floating point - Usually easy, but be aware of problems: roundoff, differencing similar numbers, comparisons
- ASCII Text - must first convert to integers or floating point and then subject to same limitations as those types

Portability

- Bytes - Most computers store bytes the same way
- Integers - byte ordering problems: little vs. big endian, fixed point problematic because of scale and offset
- Floating point - IEEE standard is most common, but heritage data may be in other forms
- ASCII Text - extremely portable and human readable on most platforms, with minor problems associated with delimiters, line end characters, and file transfer

Scientific Data Storage

- Text vs. Binary, Public vs. Private
- Issues:Numerical Precision
  - Numerical Range
  - Portability
  - Efficiency
  - Self-Documenting
  - Power & Extensibility
- How do the Various Formats Rate?

How do the Various Formats Rate?

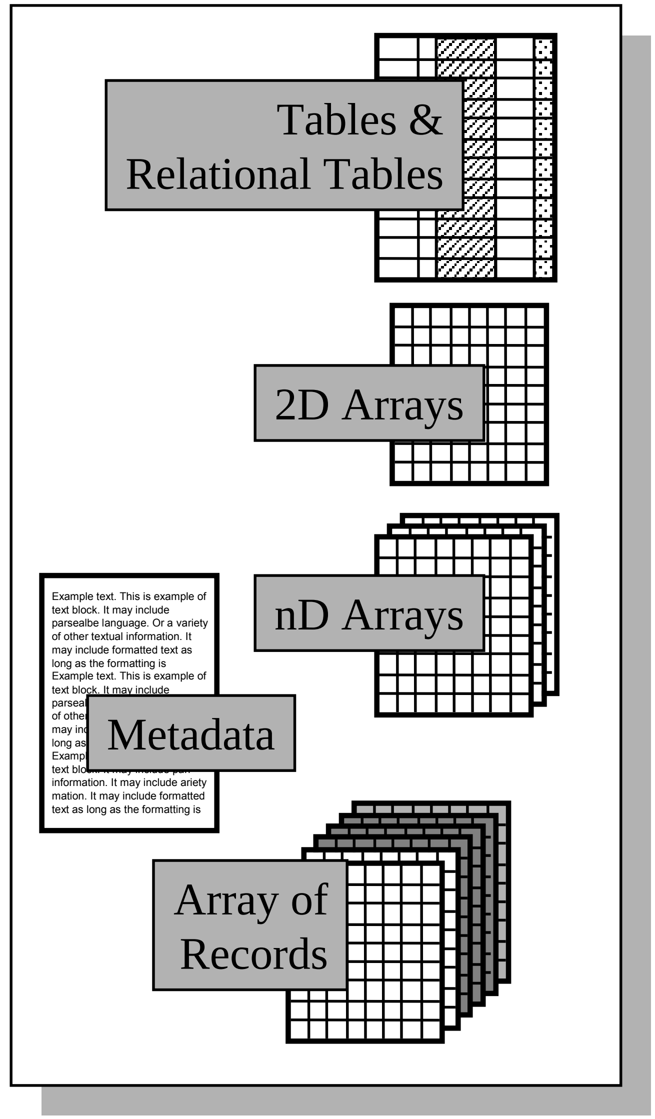
	Text	Binary Integer	Binary Float
Precision	Variable	Fair	Good
Range	Infinite	Poor	Excellent
Portability	Excellent	Fair	Poor
Efficiency	Poor	Excellent	Excellent

	Private Binary	Private Text	Standard Binary
Portability	Poor	Excellent	Excellent
Efficiency	Excellent	Poor	Excellent
Self-Document	Poor	Good	Excellent
Power & Extensibility	Good	Good	Varies

- *What the World Needs is a Powerful, Extensible, Self-Documenting, Standard Binary Floating Point File Format for Technical Data*

Kinds of Science Data

- Science Data Types
- Metadata
- Images
- Atomic Types
  - Integers
  - Floating Point
  - ASCII Text



How are my Numbers Organized?

- Dimensionality, Data Locations, and Data Values
- Column Data: *List of Data Locations and Values*

X	Y	Velocity
0.5	0.5	0.0350
0.5	1.0	0.0714
0.5	1.5	0.3853
1.0	0.5	0.4911
1.0	1.0	0.2422
1.0	1.5	0.9207
1.5	0.5	0.5744
1.5	1.0	0.3305
1.5	1.5	0.8485

- 2D Matrix Data: *Locations Implicit in Matrix*

		X		
		0.5	**1.0**	**1.5**
	0.5	0.0350	0.4911	0.5744
Y	**1.0**	0.0714	0.2422	0.3305
	1.5	0.3853	0.9207	0.8485

- 3D Matrix Data: *Same as 2D*

Z=6.0	0.5	1.0	1.5
			2.4064
			2.5157
			2.7190

Z=3.0	0.5	1.0	1.5
			1.7672
			1.7253
			1.7757

Z=0.0	0.5	1.0	1.5
0.5	0.0350	0.4911	0.5744
1.0	0.0714	0.2422	0.3305
1.5	0.3853	0.9207	0.8485

- Polygonal Data: *List of Objects*

Polygon Name	Polygon Position	Vertices	Temp.	Stress
A	(0.5,0.5,0.0)	(vertex info)	72.2	0.034
B	(0.5,1.0,0.5)	•	74.8	0.056
C	(1.0,0.5,0.5)	•	71.3	0.089
•	•	•	•	•

Column Data

- Column, Record, Flat File or Table Data
  - Records, fields

Time	Free_Response	Controlled_Response
0.00	0.000	0.0000
0.02	-0.0001	-0.0001
0.04	-0.0012	-0.0012
0.06	-0.0039	-0.0039
0.08	-0.0081	-0.0081
0.1	-0.0137	-0.0137
0.12	-0.0211	-0.0211
0.14	-0.0305	-0.0305
0.16	-0.042	-0.042
0.18	-0.0554	-0.0553
0.2	-0.0702	-0.07
0.22	-0.0863	-0.0856
0.24	-0.1038	-0.1022
0.26	-0.1232	-0.12
0.28	-0.1447	-0.1387
0.3	-0.1675	-0.1577
0.32	-0.1909	-0.1758
0.34	-0.2139	-0.1924
0.36	-0.2356	-0.2067
0.38	-0.255	-0.2182
0.4	-0.2719	-0.227

Column1	Column2	Column3
4727	1097	0
4470	1064	1
4470	1047	1
4501	1014	1
4501	964	1
4449	931	1
4464	948	1
4438	1031	1
4433	948	1
4407	956	1
4396	1014	1
4381	1196	1
4349	1717	1
[illegible]	[illegible]	1
1988	1923	1
1993	1932	1

\# Data recorded on 01/02/91 at 694 stations. -99 means data not avail.
\# X and Y is Lat-Long mapped onto polar sterographic projection

X-coord	Y-coord	Temp(F)	Dewpt(F)	Press(Mb)	U(m/s)	V(m/s)
4158095.30	2769728.30	15.00	5.00	1020.60	-0.31	2.55
4206175.00	2711076.00	17.00	5.00	1022.30	-0.52	4.09
4157729.80	2479427.00	24.00	7.00	1025.70	-1.21	4.47
3925337.00	2395113.80	27.00	11.00	1026.70	-1.06	5.04
3975751.30	2386686.80	22.00	7.00	1025.40	-0.89	4.02
4151424.80	2401246.50	25.00	10.00	1026.70	0.68	2.48
4089275.00	2322289.50	27.00	4.00	1027.10	2.24	2.13
4105378.00	2298648.30	30.00	14.00	1027.80	3.60	0.22
4102765.80	2221281.00	25.00	4.00	1026.40	4.26	1.82
4157063.30	2257515.30	32.00	8.00	1029.10	3.03	4.16
4135176.50	2259278.80	30.00	10.00	1028.40	4.71	2.07

Column Data

- Univariate
  - One Parameter
- Bivariate
- Trivariate
- Hypervariate
  - Multi-parameter

Univariate

Temperature Variation
53.379
56.132
58.803
74.001
100.896
105.082
74.425
61.110
90.954
74.488
50.073
52.929
73.256
68.178
82.253
102.230
69.395
82.663
103.075
70.152
101.360
66.188

Hypervariate

Alpha Channel	Data Age	z-fact	Temperature Bias	Temperature Drift	Temperature Aging
0.9950	0.1987	0.1297	107.289	57.677	53.379
0.9801	0.3894	0.1739	106.358	69.561	56.132
0.9553	0.5646	0.2167	104.817	80.477	58.803
0.9211	0.7174	0.4607	102.682	89.991	74.001
0.8776	0.8415	0.8924	99.973	97.724	100.896
0.8253	0.9320	0.9596	96.718	103.366	105.082
0.7648	0.9854	0.4675	92.950	106.694	74.425
0.6967	0.9996	0.2538	88.705	107.573	61.110
0.6216	0.9738	0.7328	84.026	105.971	90.954
0.5403	0.9093	0.4685	78.961	101.949	74.488

Categorical Data

- Non-numeric
  - Data broken into groups
  - Examples: color, educational level, race, station, age group
- Numeric data can be divided into categories:
  - Low
  - Medium
  - High
- Often tabular column data with relations to numeric paramters

Test Site	Source	z-fact
alpha	red	0.1297
gamma	blue	0.1739
epsilon	blue	0.2167
delta	blue	0.4607
nu	blue	0.8924
omicron	blue	0.9596
lambda	blue	0.4675
kappa	blue	0.2538
delta	red	0.7328
nu	red	0.4685
omicron	red	0.0766
pi	red	0.1224
sigma	red	0.4487
tau	green	0.3672
gamma	red	0.5931
kappa	red	0.9138
delta	blue	0.3868
nu	blue	0.5997
omicron	green	0.9274

Arrays of Data

- Scales and Grids
- Numeric Types

5.444	7.323	9.629	11.170	12.655	13.687	14.453
5.119	7.170	8.985	10.786	12.467	13.742	14.693
4.458	6.180	8.355	10.584	12.381	13.777	14.852
4.144	5.485	7.459	9.768	11.900	13.605	14.910
3.706	4.821	6.489	8.466	10.817	13.017	14.750
3.457	4.145	5.637	7.342	9.451	11.882	14.166
3.044	3.397	4.697	6.352	8.182	10.364	12.903
2.763	2.790	3.698	5.181	6.960	8.863	11.123
3.049	2.520	2.905	3.994	5.607	7.395	9.330

Grids

- Uniform grids and no grids

	0.5	1	1.5
0.5	0.0350	0.4911	0.5744
1.0	0.0714	0.7477	0.3305
1.5	0.3853	0.9207	0.8485

- Non-uniform grids

	0.5	0.7	1.8
0.5	0.0350	0.4911	0.5744
1.0	0.0714	0.7477	0.3305
1.1	0.3853	0.9207	0.8485

- Warped grids

0.0350 (0.5, 0.5)	0.4911 (1.0, 0.7)	0.5744 (1.5, 0.5)
0.0714 (0.7, 1.0)	0.7477 (1.0, 1.0)	0.3305 (1.3, 1.0)
0.3853 (0.5, 1.5)	0.9207 (1.0, 1.2)	0.8485 (1.5, 1.5)

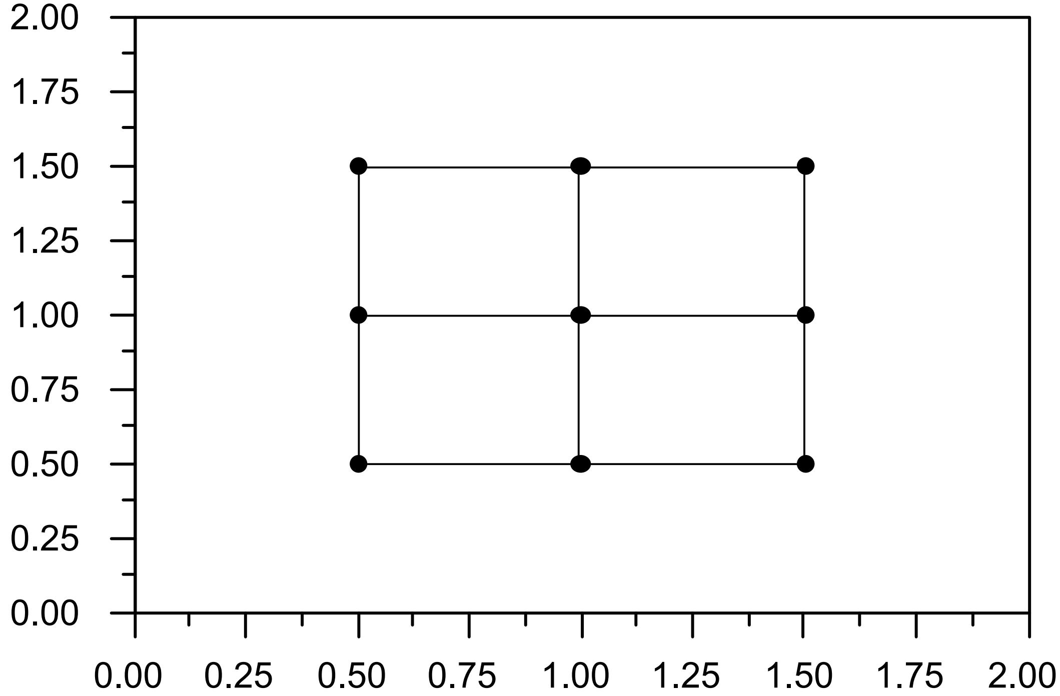
- Sparse grids

	0.5	1	1.5
0.5	NaN	0.4911	0.5744
1.0	0.0714	0.7477	NaN
1.5	0.3853	NaN	0.8485

Grids

- Uniform grids and no grids

	0.5	1	1.5
0.5	0.0350	0.4911	0.5744
1.0	0.0714	0.7477	0.3305
1.5	0.3853	0.9207	0.8485

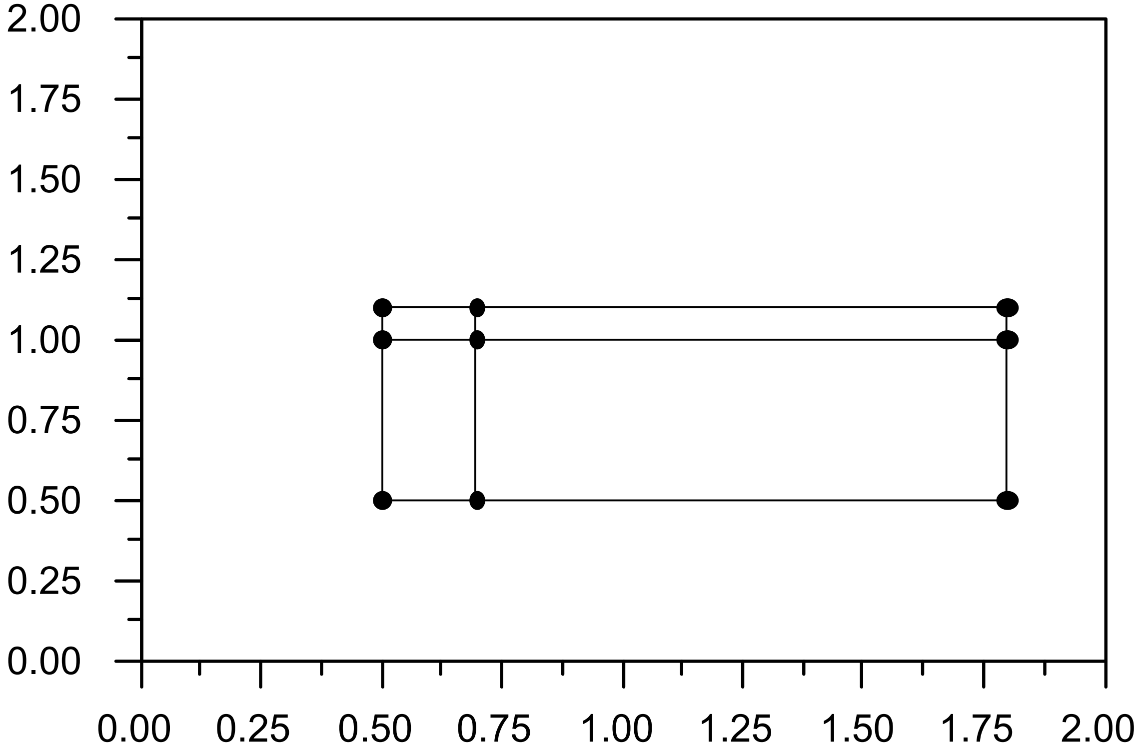


X	Y	
0.5	0.5	0.0350
0.5	1	0.4911
0.5	1.5	0.5744
1	0.5	0.0714
1	1	0.7477
1	1.5	0.3305
1.5	0.5	0.3853
1.5	1	0.9207
1.5	1.5	0.8485

Grids

- Non-uniform grids

	0.5	0.7	1.8
0.5	**0.0350**	**0.4911**	**0.5744**
1.0	**0.0714**	**0.7477**	**0.3305**
1.1	**0.3853**	**0.9207**	**0.8485**

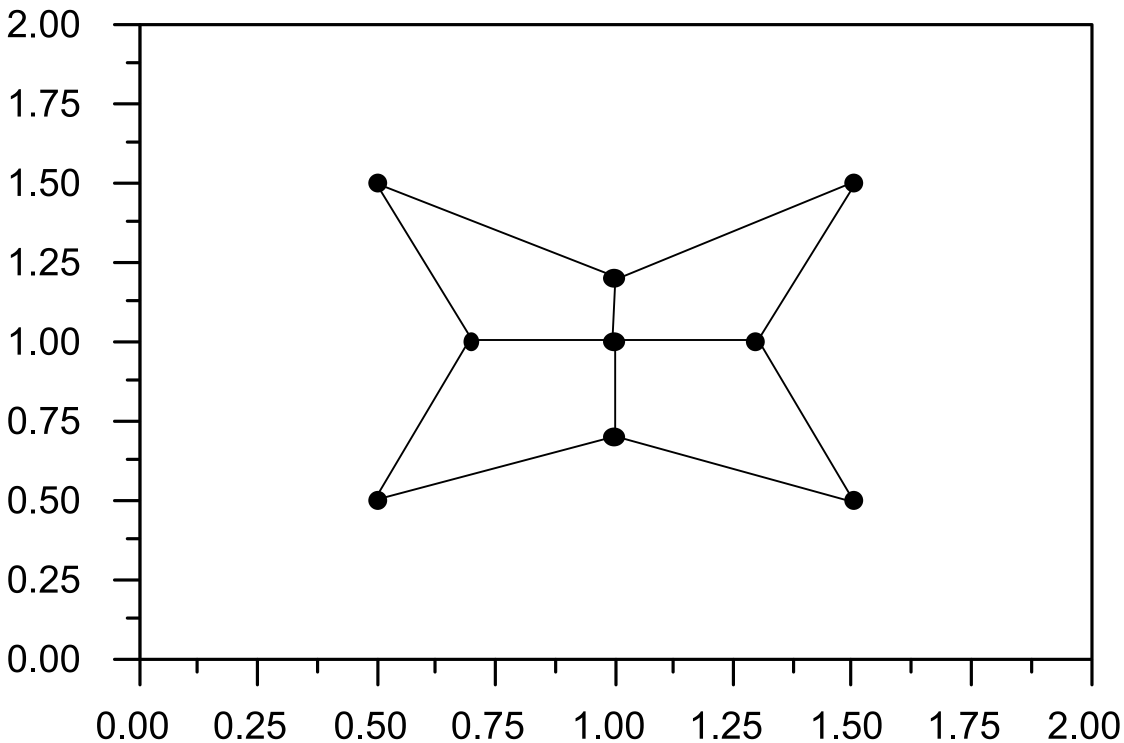


X	Y	
0.5	**0.5**	**0.0350**
0.5	**1**	**0.4911**
0.5	**1.1**	**0.5744**
0.7	**0.5**	**0.0714**
0.7	**1**	**0.7477**
0.7	**1.1**	**0.3305**
1.8	**0.5**	**0.3853**
1.8	**1**	**0.9207**
1.8	**1.1**	**0.8485**

Grids

- Warped grids

0.0350 (0.5, 0.5)	0.4911 (1.0, 0.7)	0.5744 (1.5, 0.5)
0.0714 (0.7, 1.0)	0.7477 (1.0, 1.0)	0.3305 (1.3, 1.0)
0.3853 (0.5, 1.5)	0.9207 (1.0, 1.2)	0.8485 (1.5, 1.5)

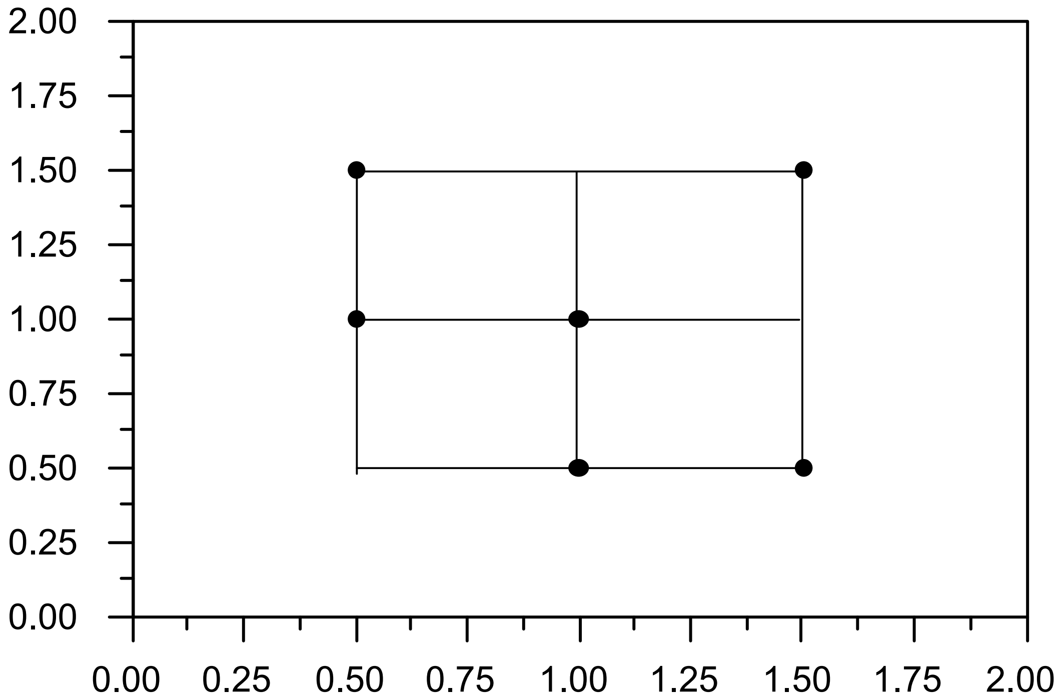


X	Y	
0.5	0.5	0.0350
1	0.7	0.4911
1.5	0.5	0.5744
0.7	1	0.0714
1	1	0.7477
1.3	1	0.3305
0.5	0.15	0.3853
1	1.2	0.9207
1.5	1.5	0.8485

Grids

- Sparse grids

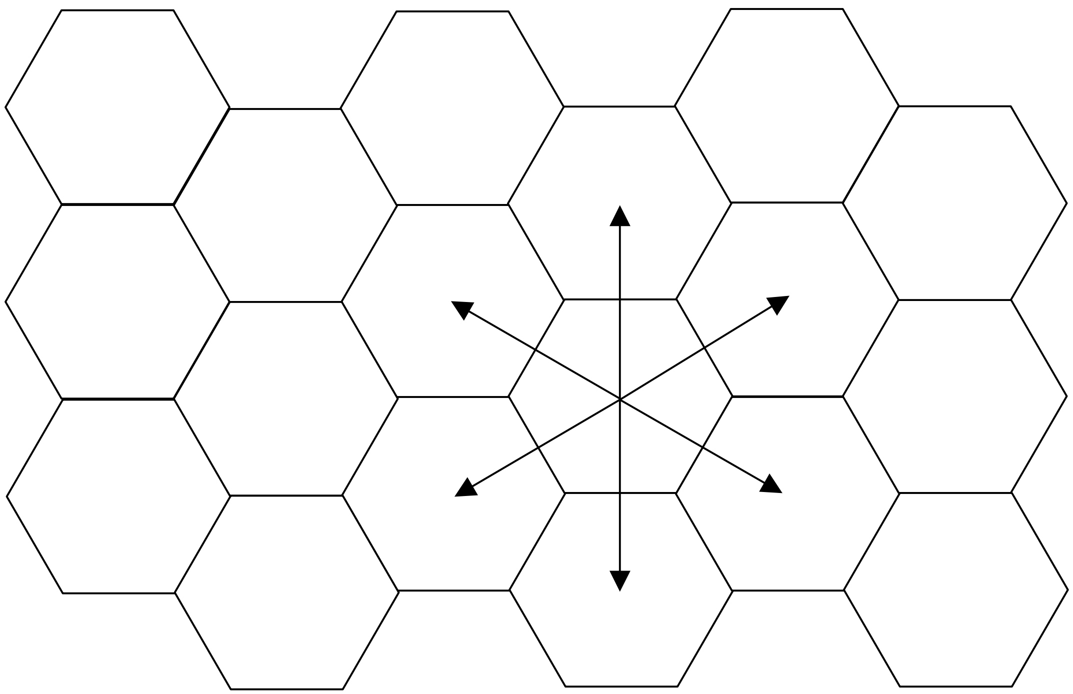
	0.5	1	1.5
0.5	NaN	0.4911	0.5744
1.0	0.0714	0.7477	NaN
1.5	0.3853	NaN	0.8485

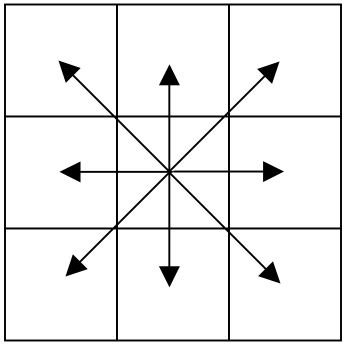


X	Y	
0.5	0.5	NaN
0.5	1	0.4911
0.5	1.5	0.5744
1	0.5	0.0714
1	1	0.7477
1	1.5	NaN
1.5	0.5	0.3853
1.5	1	NaN
1.5	1.5	0.8485

Grids

- Hex grids





Grids & Scales

- Scales Support Gridding

Indices

Uniform Gridding

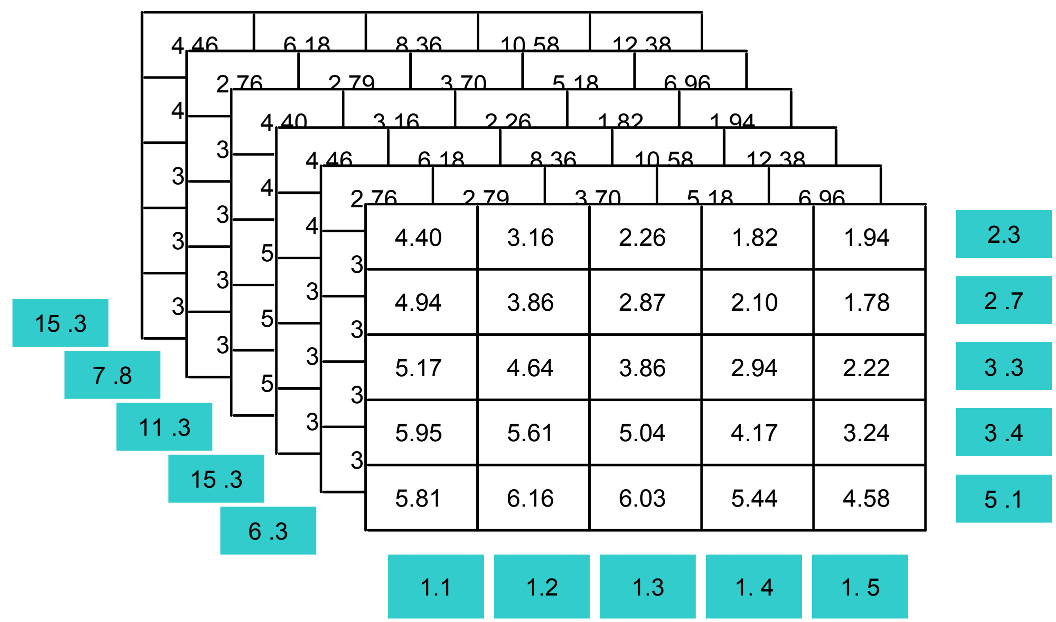
Etc…

	1.1	1.2	1.3	1. 4	1. 5	1. 6	1. 7
2.3	5.444	7.323	9.629	11.170	12.655	13.687	14.453
2 .7	5.119	7.170	8.985	10.786	12.467	13.742	14.693
3 .3	4.458	6.180	8.355	10.584	12.381	13.777	14.852
3 .4	4.144	5.485	7.459	9.768	11.900	13.605	14.910
5 .1	3.706	4.821	6.489	8.466	10.817	13.017	14.750
6 .3	3.457	4.145	5.637	7.342	9.451	11.882	14.166
7 .8	3.044	3.397	4.697	6.352	8.182	10.364	12.903
11 .3	2.763	2.790	3.698	5.181	6.960	8.863	11.123
15 .3	3.049	2.520	2.905	3.994	5.607	7.395	9.330

Volumes of Data

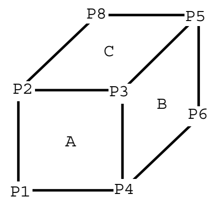
- Large I/O and processing requirements
- Specialized Visualization

And Scales



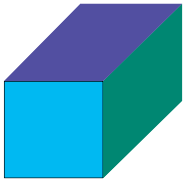
Visualization Polygonal Data

- Positioning of polygons in 3-space



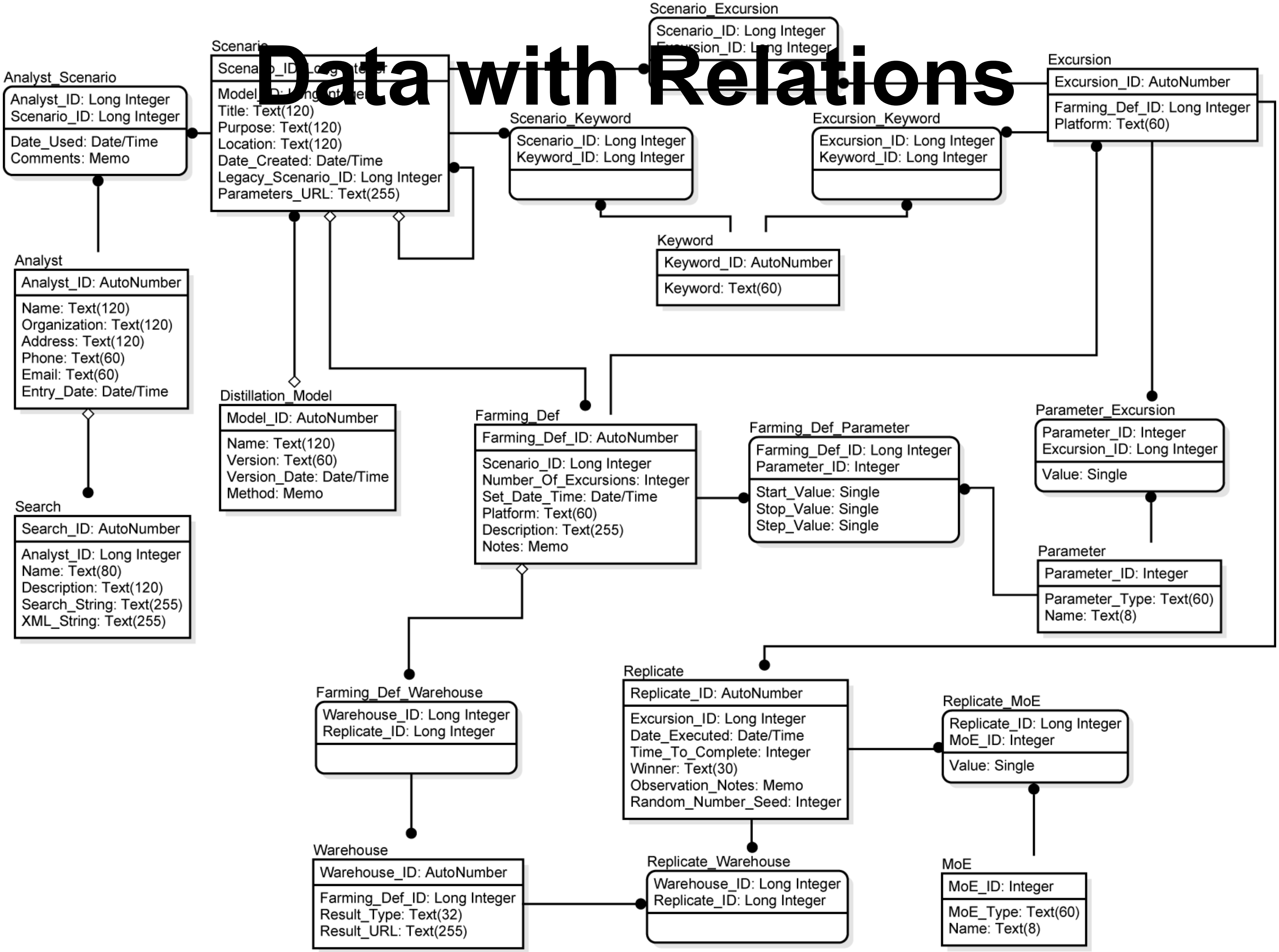
Polygon Name	Nodes	Nearest Neighbors	Temperature	Stress
A	P1,P2,P3,P4	B,C,D,E	34.7	.023
B	P4,P3,P5,P6	A,C,E,F	23.1	.028
C	P2,P3,P5,P8	A,B,D,F	24.5	.024
D	P1,P2,P7,P8	A,C,E,F	29.4	.033
E	P1,P4,P6,P7	A,B,D,F	28.6	.023
F	P5,P6,P7,P8	B,C,D,E	31.9	.031

- Polygons may be colored according to data value



- Requires significant computational resources
- Animation

Data with Relations



Files of Data

- Proprietary Formats
- Images Files
- Science Data Formats
- Multi-object files
- Linking Metadata to Data

Collections of Data

- Databases
- Files and Directory Structures
- Tapes
- Hybrid Systems
- Boxes in Corners

Metadata

- Metadata is Data
- One person’s metadata is another person’s data
- Types of Metadata
  - Structural
  - Attribute - Search and Labeling
  - Descriptive

- XML

Metadata Example

Human Readable

…Image of object GL 388 taken on April 18, 1990 with the IRTF telescope, using the ProtoCAM. The image is centered on 10 degrees, 16 minutes, 53 seconds right ascension, 20 degrees, 7 minutes, 21 seconds declination, using the 1950 epoch…

Computer Readable

```
DATE    = '26-04-90' /
ORIGIN  =    'UH IFA' / INSTITUTION WRITING THE DATA
TELESCOP=  'NASA IRTF' / DATA ACQUISITION TELESCOPE
INSTRUME=   'ProtoCAM' / DATA ACQUISITION INSTRUMENT
OBSERVER=       'BBLW' / OBSERVER NAME/IDENTIFICATION
DATE_OBS=   '18/04/90' / DATE OF ACQUISITION ('dd/mm/yy')
TIME_OBS= '20:33:26.77' / TIME OF ACQUISITION(hh:mm:ss.ss)
ITIME   =        20.00 / INTEGRATION TIME IN SECONDS
FILTER  =            0 / 0=BROADBAND 1=CVF
WAVE_LEN=         2.20 / WAVELENGTH IN MICRONS
PLATE_SC=         0.25 / PLATESCALE
RA      = '10:16:53.92' / RIGHT ASCENSION in degree
DEC     = '20:07:21.8' / DECLINATION in degree
EPOCH   =       1950.0 / EPOCH
AIRMASS =        1.003 / AIRMASS
OBJECT  =     'GL 388'
COMMENT =  'k=4.6 l=4.6'
VGATE   =        -2.30 / SBRC ARRAY GATE VOLTAGE
```

XML

Case Studies:
ASCI and EOSDIS

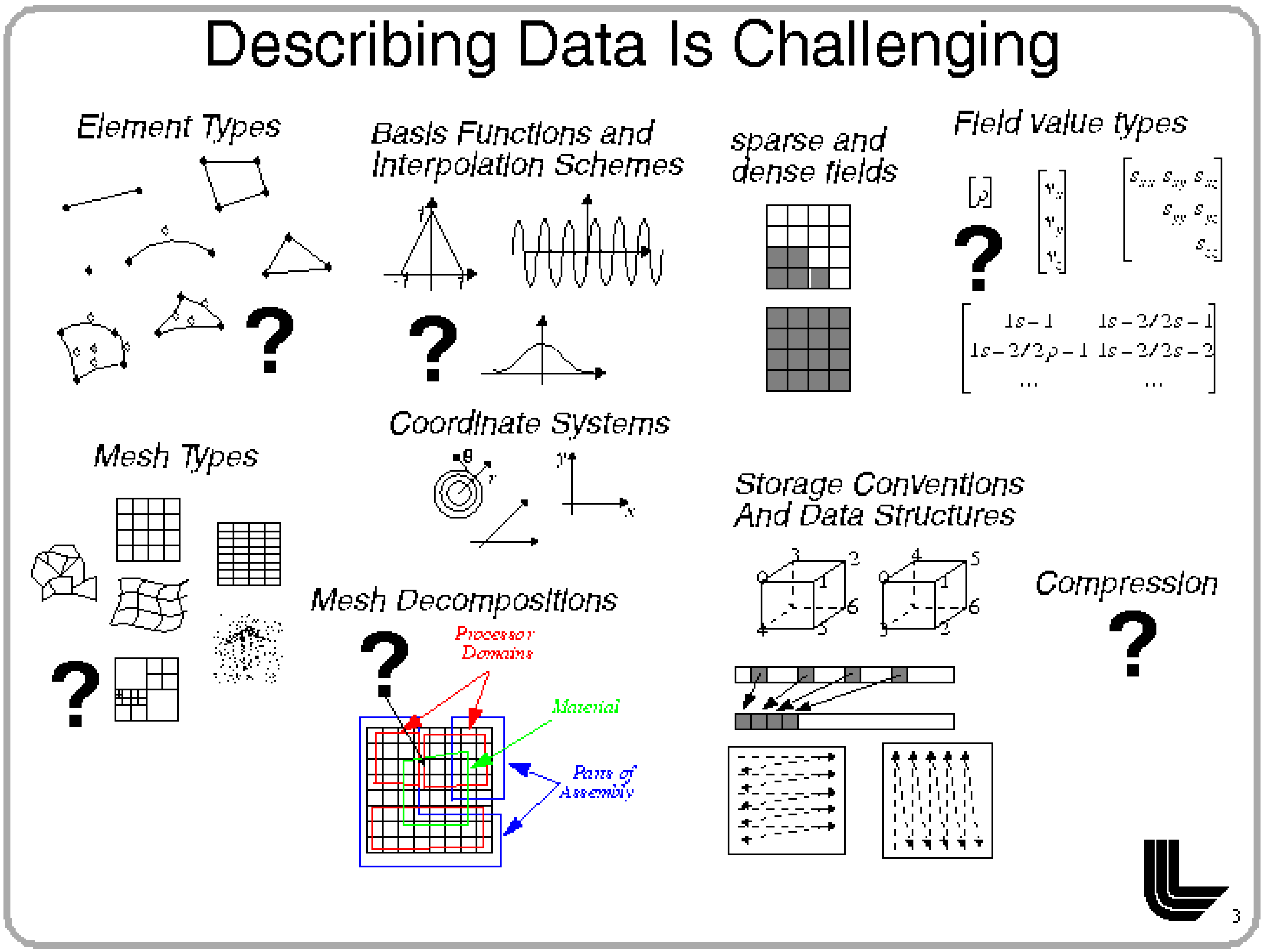
ASCI

- Accelerated Strategic Computing Initiative
- Comprehensive Test Ban Treaty, 1992
- ASCI's vision:
  - “to shift promptly from nuclear test-based methods to computational-based methods of ensuring the safety, reliability, and performance of our nuclear weapons stockpile.”

ASCI Data Requirements

- Computational mechanics: meshes & fields
- Sound data model w. robust data abstractions
- Common format allows
  - common tools
  - sharing
- Common appl'n programming interface (API)
  - shield apps from model complexities
  - standardize data organization and semantics

ASCI Datatypes



EOSDIS: Understanding Global Climate Change

EOSDIS Processing Levels

EOSDIS Example: Library Analogy

Example Categories for Granule- and Collection-Level Metadata

Granule

Platform, Instrument, Sensor

Spatial and Temporal

Orbit Parameters

Browse

QA Data Statistics

Production History

Collection

Platform, Instrument, Sensor

Delivered Algorithm Package

Guide

Bibliographic Reference

Papers/Documents

Keyword

Biases

- Disciplines
  - Geospatial, simulation
- Data structures
  - large multidimensional structures, multi-layered structures, meshes, some indexed structures
- Geometry
  - space and time
- Operations
  - visualization, partial access, filtering, integration

What is Scientific Data?

What is scientific data?

- A variety of data types and structures
- Large data structures
- Many objects
- Metadata: parameters, variables, legacy in a variety of forms