Feature Detection in Data Analysis

University of Houston, Spring 2013

Instructor: Guoning Chen

Course Information

- **Location:** SEC 201
- **Time:** 10am~11:30am Tuesday and Thursday
- **Office Hours**: 11:30am~12:30pm Tuesday and Thursday, or by appointment.
- **Office:** PGH 566
- **Email:** chengu@cs.uh.edu
- **Course webpage**: http://www2.cs.uh.edu/~chengu/Teaching/Spring2013/Analysis_Spring2013.html
- **TA:** Cheng Xiao

Course Information

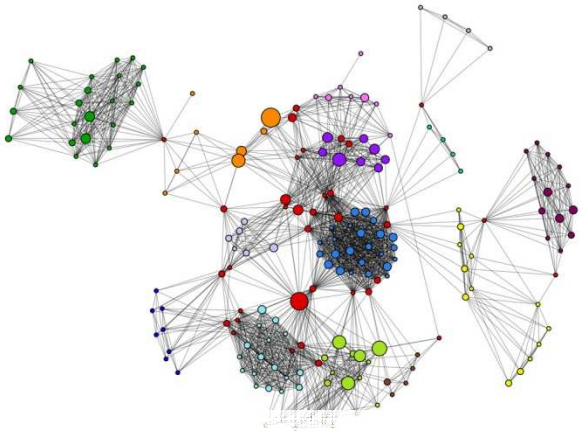
- **Prerequisites**:
  - Algorithms and data structures
  - Knowledge and experience in programming (we will use C/C++, other languages are possible, please talk to me)
  - Computer graphics and OpenGL (will provide a review in next lecture), or other library with graphics functionality
  - Mathematics (as a start you will need linear algebra, numerical analysis, and calculus)

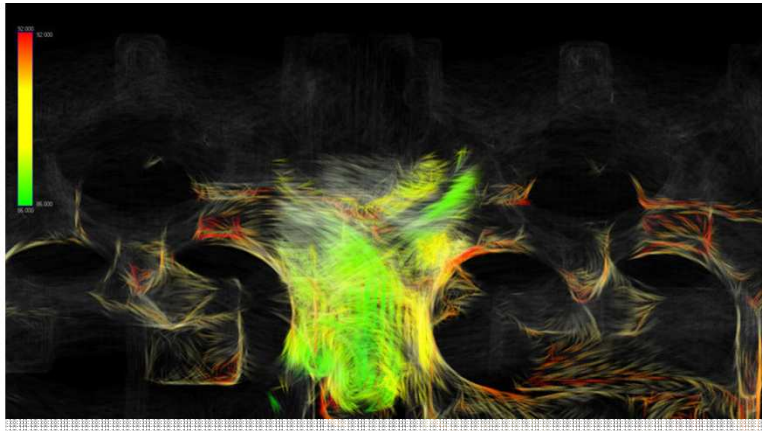
- **Textbook:** (recommended)
  - Topology-Based Methods in Data Analysis and Visualization: Theory, Algorithms, and Applications (Mathematics and Visualization) II, R. Peikert, H. Hauser, H. Carr, and R. Fuchs (Eds), Springer, 2012.
  - Topology-Based Methods in Data Analysis and Visualization: Theory, Algorithms, and Applications (Mathematics and Visualization), V. Pascucci, X. Tricoche, H. Hagen, and J. Tierny (Eds), Springer, 2010.
  - Visualization and Mathematics III (Mathematics and Visualization) (v.3) by Hans-Christian Hege and Konrad Polthier (Aug 13, 2003), Springer
  - Topology-Based Methods in Visualization II (Mathematics and Visualization) (v.2) by Hans-Christian Hege, Konrad Polthier and Gerik Scheuermann (Dec 18,2008), Springer.

Course Information

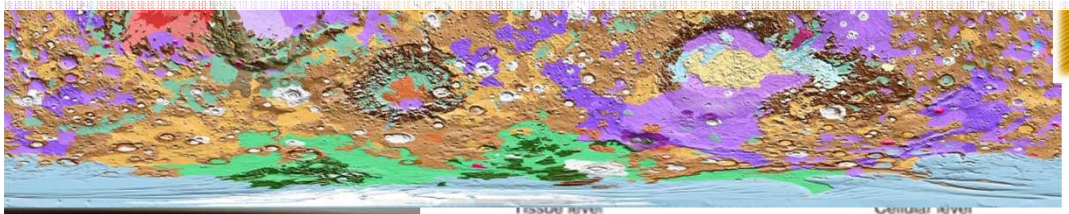
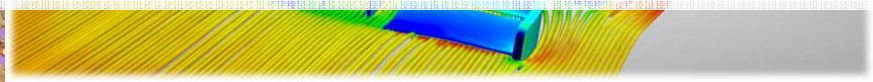
- **Grading:** 3 assignments (50%) + 1 final project and presentation (40%) + course participation (10%)
- **Late Policy**: Late assignments will be marked off 20% for each weekday that it is late.
- **Academic Dishonesty:** Do your own work! No code copy!

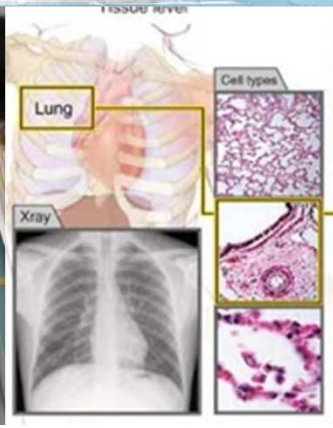
Background

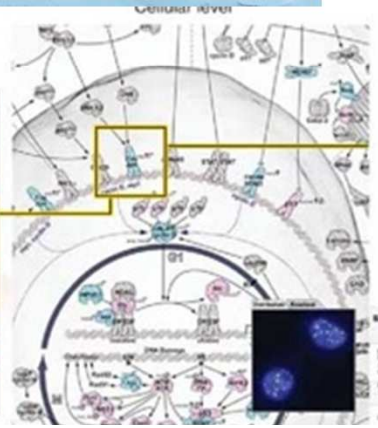
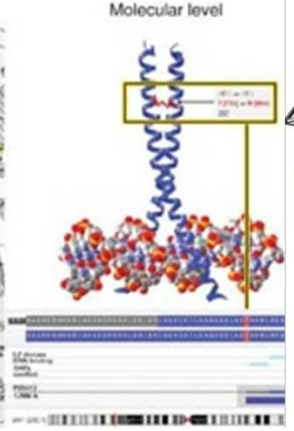


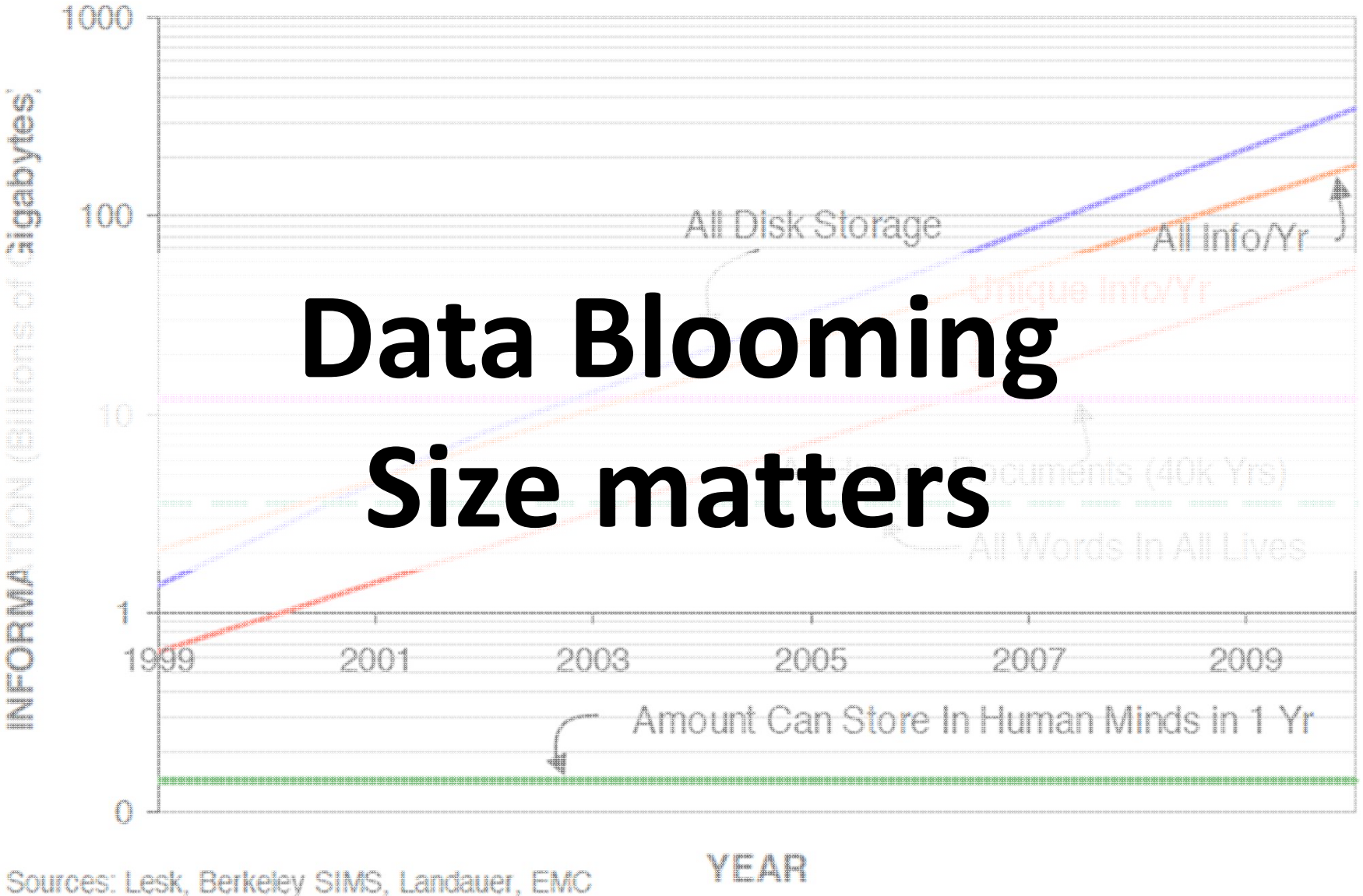


Data is generated everywhere



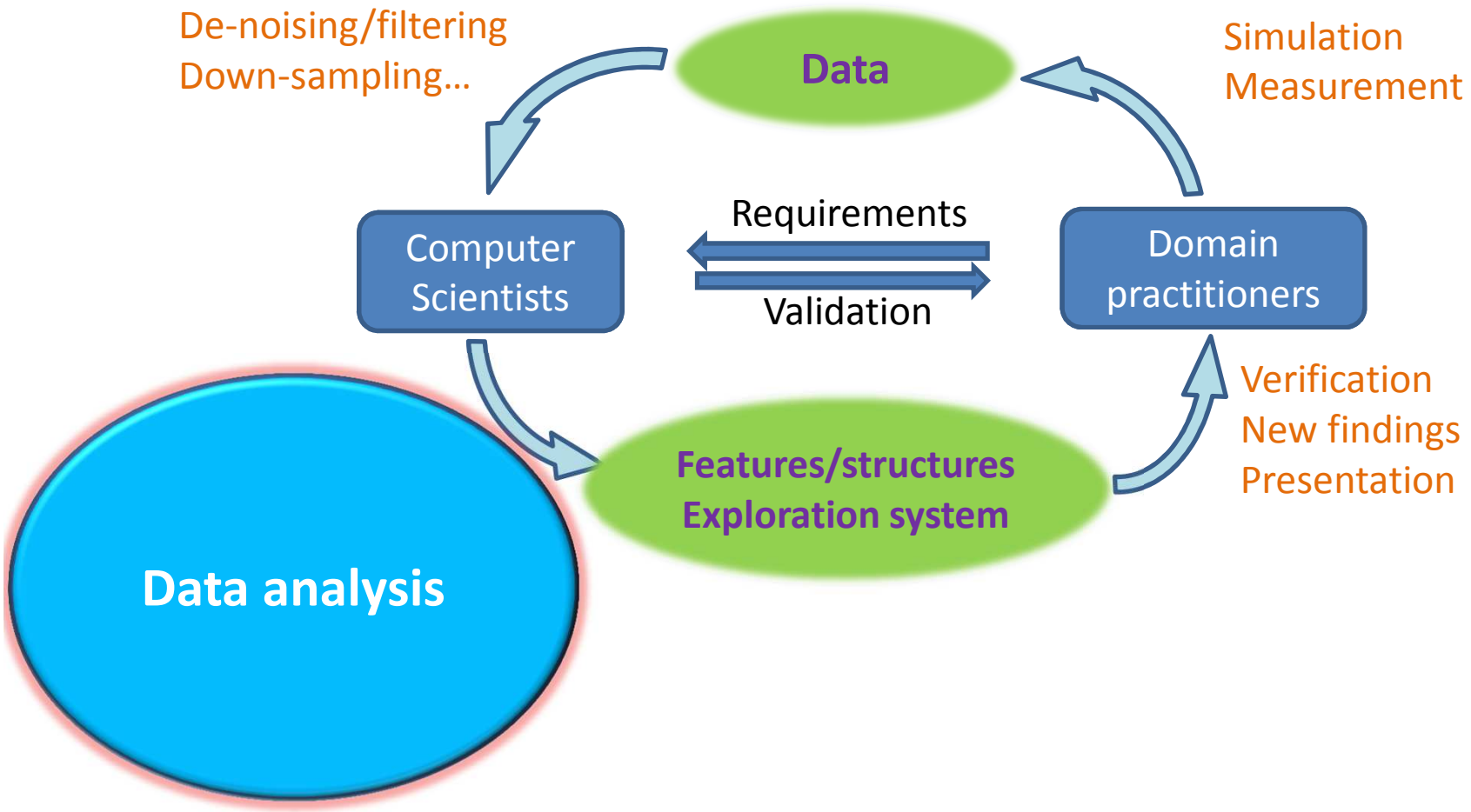





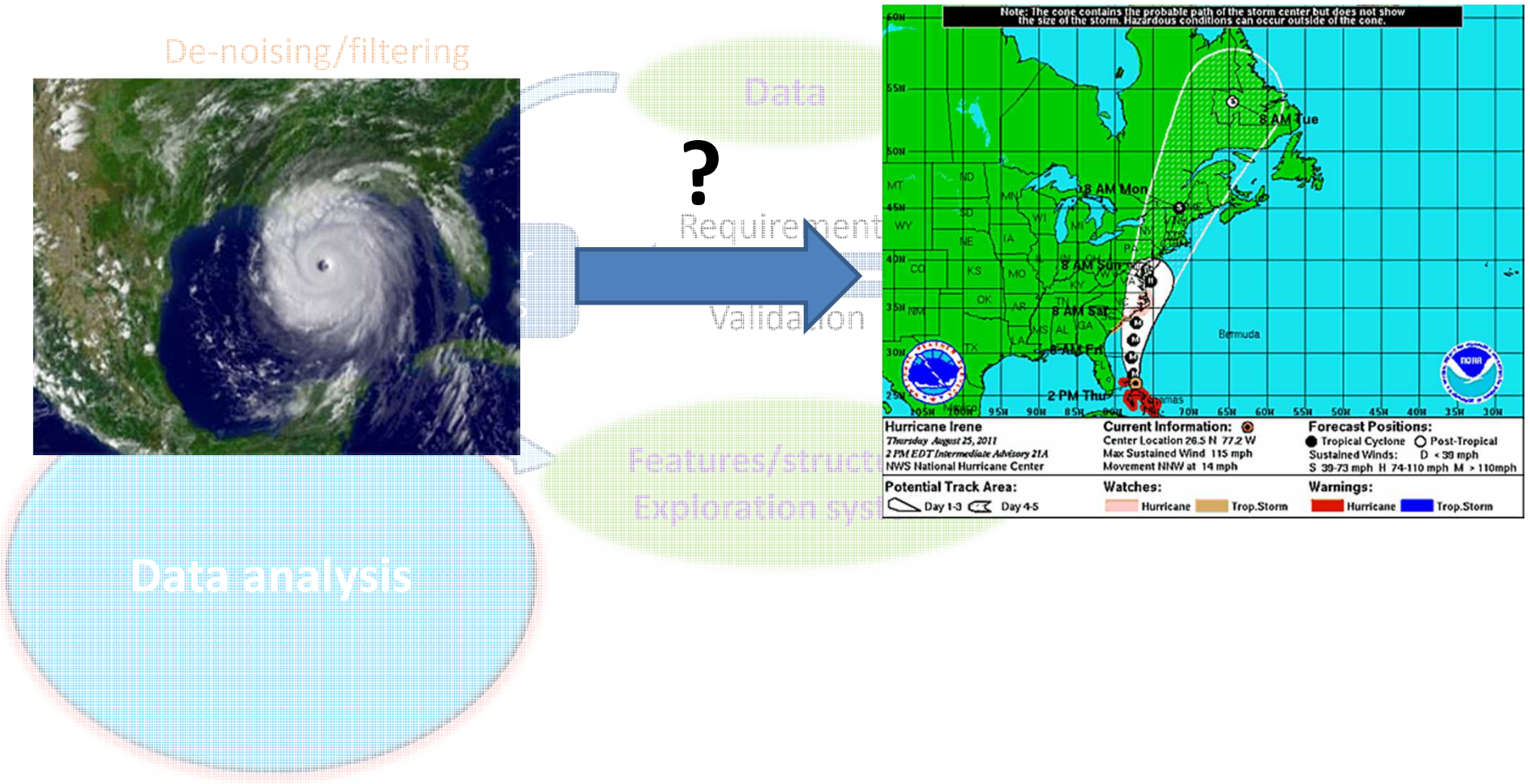
Data Blooming
Size matters

Sources: Lesk, Berkeley SIMS, Landauer, EMC

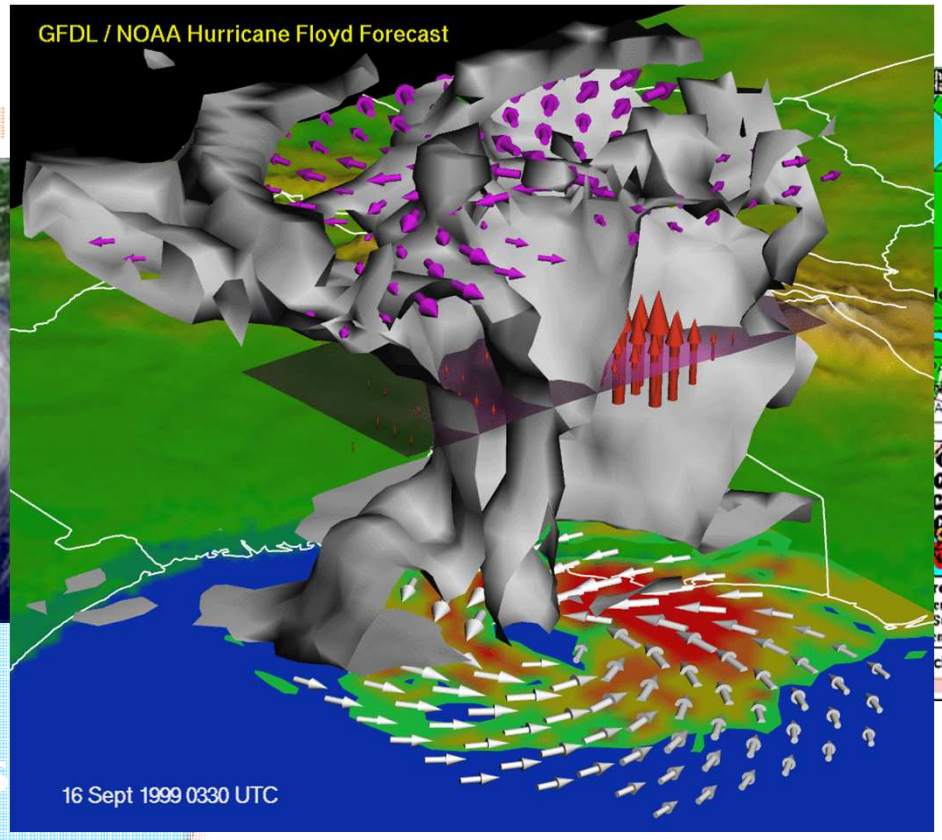
Data Analysis is important!

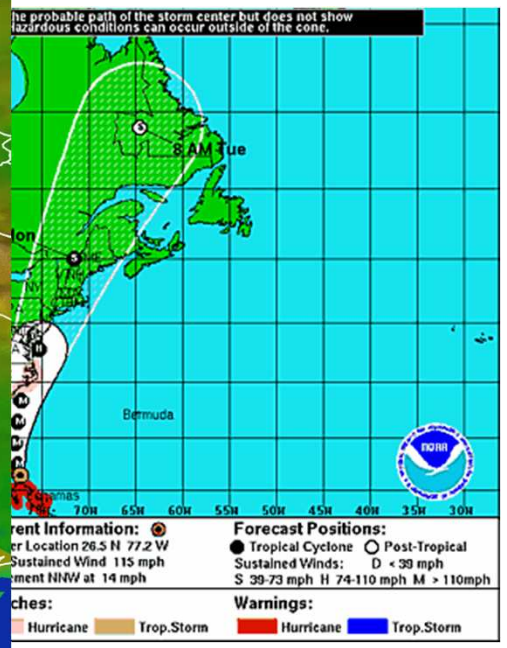


Data Analysis is important!

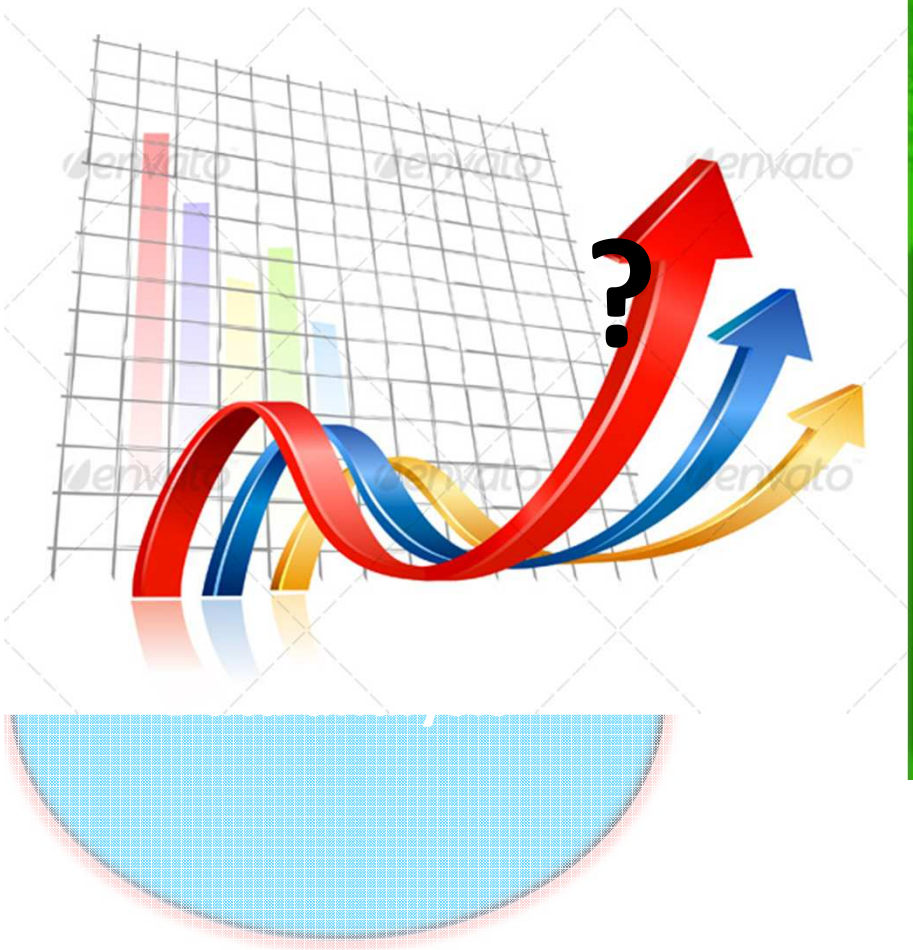


Data Analysis is important!

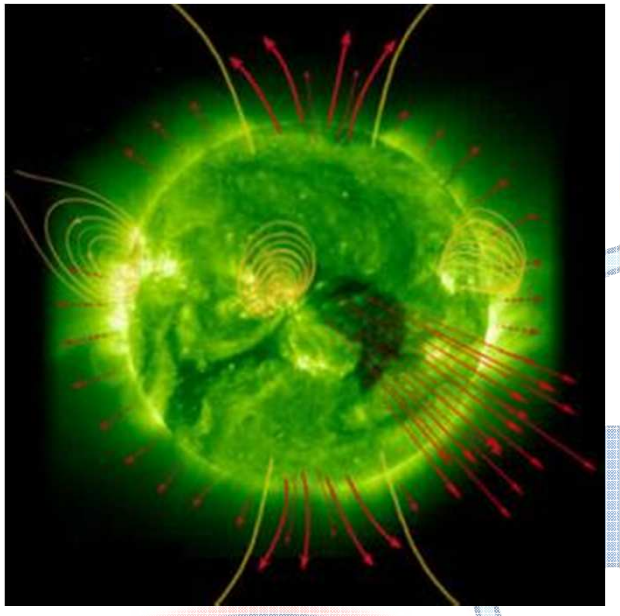


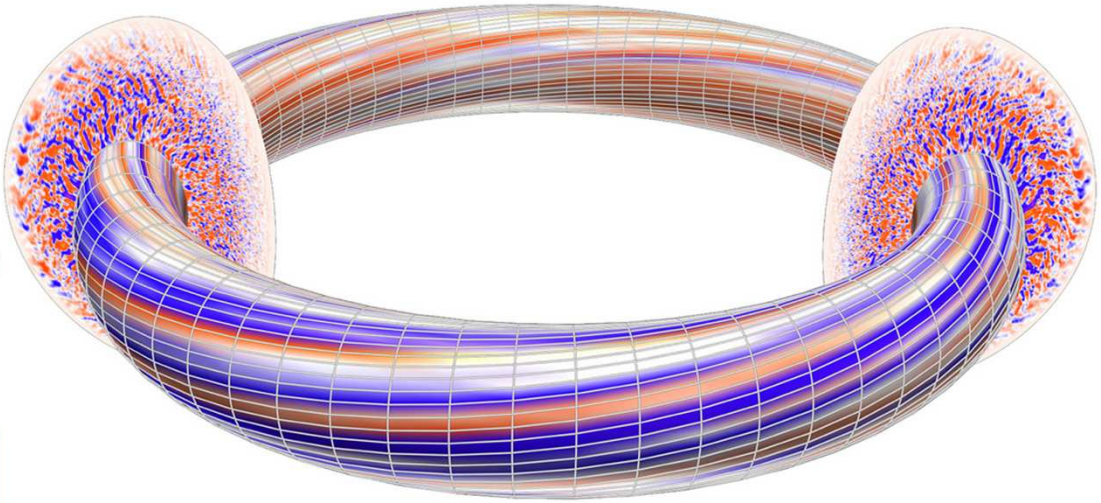


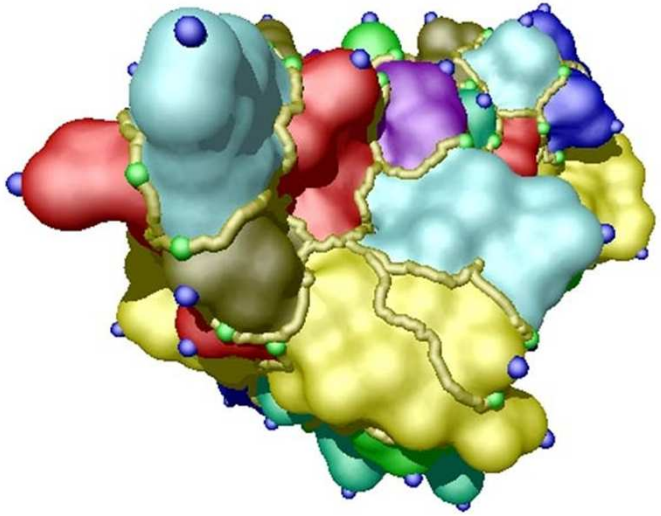
Data Analysis is important!

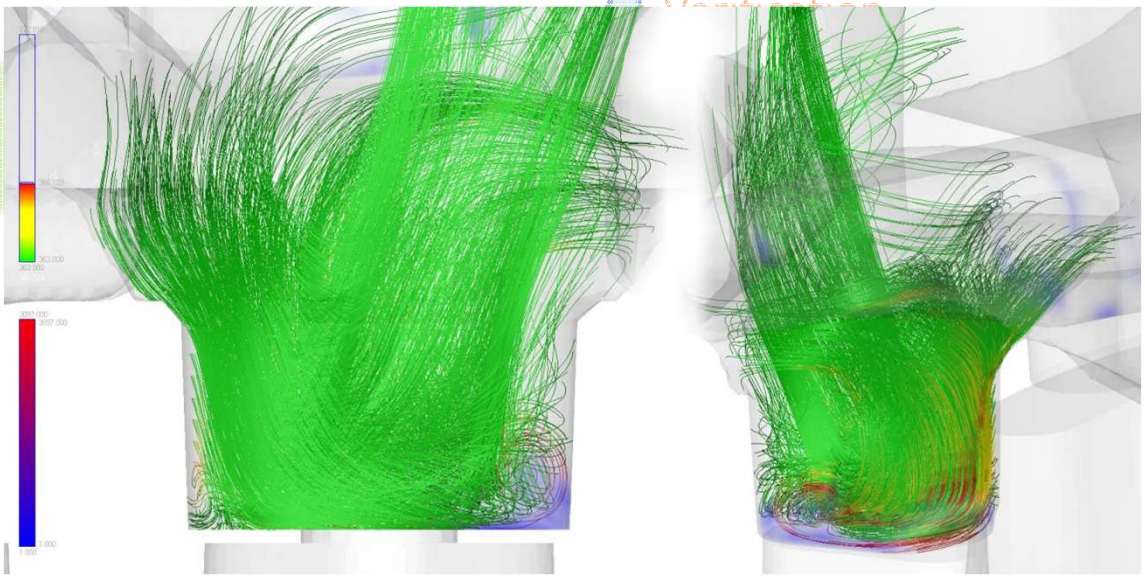


Data Analysis is important!









What is Data Analysis?

- From wiki:
  - **Analysis of data** is a process of inspecting, cleaning, transforming, and modeling **data** with the **goal** of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facts and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

Different Data Analysis

- Depending on the type of data, there is **no a universal analysis framework** that works for all data!

- Structured vs. unstructured or semi-structured data analysis
  - *structured data*. Data that resides in fixed fields within a record or file. Relational databases and spreadsheets are examples of *structured data*. (clustering by attribute values)
  - *unstructured data*. Information that does not have a data model. Text data such as books, journals, documents, emails, social media data, and audio are examples of *unstructured data*.
  - Analysis techniques: **database techniques**, **data mining** and **text analytics**, respectively.

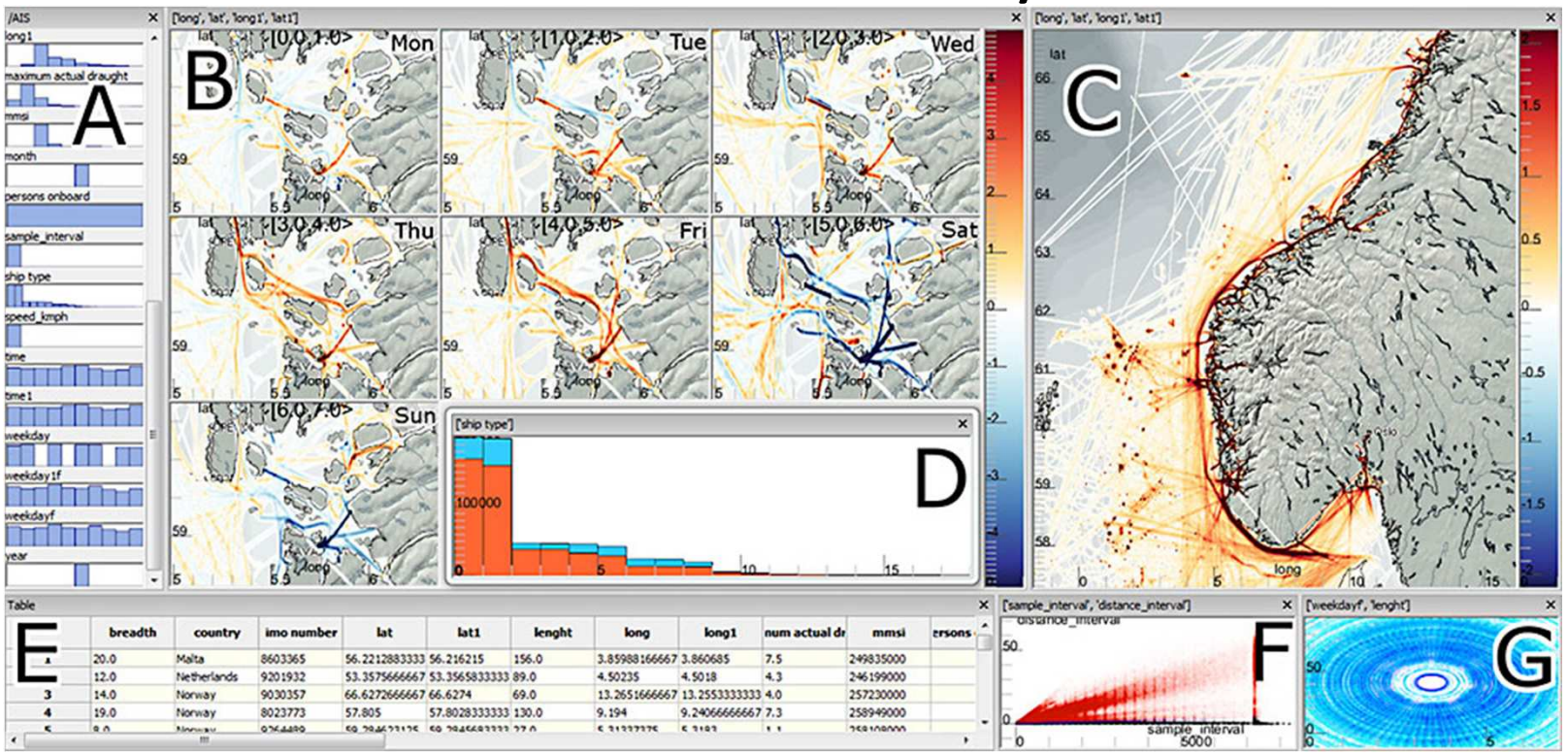
Image Analysis

- The data is already in the image form
- Digital image analysis
- Techniques: pattern recognition, digital geometry, and signal processing.

Data analysis has close link to data visualization

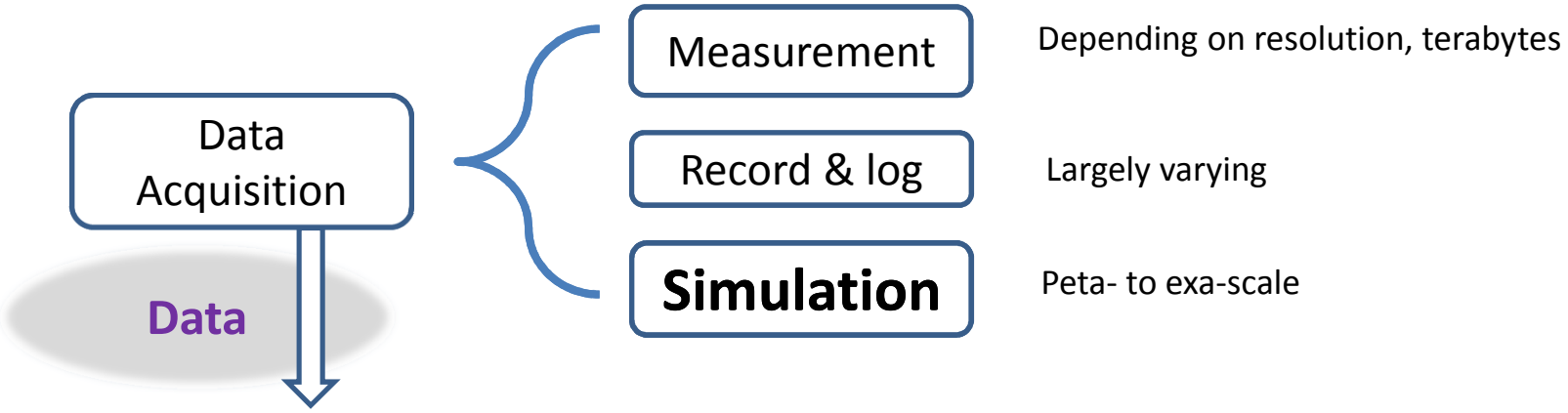
- Either the analysis results need to be visualized for efficient interpretation
  - A has relation with B, B has relation with C, C has relation with D, both D and C have relations with A
- Or the analysis is performed via visualization interface (visual exploration)

Data visual analytic

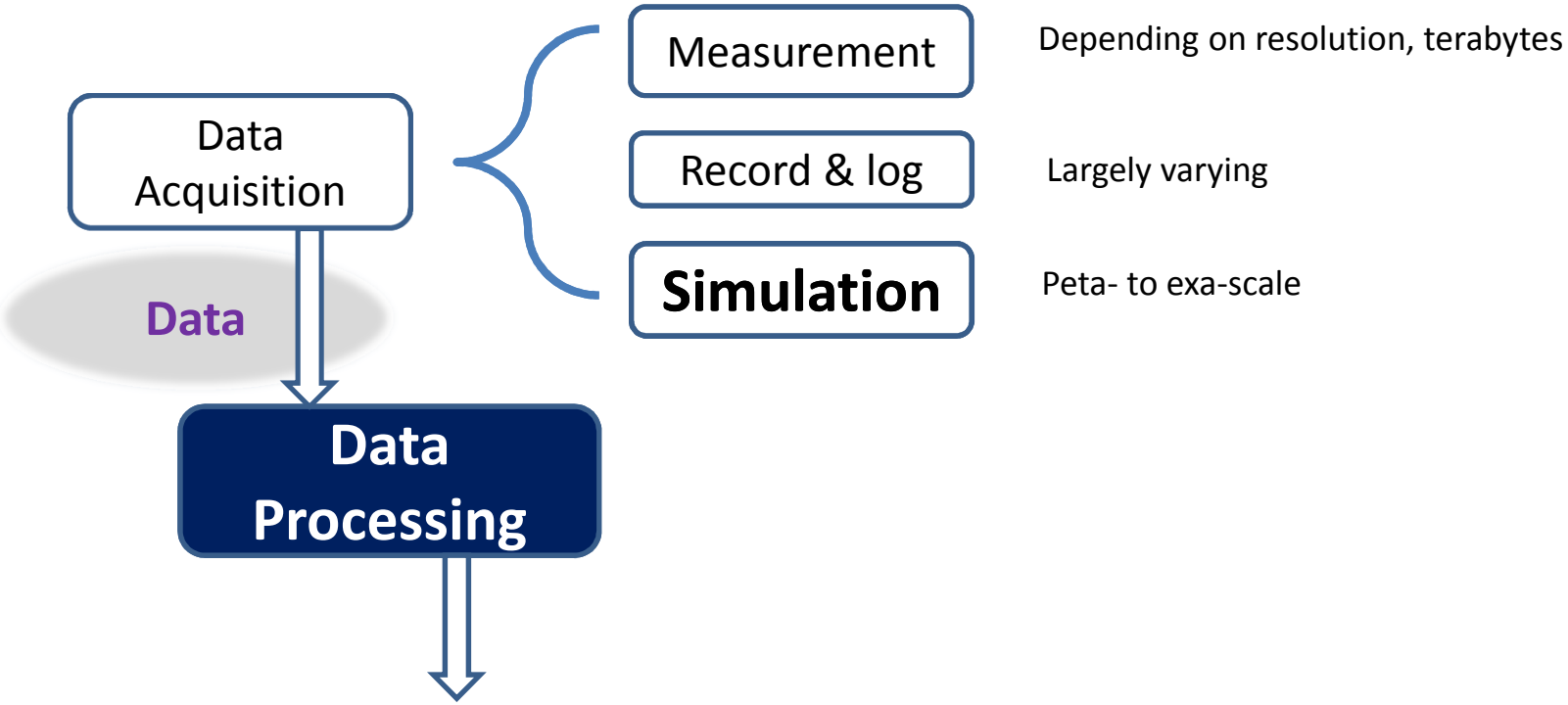


Source: http://www.ii.uib.no/vis/publications/publication/2010/lampe10differenceViews

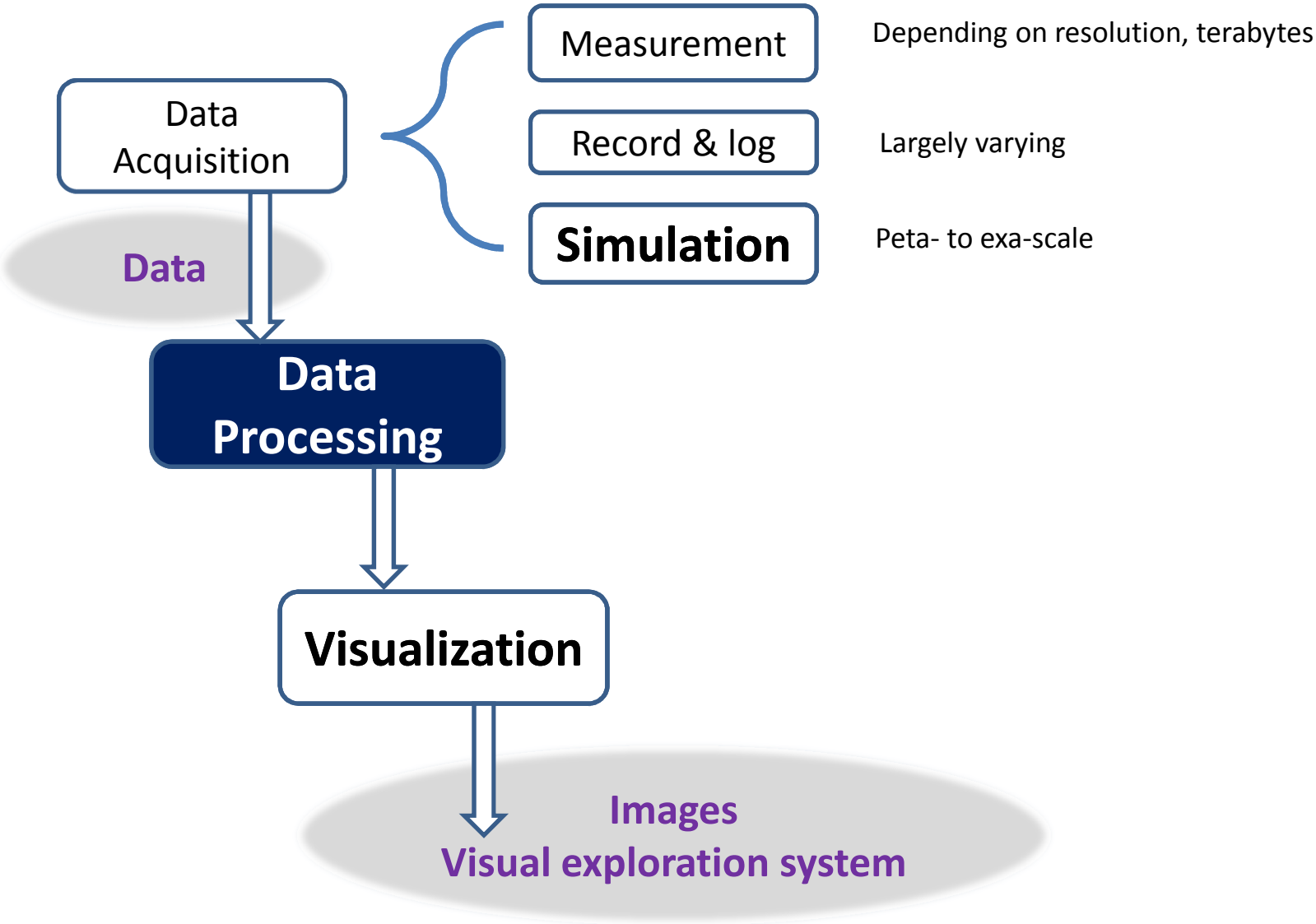
Data Visual Analytic Pipeline



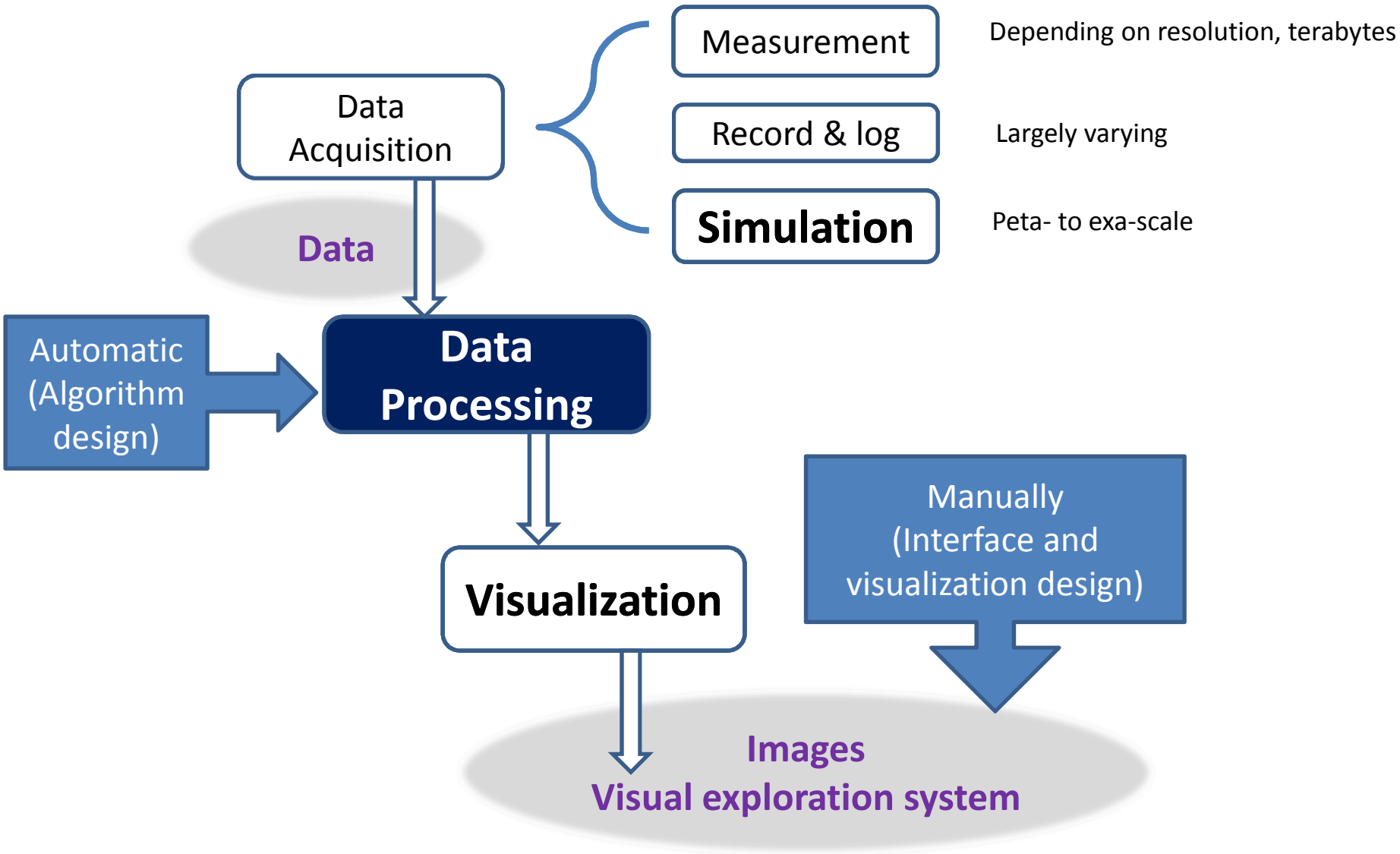
Data Visual Analytic Pipeline



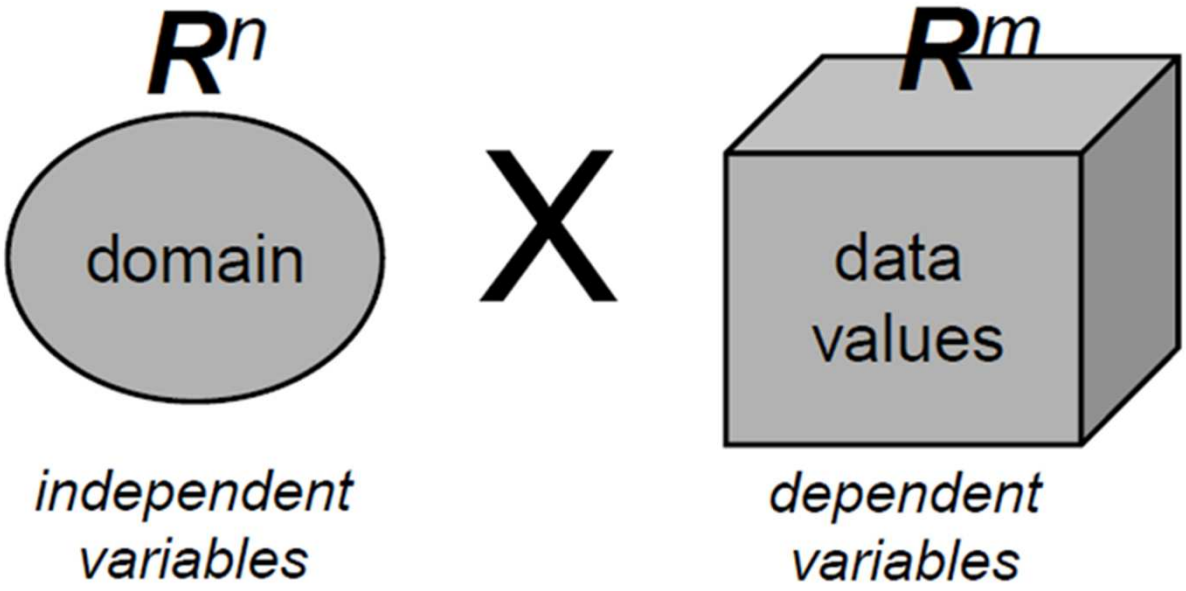
Data Visual Analytic Pipeline



Data Visual Analytic Pipeline



Data we are discussing in the class



Source: VIS, University of Stuttgart

Scientific data	3**D**+time (n<4)	Scalar/vector/tensor
Information data	n**D** (n>3)	Heterogeneous

SciVis vs. InfoVis

- Scientific visualization is mostly concerned with:
  - 2, 3, 4 dimensional, spatial or spatio-temporal data
  - discretized or sampled data (for continuous function)
- Information visualization focuses on:
  - high-dimensional, abstract data
  - discrete data
  - financial, statistical, etc.
  - visualization of large trees, networks, graphs
  - data mining: finding patterns, clusters, voids, outliers

By Ronald Peikert

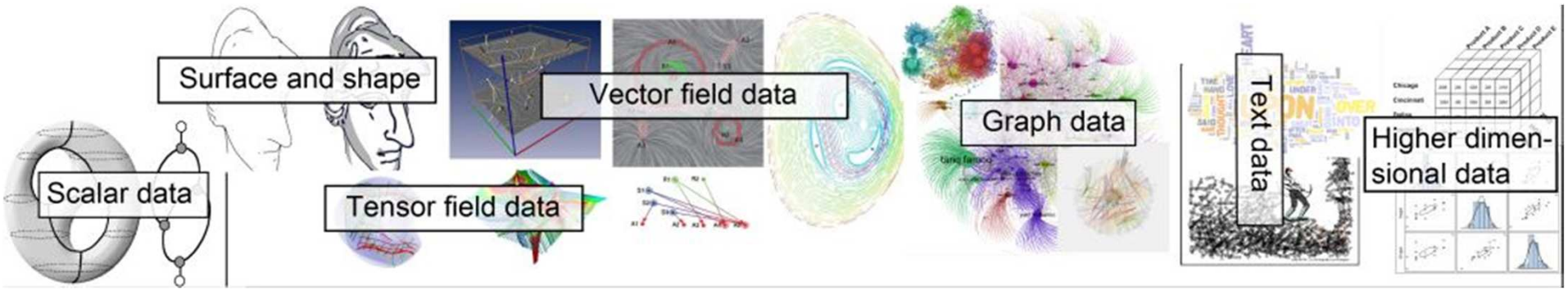
Goals and Topics

Goals of this Course

- Understand the different characteristics of different data types.
- Familiar with classical techniques for the analysis of various data types.
- Able to develop the customized analysis techniques and systems for the practical and your research needs.

Topics to Be Covered in this Course

- What will be covered?
  - Scalar data analysis for geometric data
  - Vector-valued data analysis for flow data
  - Tensor data analysis for medical and flow data
  - Graph data analysis
  - Text analytic
  - Higher dimensional data analysis



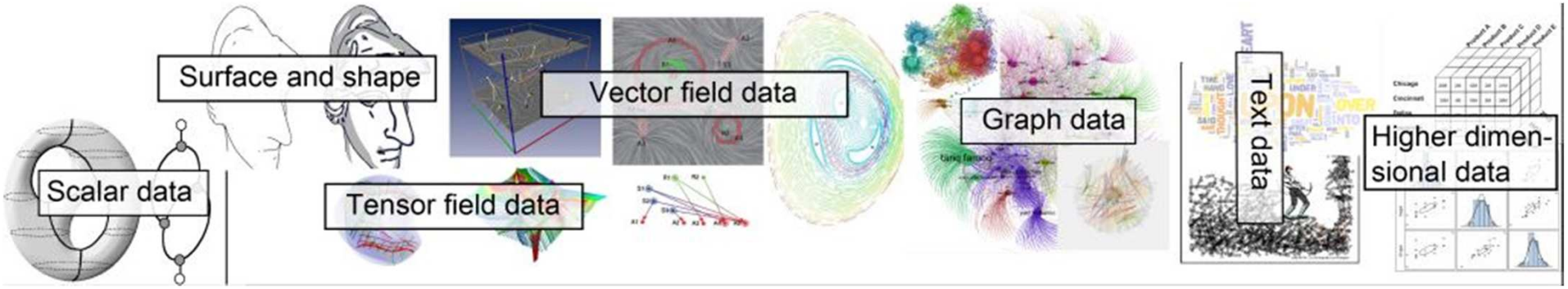
Topics to Be Covered in this Course

- What will be covered?
  - Scalar data analysis for geometric data
  - Vector-valued data analysis for flow data
  - Tensor data analysis for medical and flow data
  - Graph data analysis
  - Text analytic
  - Higher dimensional data analysis

Scientific data

Information data

- What types of data do you have?

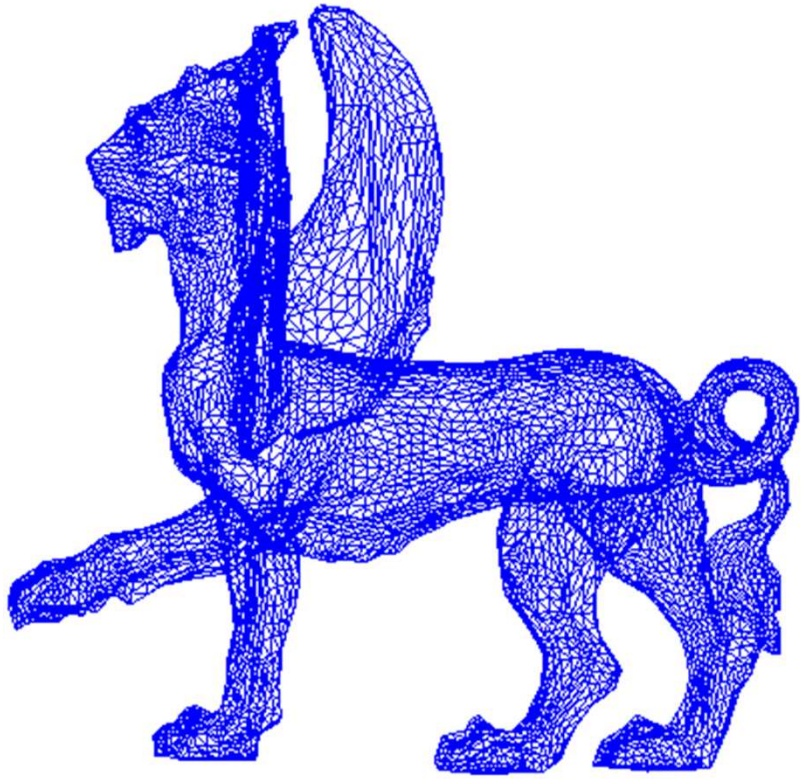


Surfaces

- Shape visualization
- Attributes:
  - Shading/lighting
  - Colors
  - Transparency
  - Silhouette
  - Feature curves
- Why do we care about it?

Surfaces

- Mesh



Data Structures

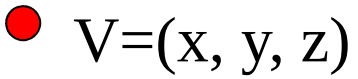
- Vertex
- Edge
- Face
- Corner
- Polyhedron

Polyhedron

- Vertex
  - vertex **vlist;
  - int num_verts, max_verts;
- Face **flist
  - face **flist;
  - int num_faces, max_faces;
- Corner **clist
  - corner **clist;
  - int num_corners, max_corners;
- Edge **elist
  - face **elist;
  - int num_edges, max_edges;

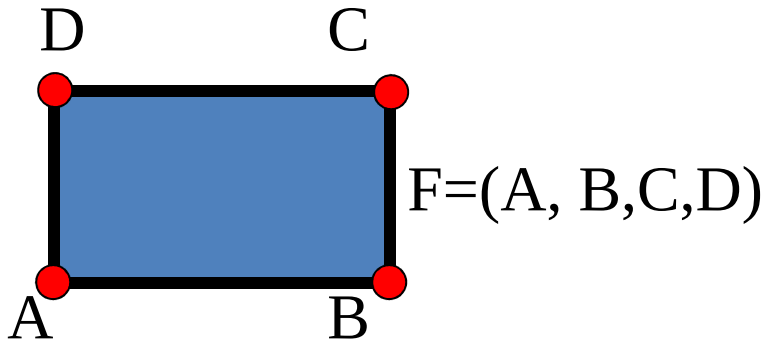
Vertex
(required)

- (x, y, z) - necessary
- (nx, ny, nz) – optional
- index – almost always needed
- List of faces incident to the vertex - almost always needed
- List of edges incident to the vertex – almost always needed
- Other types of data - optional



Face
(required)

- List of vertices - necessary
- index – almost always needed
- List of edges – almost always necessary
- Other types of data - optional



PLY format

- Header
  - Elements
  - Properties
- Data

PLY format of a Cube

```
ply
format ascii 1.0
comment created by platoply
element vertex 8
property float32 x
property float32 y
property float32 z
element face 6
property list uint8 int32 vertex_indices
end_header
-1 -1 -1
1 -1 -1
1 1 -1
-1 1 -1
-1 -1 1
1 -1 1
1 1 1
-1 1 1
4 0 1 2 3
4 5 4 7 6
4 6 2 1 5
4 3 7 4 0
4 7 3 2 6
4 5 1 0 4
```

header

Vertex list

Face list

Data

PLY format (header)

```
ply
format ascii 1.0
comment created by platoply
element vertex 8
property float32 x
property float32 y
property float32 z
element face 6
property list uint8 int32 vertex_indices
end_header
```

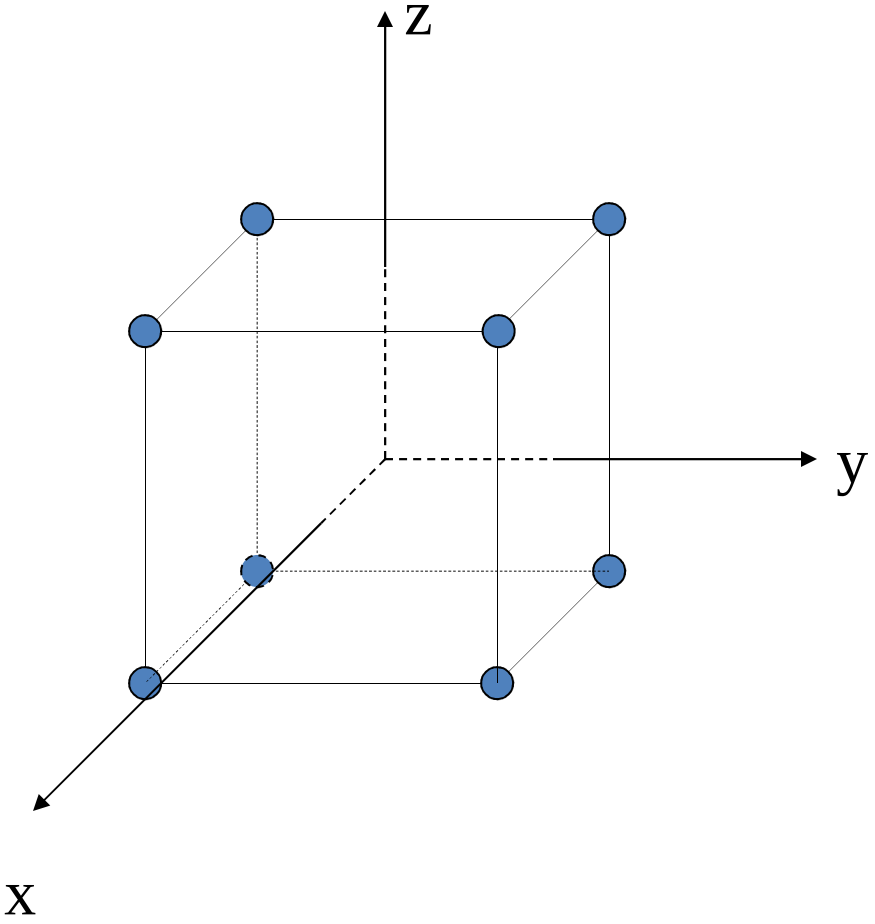
PLY format

-1 -1 -1
1 -1 -1
1 1 -1
-1 1 -1
-1 -1 1
1 -1 1
1 1 1
-1 1 1

Vertex list

4 0 1 2 3
4 5 4 7 6
4 6 2 1 5
4 3 7 4 0
4 7 3 2 6
4 5 1 0 4

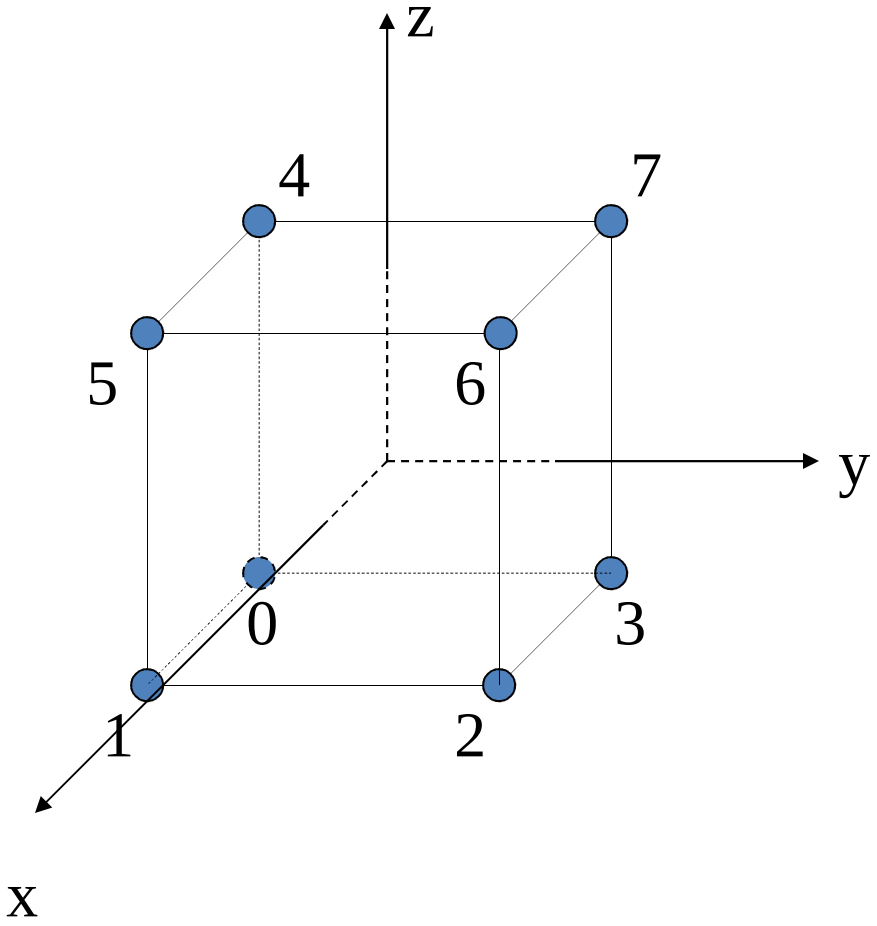
Face list



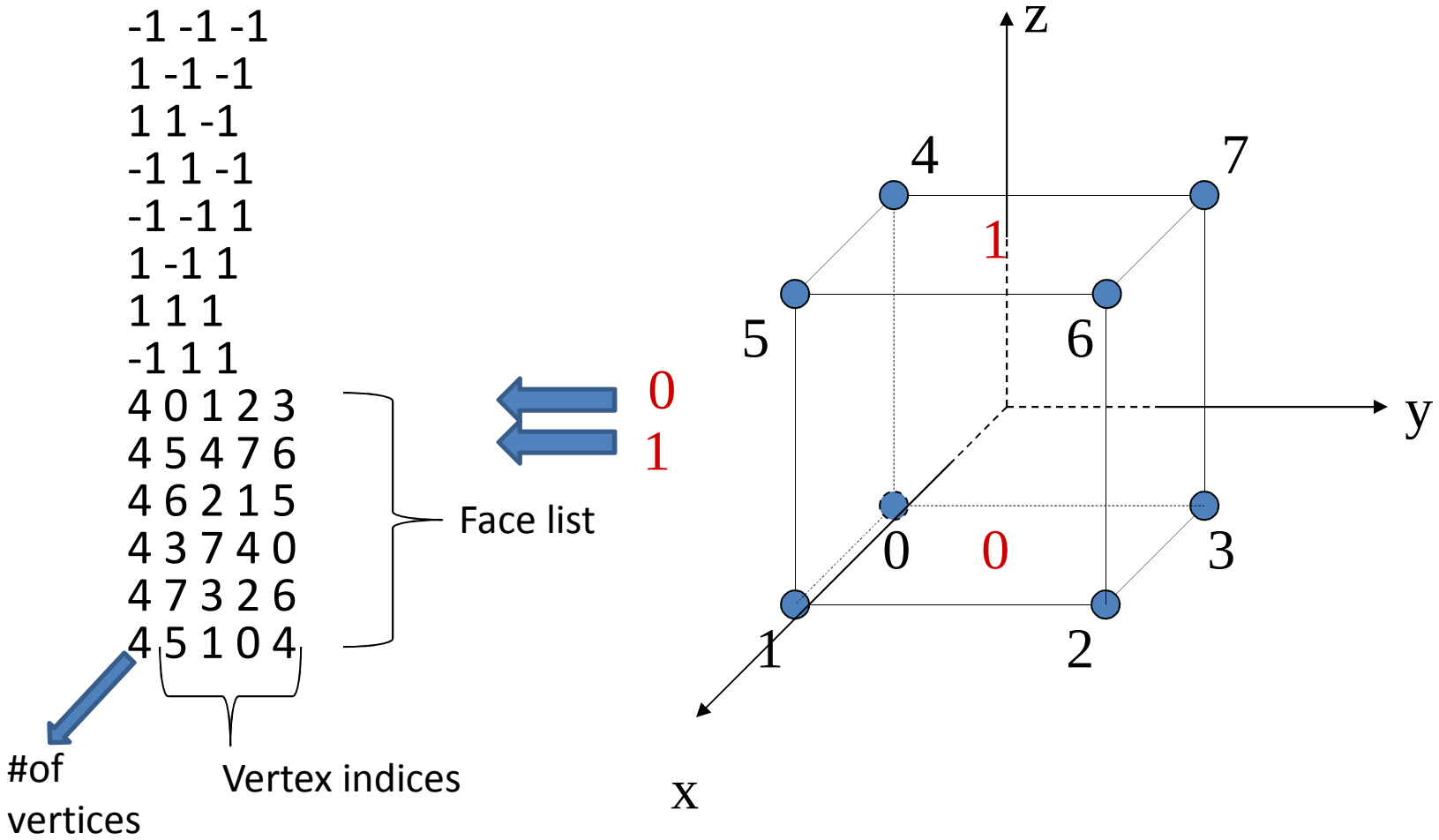
PLY format

X Y Z

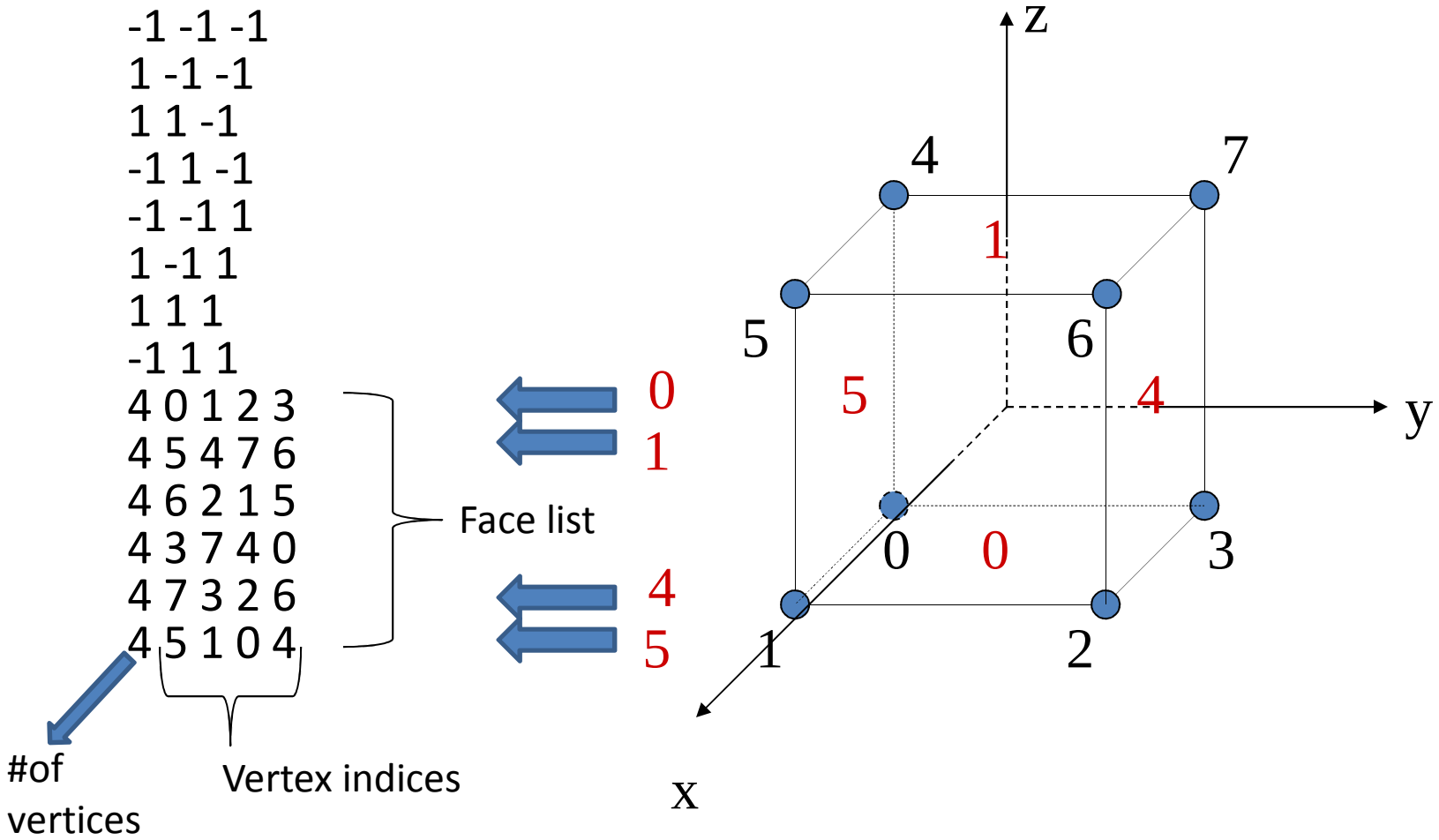
-1 -1 -1
1 -1 -1
1 1 -1
-1 1 -1
-1 -1 1
1 -1 1
1 1 1
-1 1 1
4 0 1 2 3
4 5 4 7 6
4 6 2 1 5
4 3 7 4 0
4 7 3 2 6
4 5 1 0 4



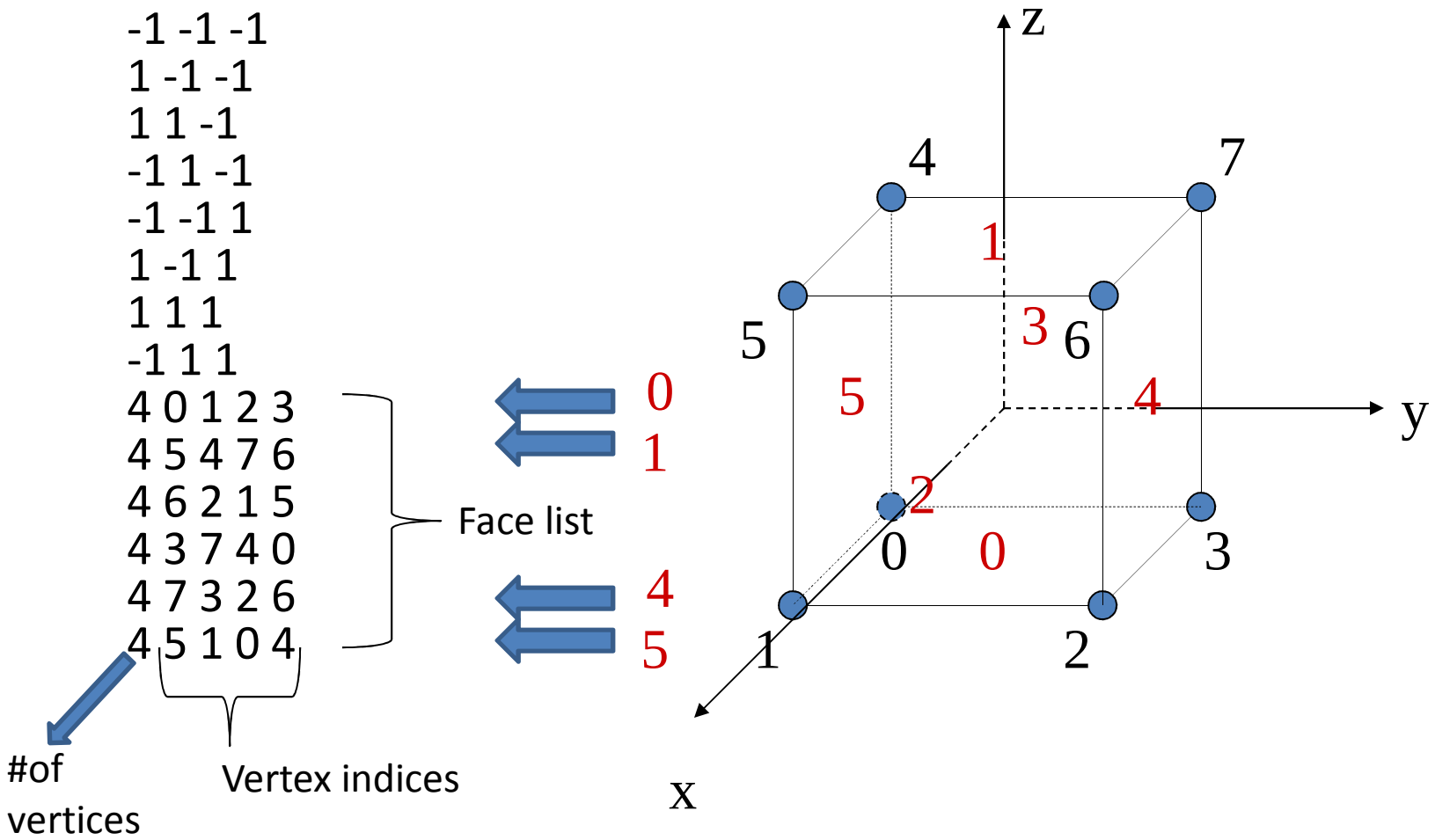
PLY format

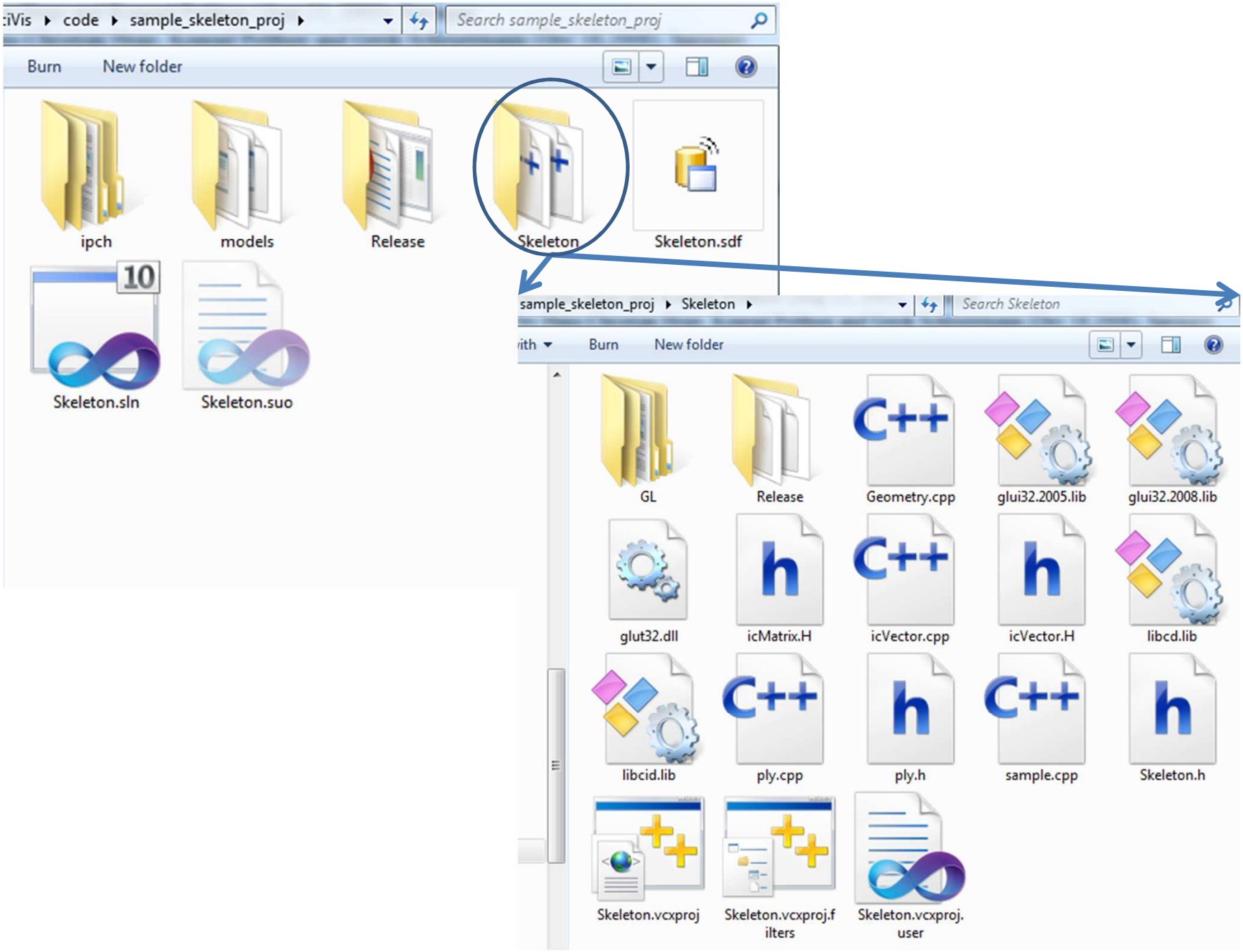

-1 -1 -1
1 -1 -1
1 1 -1
-1 1 -1
-1 -1 1
1 -1 1
1 1 1
-1 1 1
4 0 1 2 3
4 5 4 7 6
4 6 2 1 5
4 3 7 4 0
4 7 3 2 6
4 5 1 0 4
0
1
Face list
#of vertices
Vertex indices
z
y
x
4
7
5
6
0
3
1
2
1
0

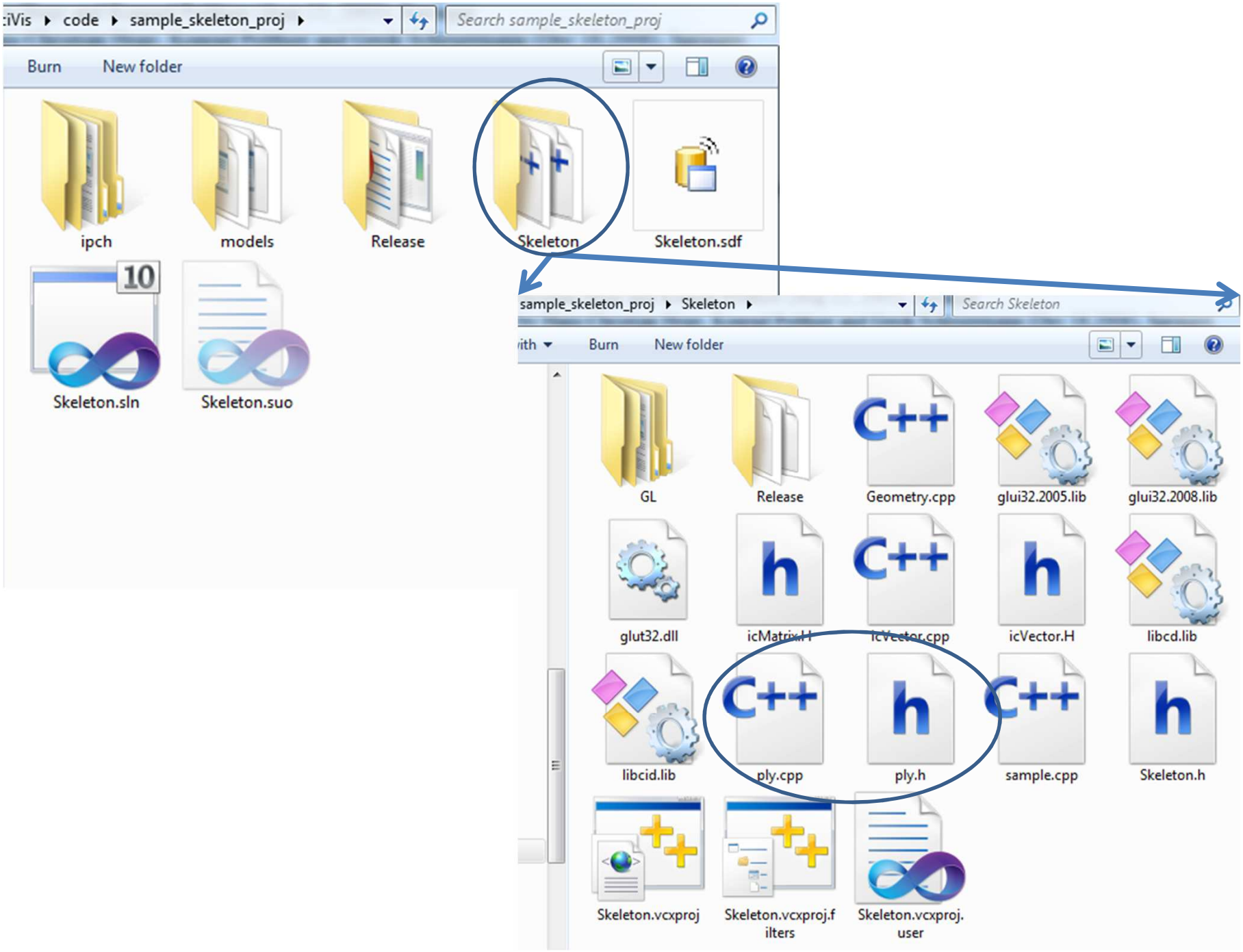
PLY format

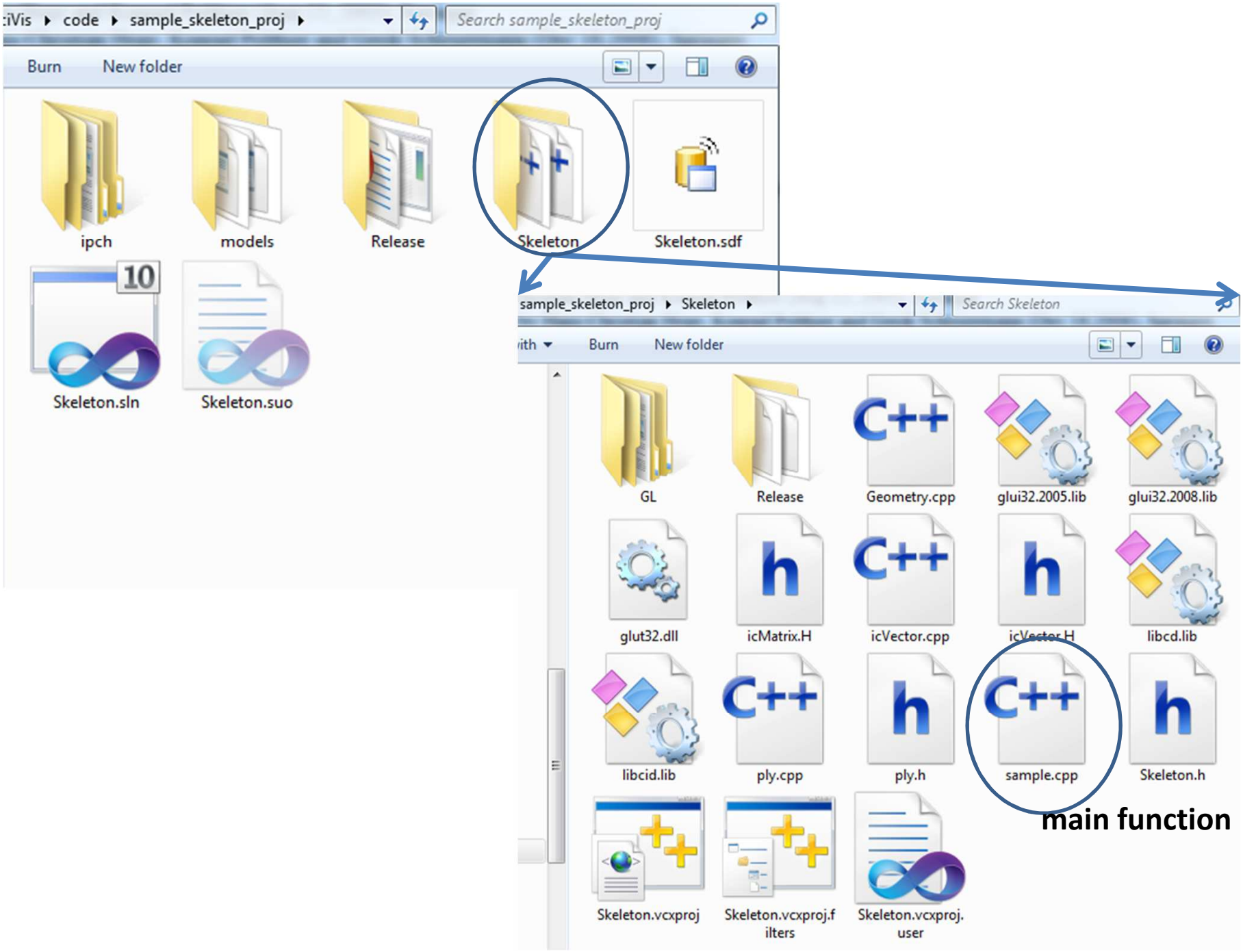

-1 -1 -1
1 -1 -1
1 1 -1
-1 1 -1
-1 -1 1
1 -1 1
1 1 1
-1 1 1
4 0 1 2 3
4 5 4 7 6
4 6 2 1 5
4 3 7 4 0
4 7 3 2 6
4 5 1 0 4
Face list
0
1
4
5
#of vertices
Vertex indices
z
y
x
4
7
5
6
0
3
1
2
1
5
4
0
0

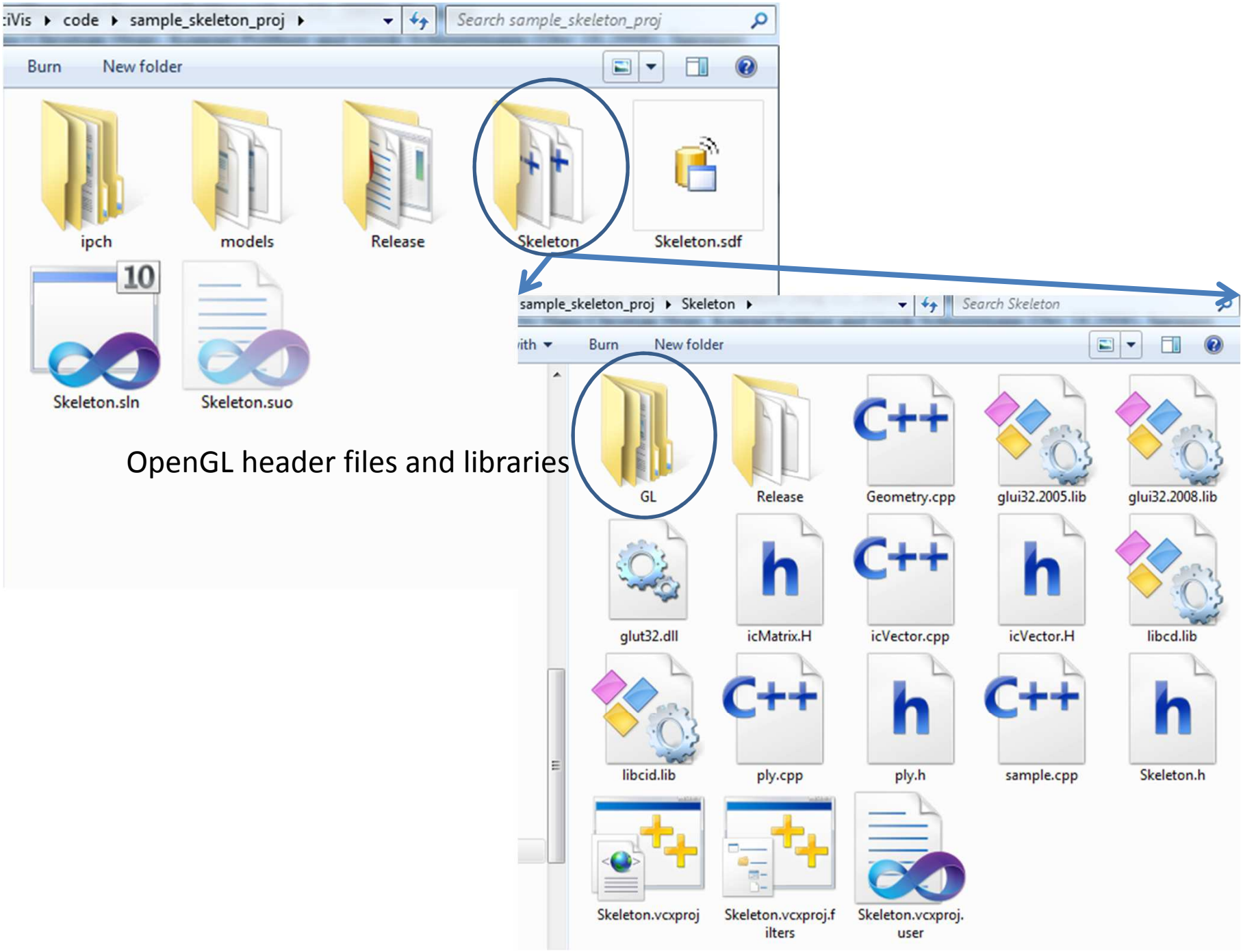
PLY format


-1 -1 -1
1 -1 -1
1 1 -1
-1 1 -1
-1 -1 1
1 -1 1
1 1 1
-1 1 1
4 0 1 2 3
4 5 4 7 6
4 6 2 1 5
4 3 7 4 0
4 7 3 2 6
4 5 1 0 4
Face list
0
1
4
5
#of vertices
Vertex indices
z
y
x
4
7
5
6
1
2
0
3
1
3
5
4
2
0

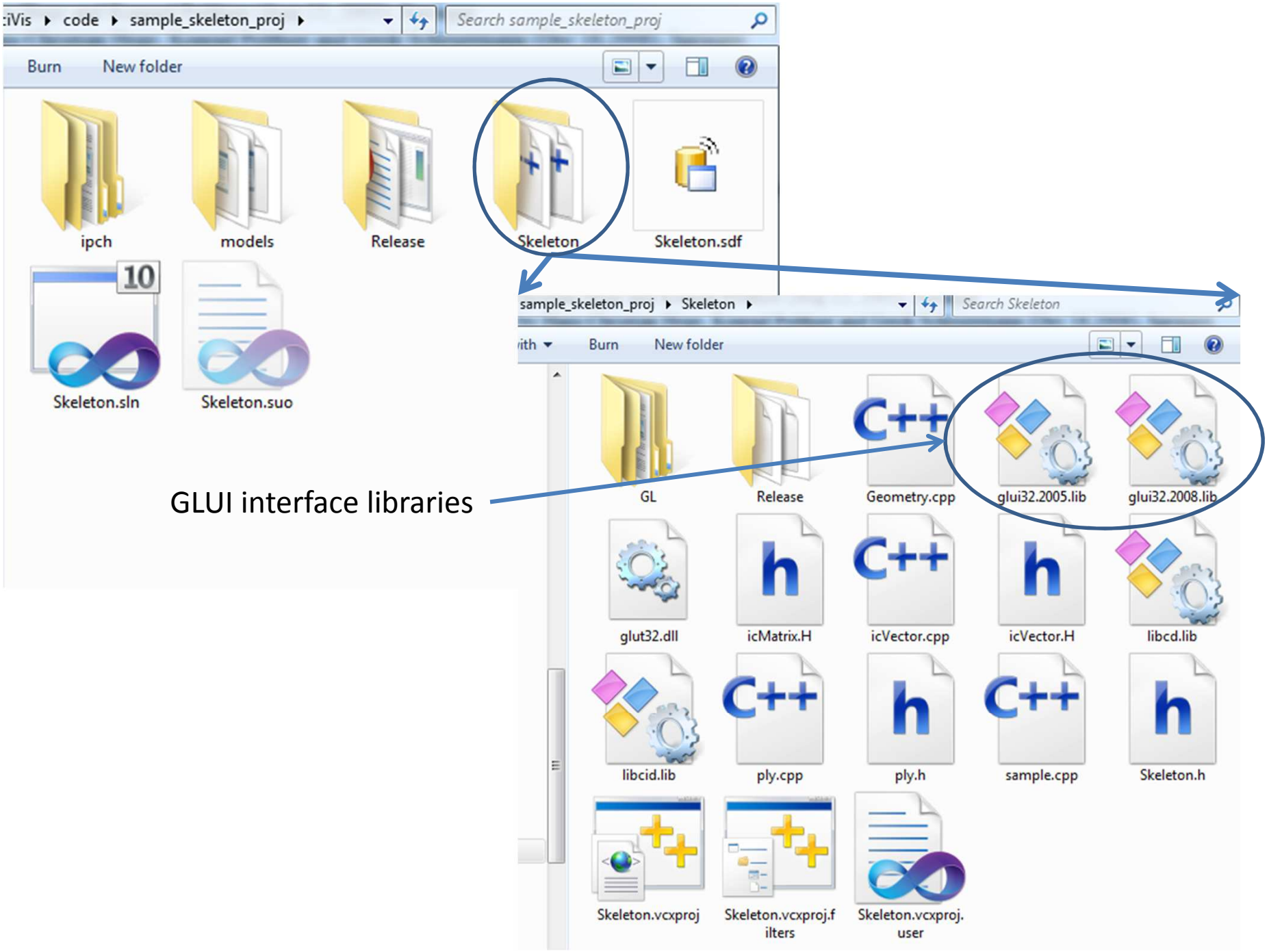

iVis ▸ code ▸ sample_skeleton_proj ▸
Search sample_skeleton_proj
Burn
New folder
ipch
models
Release
Skeleton
Skeleton.sdf
Skeleton.sln
Skeleton.suo
sample_skeleton_proj ▸ Skeleton ▸
Search Skeleton
Burn
New folder
GL
Release
Geometry.cpp
glui32.2005.lib
glui32.2008.lib
glut32.dll
icMatrix.H
icVector.cpp
icVector.H
libcd.lib
libcid.lib
ply.cpp
ply.h
sample.cpp
Skeleton.h
Skeleton.vcxproj
Skeleton.vcxproj.filters
Skeleton.vcxproj.user

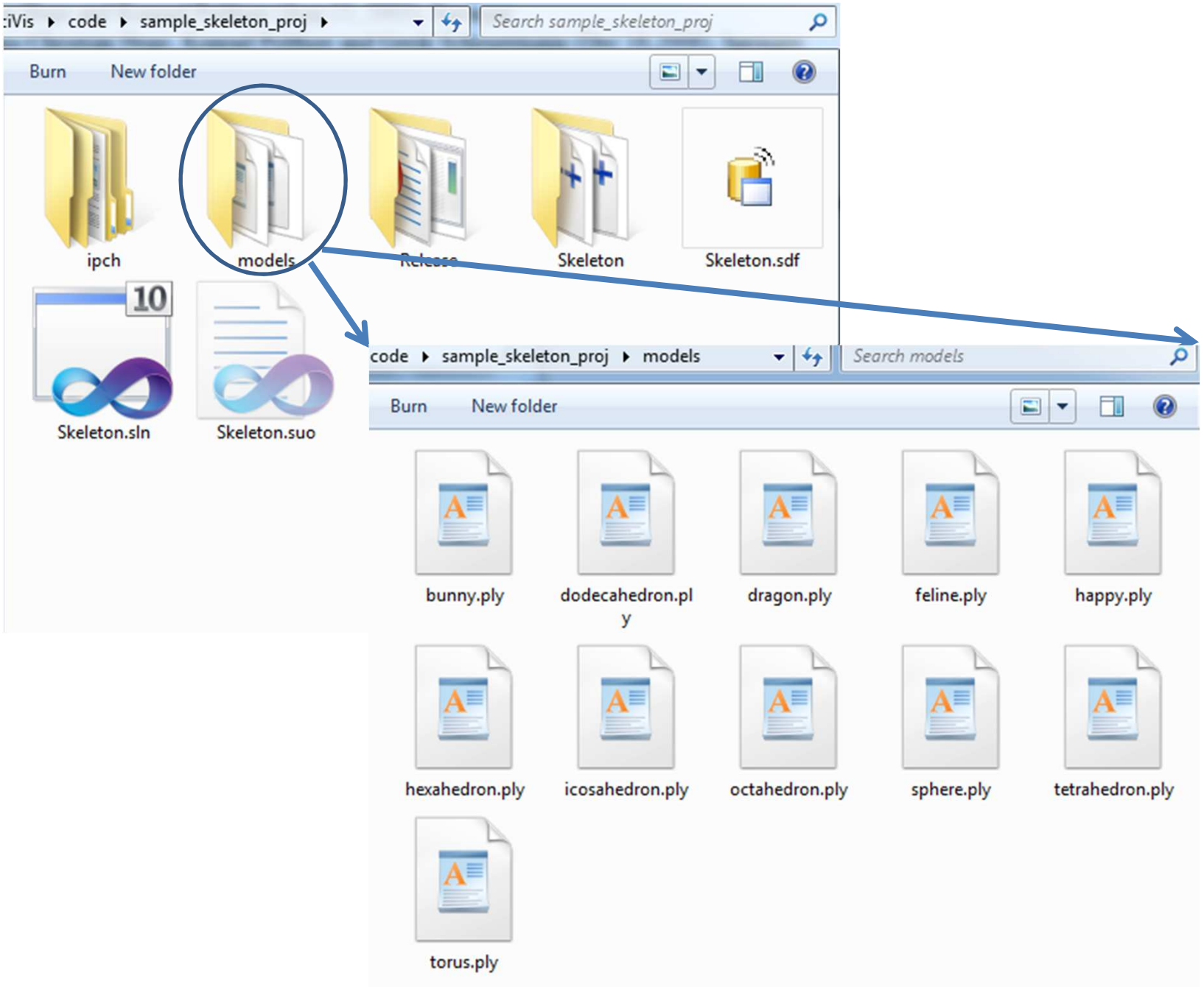

iVis ▸ code ▸ sample_skeleton_proj ▸
Search sample_skeleton_proj
Burn
New folder
ipch
models
Release
Skeleton
Skeleton.sdf
Skeleton.sln
Skeleton.suo
sample_skeleton_proj ▸ Skeleton ▸
Search Skeleton
Burn
New folder
GL
Release
Geometry.cpp
glui32.2005.lib
glui32.2008.lib
glut32.dll
icVector.H
libcd.lib
libcid.lib
ply.cpp
ply.h
sample.cpp
Skeleton.h
Skeleton.vcxproj
Skeleton.vcxproj.filters
Skeleton.vcxproj.user


iVis ▸ code ▸ sample_skeleton_proj ▸
Search sample_skeleton_proj
Burn
New folder
ipch
models
Release
Skeleton
Skeleton.sdf
Skeleton.sln
Skeleton.suo
sample_skeleton_proj ▸ Skeleton ▸
Search Skeleton
Burn
New folder
GL
Release
Geometry.cpp
glui32.2005.lib
glui32.2008.lib
glut32.dll
icMatrix.H
icVector.cpp
icVector.H
libcd.lib
libcid.lib
ply.cpp
ply.h
sample.cpp
Skeleton.h
Skeleton.vcxproj
Skeleton.vcxproj.filters
Skeleton.vcxproj.user
main function



OpenGL header files and libraries


iVis ▸ code ▸ sample_skeleton_proj ▸
Search sample_skeleton_proj
Burn
New folder
ipch
models
Release
Skeleton
Skeleton.sdf
Skeleton.sln
Skeleton.suo
GLUI interface libraries
sample_skeleton_proj ▸ Skeleton ▸
Search Skeleton
Burn
New folder
GL
Release
Geometry.cpp
glui32.2005.lib
glui32.2008.lib
glut32.dll
icMatrix.H
icVector.cpp
icVector.H
libcd.lib
libcid.lib
ply.cpp
ply.h
sample.cpp
Skeleton.h
Skeleton.vcxproj
Skeleton.vcxproj.filters
Skeleton.vcxproj.user


iVis ▸ code ▸ sample_skeleton_proj ▸
Search sample_skeleton_proj
Burn
New folder
ipch
models
Release
Skeleton
Skeleton.sdf
Skeleton.sln
Skeleton.suo
code ▸ sample_skeleton_proj ▸ models
Search models
Burn
New folder
bunny.ply
dodecahedron.ply
dragon.ply
feline.ply
happy.ply
hexahedron.ply
icosahedron.ply
octahedron.ply
sphere.ply
tetrahedron.ply
torus.ply

Acknowledge

- Part of the materials are provided by
  - Prof. Eugene Zhang at Oregon State University