HDF5 as foundation for big data of manifold types in scientific visualization

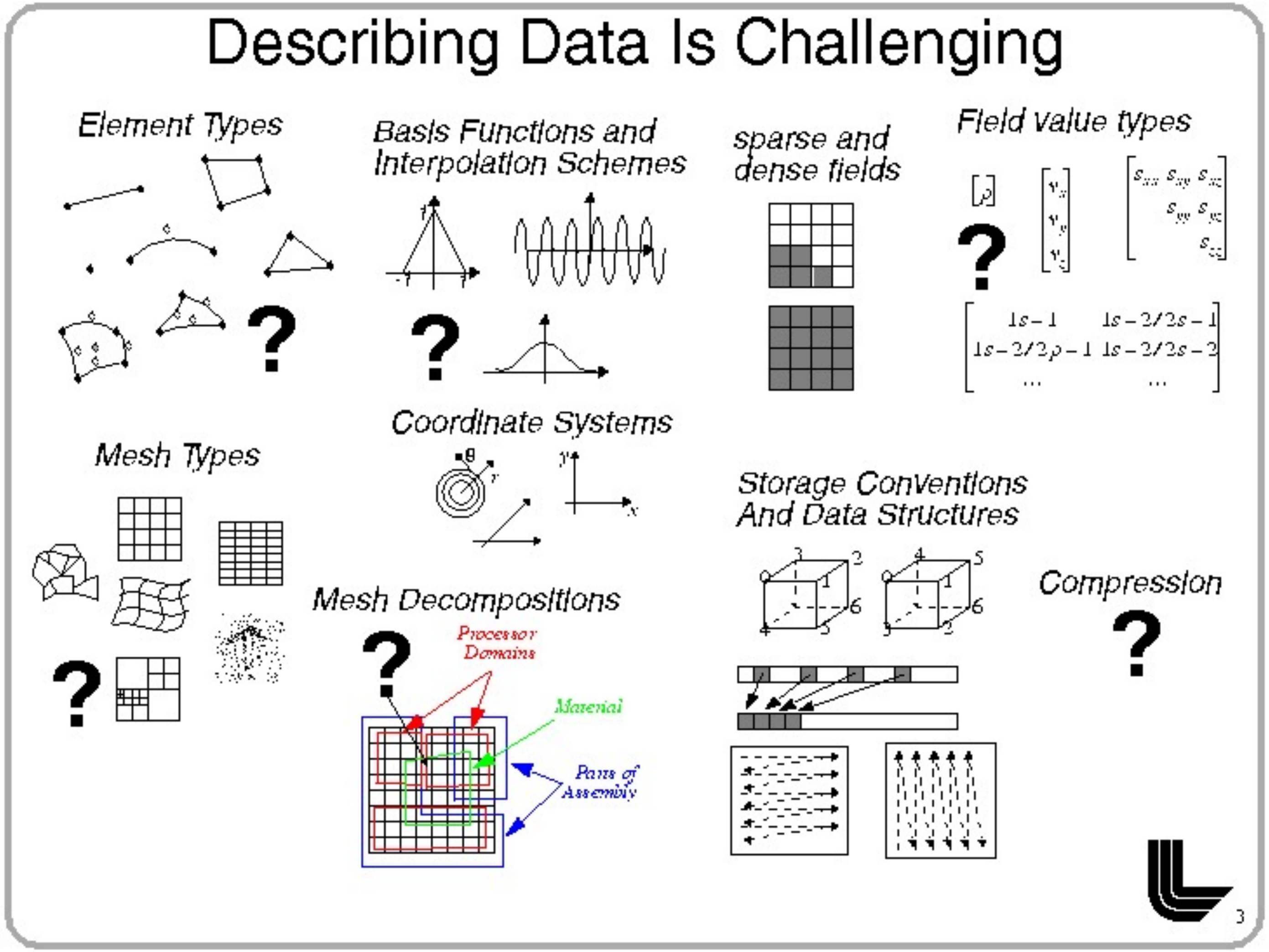
Dr. Werner Benger, Airborne HydroMapping GmbH

Abstract

Scientific visualization encounters a multitude of data types generated from observations or numerical simulations, ranging from simple images, point clouds to complex hierarchical multigrid structures of unbound complexity. Many, if not each, of these data types comes with their own specific file formats. Frequently, there is even a multitude of different file formats for the same data type (e.g., for images). The wheel is re-invented over and over, as everyone has their own notion of “roundness”. By the means of HDF5 this gordian knot can be untied – not by “another file format”, but via its capabilities to clearly distinguish between semantic, syntactical, and technical-internal properties of a dataset. By dealing with data sets on a higher level of abstraction – i.e., a purely semantic level – the burden of lower levels (such as byte order, compression schemes, coordinate systems…) become technical details while at the same time allowing data to be completely self-descriptive – rather than requiring an email in addition to a data file as requirement to interpret its contents. This presentation demonstrates the F5 scheme to layout scientific data to formulate their topological, geometrical and other properties as well as inter-relationships between them. This model allows to cover a wide range of data types under a common abstraction scheme, utilizing as HDF5 powerful and scalable basis.

The Problem:

A jungle of data types



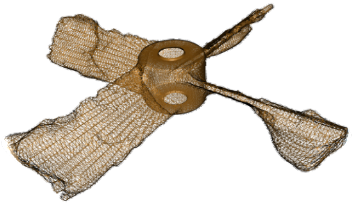
Graphics thanks to Mark Miller, LLNL

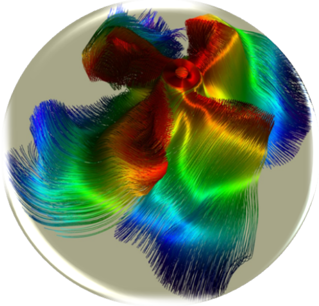
On Safari in the File Format Jungle—Why Can't You Visualize My Data?

Computing in Science & Engineering (Volume: 11, Issue: 6, Nov.-Dec. 2009)

The Tool Approach
“Specific Problems need specific Solutions”

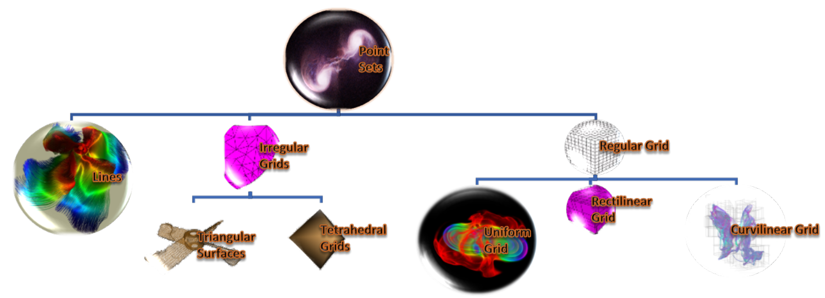
Point Cloud Data	Triangular Mesh	Lines	?
Display as Points	Display as Surface	Display as Lines	Display as ??



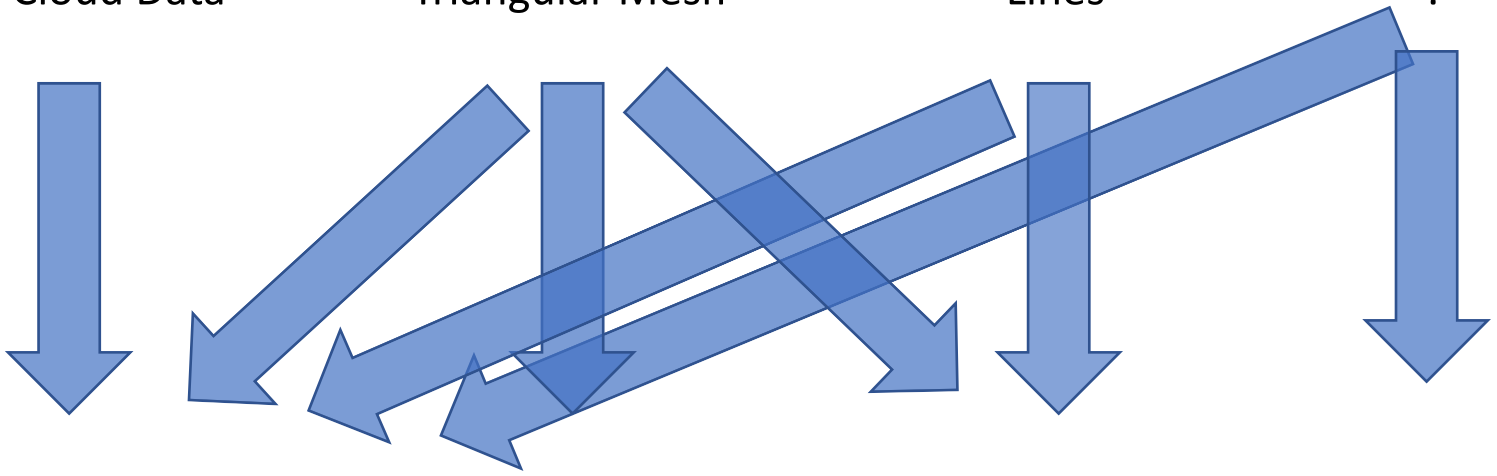


The Framework Approach

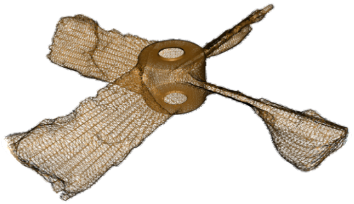
Generic Methods to avoid reinventing the wheel

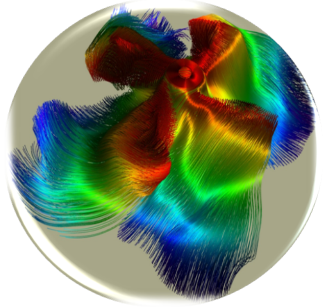


Point Cloud Data | Triangular Mesh | Lines | ?



Display as Points | Display as Surface | Display as Lines | Display as ??





Challenge of a Common Data Model

“The proper abstractions for scientific data are known. We just have to use them.”

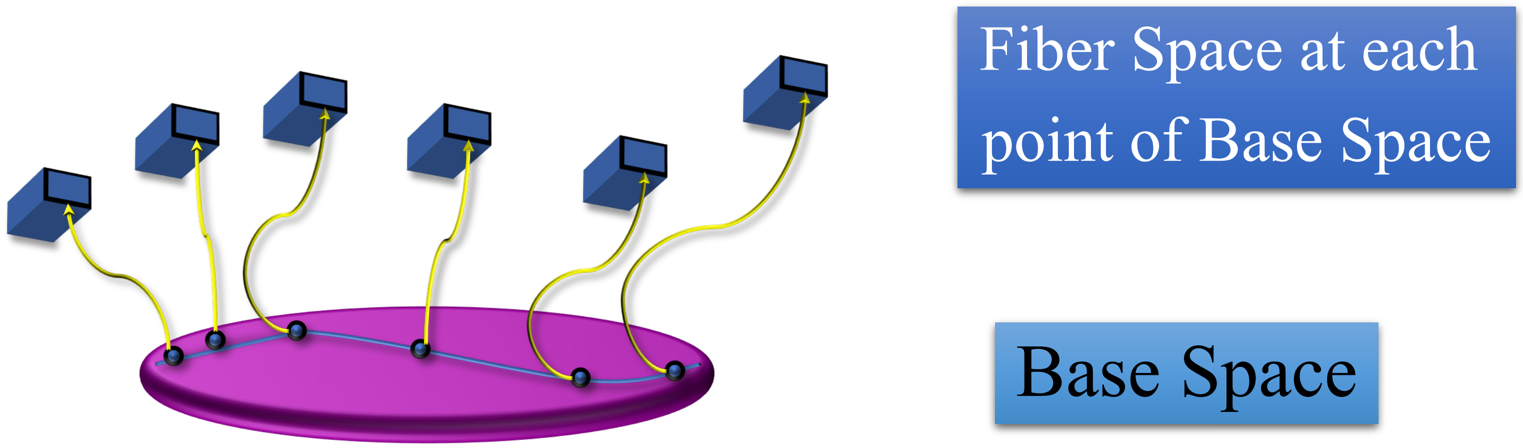
D. M. Butler & S. Bryson

Vector-Bundle Classes form Powerful Tool for Scientific Visualization

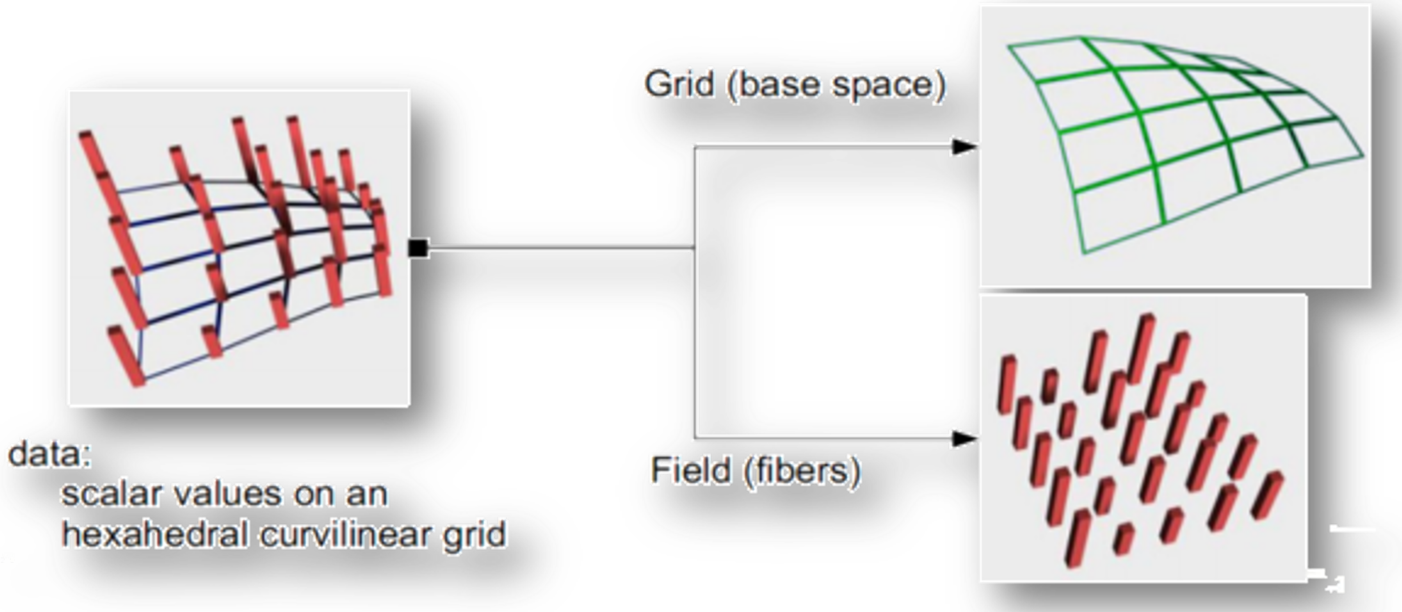
Computers in Physics, Vol. 6, No 6., Nov/Dec 1992

Fiber Bundle Data Model

- Is a generic approach to handle a wide range of data types used for scientific visualization
- Basic concept: Base space maps to fibers



Data organization in the Fiber Bundle



Base-Space operations:
- Dealing with large data
- Differential operators

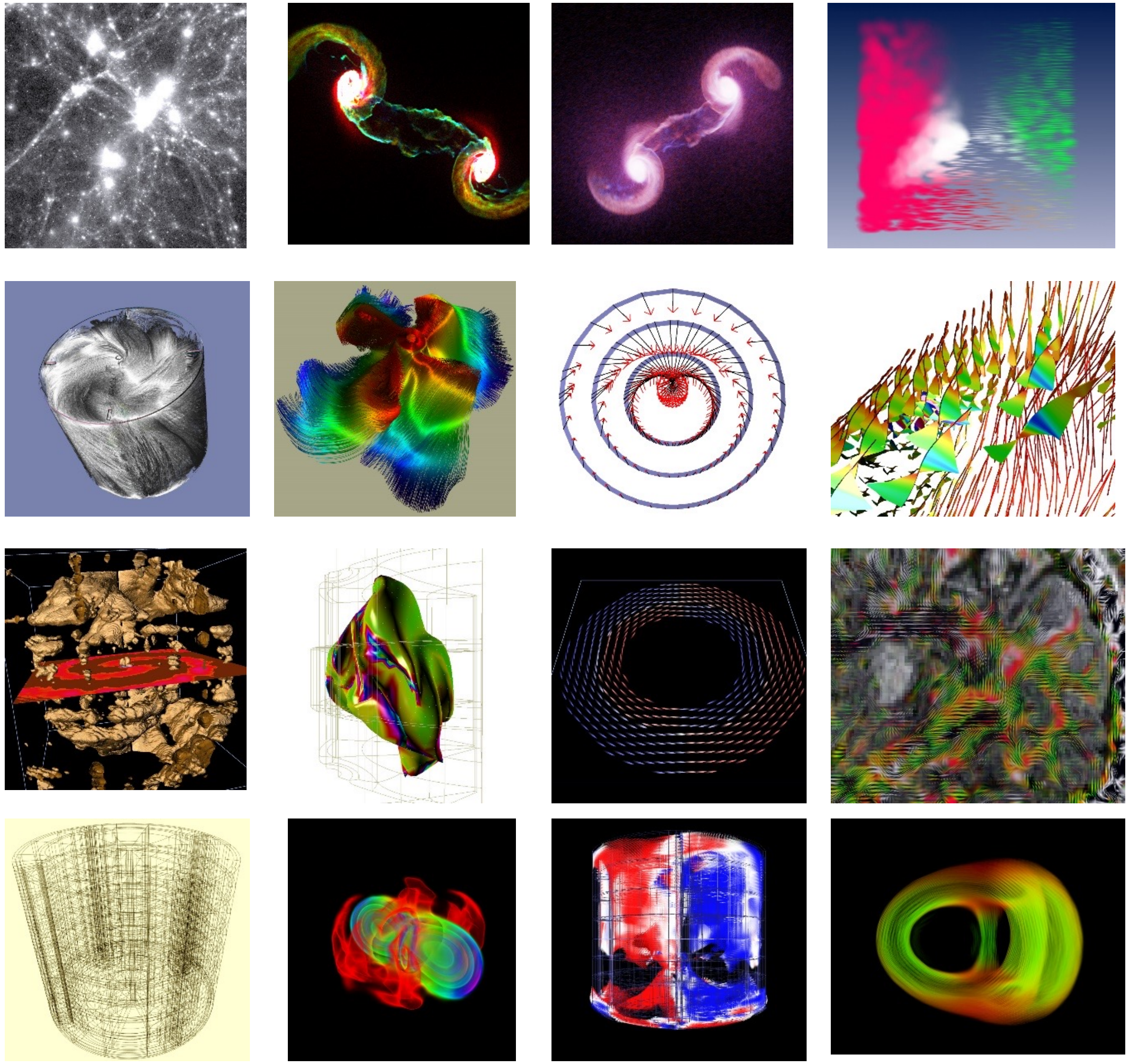
→ Operations on the manifold (possibly a vector space)

Fiber-Space elements:
- This is where data fields „live“
- Zoo of multivectors and tensors

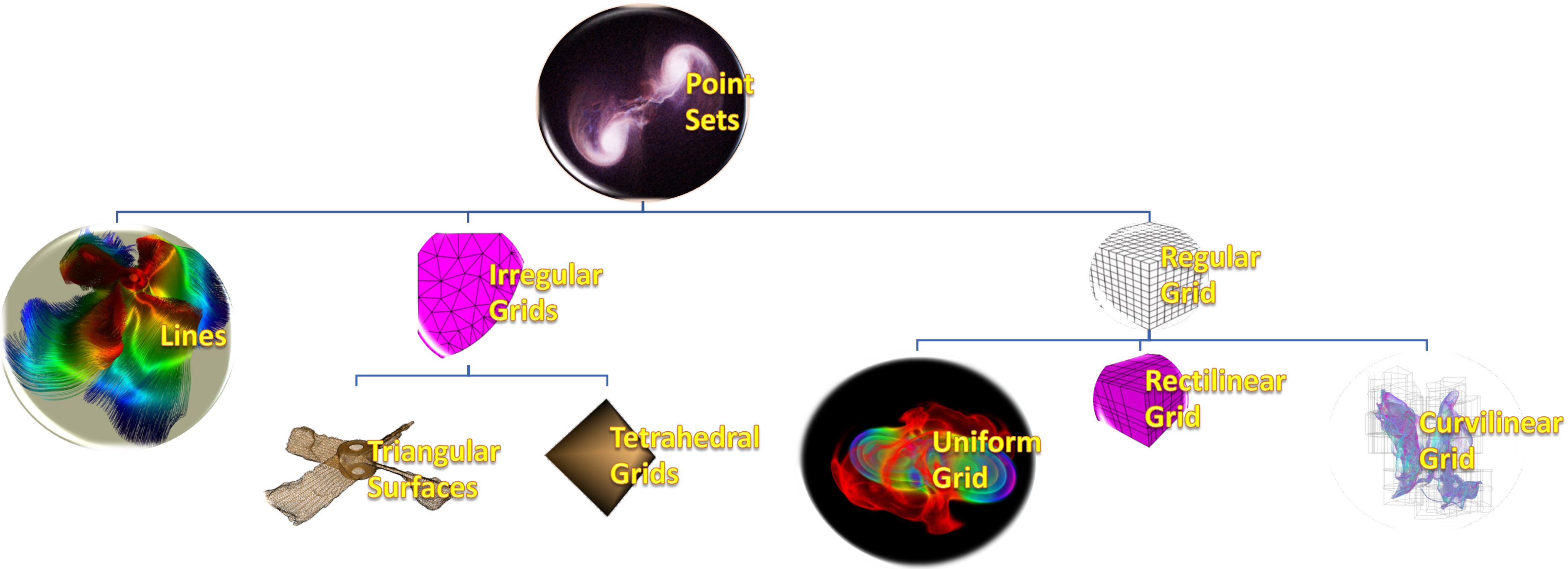
→ Operations on the tangential space and powers of it (always vector space)

Fiber: 0D 1D 3D 6D

Base: 0D 1D 2D 3D

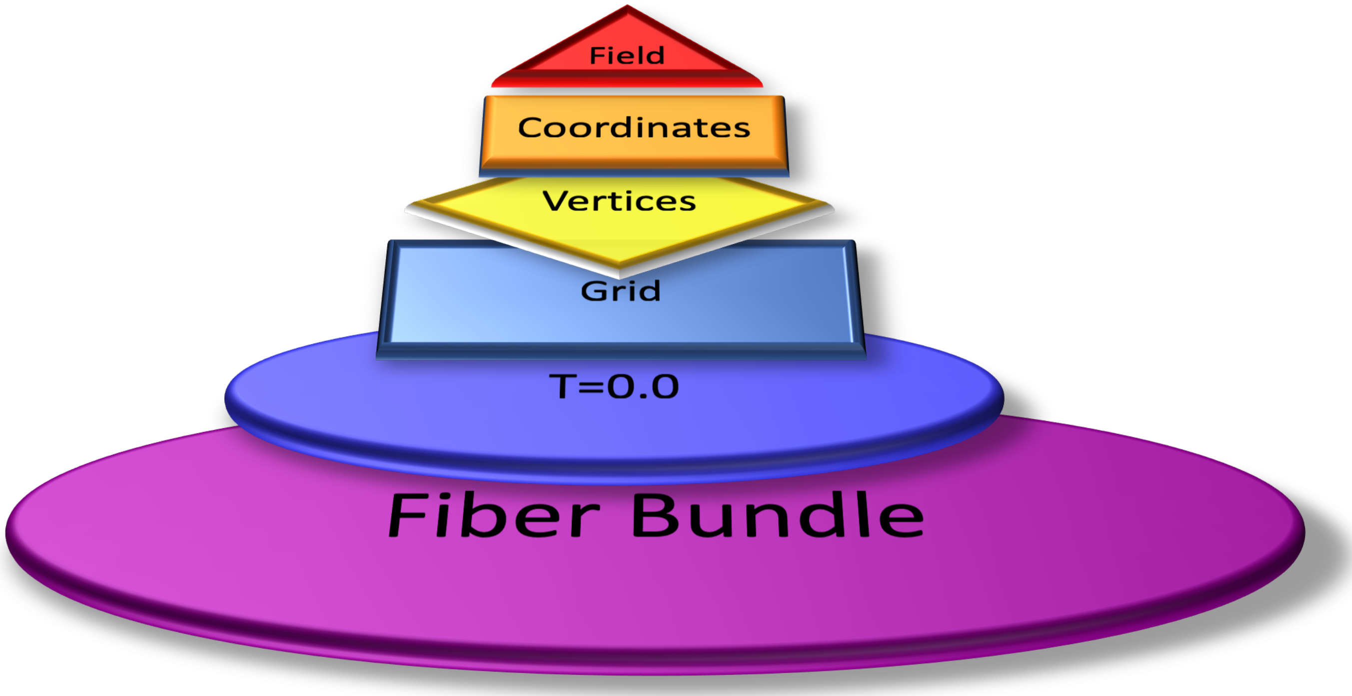


Hierarchy of discretization schemes („Grids“)



The F5 Fiber Bundle Data Model

Casts data into hierarchy of five levels:



1. **Field**
2. **Representation**
3. **Skeleton**
4. **Grid object**
5. **(Time) Slice**
6. **Bundle**

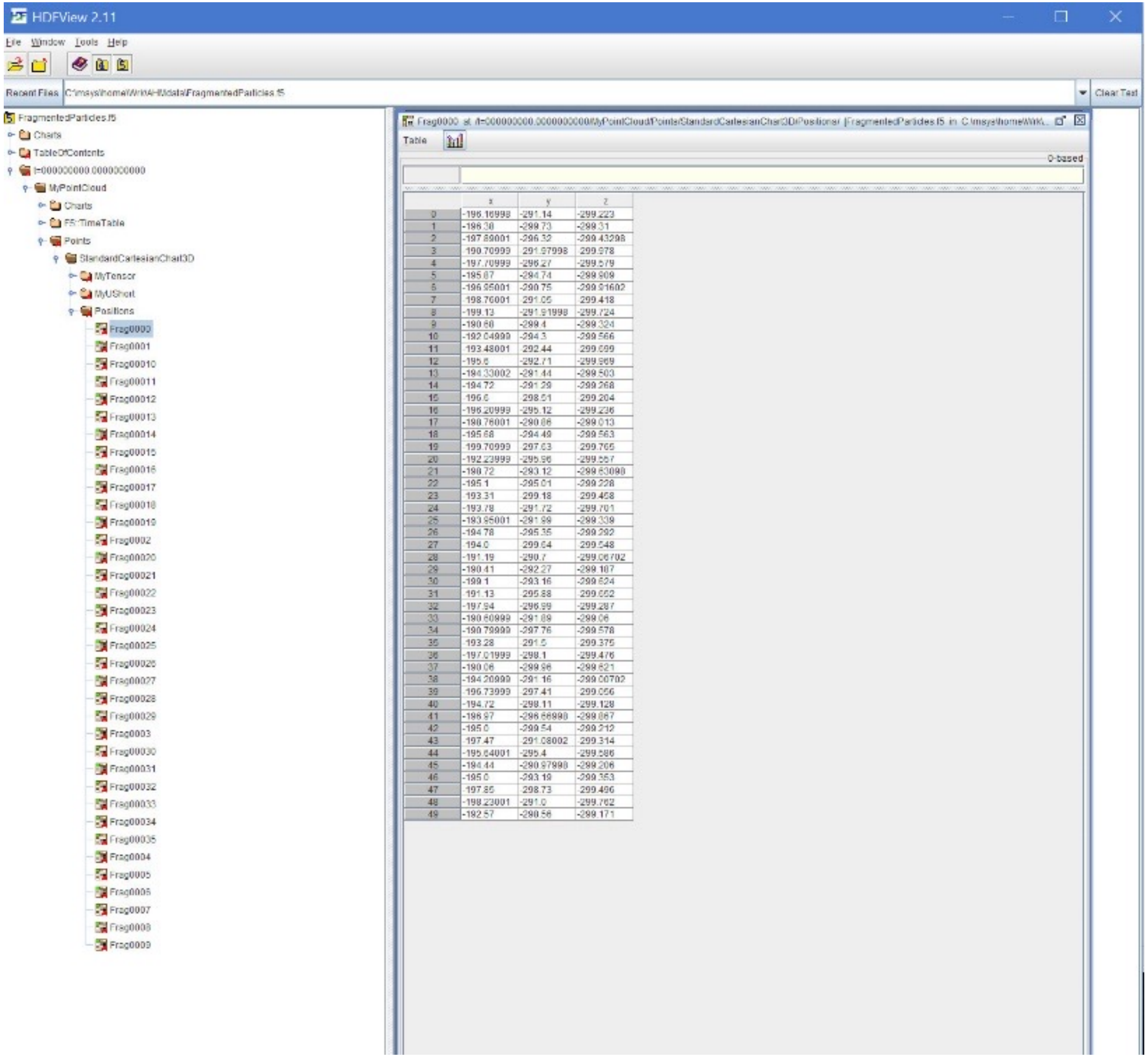
Visualization of general relativistic tensor fields via a fiber bundle data model, PhD thesis, Benger (2004)

Maps to HDF5 as hierarchy:
/Time/Grid/Top/Rep/Field/

```
h5ls -r minkowski-0000.f5
/T=0/minkowski/Points/StandardCartesianChart3D Group
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV Group
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV/gxx Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV/gxy Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV/gxz Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV/gyy Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV/gyz Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::CURV/gzz Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::LAPSE Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC Group
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC/gxx Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC/gxy Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC/gxz Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC/gyy Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC/gyz Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/ADMBASE::METRIC/gzz Dataset {43, 37, 19}
/T=0/minkowski/Points/StandardCartesianChart3D/Positions Group
/T=0.11904761904761904101057723437 4 Group
/T=0.119047619047619041010577234374/minkowski Group
```

HDFView
Browsing HDF5 Files

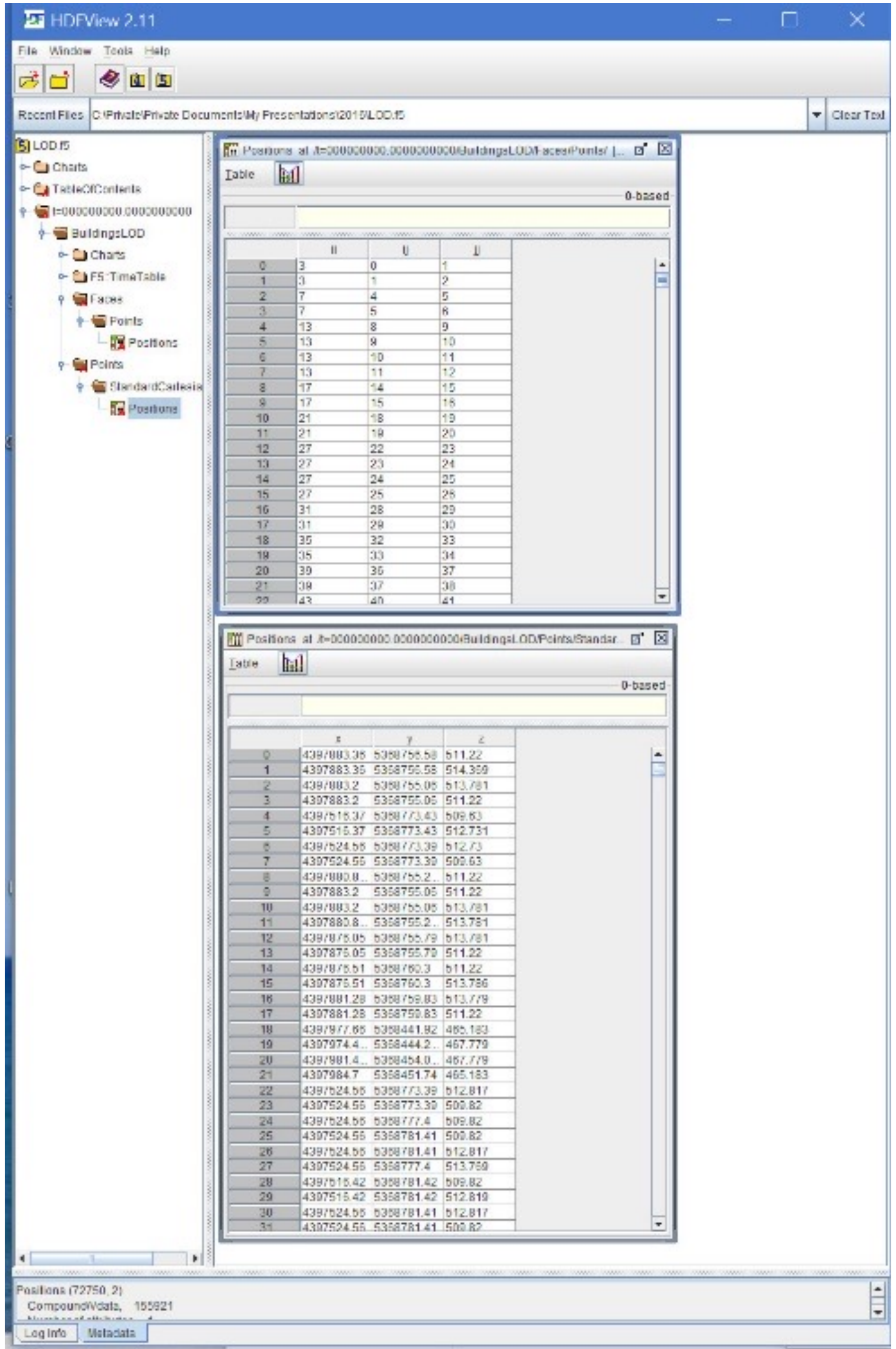
A point cloud with
Coordinates, fragmented



HDFView
Browsing HDF5 Files

Triangular Mesh

- Coordinates
- Triangle Indices



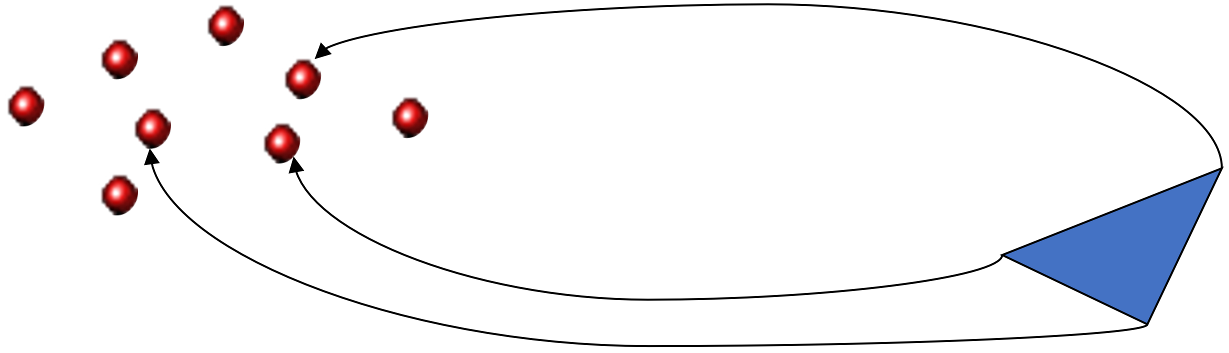
Relative Representations

Triangular Surface: 600331 Triangles, 314442 Points

```
/T=1                             Group
/T=1/adcirc                      Group
/T=1/adcirc/Points               Group
/T=1/adcirc/Points/StandardCartesianChart3D Group
/T=1/adcirc/Points/StandardCartesianChart3D/Positions Dataset {314442}
/T=1/adcirc/Connectivity Group
/T=1/adcirc/Connectivity/Points Group
/T=1/adcirc/Connectivity/Points/Positions Dataset {600331}
```

Index Depth = 0

Index Depth = 1



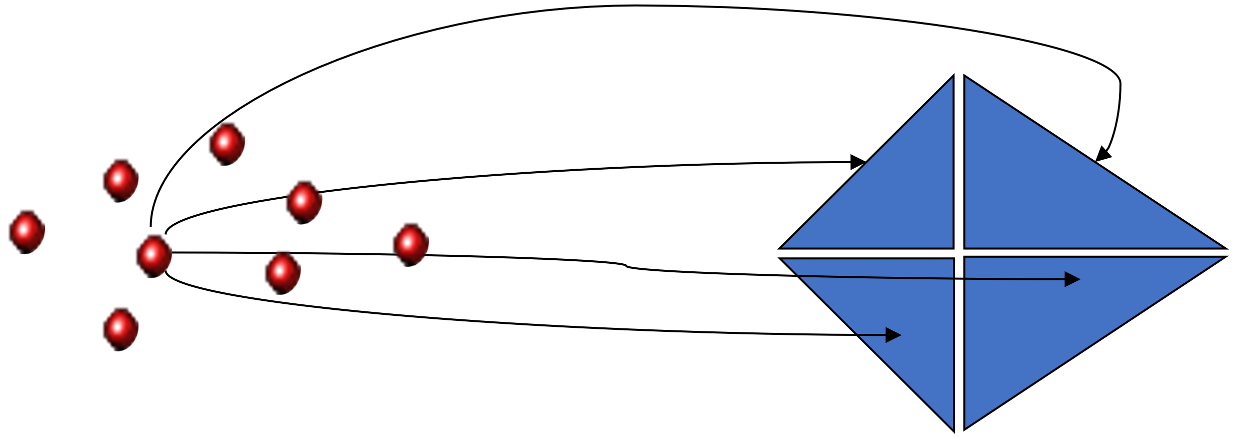
Relative Representations

Inverse Information: Triangles per Vertex

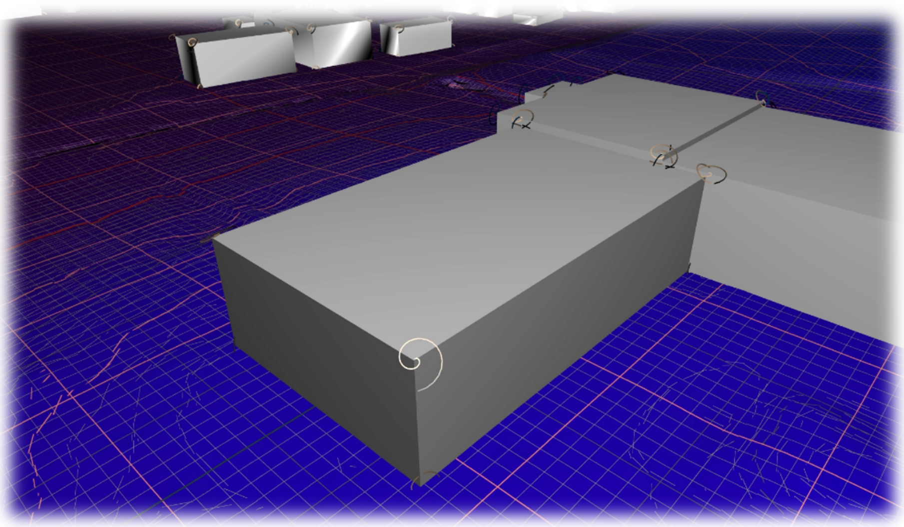
```
/T=1                         Group
/T=1/adcirc                  Group
/T=1/adcirc/Points           Group
/T=1/adcirc/Points/StandardCartesianChart3D Group
/T=1/adcirc/Points/StandardCartesianChart3D/Positions Dataset {314442}
```

/T=1/adcirc/Points/Connectivity/Positions Dataset {314442}

```
/T=1/adcirc/Connectivity Group
/T=1/adcirc/Connectivity/Points Group
/T=1/adcirc/Connectivity/Points/Positions Dataset {600331}
```



Fields on Relative Representations

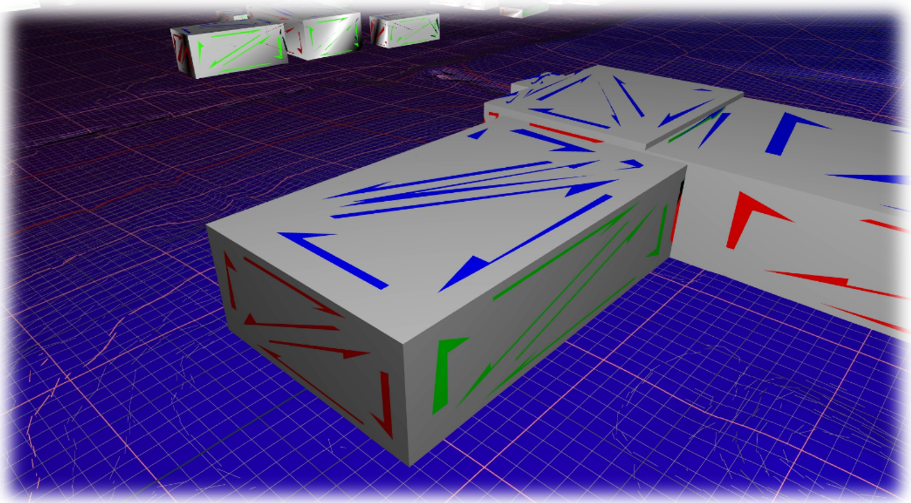


Triangular Surface: 600331 Triangles, 414442 Points

plus data on the vertices

```
/T=1                         Group
/T=1/adcirc/Connectivity Group
/T=1/adcirc/Connectivity/Points Group
/T=1/adcirc/Connectivity/Points/Positions Dataset {600331}
/T=1/adcirc/Points           Group
/T=1/adcirc/Points/StandardCartesianChart3D Group
/T=1/adcirc/Points/StandardCartesianChart3D/Positions Dataset {314442}
/T=1/adcirc/Points/StandardCartesianChart3D/elevation Dataset {314442}
/T=1/adcirc/Points/StandardCartesianChart3D/vector Dataset {314442}
```

Fields on Relative Representations

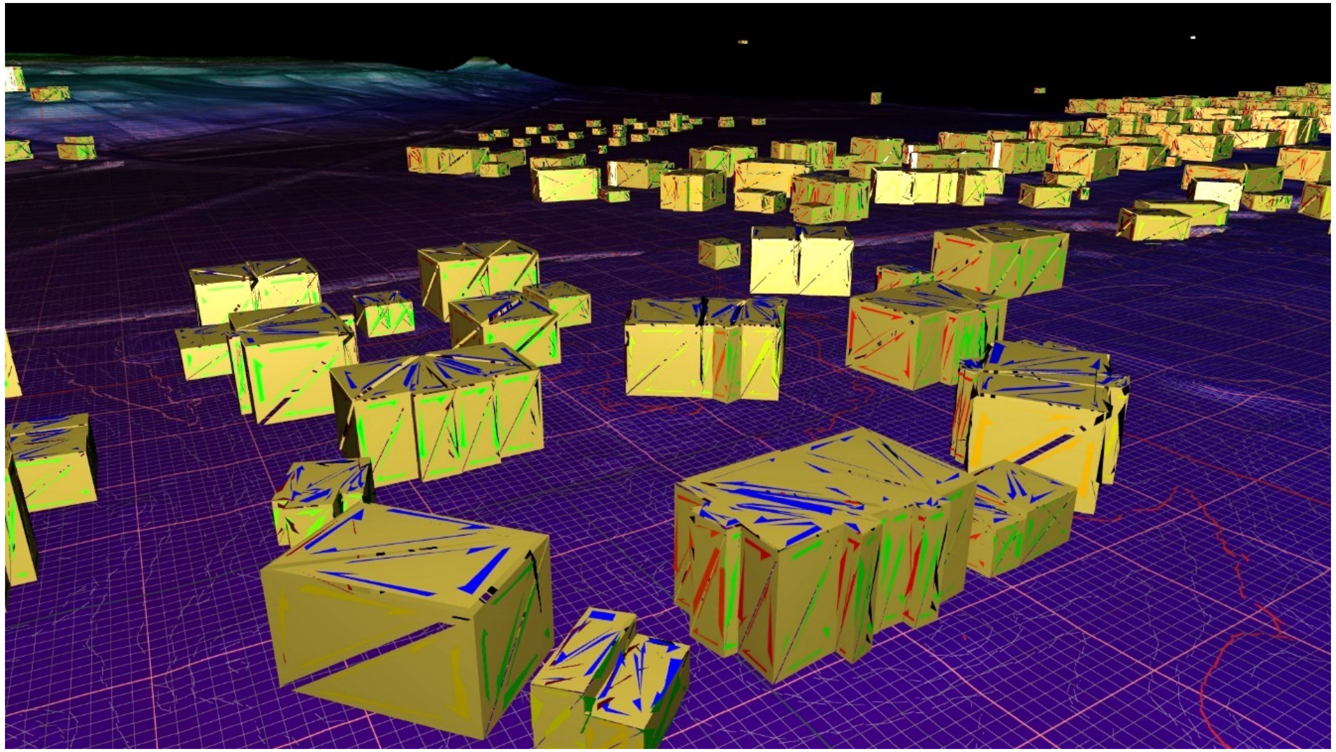


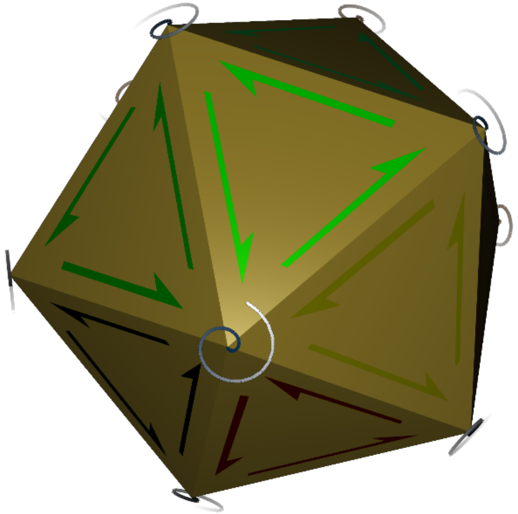
Triangular Surface: 600331 Triangles, 414442 Points

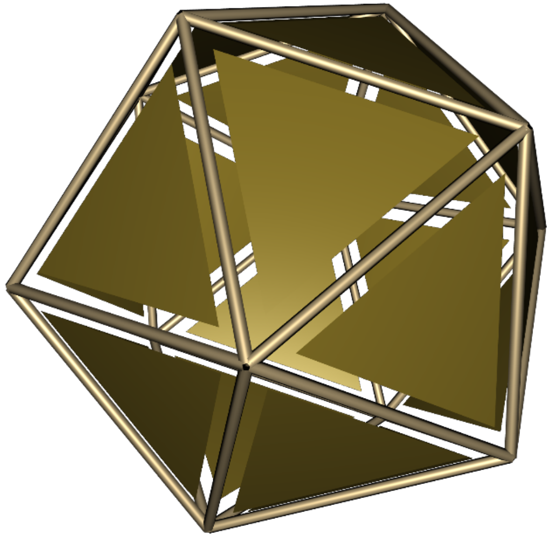
plus data on the triangles

```
/T=1                          Group
/T=1/adcirc/Connectivity Group
/T=1/adcirc/Connectivity/Points Group
/T=1/adcirc/Connectivity/Points/Positions Dataset {600331}
/T=1/adcirc/Connectivity/StandardCartesianChart3D/elevation Dataset {600331}
/T=1/adcirc/Connectivity/StandardCartesianChart3D/vector Dataset {600331}
/T=1/adcirc/Points       Group
/T=1/adcirc/Points/StandardCartesianChart3D Group
/T=1/adcirc/Points/StandardCartesianChart3D/Positions Dataset {314442}
```

Data Fields on topological Skeletons Vertices, Edges, Faces, Sets of Edges…







Unstructured Meshes

Skeletons and Secondary Representations

```
/T=1                         Group
/T=1/adcirc                  Group
/T=1/adcirc/Points           Group
/T=1/adcirc/Points/StandardCartesianChart3D Group
/T=1/adcirc/Points/StandardCartesianChart3D/Positions Dataset
/T=1/adcirc/Points/Faces Group
/T=1/adcirc/Points/Faces/Positions Dataset

/T=1/adcirc/Edges Group
/T=1/adcirc/Edges/Faces Group
/T=1/adcirc/Edges/Faces/Positions Dataset

/T=1/adcirc/Faces Group
/T=1/adcirc/Faces/Edges Group
/T=1/adcirc/Faces/Edges/Positions Dataset

/T=1/adcirc/Connectivity Group
/T=1/adcirc/Connectivity/Edges Group
/T=1/adcirc/Connectivity/Edges/Positions Dataset
```

Faces per Vertex

Faces per Edge

Edges per Face

Edges per Cell

Unstructured Meshes

Skeletons and Primary Representations

Index Depth 0

```
/T=1                          Group
/T=1/adcirc                   Group
/T=1/adcirc/Points            Group
/T=1/adcirc/Points/StandardCartesianChart3D Group
/T=1/adcirc/Points/StandardCartesianChart3D/Positions Dataset
```

Index Depth = 1

```
/T=1/adcirc/Edges Group
/T=1/adcirc/Edges/Points Group
/T=1/adcirc/Edges/Points/Positions Dataset
```

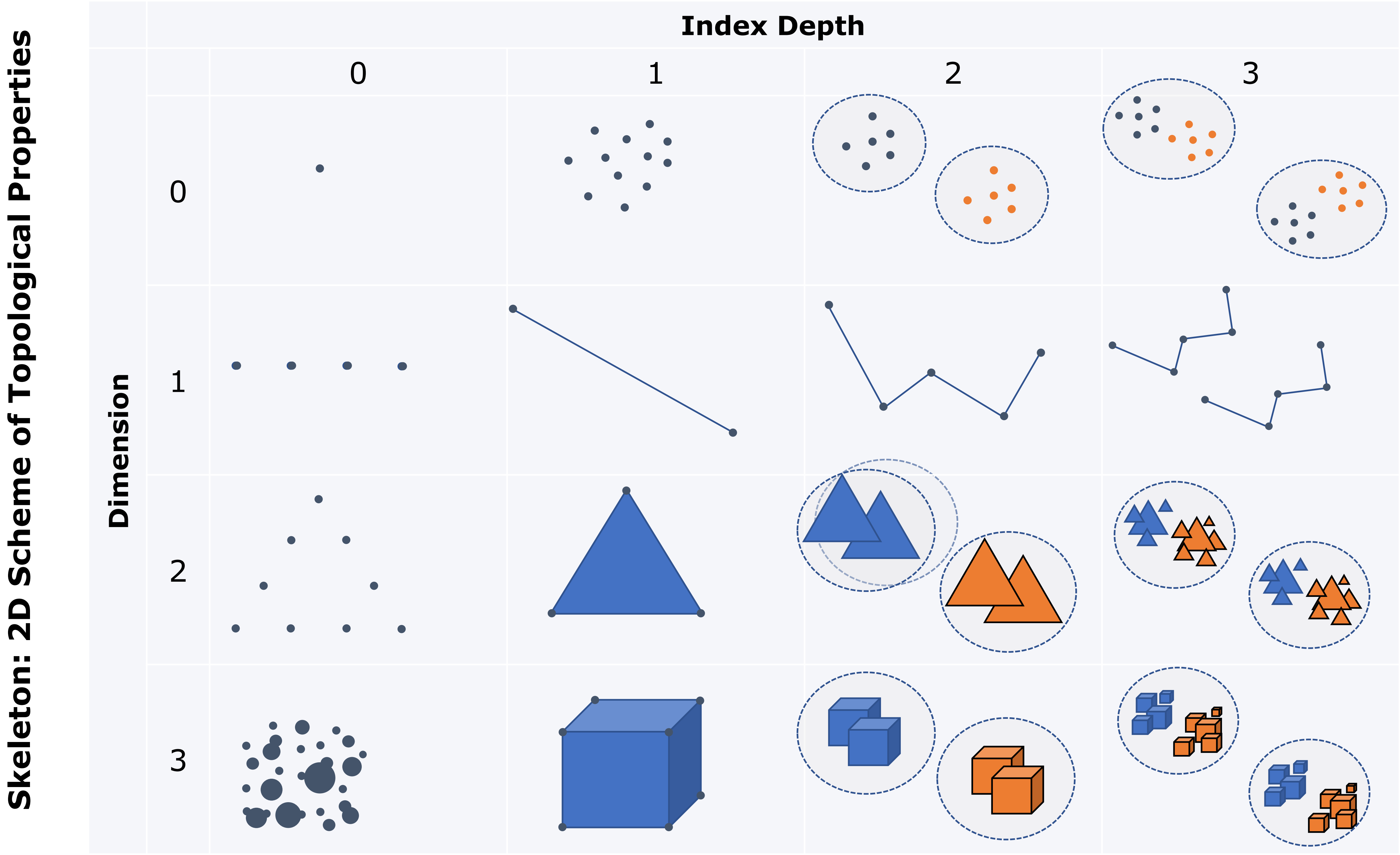
Dimensionality 1

```
/T=1/adcirc/Faces Group
/T=1/adcirc/Faces/Points Group
/T=1/adcirc/Faces/Points/Positions Dataset
```

Dimensionality 2

```
/T=1/adcirc/Connectivity Group
/T=1/adcirc/Connectivity/Points Group
/T=1/adcirc/Connectivity/Points/Positions Dataset
```

Dimensionality 3

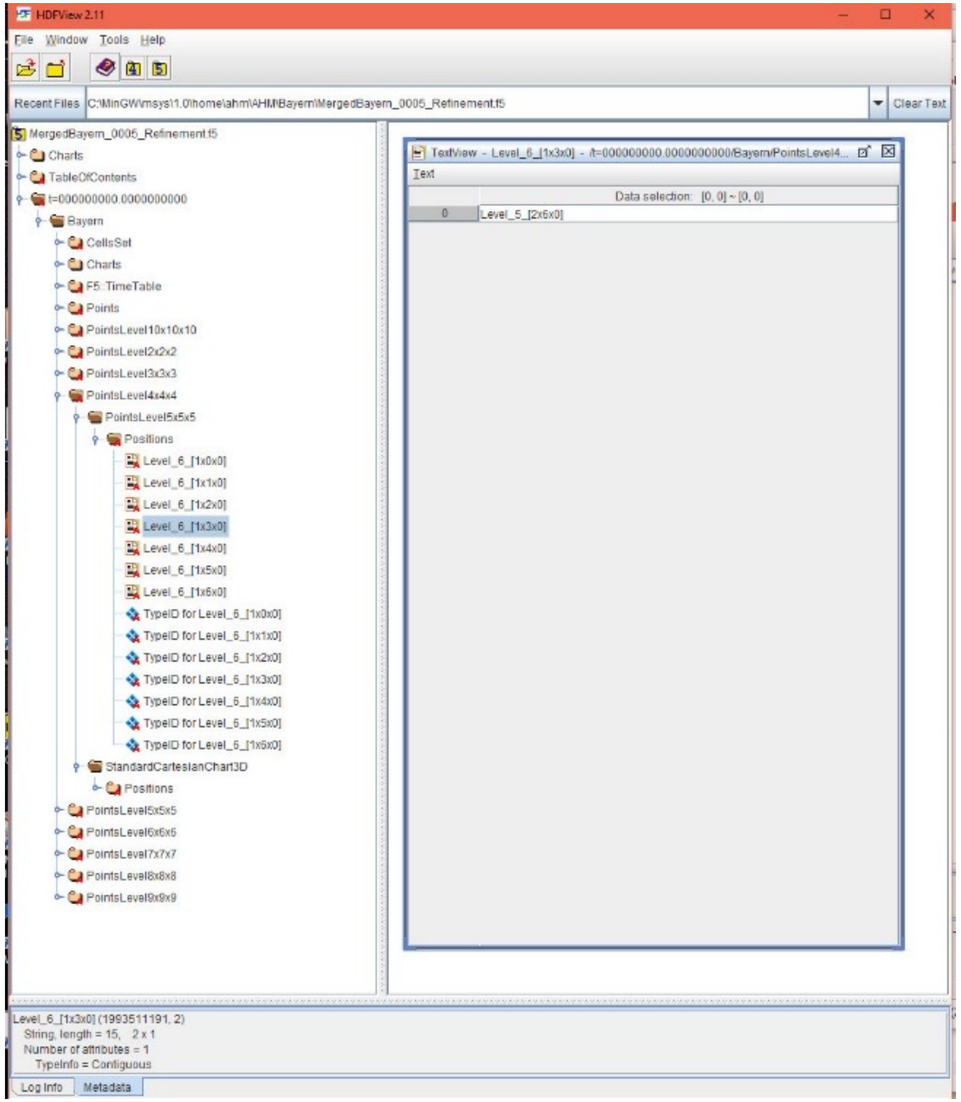

Skeleton: 2D Scheme of Topological Properties
Index Depth
Dimension
0
1
2
3
0
1
2
3

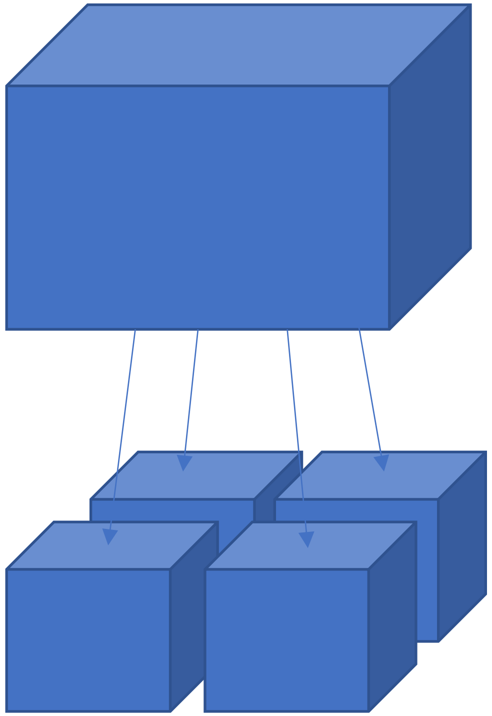
Hierarchical Skeletons

- Third identification parameter on Skeletons:

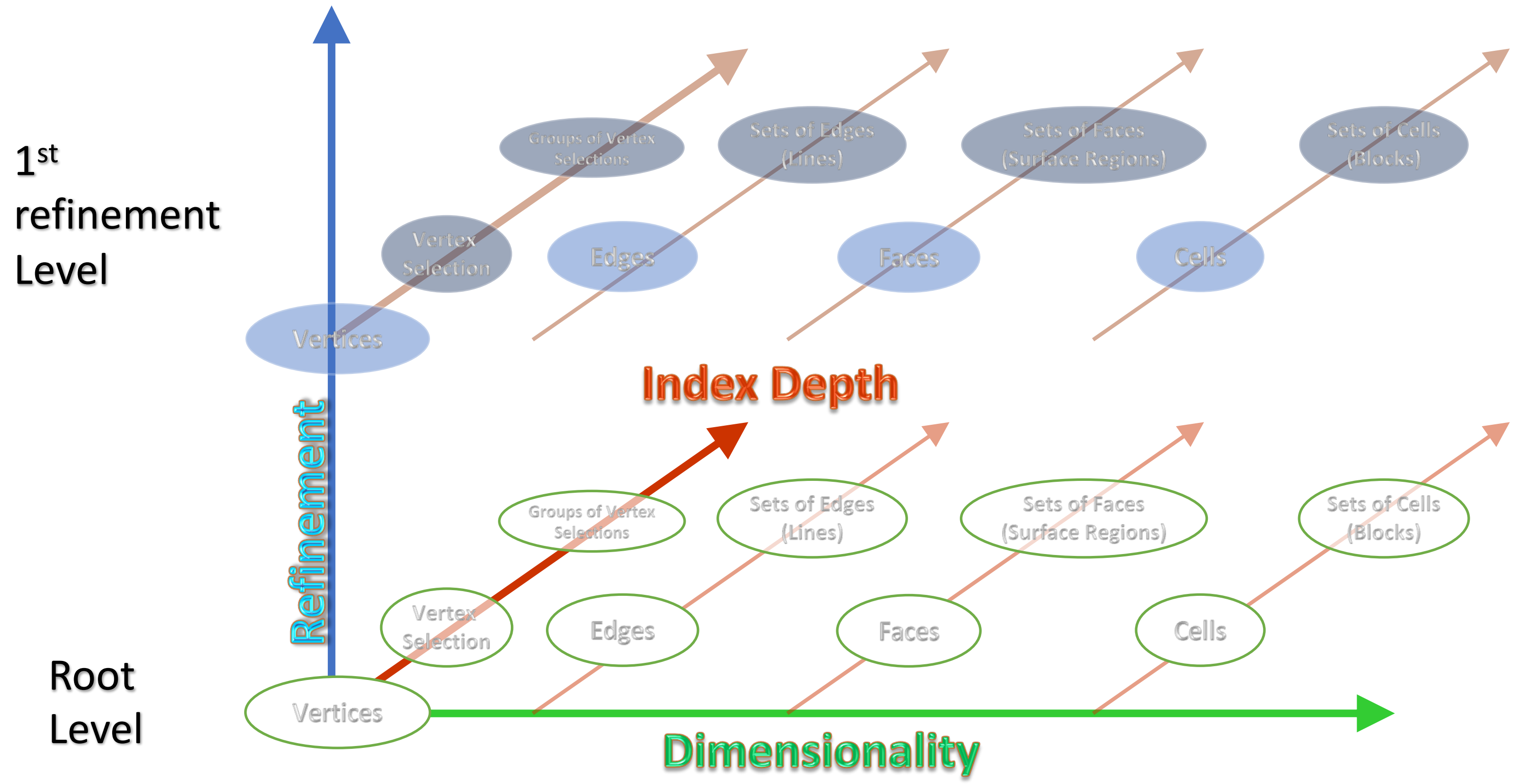
  "refinement" (integer valued or n-dimensional set of integer valued group attribute)
- Allows to formulate "a grid within a grid"
- Replicates topological structure on different refinement levels

Hierarchical References in F5

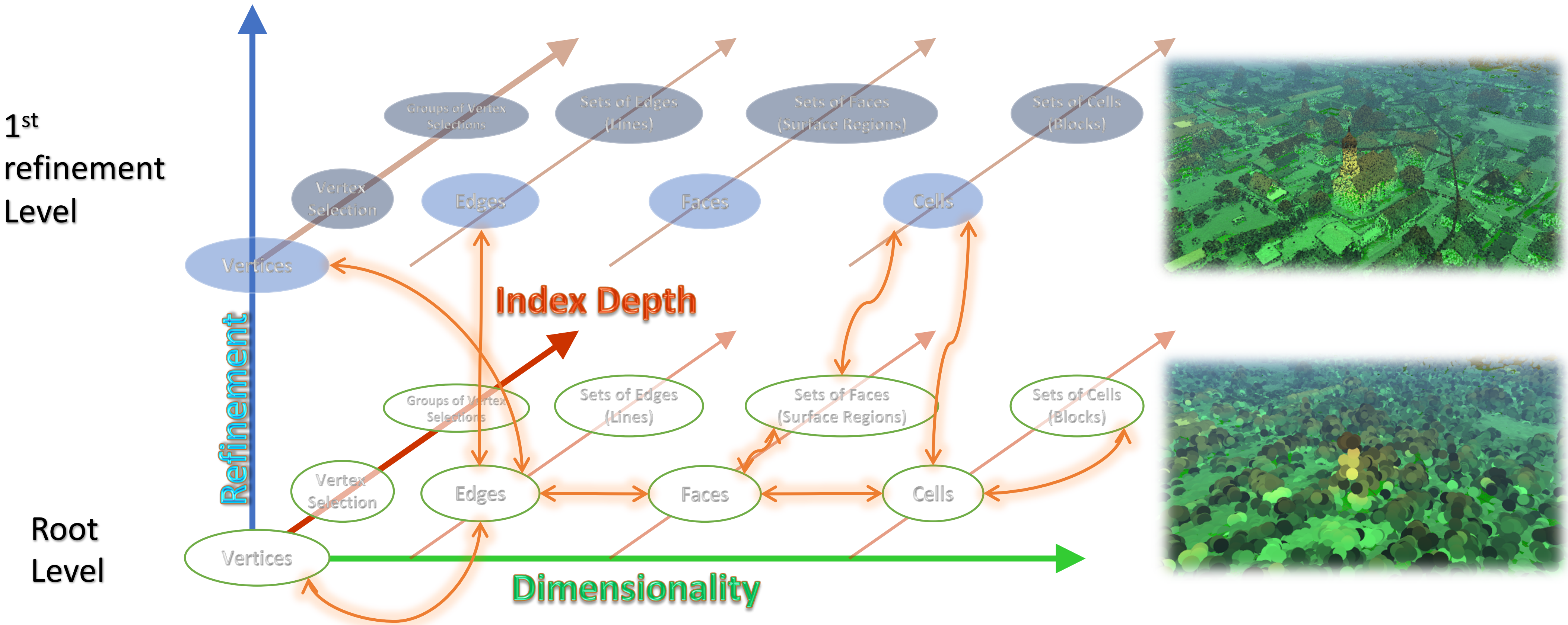


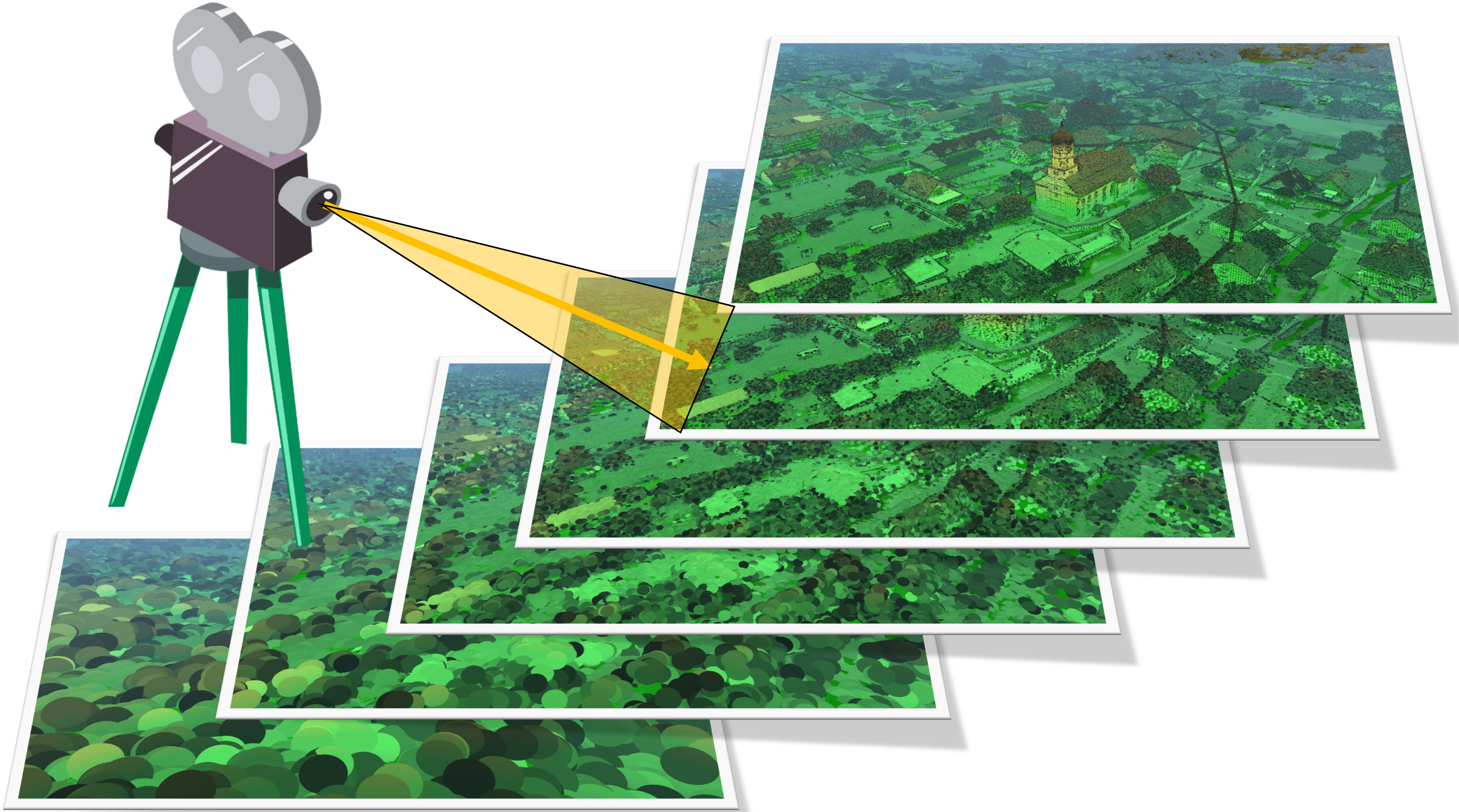


Skeletons as Fibers of a 3-dimensional parameter space

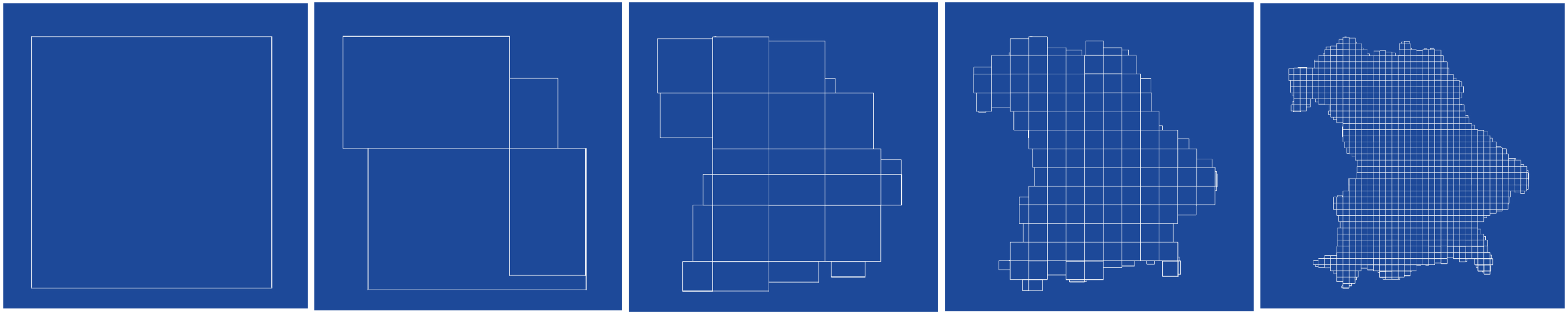


Relative Representations as Connectors between Skeletons





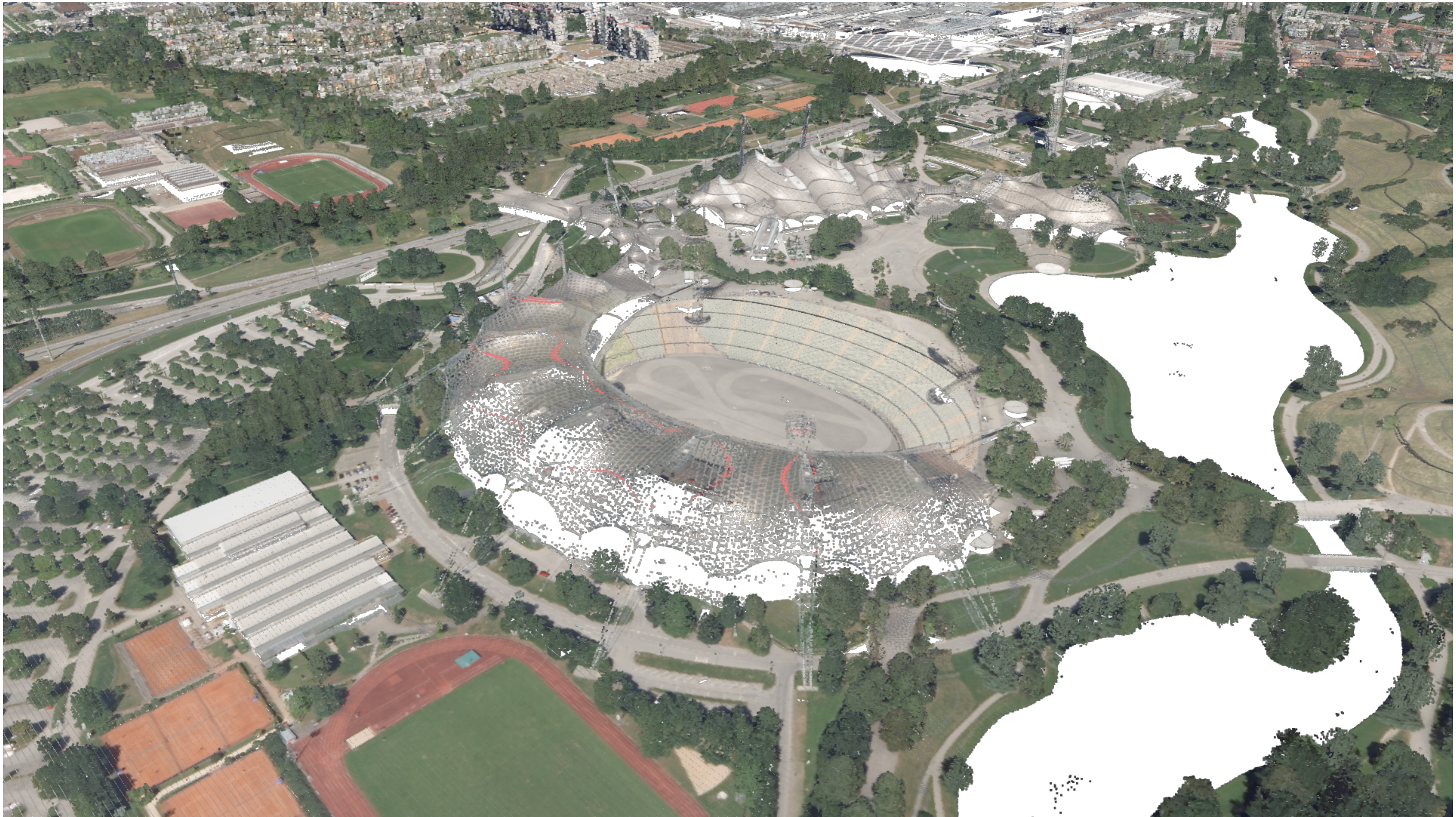
Bavaria AMR Scheme

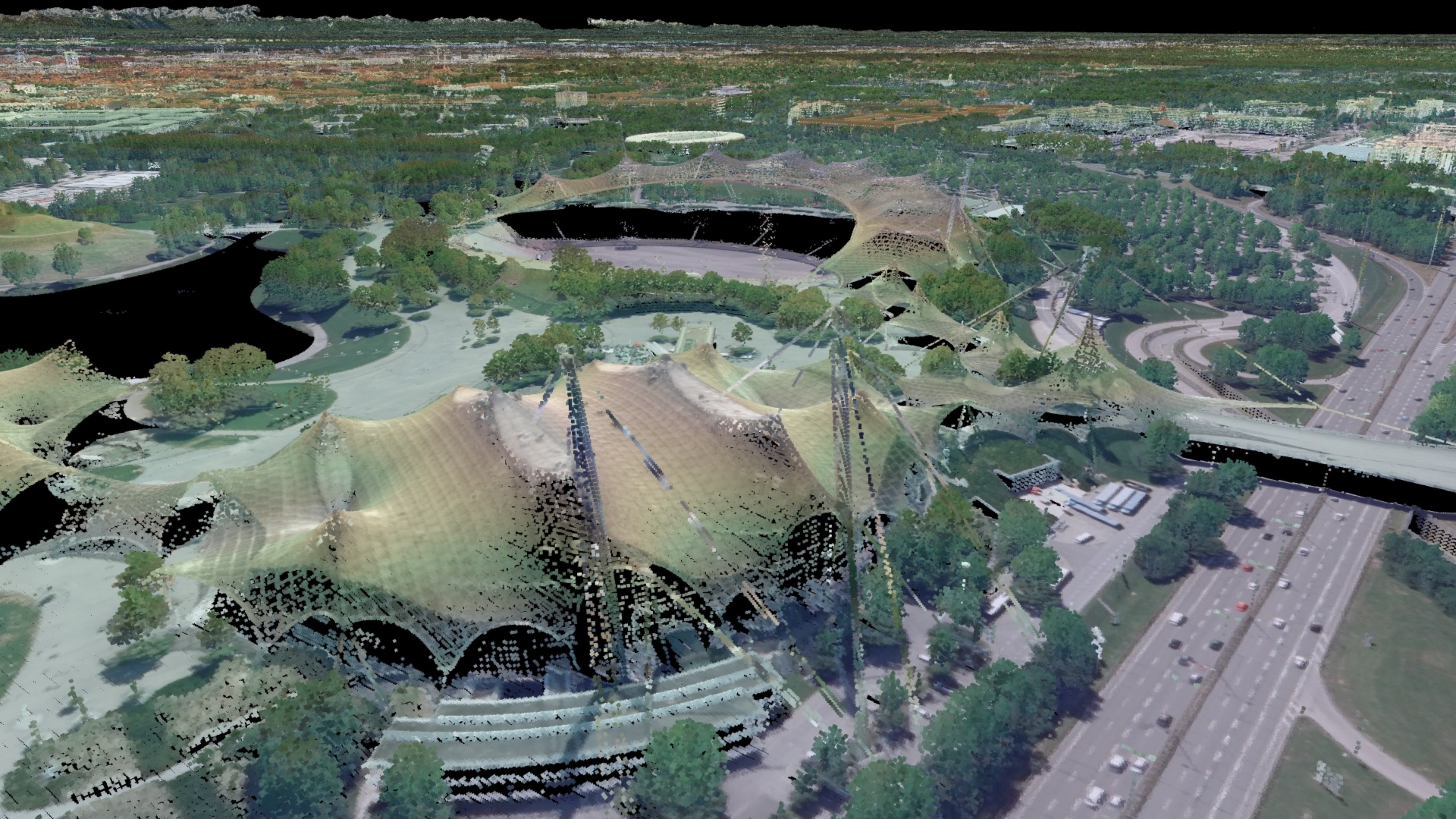


LIDAR: 12 Levels of Refinement, 250m on finest level

509.802.511.600 points,
8548174847160 bytes HDF5 (8.5TB)

Bavaria: Finest Data – Level 11





Summary

- The F5 scheme covers a wide category of data types for scientific visualization (i.e., data with spatio-temporal content)

- Maps directly to HDF5 as a hierarchy of groups expressing the semantics of a data set
- Simple data types modelled simple (e.g., just one dataset)
- Complex relationships are expressed using reuseable schemes
- Highly efficient for large data

Further Information

- The F5 data model and C library:

  www.fiberbundle.net

- Airborne HydroMapping and the HydroVish framework:

  www.ahm.co.at

- Contact: Werner Benger, w.benger@ahm.co.at