Scientific Data Model

Han-Wei Shen
The Ohio State University
What is a Data Model?

How do you describe the data represented by this image?
Data Model

- Describe the objects represented by the data
Data Model

• Describe the objects represented by the data
  – Structures of the objects
Data Model

- Describe the objects represented by the data
  - Structures of the objects
  - Properties of the objects
Data Model

- Describe the objects represented by the data
  - Structures of the objects
  - Properties of the objects
  - Relationships between the objects
Scientific Data Model

- Data set – a single or multiple valued function

\[ y_1 = f_1(x_1, x_2, x_3, \ldots, x_n) \]
\[ y_2 = f_2(x_1, x_2, x_3, \ldots, x_n) \]
\[ y_m = f_m(x_1, x_2, x_3, \ldots, x_n) \]

Each dependent variable \( y_i \) can have a tensor rank \( k \)
- \( k = 0 \): scalar; \( k = 1 \): vector; \( k = 2 \): 2D matrix, etc.

Data Model

\[ m \] dependent variables \( x_i \) (i=1..m)
\[ n \] independent variable \( v_j \) (j = 1..n)

Temperature
Pressure
Cloud density
...
Scientific Data Model

- Data set – a single or multiple valued function

- **Independent variables** (dimensions)
  - Spatial coordinates (longitude, latitude, height)
  - Time
  - Zone ID
  - ...

- **Dimensionality** - number of independent variables

- **Dependent variables**
  - The function values of independent variables
  - The number of values associated with each dependent variable can be described by its tensor rank
    - 0: scalar
    - 1: vector
    - 2: n x n matrix ...

- Data set – a single or multiple valued function
Domain Discretization

Continuous Domain

compute values
Scientific Data Set

Scientific Data Set = Domain Structure + Attributes

Domain Structure
- Topology: property invariant under transformation
- Geometry: instantiation of topology with specific positions
- Consists of Points and Cells, which define the Mesh

Attributes
One or multiple values (scalars, vectors, tensors) defined at points or cells
Cells are the fundamental building blocks of scientific data sets.
Cells define how points are connected together to form the basis for interpolation.
Cells can be of different dimensionality:
- 0 D: Vertices
- 1 D: Line; Polyline
- 2 D: Triangle; Quadrilateral; Polygon
- 3 D: Tetrahedron; Hexahedron; Voxel
Cell Types

1D
- Vertices
- Line
- Polyline

2D
- Triangle
- Quad
- Polygon

3D
- Tetrahedron
- Cube
- Hexahedron
- Pyramid
Attributes

- Scalars (e.g. density), Vectors (e.g. momentum), Tensors (e.g. stress tensor)
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structures)
  • Structured Grid
  • Unstructured Grid
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structures)
  – Structured Grid
    – Consisting of a collection of points and cells arranged on a regular lattice
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structures)
  – Structured Grid
    – Consisting of a collection of points and cells arranged on a regular lattice
    – Every point in the structured grid can be indexed by \((i,j)\) in 2D, \((i,j,k)\) in 3D, etc.
Scientific Dataset Types

- Data sets are categorized into different types based on their underlying grid (domain structures)
  - Structured Grid
    - Consisting of a collection of points and cells arranged on a regular lattice
    - Every point in the structured grid can be indexed by (i,j) in 2D, (i,j,k) in 3D, etc.
    - The position of the points, and hence the geometry of the cells, can be either implicitly defined (Cartesian grid), or explicitly specified (rectilinear or curvilinear grid)
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structures)
  – Structured Grid
    • Cartesian mesh
Scientific Dataset Types

- Data sets are categorized into different types based on their underlying grid (domain structures)
  - Structured Grid
    - Cartesian mesh
    - Rectilinear mesh
Scientific Dataset Types

- Data sets are categorized into different types based on their underlying grid (domain structures)
  - Structured Grid
    - Cartesian mesh
    - Rectilinear mesh
    - Curvilinear mesh
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structure)
  – Unstructured Grid
    • Also called irregular grid data
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structure)
  – Unstructured Grid
    • Also called irregular grid data
    • Unstructured grid points are irregularly distributed in space
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structure)
  – Unstructured Grid
    • Also called irregular grid data
    • Unstructured grid points are irregular located in space
    • It is often a result of space tessellation with simple shapes
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structure)
  – Unstructured Grid
    • Also called irregular grid data
    • Unstructured grid points are irregular located in space
    • It is often a result of space tessellation with simple shapes
    • Explicit connectivity information to form cells is necessary
Scientific Dataset Types

- Data sets are categorized into different types based on their underlying grid (domain structure)
  - Unstructured Grid
    - Polygonal mesh
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structure)
  – Unstructured Grid
    • Polygonal mesh
    • Tetrahedral mesh
Scientific Dataset Types

• Data sets are categorized into different types based on their underlying grid (domain structure)
  – Unstructured Grid
    • Polygonal mesh
    • Tetrahedral mesh
    • Hybrid Mesh