Overview of file formats for unstructured data

Anastasiia Novikova

Scientific Computing
Department of Informatics
University of Hamburg

2017-09-25
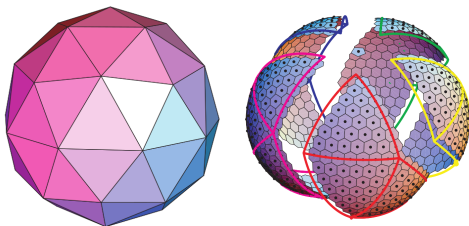
Motivation

Icosahedral grid
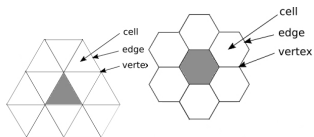

Figure: Icosahedral grid [2]

Figure: Ico grid and variables

File Formats

- CGNS
- ACRi CFD
- UGRID
- ADIOS
- NetCDF
- VTK
- PVTK
- ParaViewGeo

CFD General Notation System (CGNS)

- Goal: to provide a standard for recording and recovering computer data associated with the numerical solution of the equations of fluid dynamics.
- Current version: 3.3.0 (2016), was started in 1994.
- Formats: ".cgns" ADF (Advanced Data Format) based on CFD, and "_hdf5.cgns" HDF5. Both are binary formats, based on a simple tree structure.
- File addressing: 32 or 64 bits
- Dimentions: upto 3
- Shapes:
    - 0-D point
    - 1-D line
    - 2-D triangle, quadrangle
    - 3-D tetrahedron, pyramid, pentahedron, hexahedron
- Features: zones, portability, parallelity, compression with HDF5

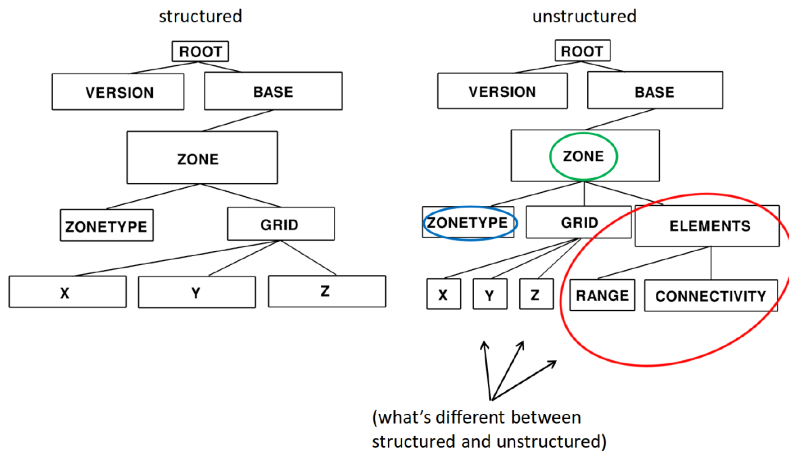# CFD General Notation System (CGNS)

Figure: CGNS File Mapping [3]

ACRi CFD

- Current version: 5.5 (2015), was started in 1979.
- Files:
  - mandatory commands file (".inp")
  - mandatory vertex file (".xyz")
  - mandatory element connectivity file (".cnc")
  - (optional) auxiliary connections (local refinement or split connectivity, periodic pairs) file (".blk", ".per")
- Modes:
  - MODE 1: Vertex Connectivity for Quad or Hex Elements
  - MODE 2: Vertex Connectivity for Mixed Hybrid Elements
  - MODE 3: Connectivity for Quad or Hex Elements with Split Sides
- Types:
  - Unstructured (Regular, Hybrid, Split)
- Dimentions: upto 3
- Shapes:
  - 2-D triangle, quadrangle
  - 3-D tetrahedron, pentahedron, hexahedron

Adaptable I/O System (ADIOS)

- Current version: 1.0 (2009).
- Formats:
    - binary packed file (".bp")
    - NetCDF4 and HDF5
- File addressing: 32(max. 2 GB) or 64 bits
- Dimentions: Unlimited
- Features: process groups, parallelity
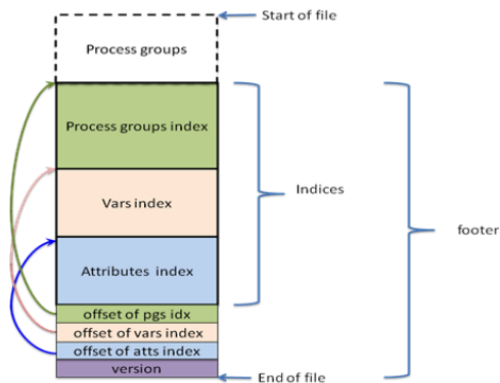
Adaptable I/O System (ADIOS)

Figure: ".bp" File Structure [1]

Network Common Data Form (NetCDF)

- Formats:
    - Classic, 64-bit, NetCDF4 ("*.nc" "*.cdf" "*.elev" "*.ncd")
- File addressing: 32 or 64 bits
- Dimentions: unlimited
- Shapes: use UGRID conventions

Comparisson of NetCDF and HDF5

Feature	NetCDF			HDF5
	Classic	64-bit	NetCDF4	
compresion	no	no	yes	yes
max. file size	8 EiB	8 EiB	unlimited	unlimited
max. dataset size	2 GiB	4 GiB	unlimited	unlimited
signed integer	8-bit	8-bit	8-bit	8-bit
	16-bit	16-bit	16-bit	16-bit
	32-bit	32-bit	32-bit	32-bit
				64-bit
unsigned integer			8-bit	8-bit
			16-bit	16-bit
			32-bit	32-bit
			64-bit	64-bit
floating point	32-bit	32-bit	32-bit	32-bit
	64-bit	64-bit	64-bit	64-bit
				+ prog. lang. specific
				+ hardware specific
compound	no	no	yes	yes
chunking			yes	yes
other data types				time
				references
string	no	no	yes	yes
tree depth	1	1	unlimited	unlimited
unlimited dimensions	yes (only 1)	yes (only 1)	yes	yes

UGRID

- Current version: 1.0 (2016), was started in 2013.
- Format: NetCDF
- Dimentions: upto 3
- Types of elements:
    - 0-D nodes
    - 1-D edge
    - 2-D face: triangle, quadrangle
    - 3-D volume: tetrahedron, pyramid, wedge, hexahedron

Visualization Toolkit (VTK)

- Current version: 8.0 (2017)
- Formats:
    - ImageData (".vti") — Serial vtkImageData (structured).
    - RectilinearGrid (".vtr") — Serial vtkRectilinearGrid (structured).
    - StructuredGrid (".vts") — Serial vtkStructuredGrid (structured).
    - UnstructuredGrid (".vtu") — Serial vtkUnstructuredGrid (unstructured).
    - PolyData (".vtp") — Serial vtkPolyData (unstructured).
- Shapes: 19
- Features: byte order automatical swap by software

Visualization Toolkit (VTK)

```
# vtk DataFile Version 2.0                    ⌐(1)
Really cool data              ⌐(2)
ASCII | BINARY            ⌐(3)
DATASET type
                               ⌐(4)
...
POINT_DATA n
...                            ⌐
CELL_DATA n                    (5)
...
```

Part 1: Header

Part 2: Title (256 characters maximum, terminated with newline \n character)

Part 3: Data type, either ASCII or BINARY

Part 4: Geometry/topology. *Type* is one of:
```
STRUCTURED_POINTS
STRUCTURED_GRID
UNSTRUCTURED_GRID
POLYDATA
RECTILINEAR_GRID
FIELD
```

Part 5: Dataset attributes. The number of data items n of each type must match the number of points or cells in the dataset. (If *type* is FIELD, point and cell data should be omitted.

Figure: VTK File Structure [4]

Parallel Visualization Toolkit (PVTK)

- Formats:
    - PImageData (".pvti") — Parallel vtkImageData (structured).
    - PRectilinearGrid (".pvtr") — Parallel vtkRectilinearGrid (structured).
    - PStructuredGrid (".pvts") — Parallel vtkStructuredGrid (structured).
    - PUnstructuredGrid (".pvtu") — Parallel vtkUnstructuredGrid (unstructured).
    - PPolyData (".pvtp") — Parallel vtkPolyData (unstructured).
- Shapes: 19
- Features: compression

ParaViewGeo

- Current version: 1.4.13
- Format: ParaView Data (".pvd") based on XML
- Shapes: like VTK

Summary

Type	CGNS	VTK	ParaView	ACRi CFD	UGRID	PETSc	ADIOS
Structured grid	+	+	+	+	+	+	+
Unstructured grid	+	+	+	+	+	+	+
Icosahedral	+	+	+	NA	+	NA	NA
HDF5 implementation	+	+	+	-	+(".nc")	+(".nc")	+

Table: ICO support

Bibliography

📄 *ADIOS 1.0 User's Manual*.
http://users.nccs.gov/~pnorbert/ADIOS-UsersManual-1.0.pdf.

📄 David A. Randall et al. "Climate modeling with spherical geodesic grids." In: *Computing in Science & Engineering*. Vol. 4. IEEE, 2002, pp. 32 –41. DOI: 10.1109/MCISE.2002.1032427.

📄 C. L. Rumsey et al. *Recent Updates to the CFD General Notation System (CGNS)*. https://cgns.github.io/CGNS_docs_current/slides/AIAA-2012-1264-CGNSstatus_slides.pdf. 2012.

📄 *VTK User's Guide*. 11th. Kitware Inc., 2010. ISBN: 978-1-930934-23-8. URL: https://www.vtk.org/vtk-users-guide.