Geophysical Research Abstracts
Vol. 13, EGU2011-12419-1, 2011
EGU General Assembly 2011

Using HDF5 for Cross Disciplinary Data Archival and Highly Performing Processing of Multi-Variant Observational and Computational Data with Simple or Complex Topology

Marcel Ritter (1), Werner Benger (2,3), Frank Steinbacher (1), Markus Aufleger (1), Joseph Stoeckl (3), Donna Delparte (4), Mike Folk (5), and Quincey Koziol (5)

(1) Unit for Hydraulic Engineering, Department for Infrastructure, University of Innsbruck, Austria, (2) Center for Computation & Technology, Louisiana State University, USA, (3) Department of Astrophysics, University of Innsbruck, Austria, (4) Geography and Environmental Studies, University of Hawaii at Hilo, USA, (5) The HDF Group, Champaign, IL, USA

Delivering observational data in "the best manner", best for the specific purpose, is an cross-disciplinary problem occurring in geosciences same as in related fields such as astrophysics or more distant fields like medical imaging. Especially, storing, collecting, searching and retrieving large and complex geo-datasets is a challenge that is faced by many organizations conducting multi-disciplinary research. The incompatibilities of data models in the geosciences and numerical community are a known issue [Nativi04]. Despite application-specific particular needs common problems exist: descriptiveness, ability to exchange data across software tools, high-performing data access, flexibility and extendability on meta data storage, support for numerical types beyond standard float and double precision, compression or partitioning.

We propose to store data sets from geosciences in the HDF5 format, which enables interdisciplinary collaborations and reuse of data processing techniques over different scientific domains.

HDF5 was originally developed for the high performance computing community and has been successfully applied to many different scientific applications over years. HDF5 provides a self-describing and sustainable way of storing and accessing data, both observational and numerical. Its flexibility is comparable to XML, but as a binary format allows high performance and compact storage while always being convertible to XML or similar text formats. It contains all meta information about types, such as precision or endianness and allows for custom types. HDF5 is already used in geoscience applications by NASA's HDF5-EOS and in astrophysics applications by the HDF5-SRB-FLASH format.

On top of the generic capabilities of HDF5 the F5 model [Benger04][Ritter09] offers a data model similar to the one envisioned by Butler [Butler89] for data that carry intrinsic geometric information, covering cross-disciplinary properties of spatio-temporal data. The F5 model has been applied successfully for data stemming from computer simulations, ranging from 500GB to 9TB, addressing data from Computational Fluid Dynamics [Benger09], numerical relativity [Ritter10] and medical visualization.

At the University of Hawaii data exchange challenges are being faced by creating a Hawaii Geospatial Data Repository to support critical statewide research into ecological and anthropogenic relationships that addresses the unique environmental and socio-economic issues for the Hawaiian Islands. The data repository will be a catch of all existing datasets and facilitate the input of new data from sensor networks, field data collection as well as remotely sensed Light Detection And Ranging (LIDAR) and image data formats. This data from diverse sources and disciplines needs to perform across software platforms and be accessible for researchers.

We demonstrate the formulation of observational multi-variant and multi-channel data from full waveform LIDAR scanning [Steinbacher10] in F5. These data are represented on geo-referenced Cartesian coordinates and local laser coordinates. Point and line topologies are used to hold multi-variant data. Two time dependencies are captured on a micro and macro level. The open, easy to read and write F5 format captures the huge observational data including partitioning and data compression. Finally, data could be loaded into VISH [Benger07] which supports generic data types for scientific visualization.

Using the widely used generic file format HDF5 opens new perspectives for data handling and exchange across disciplines and applications. For example, with the support for parallel processing as enabled by HDF5 tools, such as the astronomical MOSAIC parallel image processing tool, these may directly become applicable to geoscience data.

Bibliography:

[Benger04] Werner Benger, Tensor Field Visualisation via a Fiber Bundle Data Model,

Department of Mathematics and Computer Science, University of Berlin, 2004.
[Benger07] Werner Benger, Georg Ritter, and Rene Heinzl. The Concepts of VISH. In 4th High-End Visualization Workshop, Obergurgl, Tyrol, Austria, June 18-21, pages 26-39. Berlin, Lehmanns Media-LOB.de, 2007.
[Benger09] Fiberbundle-based visualization of a stir tank Fluid. Benger W., Ritter M., Acharya S., Roy S., Jijao F. At WSCG, Plzen. Czech Republic. 2009
[Butler89] David M. Butler and M. H. Pendley, A visualization model based on the mathematics of fiber bundles, Computers in Physics 3, 1989, no. 5, 45-51
[FiberBundle11] Werner Benger, The Fiber Bundle HDF5 Library, http://www.fiberbundle.net/ 2011
[HDF5] Homepages of The HDF Group (THG), THG Home Page Information, Support, and Software from The HDF Group, http://hdfgroup.com, Champaign, 2011
[Nativi04] Nativi, S., Blumenthal, B., Habermann, T., Hertzmann, D., Raskin, R., Caron, J., Domenico, B., Ho, Y., Weber, J.: Di[U+FB00]erences among the data models used by the geographic information systems and atmospheric science communities. In: Proceedings American Meteorological Society - 20th Interactive Image Processing Systems Conference. 2004
[Ritter09] Introduction to HDF5 and F5. Ritter M. (2009). CCT Technical Report Series, Baton Rouge, Louisiana, CCT-TR-2009-13.
[Ritter10] Computing Geodesics in Numerical Space Times. Marcel Ritter. Master's Thesis. University of Innsbruck. 2010
[Steinbacher11] Airborne Hydromapping and HydroConnect Shallow water bathymetry – pioneering underwater insights. Frank Steinbacher, Martin Pfennigbauer, Robert Klar and Markus Aufleger. International Association for Hydro-Environment Engineering and Research (IAHR) 2011. Brisbane.