HDF Product Designer: A tool for building HDF5 containers with granule metadata

Lindsay Powers
Aleksandar Jelenak, Joe Lee, Ted Habermann
The HDF Group
Data Producer’s Conundrum

HDF Features
- Datatypes
- Groups
- Attributes
- Dimension scales
- Compression
- Chunking
- Scale/offset
- Etc.

Interoperability
- Conventions
- Metadata
- Software

Project Requirements
- Science objectives
- Data processing, discovery & distribution
- Data documentation
- User engagement, preparedness, feedback

HDF Product Designer
Key Goals

• Facilitate creation of interoperable and standards-compliant data products in HDF5 as early as possible in the project development process
• Support multiple computing platforms without requiring the full software stack of development tools and libraries
• Easy and intuitive editing (create, update, move, copy, delete) of HDF5 objects
• Collaborative approach to product design (project, team, organization)
• Incorporation of best practices and standards from targeted data user communities
• Integration of compliance and interoperability tests into the design workflow
• Design import from existing files
• Design export as HDF5 files, HDF5/JSON, or as source code in several programming languages
Features

- Projects
- Designs
- CRUD operations on HDF5 objects
- Conventions support
- Validation services
- Collaborative workflow
Project

- Organizational and collaborative space
- One or more users
- Zero or more designs
- Every user must belong to at least one project
- All members of a project have access to its designs
- User project roles:
  - Manager (not implemented yet)
  - Designer
  - Value Editor (not implemented yet)
  - Viewer
Design

- Represents content to be stored in one HDF5 file
- Not actual HDF5 file
- Versioned
  - Simple timeline of checkpoints (saved versions)
  - Each version must have unique label
  - Only the current working version (label: HEAD) can be edited
- Many import source types
- Many export types including source code
CRUD Operations

- Create, read, update, delete, copy, move
- Available on designs and HDF5 objects
- Editable properties:
  - Datatype
  - Rank, shape, max/unlimited dimension sizes
  - Storage (compact, contiguous, chunked)
  - Fill value
  - Compression
  - Attribute value
Conventions

• Currently Supported:
  • NetCDF User Guide Attribute Conventions (NUG)
  • Attribute Convention for Data Discovery (ACDD)
  • Climate and Forecast convention (CF)
  • HDF-EOS (partial)

• Implemented as CLIPS expert system rules
Validation Services

• A set of online services for interoperability testing
• The level of support for conventions varies between different software tools so it is important to verify using actual file
• Input is HDF5 template file
• Output is typically displayed in a web browser
Validation Services

• Currently available:
  • netCDF CDL
  • Get as netCDF3 file
  • CF (NCO’s ncdismember)
  • ACDD (THREDDS UDDC service)
  • ISO metadata (THREDDS ISO service)
  • OPeNDAP Data Access Form
  • THREDDS Dataset Access Page
System Architecture

HDF5 JSON

HDF4 Map XML

NcML

netCDF CDL

HDF5

Desktop App

RESTful Server

Hyrax, THREDDS Server

Data Store

Flexible Output

HDF5 File Template

Product Documentation

Fortran

IDL

MATLAB

Python
Collaboration

Individuals
Teams
Projects
Programs
Thank you!

Questions?

Contact: ajelenak@hdfgroup.org

This work was supported under the NASA Earth Observing System Data and Information Systems (EOSDIS) Evolution and Development (EED-2) Program under prime contract number NNG15HZ39C. Any opinions, findings, or conclusions expressed in this material are those of the author and do not necessarily reflect the views of NASA.