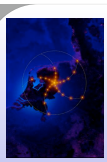


LOFAR & HDF5: Toward a New Radio Data Standard.

K. R. Anderson[1], A. Alexov[1], L. Bähren[1], J-M. Grießmeier[2], M. Wise[3],G.A. Renting[3]

1: University of Amsterdam (UvA), 2: Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E),
3: Netherlands Institute for Radio Astronomy (ASTRON)

WWW.LOFAR.ORG

For decades now, scientific data volumes have experienced relentless, exponential growth. As a result, legacy astronomical data formats are straining under a burden not conceived when these formats were first introduced. With future astronomical projects ensuring this trend, ASTRON and the LOFAR project is exploring the use of the Hierarchical Data Format, version 5 (HDF5), for LOFAR radio data encapsulation. Most of LOFAR's standard data products will be stored natively using the HDF5 format. In addition, HDF5 analogues for traditional radio data structures such as visibility data and spectral image cubes are also being developed. The HDF5 libraries allow for the construction of potentially distributed, entirely unbounded files. The nature of the HDF5 format further provides the ability to custom design a data encapsulation framework. The LOFAR project has designed several data formats that will accommodate and house all LOFAR data products, the primary styles and kinds of which are presented in this paper. With proper development and support, it is hoped that these data formats will be adopted by other astronomical projects as they, too, attempt to grapple with a future filled with mountains of data.

The LOFAR Radio Telescope

- The "LOw Frequency ARray"
- Officially opened by Queen Beatrix, June 12, 2010.
- Currently, the "largest radio telescope in the world."
- 25,000 networked, passive phased array antennae, 36 fields in The Netherlands
- International stations, constructed or planned: Germany(5), Sweden(2), France(2), and the United Kingdom(1)
- International Baselines to 1500 km; unlimited potential expansion.
- Low Band Antenna (LBA) bandpass:10-90 MHz
- High Band Antenna (HBA) bandpas:110-250 MHz.
- Data Correlation: IBM Blue Gene/P supercomputer, Groningen, NL.

The LOFAR "Superterp", Exloo, NL

The Blue Gene/P supercomputer at Groningen, aka "STELLA"

LOFAR Data: Variety, Complexity, Volume

LOFAR – Multiple Observational Modes

Datasets produced by LOFAR observations will vary tremendously in size. Images, Beam-formed data, Transient Buffer board (TBB) time-series data are expected to produce large files, with the beam-formed and TBB potentially forming files of several tens of terabytes.

Sample Beam-Formed data sizes

Exposure Time	Number of Subbands	Number of Stations	File Size Known Mode	File Size Search Mode
1 min	248	5	11.2GB	56GB
1 min	248	20	11.2GB	244GB
10 min	248	5	112GB	560GB
10 min	248	10	112GB	1.1TB
10 min	248	20	112GB	2.2TB
10 min	248	30	112GB	3.3TB
20 min	248	5	224GB	1.1TB
30 min	248	5	336GB	1.7TB
1 hr	248	5	672GB	3.4TB
1 hr	248	10	672GB	6.7TB
1 hr	248	20	672GB	13.4TB
1 hr	248	30	672GB	26.8TB
2 hr	248	5	1.3TB	6.7TB
12 hr	248	5	8.0TB	40.3TB
12 hr	248	15	24.0TB	120.1TB

- LOFAR's observational modes are capable of producing highly complex, large volume datasets *(see table, left)*
- Data rates up to 10Tb/sec
- File sizes, 10's to 100's terabytes
- Hierarchical, multidimensional data
- Not a job for FITS!

BoF No. 5! HDF5! Wed, Nov 10 18:30 h

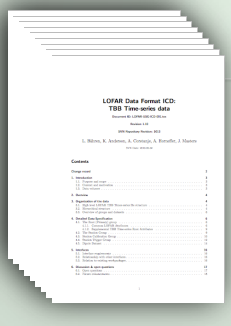
LOFAR Data Format Specifications -- HDF5

A viable solution was needed for potentially massive LOFAR data products. HDF5 provides a framework allowing users to essentially design their own files to appropriately accommodate a variety of data, something LOFAR will produce. For over a year now, the LOFAR project has been engaged in developing and designing a complete set of specifications for all LOFAR observational data. This has necessarily required differing file designs for differing data, with a certain structural parallelism maintained across all file designs.

ATTEND! *"Towards HDF5"* Birds of a Feather Session No.5 Wed, Nov 10 18:30 h

These **Interface Control Documents[4] (ICD)** provide detailed descriptions of all expected LOFAR Data Products.

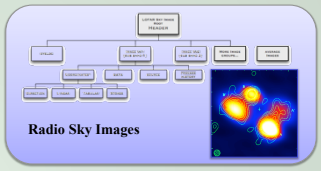
- **Radio Sky Images**
- **Transient Time Series**
- **Beam-Formed data**
- **Dynamic Spectra**
- **UV Visibility**
- **Rotation Measure Synthesis**
- **Near-field Imaging**

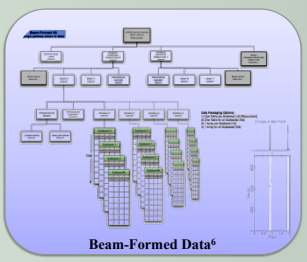


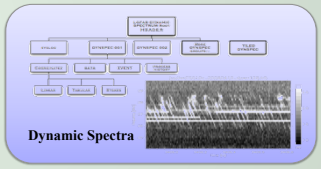
HDF5 Home Page: www.**hdfgroup.org**

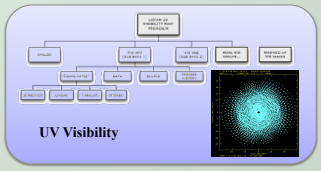
*4 Documents available from the **LOFAR Wiki ("Data Products" link), http://usg.lofar.org/wiki***

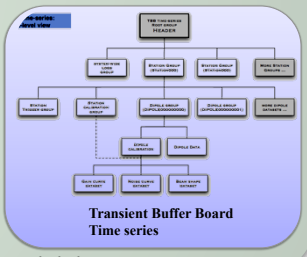
Schemata: LOFAR hierarchical data formats[5]









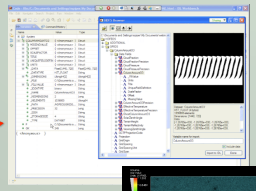


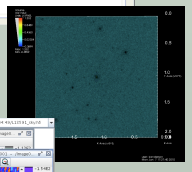
5 RM Synthesis and Near Field imaging under development
6 See Alexov, A., The LOFAR Pulsar Data Pipeline, ISKAF, 2010

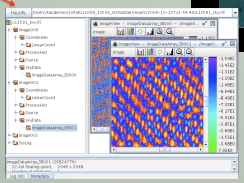
Toolsets, Libraries, and Packages

Since inception, the body of libraries and tools available to work with HDF5 files has grown substantially. In addition to the HDF5 library itself, these packages and interfaces are also available:

- **IDL**
- **Visit**
- **HDFView**
- **DAL (LOFAR)**
- **Pydal (LOFAR)**
- **h5.py**
- **Pytables**







The LOFAR project is developing the C++ **Data Access Library (DAL)**, which will provide full scope constructors for creating and accessing LOFAR data products.

Summary And Future Considerations

In order that adoption of HDF5 in astronomy prove useful in the real world, the LOFAR project is committing resources to help develop the next generation of astronomical tools for LOFAR data.

The large effort is the development of the **Data Access Library (DAL)**, which ultimately will provide interfaces through FITS, the Casa/AIPS++ Measurement Sets and HDF5. A python interface to the DAL, **pydal**, will also be forth coming. All LOFAR products will be accessible through DAL tools.

Future work involves developing an interface for **DS9** and HDF5 LOFAR data files – yes, you will be able to open and examine LOFAR data with the de facto standard astronomical image viewer.

The large effort by LOFAR to design an HDF5 radio data standard is driven in large part by two considerations. One, there is no effective standard in radio astronomy data. Two, as indicated above, the expected data volumes produced by LOFAR will, in many cases, swamp current file technology. It is this reality that has led LOFAR on this work, and it is hoped that these efforts toward building a radio data standard will be recognized as essential to the future of radio astronomy.

LOFAR · University of Amsterdam · NWO · ASTRON