

Data and Metadata standardization in the Madrigal Database

Bill Rideout

MIT Haystack Observatory

brideout@mit.edu

2023 IHDEA Meeting

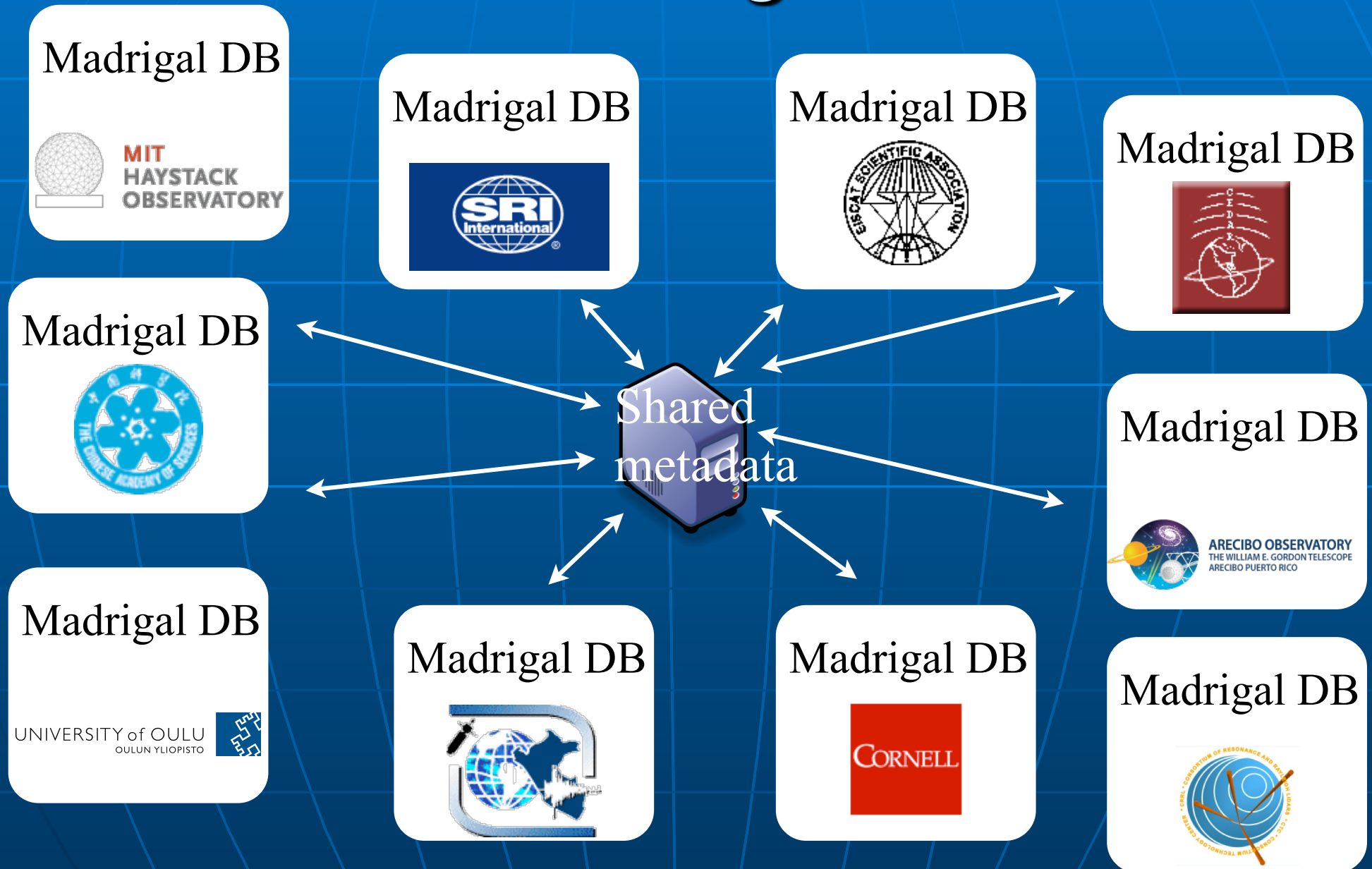
APL

Outline

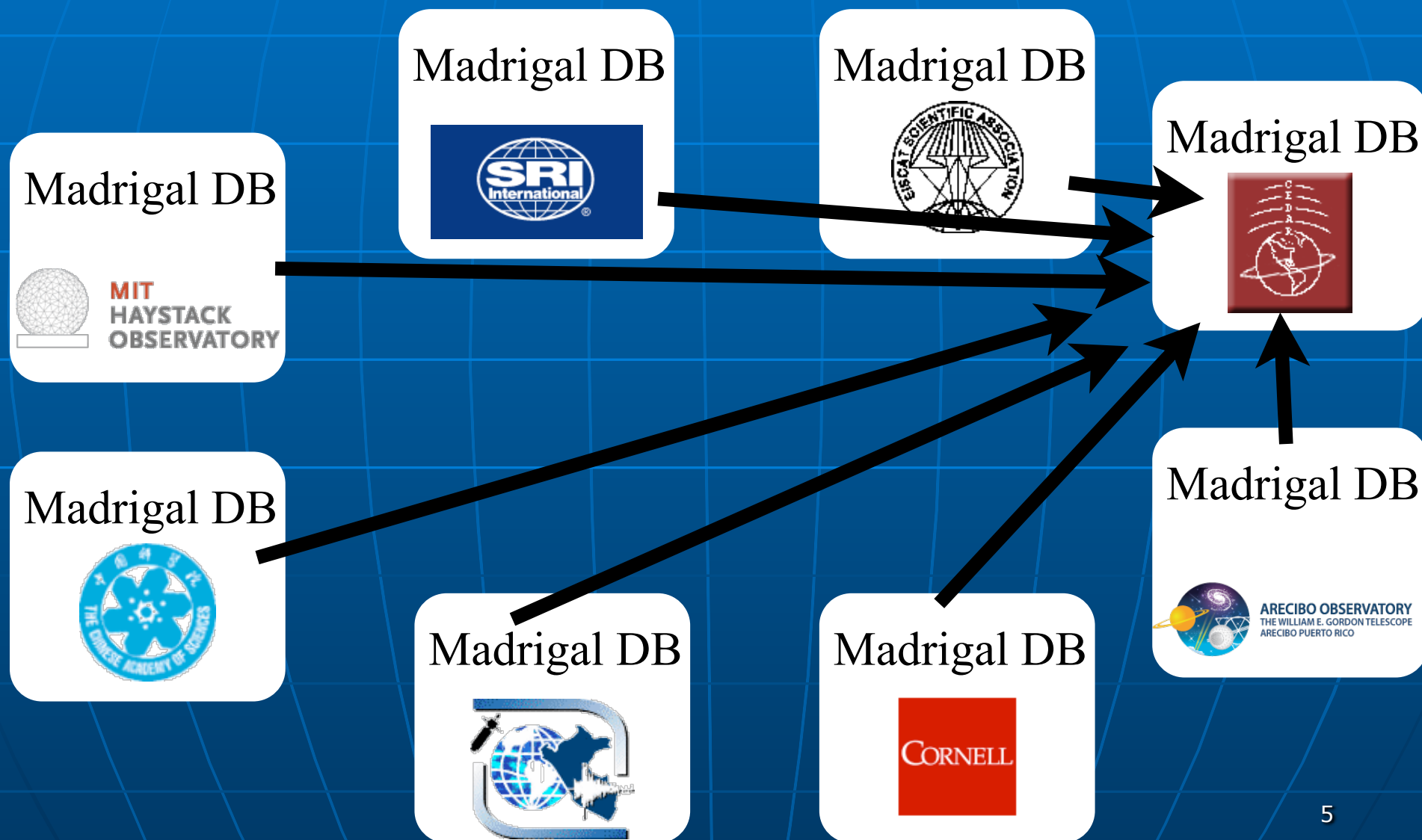
- What is Madrigal?
- Standardization of Madrigal data
- Standardization of Madrigal metadata
- Accessing data and metadata
- Lessons learned
 - Reproducibility of datasets
 - Attribution of datasets

What is Madrigal?

Madrigal is a distributed database with a focus on ground-based



Cedar Madrigal archive imports all data weekly



The Madrigal database stores data from a wide variety of upper atmosphere research instruments

Incoherent Scatter Radar



TEC via GPS



MF Radar



Examples of number of instruments in Madrigal:

- Incoherent scatter radars: 22
- MST radars: 3
- MF radars: 16
- Meteor radars: 11
- FPI: 32
- Michelson Interferometers: 6
- Lidars: 9
- Photometers: 7

Other examples:

- GPS TEC
- DMSP

Madrigal is open-source

[CEDAR Home](#) [Access data](#) [Access metadata](#) [Run models](#) [Documentation](#) [Other Madrigal sites](#) [OpenMadrigal](#)

Welcome to the Madrigal3 CEDAR Database

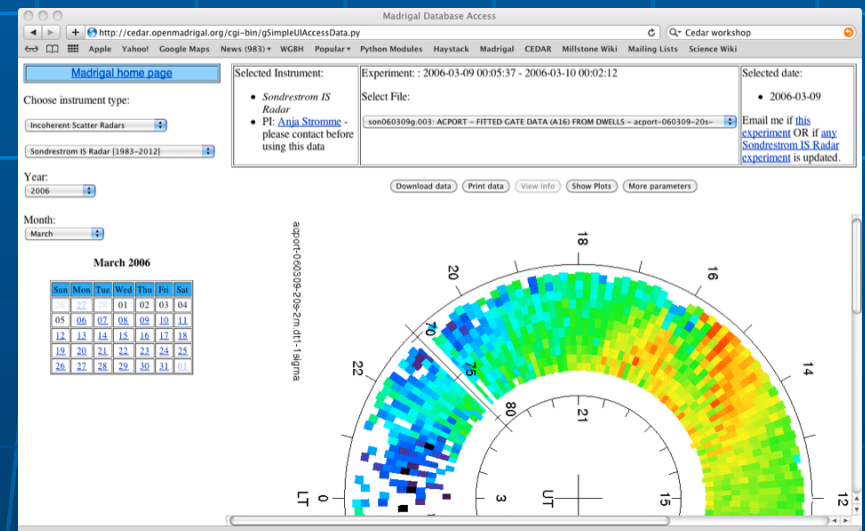
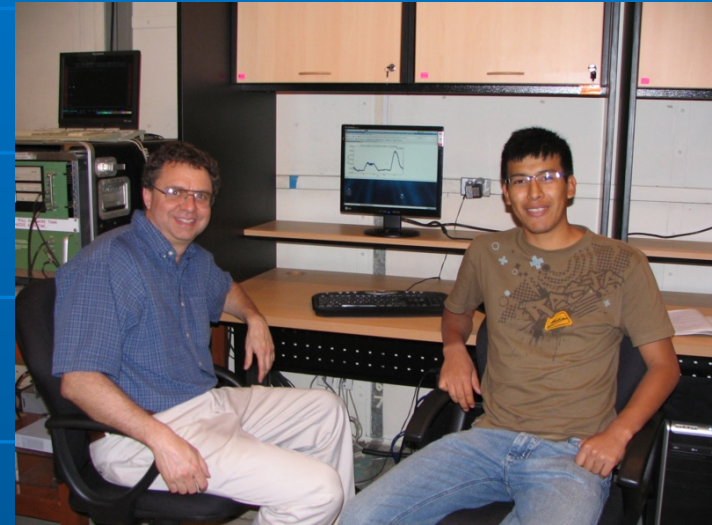
Madrigal is an upper atmospheric science database used by groups throughout the world. Madrigal is a robust, World Wide Web based system capable of managing and serving archival and real-time data, in a variety of formats, from a wide range of upper atmospheric science instruments. Data at each Madrigal site is locally controlled and can be updated at any time, but shared metadata between Madrigal sites allow searching of all Madrigal sites at once from any Madrigal site.

To see a list of all Madrigal sites, use the *Other Madrigal sites* pull down menu. Data can also be accessed directly, using *APIs* which are available for several popular programming languages (Matlab, python, and IDL). A Subversion archive of all Madrigal software and documentation is available from the *Open Madrigal* Web site. The latest version of Madrigal and the remote APIs may also be downloaded from there.

Use of the Madrigal Database is generally subject to the CEDAR Rules-of-the-Road . Prior permission to access the data is not required. However, the user is required to establish early contact with any organization whose data are involved in the project to discuss the intended usage. Data are often subject to limitations which are not immediately evident to new users. Before they are formally submitted, draft copies of all reports and publications must be sent to the contact scientist at all data supplying organizations along with an offer of co-authorship to scientists who have provided data. This offer may be declined. The Database and the organizations that contribute data must be acknowledged in all reports and publications and whenever this data is made available through another database. If you have any questions about appropriate use of these data, contact brideout@haystack.mit.edu

If you want to use the old Madrigal 2 version of the CEDAR Madrigal database, it is still temporarily available at <http://madrigal.haystack.mit.edu>. If you are using the old version because of a problem with Madrigal3, please contact brideout@haystack.mit.edu to describe the issue.

[Link to Subversion \(source code\)](#)



Standardization of Madrigal data

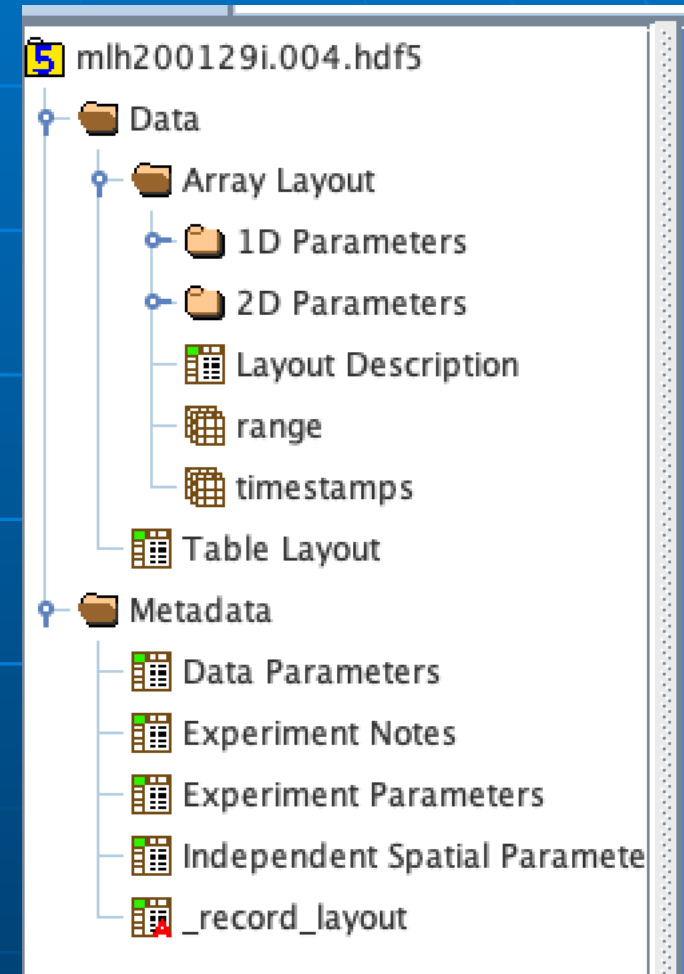
- Madrigal's underlying data format is well-defined Hdf5
 - Required data fields meant that the data should be fully self-defining
 - Standard parameters/units, definitions required in the file
 - Simple Python API for creating files
- Other export formats - Ascii, netCDF4

Other features of Madrigal HDF5 standard

- Standardized parameters allows the existence of derivation engine
- All parameters have corresponding error parameters
- Independent parameters built into standard

Cedar file format: Structured Hdf5

- Table Layout always given
- Array Layout default if independent parameters given (other than time)
- Array layout can be turned off if independent parameters non-repeating so that array would be too sparse.



- Standardization of Madrigal metadata

Madrigal Metadata Model

Madrigal site

(typically a facility with scientists and a Madrigal installation)

Madrigal DB



Data shared among all Madrigal sites

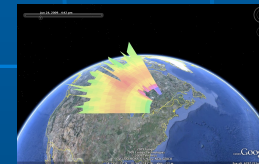
Instruments

(ground-based, typically with a set location)



Experiments

(typically of limited duration, with a single contact)



Experiment Files

(represents data from one analysis of the experiment)



Records

(measurement over one period of time)



Data unique to one Madrigal site

Madrigal metadata

- Internally stored as comma-delimited text
- Externally accessible via python API
 - API has been used to make Madrigal export other metadata standards
- Same API can be used to search for and access data

Data reproducibility

- Madrigal files are never deleted
 - Newer files may make older files by marked as “history”
- Every file has citable url
- Python API allows creation of a citation to multiple files

Millstone Home Access data ▾ Access metadata ▾ Run models ▾ Documentation Other Madrigal sites ▾ OpenMadrigal

Long Duration Run: 2023-01-24 14:04:50-2023-01-24 20:11:07

PI: Phil Erickson - please contact before using this data Email me if [this experiment](#) OR if any Millstone Hill IS Radar experiment is updated. Show non-default files:

Select file:

mlh230124g.001.hdf5: Combined basic parameters file - all antennas and modes - Preliminary - final calibration pending

Plots/docs Download file ▾ Print file ▾ View file info Cite this file

Use the citation below to reference this data file. The url is meant to be a permanent url. This data file may later be updated, but this link will still point to the original version of the data file.

Phil Erickson, MIT/Haystack Observatory. (2023) Data from the CEDAR Madrigal database. Available from <https://w3id.org/cedar?experiment=mlh230124g.001>

Data attribution

- PI's want to know when their data is used vs. free access to data
 - Madrigal has no passwords, but requires a name and email
 - Enforced by cookies in browser, arguments in API
- PI given access to Madrigal usage logs