How to use the Madrigal database for atmospheric science

Bill Rideout
MIT Haystack Observatory
brideout@haystack.mit.edu

ISR workshop
Kangerlussuaq, Greenland
July 18, 2011

Outline

- What is Madrigal?
- How is Madrigal different from an ftp site?
- How do I use Madrigal?
  - The website
  - Script data access
- What's coming soon with Madrigal?

What is Madrigal?

Open-source, standard-based local databases that share metadata

Madrigal is a Science database.

The Madrigal database stores data from a wide variety of upper atmosphere research instruments in the Cedar database format.

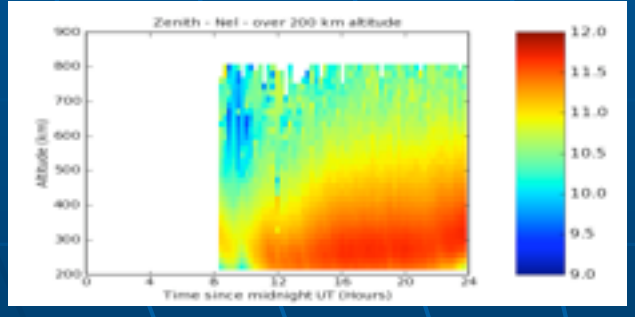
Incoherent Scatter Radar

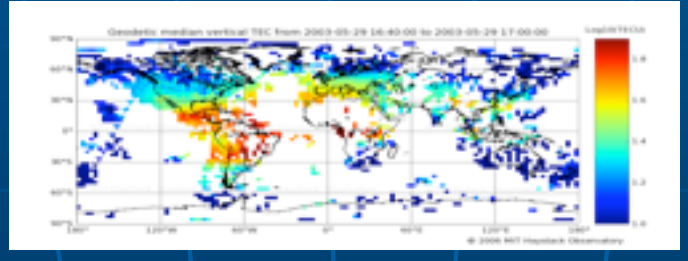
TEC via GPS

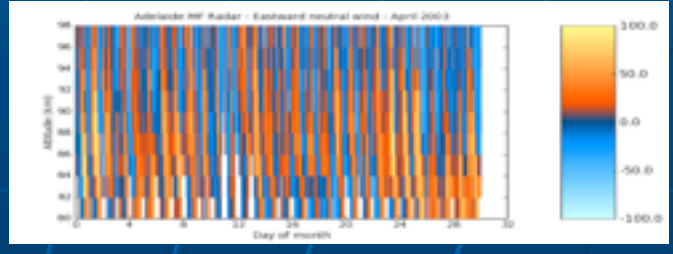
MF Radar

Cedar database format

Loading programs can be written in Python, C, or Tcl







Other instrument types in Madrigal: Meteor radar, Digisonde, Fabry-Perot, Geophysical indices

Madrigal is open-source

Madrigal is open-source

www.openmadrigal.org

Madrigal is open-source

www.openmadrigal.org

How is Madrigal different from an ftp site?

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data
- Madrigal has a derivation engine

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data
- Madrigal has a derivation engine
- How does this help a scientist?

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data
- Madrigal has a derivation engine
- How does this help a scientist?

Cedar/Madrigal Database

Cedar/Madrigal Database

- All parameters in file defined
  - http://cedarweb.hao.ucar.edu/documents/parameters_list.txt

Cedar/Madrigal Database

- All parameters in file defined
  - http://cedarweb.hao.ucar.edu/documents/parameters_list.txt
- Ranges of parameters for each instrument

Cedar/Madrigal Database

- All parameters in file defined
  - http://cedarweb.hao.ucar.edu/documents/parameters_list.txt
- Ranges of parameters for each instrument
- Data stored in one or two 16 bit ints
  - Additional increment parameters
  - Madrigal 3.0 will switch seamlessly to HDF5

Cedar/Madrigal Database

- All parameters in file defined
  - http://cedarweb.hao.ucar.edu/documents/parameters_list.txt
- Ranges of parameters for each instrument
- Data stored in one or two 16 bit ints
  - Additional increment parameters
  - Madrigal 3.0 will switch seamlessly to HDF5
- Error parameters always available

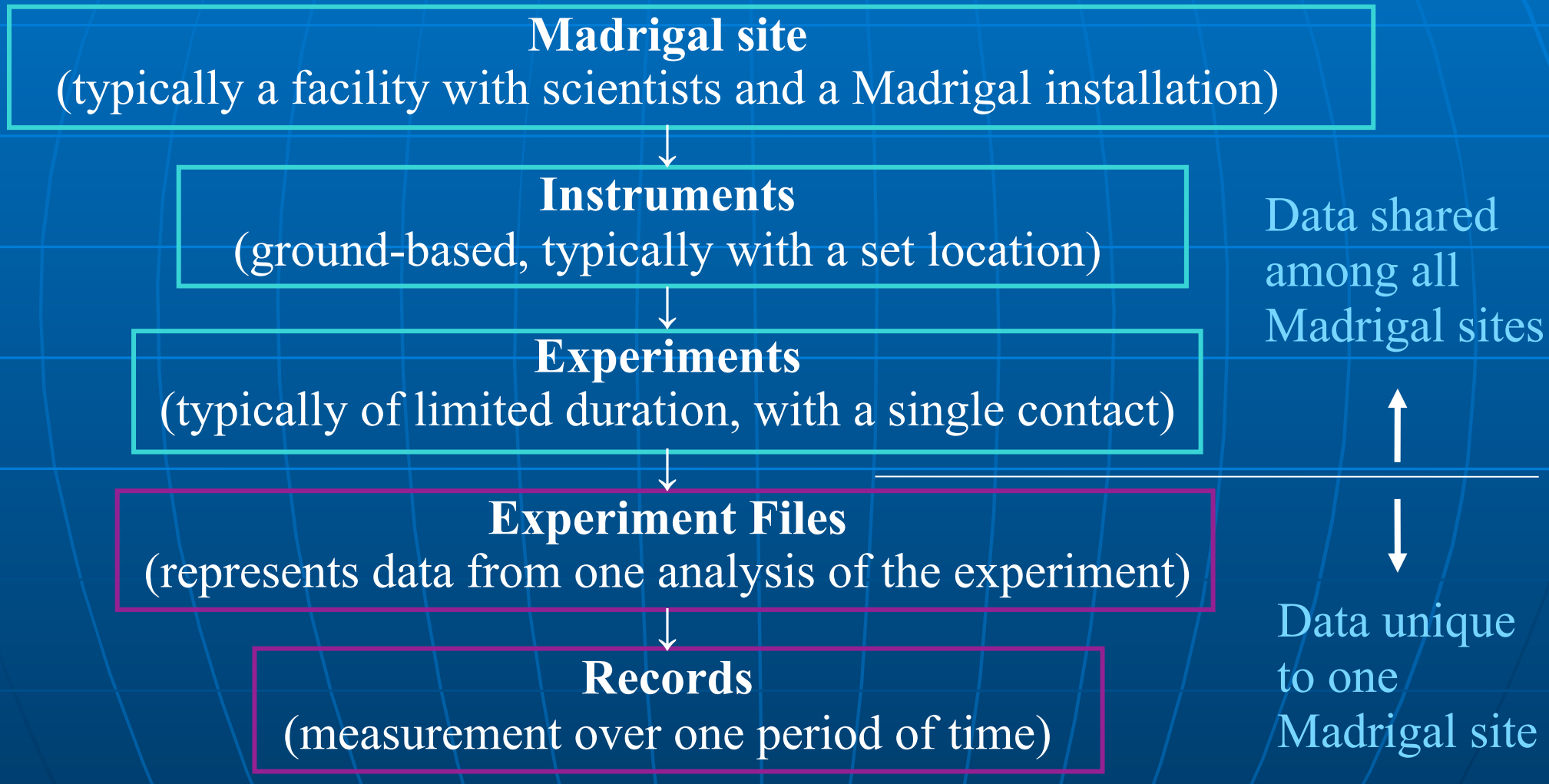
Cedar/Madrigal Database

- All parameters in file defined
  - http://cedarweb.hao.ucar.edu/documents/parameters_list.txt
- Ranges of parameters for each instrument
- Data stored in one or two 16 bit ints
  - Additional increment parameters
  - Madrigal 3.0 will switch seamlessly to HDF5
- Error parameters always available
- File format defined in
  - http://cedarweb.hao.ucar.edu/cgi-bin/cedar_file_access.pl?filename=documents/cedar_fmt.pdf

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data
- Madrigal has a derivation engine
- How does this help a scientist?

Madrigal Data Model



How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data
- Madrigal has a derivation engine
- How does this help a scientist?

Madrigal Derivation Engine

Madrigal Derivation Engine

- Derived parameters appear to be in file

Madrigal Derivation Engine

- Derived parameters appear to be in file
- Engine determines all parameters that can be derived

Madrigal Derivation Engine

- Derived parameters appear to be in file
- Engine determines all parameters that can be derived
- Easy to add new derived parameters using code written in C or Fortran

Classes of derived parameters

Classes of derived parameters

- Space, time
  - Examples: Local time, shadow height

Classes of derived parameters

- Space, time
  - Examples: Local time, shadow height
- Geophysical
  - Examples: Kp, Dst, Imf, F10.7

Classes of derived parameters

- Space, time
  - Examples: Local time, shadow height
- Geophysical
  - Examples: Kp, Dst, Imf, F10.7
- Magnetic
  - Examples: Bmag, Mag conjugate lat and long, Tsyganenko magnetic equatorial plane intercept

Classes of derived parameters

- Space, time
  - Examples: Local time, shadow height
- Geophysical
  - Examples: Kp, Dst, Imf, F10.7
- Magnetic
  - Examples: Bmag, Mag conjugate lat and long, Tsyganenko magnetic equatorial plane intercept
- Models
  - Examples: MSIS, IRI

How is Madrigal different from an ftp site?

- Madrigal files have well-defined format and parameters
- Madrigal has metadata associated with data files
  - Allows faster searches of data
- Madrigal has a derivation engine
- How does this help a scientist?

Example data search problem

- Find out how well the a model compares with measured data depending on geophysical conditions.

Ftp site approach

Ftp site approach

- Download all the ftp data to your local computer

Ftp site approach

- Download all the ftp data to your local computer
- Write software to convert the format

Ftp site approach

- Download all the ftp data to your local computer
- Write software to convert the format
- Get geophysical data from other sources

Ftp site approach

- Download all the ftp data to your local computer
- Write software to convert the format
- Get geophysical data from other sources
- If you want to compare to models
  - Compile model code (if you're lucky)
  - Write more format conversion code

Ftp site approach

- Download all the ftp data to your local computer
- Write software to convert the format
- Get geophysical data from other sources
- If you want to compare to models
  - Compile model code (if you're lucky)
  - Write more format conversion code
- Write analysis code

Madrigal approach

Madrigal approach

- Use script globalIsprint.py - done

Madrigal approach

- Use script globalIsprint.py - done
- Use global search web interface - done

How can the Madrigal database be accessed?

User

Web interface

Web services API

- From anywhere on internet
- Python API
- Matlab API
- IDL API
- Other could be written

Live demo of Madrigal web page

- Start at any Madrigal server (e.g., http://isr.sri.com/madrigal)

Remote Access to Madrigal Data

Remote Access to Madrigal Data

- Built on web services

Remote Access to Madrigal Data

- Built on web services
- Like the web, available from anywhere on any platform

Remote Access to Madrigal Data

- Built on web services
- Like the web, available from anywhere on any platform
- Read only API

Remote Access to Madrigal Data

- Built on web services
- Like the web, available from anywhere on any platform
- Read only API
- Complete Python, Matlab, and IDL APIs written

Remote Access to Madrigal Data

- Built on web services
- Like the web, available from anywhere on any platform
- Read only API
- Complete Python, Matlab, and IDL APIs written
- More APIs available on request or via contribution

Madrigal Web Services

Madrigal Web Services

- Simple delimited output via CGI scripts

Madrigal Web Services

- Simple delimited output via CGI scripts
- Not based on SOAP or XmlRpc since no support in languages such as Matlab

Madrigal Web Services

- Simple delimited output via CGI scripts
- Not based on SOAP or XmlRpc since no support in languages such as Matlab
- CGI arguments and output fully documented at http://www.haystack.edu/madrigal/remoteAPIs.html

Simple Python example

```
    # create the main object to get all needed info from
    Madrigal
madrigalUrl = "http://www.haystack.mit.edu/madrigal"
testData = madrigalWeb.madrigalWeb.MadrigalData(madrigalUrl)

# get all MLH experiments in 1998
expList = testData.getExperiments(30, 1998,1,1,0,0,0,1998,
                                  12,31,23,59,59)
for exp in expList:
      # print out all experiments
      print exp

# print list of all files in first experiment
fileList = testData.getExperimentFiles(expList[0].id)
      for thisfile in fileList:
          print thisfile
```

Python Remote API

- Can run on any platform with python (PC, Unix, Mac, etc)
- Fully documented with examples
- See http://madrigal.haystack.edu/madrigal/remotePythonAPI.html for documentation, more examples, and source

Live Python API demo

- See demoMadrigalWebServices.py at http://www.haystack.mit.edu/cgi-bin/madrigal_viewcvs.cgi/madroot/source/madpy/madrigalWeb/examples/

Matlab Remote API

- Methods
  - getInstrumentsWeb
  - getExperimentsWeb
  - getExperimentFilesWeb
  - getParametersWeb
  - isprintWeb
  - madDownloadFile
  - madCalculatorWeb
  - globalIsprint
- Methods match Madrigal model

Simple Matlab example

```
filename = '/usr/local/madroot/experiments
                /2003/tro/05jun03/NCAR_2003-06-05_tau2pl_60_uhf.bin';

eiscat_cgi_url = 'http://www.eiscat.se/madrigal/cgi-bin/';

% download the following parameters from the above file: ut, gdalt, ti

parms = 'ut,gdalt,ti';

filterStr = 'filter=gdalt,200,600 filter=ti,0,5000';

% returns a three dimensional array of double with the dimensions:
%
%   [Number of rows, number of parameters requested, number of records]
%
%  If error or no data returned, will return error explanation string instead.
data = isprintWeb(eiscat_cgi_url, filename, parms, filterStr);
```

Matlab Madrigal API call

Simple Matlab example, continued

- See http://madrigal.haystack.edu/madrigal/remoteMatlabAPI.html for complete documentation and more examples

Live Matlab API demo

- See demoMadrigalWebServices.m at http://www.haystack.mit.edu/cgi-bin/madrigal_viewcvs.cgi/madroot/source/madmatlab/

IDL Remote API

- Methods
  - madGetAllInstruments
  - madGetExperiments
  - madGetExperimentFiles
  - madGetExperimentFileParameters
  - madSimplePrint
  - madPrint
  - madDownloadFile
  - madCalculator
  - madGlobalPrint
- Methods again match Madrigal model
- Just added in July 2010

Madrigal application globallsprint.*

- Installed with all three remote API's.
- More robust that global search web UI.
  - Data stored locally
  - Error messages on local terminal
- Documented under Documentation-> Command line interface and in API doc.

globallsprint example

- Poker Flat
- March 10-20, 2007
- Alternating code (File kindat 5951)
- Kp above 4
- Alt between 240 and 260 and
- Ne > 2e11

Example command line (python version)

```
./globalIsprint.py \
--url=http://isr.sri.com/madrigal \
--parms=year,month,day,hour,min,sec,elm,azm,gdalt,gdlat,glon,kp,ne,te,ti \
--output=demo.txt \
--user_fullname="Bill Rideout" \
--user_email=brideout@haystack.mit.edu \
--user_affiliation=MIT \
--startDate=02/01/2007 --endDate=02/28/2007 \
--inst="Poker*" \
--kindat=5951 \
--filter=ap3,15, \
--filter=gdalt,240,260 \
--filter=ne,2e11, \
--filter=te,1000, \
--verbose
```

What's coming soon for Madrigal?

- NSF funded development
  - Release 2.6
  - Release 3.0
- Open source development with Jicamarca

NSF funded development for next release

- Transition of CEDAR database to Madrigal platform
  - Ability of Madrigal to automatically import data from other Madrigal sites
  - Importing of existing non-Madrigal data into Madrigal
- Ability to extend Madrigal with external hard drives

Next release development, continued

- Work with Jicamarca to integrate new UI into Madrigal
- Export HDF5 format
- FTP like interface added?
- Users can register interest in experiment
- Experiment PI, analyst now shared metadata
- IDL API added to Matlab, python

Following release (Madrigal 3.0)

- Conversion of CEDAR format to CEDAR HDF5/NetCDF4 format
  - Parameter definitions/data model unchanged
  - Each file will be self-describing
  - No more scale factors, integer storage
  - No more duplicate parameters with different ranges
  - Madrigal derivation engine will interface with it

Madrigal 3.0 continued

- Conversion of CEDAR format to CEDAR HDF5/NetCDF4 format
  - Conversion will be automatic using script
  - To create new CEDAR format
    - Create file old way, run convert script
    - Using python API, change one line of code
- Community standard
  - Interface to download latest standard, request new parameters

Open source development with Jicamarca

- Project is one of 3 based on Oct 2010 meeting
- Development of
  - New simple web UI
  - Export HDF5 format
- Modern open source project
  - Shared access to source control (Subversion)
- Millstone responsible for final testing, release, support

Extending/contributing to Madrigal

- Madrigal is completely open source
- See www.openmadrigal.org for CVS
- All new code is Python or C. Imported derivation methods sometimes in Fortran.
- I appreciate all contributions
  - Suggestions and ideas
  - Finding bugs
  - Code

Madrigal hands-on exercises

- Fully described on wiki at http://www.haystack.mit.edu/cgi-bin/asg_science/science.cgi/Using_Madrigal_practically_and_productively
- Web practice
- Script practice using python, Matlab, or IDL