Between a Map and a Data Rod

William Teng\textsuperscript{1,2}, Hualan Rui\textsuperscript{1,2}, Richard Strub\textsuperscript{1,2}, and Bruce Vollmer\textsuperscript{1}

\textsuperscript{1}NASA GSFC, GES DISC; \textsuperscript{2}ADNET Systems, Inc.

University of Texas at Austin: David Maidment, Tim Whiteaker, and David Arctur

GSFC/Hydrological Sciences Lab: Christa Peters-Lidard, David Mocko, Dalia Kirschbaum, Matthew Rodell

Brigham Young University: Daniel Ames

NASA ACCESS Program
NNH11ZDA001N-ACCESS
NNH13ZDA001N-ACCESS
Outline

• Motivation and background
• “Digital Divide” problem
• Solution: Pre-generated vs. on-the-fly
• Tiling, between a map and a data rod
• Summary and ongoing work
Time Series of top 1 meter soil moisture from NLDAS-2 Noah model, near the center of Texas (100W, 31N)

Courtesy of David R. Maidment
Center for Research in Water Resources
University of Texas at Austin
Select “NLDAS Grid”

List data in ASCII

Select “NLDAS Precipitation”

Plot Time Series

Latitude x

Data Curtain

Time

1Better Assessment Science Integrating Point and Nonpoint Sources
Digital Divide Problem ... Orthogonal

<table>
<thead>
<tr>
<th></th>
<th>Original GRIB Files</th>
<th>Data Rads Binary Files</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dimension lat x lon</td>
<td>Total # of Grids</td>
</tr>
<tr>
<td><strong>Noah LSM</strong></td>
<td></td>
<td>224 x 464</td>
</tr>
<tr>
<td><strong>NLDAS</strong></td>
<td></td>
<td>600 x 1440</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table shows the comparison between Original GRIB Files and Data Rads Binary Files for Noah LSM and NLDAS, detailing the dimensions, total files, file sizes, and volumes.
Data Rods: A Simple Solution for Bridging the Digital Divide

Original Data Archive

- Time
- Latitude (Y)
- Longitude (X)

Reorganized Data Archive

- One variable, one grid point, all time steps per file

Original data reorganized as ...

- All variables, all grid points, one time step per file

More than 324,000 time steps (37 years) plotted in ~ 1 second
**Generating Data Rods**

- **GRIB File**
  - Input file: Multi-variable, one time step per file

- **Parameter Subsetting WGRIB**
  - Intermediate file: One variable, one time step per file

- **Bin File**

- **Grid Subsetting**

- **Data Rods**
  - Reorganized Data Rod files: One variable, one grid point, all time steps per file

**Delivering Data Rods to Users**

- **User**
  - Specify variable, startDate, endDate, and location

- **RESTful Web Service**
  - Display data in ASCII, WaterML, or Time Series plot
  - Read user-specified file

- **Input file:** Multi-variable, one time step per file
- **Intermediate file:** One variable, one time step per file
- **Reorganized Data Rod files:** One variable, one grid point, all time steps per file

**Display data in ASCII, WaterML, or Time Series plot**
Global Level 3 (Gridded) Single Variable NASA Earth Science Data To Time Series (“Data Rods”) Using NCO, NetCDF, and Giovanni

A. Client (e.g., HydroDesktop) sends data request with spatial-temporal constraints

B. Requests time series for current year x

C. Files for current year: Stored in Giovanni Cache as single variable, single time-step, global lat-lon layers, updated as new data become available

D. NCO subsets time series of single grid point along time dimension in current year

E. Requests time series for years x-1, x-2, x-3, ...

F. Files for previous years: Stored across several file systems (for parallel I/O) as separate, month-long, global lat-lon data cubes, with 3rd dimension along time

G. NCO subsets time series of single grid point along time dimension from each cube

H. NCO concatenates all the responses. Custom software delivers resulting single time series as ASCII, WaterML1, WaterML2, or plot

I. Archived data, subset by variable, converted to “fast file” format, and stored in Giovanni Cache

J. Files for previous years in Giovanni Cache, concatenated into data cubes, and stored in other file systems

e.q., last 4 years of data at lat=42.5, lon=-124.9
Generation Time for TRMM\(^1\) On-the-Fly Data Rods vs. Tiling Levels\(^2\)

![Graph showing the generation time for TRMM on-the-fly data rods versus tiling levels.](Image)

\(^1\)Tropical Rainfall Measuring Mission

\(^2\)Tiling: Dividing the data set grid into subgrids (e.g., ¼ tiling for TRMM divides its global grid into 4 equal subgrids).
# Data Rods Metrics

## 2013-01-01 to 2015-11-30

<table>
<thead>
<tr>
<th>Product</th>
<th>Protocol</th>
<th># Users</th>
<th># Files</th>
<th>Volume (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLDAS_FORA0125_RODS</td>
<td>FTP</td>
<td>8</td>
<td>17,733,371</td>
<td>20,585</td>
</tr>
<tr>
<td>NLDAS_NOAH0125_RODS</td>
<td>FTP</td>
<td>5</td>
<td>16,741,164</td>
<td>19,580</td>
</tr>
<tr>
<td>GLDAS_NOAH025_RODS</td>
<td>FTP</td>
<td>13</td>
<td>39,654,230</td>
<td>5,946</td>
</tr>
<tr>
<td>NLDAS_FORA0125_RODS</td>
<td>NLDAS_FORA</td>
<td>445</td>
<td>241,470</td>
<td>92</td>
</tr>
<tr>
<td>NLDAS_NOAH0125_RODS</td>
<td>NLDAS_NOAH</td>
<td>286</td>
<td>187,923</td>
<td>95</td>
</tr>
<tr>
<td>GLDAS_NOAH025_RODS</td>
<td>GLDAS_NOAH</td>
<td>487</td>
<td>62,680</td>
<td>36</td>
</tr>
<tr>
<td>NLDAS_FORA0125_RODS</td>
<td>WEB_LDAS</td>
<td>300</td>
<td>118,561</td>
<td>1,784</td>
</tr>
<tr>
<td>NLDAS_NOAH0125_RODS</td>
<td>WEB_LDAS</td>
<td>336</td>
<td>452,667</td>
<td>6,401</td>
</tr>
<tr>
<td>GLDAS_NOAH025_RODS</td>
<td>WEB_LDAS</td>
<td>392</td>
<td>79,844</td>
<td>94</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>75,271,910</td>
<td>54,613</td>
</tr>
</tbody>
</table>

**Users:** Number of distinct users

**FTP:** Get data rods via FTP

**WEB_LDAS:** Access data rods in ASCII or as Time Series plot via GES DISC Web services

**Other protocols:** Access data rods via CUAHSI HIS (HydroDesktop)
Summary and Ongoing Work

• Developed operational way to reorganize data that is optimal for user communities that are point-time series oriented.
• Solved the motivating problem presented by CUAHSI HIS: create time series of hourly data, for single grid cells for entire period of coverage.
• Key to all solutions is to reorganize data that is optimal for desired method of data access.
• Ongoing investigation into tiling of data set grids has yielded results that are very encouraging for significantly reducing the generation time for data rods.
Extras
Data Rods
Data Rods

https://www.arcgis.com/home/webmapviewer.html?webmap=93b7c28dca3b4c86863408a4a90f729f