

# ROSETTA - UNIDATA'S WEB-BASED DATA TRANSLATION TOOL PROGRESS AND FUTURE PLANS

Sean C. Arms <sup>1</sup>

Jennifer Oxelson Ganter <sup>1</sup>

Jeff Weber <sup>1</sup>

Mohan K. Ramamurthy <sup>1</sup>



<sup>1</sup>UCAR/Unidata



# Rosetta

- What is Rosetta?
  - Vision: General Purpose Data Format Translator
  - Goal: Get data into standard format

# What Many Users Like to Use

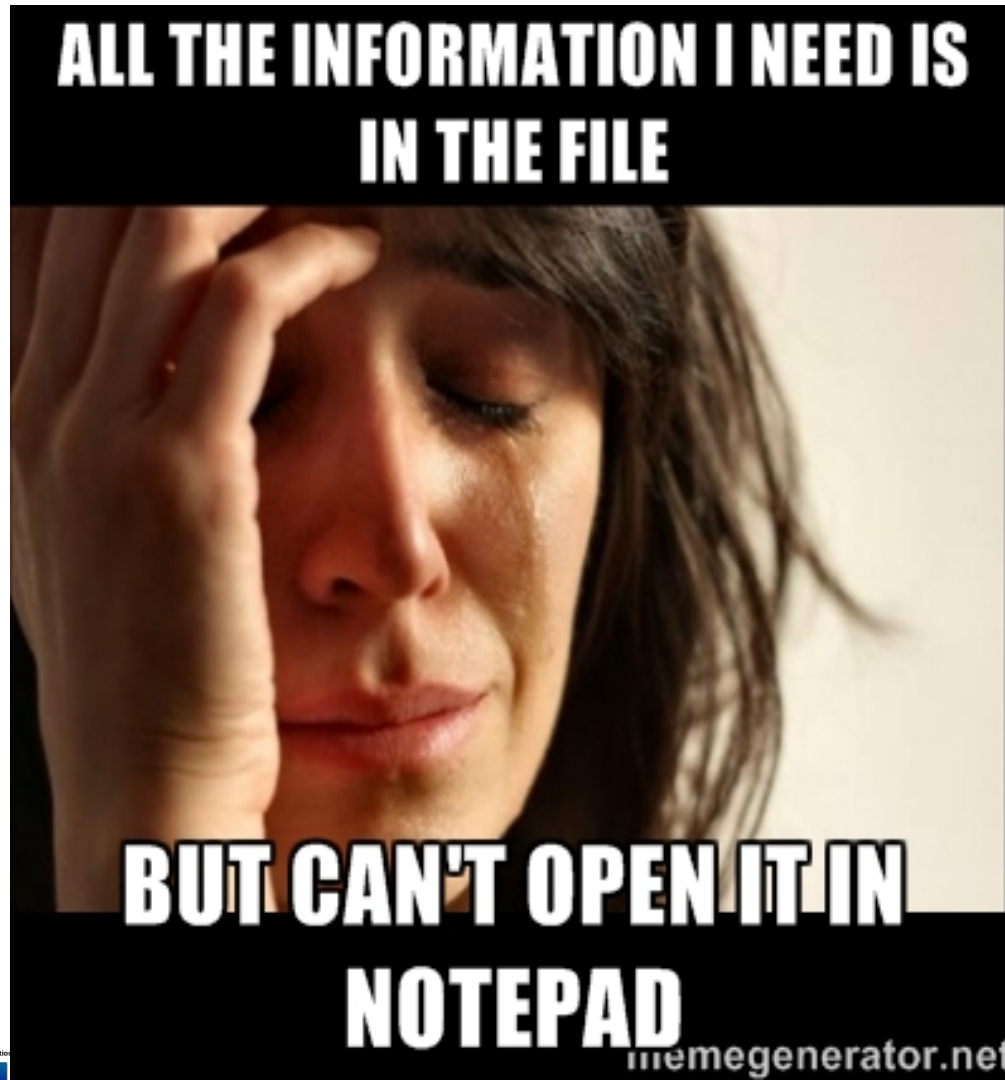
YEAR	DATE	0	0.25	0.5	0.75	1	2	3	4
2007	9/3/07	2.10	2.32	1.64	0.50	-0.39	-2.20	-3.04	-3.27
2007	9/4/07	3.63	2.42	1.44	0.44	-0.38	-2.22	-3.06	-3.30
2007	9/5/07	4.21	2.66	1.44	0.40	-0.39	-2.24	-3.08	-3.31
2007	9/6/07	2.00	2.28	1.49	0.42	-0.39	-2.24	-3.08	-3.31
2007	9/7/07	4.57	2.40	1.34	0.39	-0.39	-2.24	-3.08	-3.31
2007	9/8/07	5.32	3.11	1.53	0.40	-0.39	-2.22	-3.08	-3.31
2007	9/9/07	4.27	3.13	1.73	0.46	-0.37	-2.22	-3.08	-3.31
2007	9/10/07	3.38	2.68	1.70	0.50	-0.34	-2.21	-3.08	-3.31
2007	9/11/07	2.29	2.39	1.57	0.47	-0.27	-2.21	-3.08	-3.31
2007	9/12/07	0.26	1.48	1.31	0.42	-0.37	-2.19	-3.08	-3.31
2007	9/13/07	1.21	1.09	0.92	0.31	-0.38	-2.19	-3.08	-3.31
2007	9/14/07	0.03	0.91	0.75	0.23	-0.39	-2.19	-3.08	-3.31
2007	9/15/07	0.15	0.42	0.51	0.16	-0.39	-2.19	-3.07	-3.31
2007	9/16/07	2.47	0.48	0.30	0.09	-0.39	-2.19	-3.06	-3.31
2007	9/17/07	3.16	1.34	0.43	0.06	-0.39	-2.17	-3.05	-3.31
2007	9/18/07	2.53	1.39	0.66	0.13	-0.39	-2.16	-3.05	-3.31
2007	9/19/07	1.45	1.33	0.77	0.16	-0.39	-2.16	-3.05	-3.31
2007	9/20/07	-0.38	0.82	0.70	0.17	-0.39	-2.16	-3.05	-3.31
2007	9/21/07	-1.41	0.26	0.44	0.13	-0.39	-2.16	-3.05	-3.31

[ 2 ]

The image shows a vertical blue bar on the left side of the slide. At the bottom of this bar is the Rosetta logo, which consists of a 4x4 grid of colored squares (brown, grey, teal, and white) and the word "Rosetta" in a bold, sans-serif font.

$$\begin{bmatrix} 3 \end{bmatrix}$$

# What a Standard Format “Looks” Like



# Rosetta

- What is Rosetta?
  - Vision: General Purpose Data Format converter
  - Goal: Get data into standard format ***while providing data users with the format they want***

( 5 )



# Rosetta

- Why bother translating data into a standard format?
  - What works for the data provider may not work well for the data users scientific workflow.

[ 6 ]

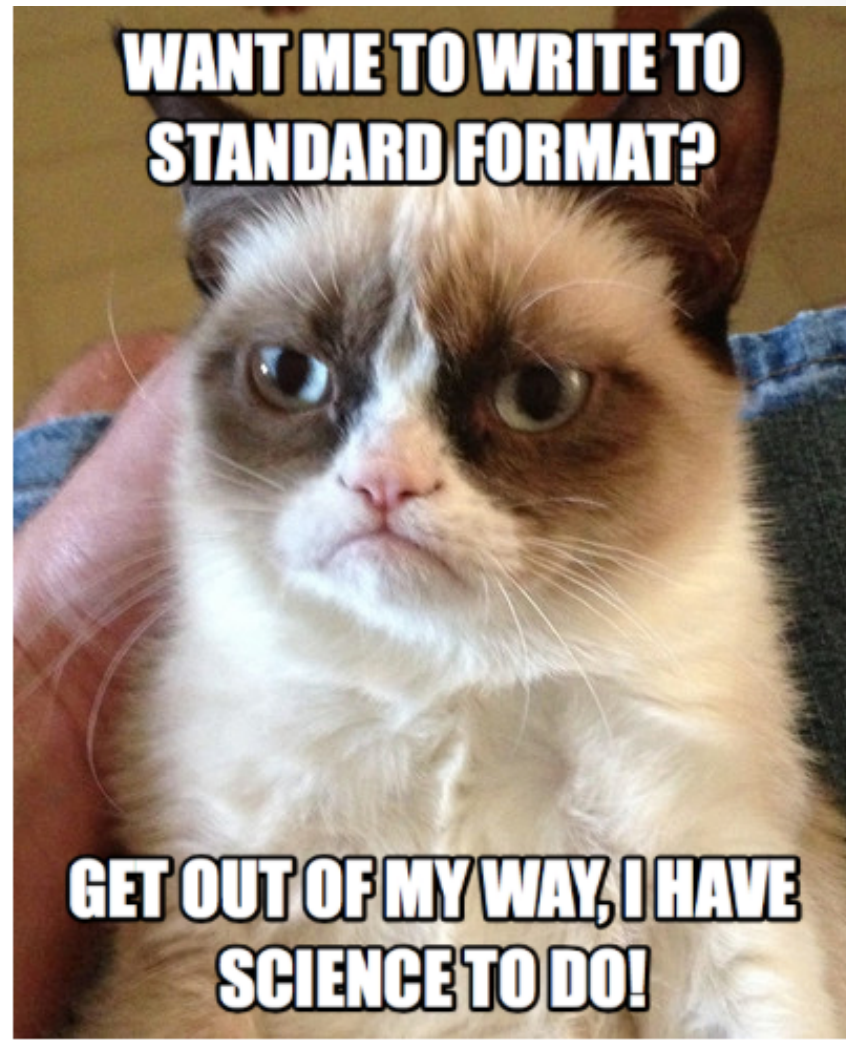
# Rosetta

- What does translating data into a standard format buy us?
  - Standard format -> a variety of output formats (programmatically!)
  - Enable subsetting, aggregation, visual data “previews”, enhancing search through automated rich metadata extraction, etc.



# Rosetta

It's really all about the end users of the data...but we need to get out of the data providers way as quickly as possible.



# Wizard Based Format Translation

Rosetta

rosetta.unidata.ucar.edu/create

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#) Other Bookmarks

## Rosetta

- Select Observation Platform
- Upload File
- Specify Header Lines
- Specify Delimiters
- Specify Variable Attributes**
- Specify Site Specific Information
- Specify General Information
- Download Converted File
- Publish

Previous Next

### Specify Variable Attributes

	column 0	column 1
0	2002/10/01,00:00,10/1/02	
1	2002/10/01	00:30
2	2002/10/01	01:00
3	2002/10/01	01:30
4	2002/10/01	02:00
5	2002/10/01	02:30
6	2002/10/01	03:00
7	2002/10/01	03:30
8	2002/10/01	04:00
9	2002/10/01	04:30
10	2002/10/01	05:00
11	2002/10/01	05:30

Quick Save

Questions or comments about Rosetta can be sent to: [support-ros](#)  
Version : 0.2-SNAPSHOT  
Build Date: 20140101.1019

**unidata**  
providing innovative data services and tools to transform the conduct

### Enter Variable Attributes

**What would you like to do with this column of data?**

☒ Assign a variable name ☐ Do not use this column of data

time

**Is this variable a coordinate variable? (examples: latitude, longitude, time)**

☒ Yes ☐ No

**What type of coordinate variable?**

Time only (hour, minute, second, and/or millisecond)

**Specify variable data type:**

☐ Integer ☐ Float (decimal) ☒ Text

**Required Metadata:**

Variable Description Time from Datalogg

Units YYYY/MM/DD

☒ show unit builder

What type of data are we building units for? date only

unit prefix: unit: YYYY/MM/DD

**Recommended Metadata:**

CF Name time

**Additional Metadata:**

Calendar Type

done cancel

# Technology Stack

- Java WebApp

- Java
- Apache Tomcat
- Spring MVC
- netCDF-  
Java(CDM)
- JavaScript
  - jQuery, SlickGrid, jWizard



{js}



github

Explore GitHub



mleibman / SlickGrid

github

Explore GitHub

Search



dominicbarnes / jWizard



EARTH CUBE

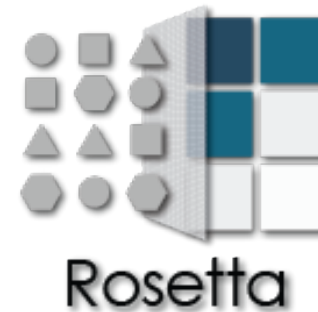
[ 10 ]



Rosetta

# Rosetta – Progress Over Previous Year

- Name change:



- NcML transaction receipt moved to json
  - Avoid file rewrites

# Rosetta – Progress Over Previous Year

- Refactor javascript wizard to ease creation of new wizards flows
  - Each wizard step is a modular javascript function call (includes validation)
  - Front end view pieced together by including jspf (jsp fragments)

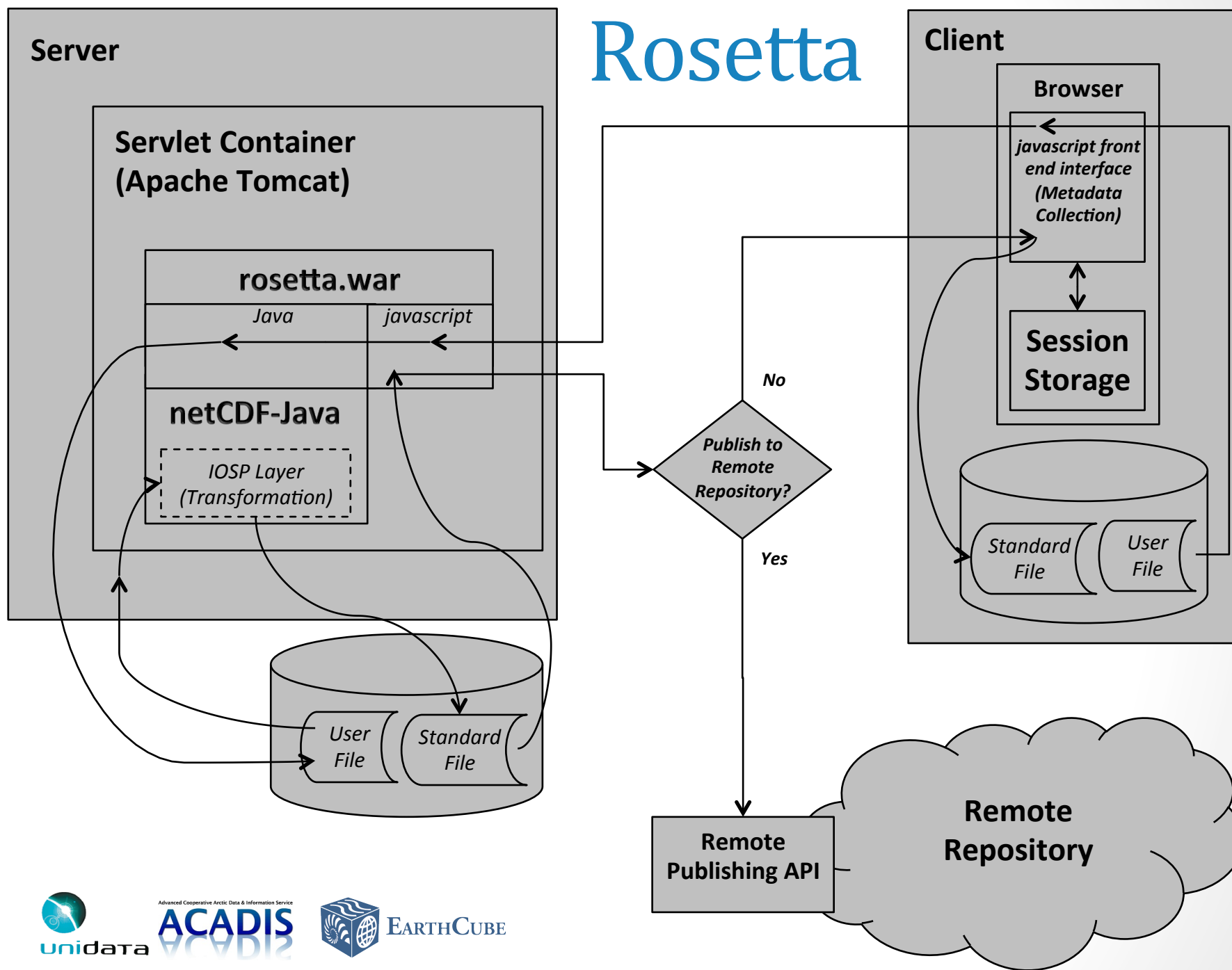
[ 12 ]

# Rosetta – Progress Over Previous Year

- Documentation infrastructure in place
  - Sphinx, jsDoc3, javadoc
  - Open to suggestions
- Publishing capabilities to remote data portals / archives RAMADDA, ACADIS Gateway

[ 13 ]

# Rosetta





# Rosetta – Current Status

- Live beta site at Unidata:  
<http://rosetta.unidata.ucar.edu>
- Source code on GitHub under Unidata account  
<http://www.github.com/Unidata/>
- Time series transforms; trajectories in progress

[ 15 ]

# Rosetta - Future

- Add WaterML to translation capabilities
- Finalize trajectory interface
- Develop profile interface
- Preview visualization using d3
  - Verifying lat/lon entry
  - Post translation sanity check

# Rosetta - Future

- Expose TDS-like services on remote data sets
  - Example: Get point time series data by specifying lat/lon location...
    - ...on a gridded data set hosted on external server exposed solely by OPeNDAP

[ 17 ]

# Rosetta

- Open to collaboration!
- Always looking for 'badASCII' example files
- Email:  
`support-rosetta@unidata.ucar.edu`
- Website:  
`www.unidata.ucar.edu/software/rosetta`

# Rosetta

- Development supported by:

NSF Award 0833450 (**AGS**): *Unidata 2013: A Transformative Community Facility for the Atmospheric and Related Sciences*

NSF Award 1016034 (**ARC**): *The Advanced Cooperative Arctic Data and Information Service (ACADIS)*

NSF Award 1343785 (**ICER**): *EarthCube Building Blocks: Integrating Discrete and Continuous Data*

