

LBA-ECO Data Breakout Group

Current Archive Scenario

Merilyn Gentry

LBA-ECO Data Coordinator

November 5, 2003

What is an archive?

A data center dedicated to:

- Preserving and protecting data and information from damage, loss, or extinction due to outdated technology
- Ensuring that the data are retrievable and available for long-term use by researchers, educators, and the public (the “20-year rule”)

Where will data be archived?

- All LBA data will be archived at LBA-DIS / CPTec, per LBA data policy
- *Selected* LBA-ECO data sets will be archived as well at the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), per NASA data policy

What is the ORNL DAAC?

- Part of a network of NASA data archive and distribution centers forming the Earth Observing System (EOS) DIS
 - Known for support of NASA field projects, ORNL has archived data from BOREAS, FIFE, SAFARI 2000, EOS Land Validation, LBA-ECO
- As part of this network, ORNL DAAC holdings must conform to certain EOSDIS standards
- ORNL DAAC has data archive standards and recommendations as well

What is Needed for Archiving LBA-ECO Data Sets at the ORNL DAAC?

Reference: Best Practices for Archiving Data

www.daac.ornl.gov/DAAC/PI/info.html

1. Metadata

- LBA project parameters (i.e., Beija-flor metadata) must comply with latest GCMD / EOSDIS standards

2. Data files

- Suggested format: tabular data in ASCII, Gridded data in ASCII Grid, Image data in binary or non-proprietary format
- Self-describing to identify key entries such as parameter names and units of measure

What is Needed for Archiving LBA-ECO Data Sets at the ORNL DAAC (continued)

3. Data Set Documentation

- Key information from B-f (e.g., investigator(s), abstract, spatial and temporal attributes, data set citation, references, etc.)
- Data collection goals and description
- Description of sample collection sites
- Description of measurement methods (e.g., calibration, calculations, software)
- Known errors and problems
- Description of data file organization
- Description of data reporting conventions (e.g., parameter names, units, codes, flags, example data records)

Documentation Tool

- Documentation should include what a user would need to know about the data 20+ years from now; i.e., the 20-year rule
- On-line tool available from the DAAC to assist PIs in providing data set documentation
- Uses existing metadata from B-f, as appropriate
- Option to produce a summary guide document (output from B-f metadata by project staff) accompanied by a more informative readme documentation file in .txt or .pdf format that could include .jpg images, formulas, figures, and tables (compiled by PI)
- Documentation tool is being simplified, per feedback from LBA-ECO

Additional Actions Needed by the Project for Archiving LBA-ECO Data

- Perform quality review of B-f metadata (e.g., data set title, abstract, investigator info, etc.)
- Perform quality reviews across data sets to ensure completeness and consistency of data and documentation; i.e., a 'Project view' as opposed to an individual PI's perspective
- Compile citations of LBA-ECO publications as a resource for the archive (Linking to pubs to data sets has not been done)
- Some LBA-ECO data sets registered in B-f should be combined as a single data set for the archive; this should be done before creating documentation for these data sets (reduces documentation effort and assists future users)

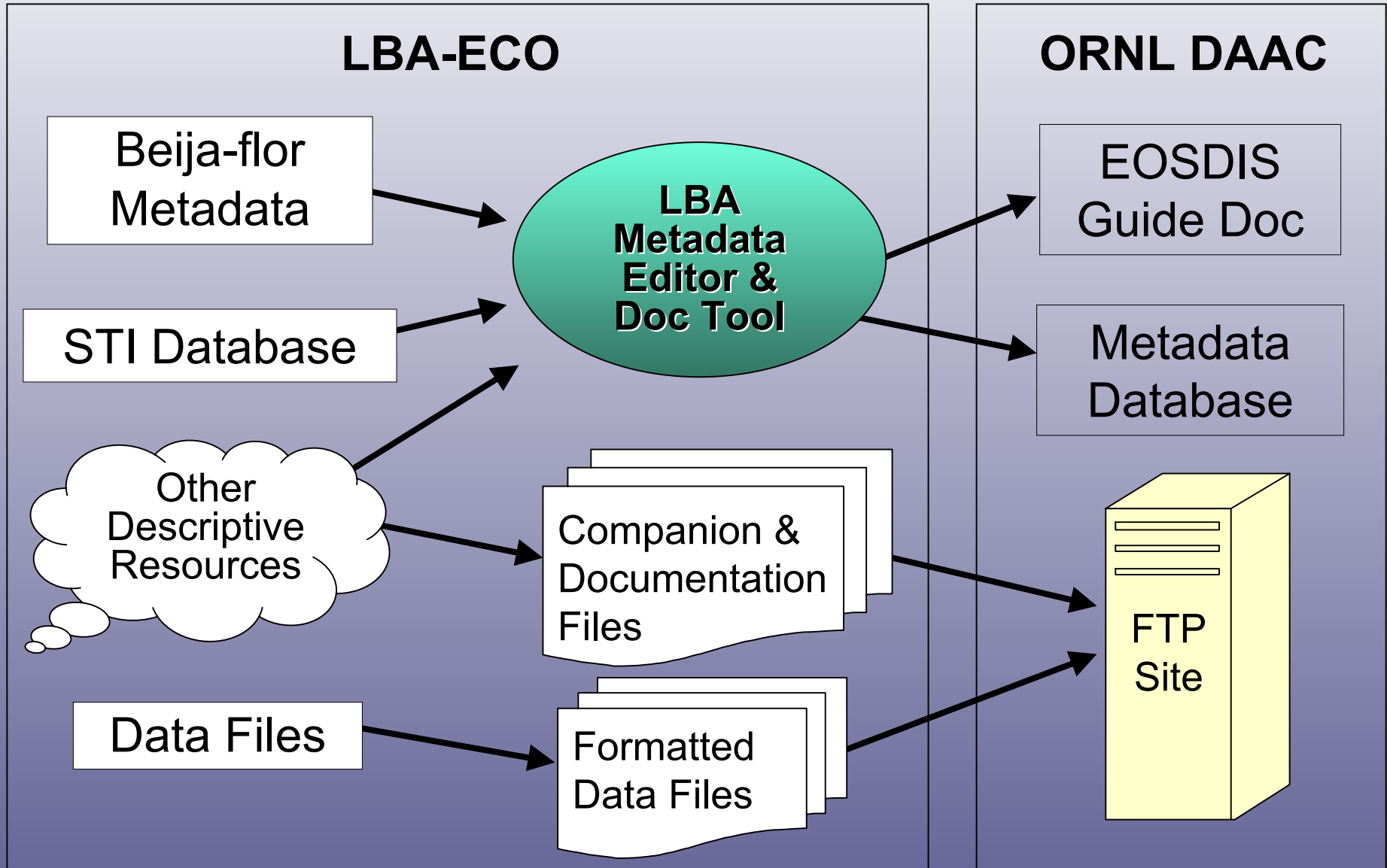
Definition of a “Data Set”?

Data Type + Study Site/Area + Time Coverage

- **AERONET (70 entries in Beija-flor = 70 data sets?
or 1? Or 3? Or 18?...)**
 - **Aerosol Optical Thickness Data (48)**
 - 18 Sites
 - Temporal coverage varies by site, range = 1993-2002
 - **Broadband Surface Irradiance Data (20)**
 - 6 Sites
 - Temporal coverage varies by site, range = 1999-2002
 - **Photosynthetically Active Radiation Data (2)**
 - One year of PAR data for each of 2 sites
- **Documentation: one document for all AERONET?
Or one for each of the 3 data types?**

Archiving LBA-ECO Data at the ORNL DAAC

Data and Information Flow



Which data sets should be archived?

Criteria to consider

- All data related to pubs in open literature should be archived
- Consider the long-term scientific merit of the data
- Seek input from science team
- Focus on Phase 1 data and prioritize teams not funded for Phase 2
- Will data be lost if not archived by LBA-ECO

Archive Evaluation Checklist

Criteria / component	GOES Smith	CASA Potter	Agri. Census Cardille
Registered in B-f	Yes	Yes	Yes
Metadata complete?	Yes	Yes	Yes
Data final?	Yes	Yes	?
Data online?	Yes	Yes	Web site
Documentation available/adequate?	Yes/Yes	Yes/?	?
Publication?	?	?	Yes
Citation in B-f?	No	No	Yes

Outstanding issues for group to consider

- Documenting/archiving models (not just model output)
- Decide on criteria for “long term archive”, then consider
 - Data in B-f that should go to long-term archive
 - Data in B-f that need not go to long-term archive
 - **Data not in B-f that should go to long-term archive**
- ...other issues from group

Some Action items...

- Reduce the long-term “documentation tool” burden so that users can complete in several 2- to 4-hour sessions (ORNL User Working Group)
- Require relevant Beija-flor entry/entries for each abstract submitted to July 2004 all-LBA science meeting
- Request any LBA-related “special issue” guest editors require relevant metadata (including long-term archive requirements) - subject to metadata peer review