

# From Storage to Archive

## Developing a Digital Preservation Infrastructure

SAA Research Forum, August 2009

Chris Jordan

# TACC and Data

- NSF and UT funded “supercomputer center”
- World-class resources and expertise in supercomputing, visualization, and data
- Since early 2008, developing services for preservation and access of research data
- Serve all disciplines at all stages of the research/data lifecycle
- Hardware/Software are the “easy” parts

# Challenges

- “Working Collections” – data is an active component of ongoing research
- Collections with widely differing characteristics (size, number, structure)
- Unique data formats from unique research processes (Provenance)
- Dynamic and uncertain funding & responsibility landscape

# Software and Practices

- Diversity of research practices and materials demands diversity in every layer of the stack:
  - Disk and Tape systems with unique configurations
  - Network access and data control mechanisms
  - Indexing, searching and presentation tools
  - Discipline- or even project-specific standards
  - Integration into distributed computational workflows
- Automation is critical wherever possible

# Policies, Standards, Trust

- OAIS and TRAC as substitutes for depth of community experience
- Useful as guide to necessary policies
- Service-Levels and SLAs are critical
- Experience and Collaboration are critical to building trust
- Developing a reputation in the communities

# Multi-Level Collaboration

- Geographical Replication – arrangements with similar institutions across the country
- Data Technologies – many R&D projects contributing pieces of the software stack
- Compute/Data Grids – TeraGrid, others provide services and need services
- Preservation Efforts – DataNet, others working on broad challenges

# Example of a Working Collection

- Institute for Classical Archeology – ongoing digs of multiple Greek colonial sites
- Heterogeneous data, dynamic relationships
- Collaboration with researchers, multiple services in front of diverse storage, periodic snapshots of collection
- Issues – Describing Provenance – Collection Structure – Dynamic Objects - Standards