

# iRODS Bioinformatics Use Cases

SC13 iRODS Reception

Chris Smith

[chris@distributedbio.com](mailto:chris@distributedbio.com)

# Bioinformatics Data Management Issues

- Data generation (from instruments, etc) is constant
- Experiment lifecycle management doesn't reach into the computing environment, even though more experimentation is now being done in silico
- POSIX filesystems are one dimensional with respect to meta-data (you get a file name)
- Scientists can't find datasets based on data properties (e.g. organism, tissue type, analysis type)
- Sharing data with collaborators is very difficult
- Scientists "need" to keep the data forever...  
... or so they say

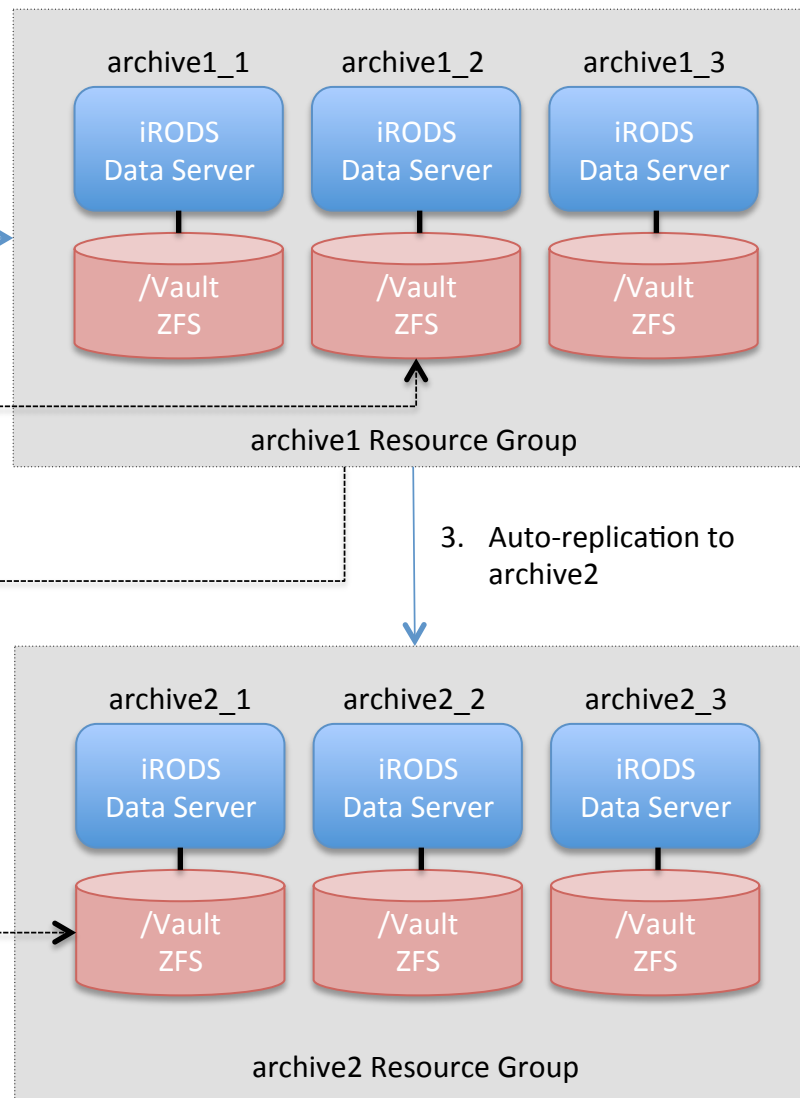
# Archival Storage



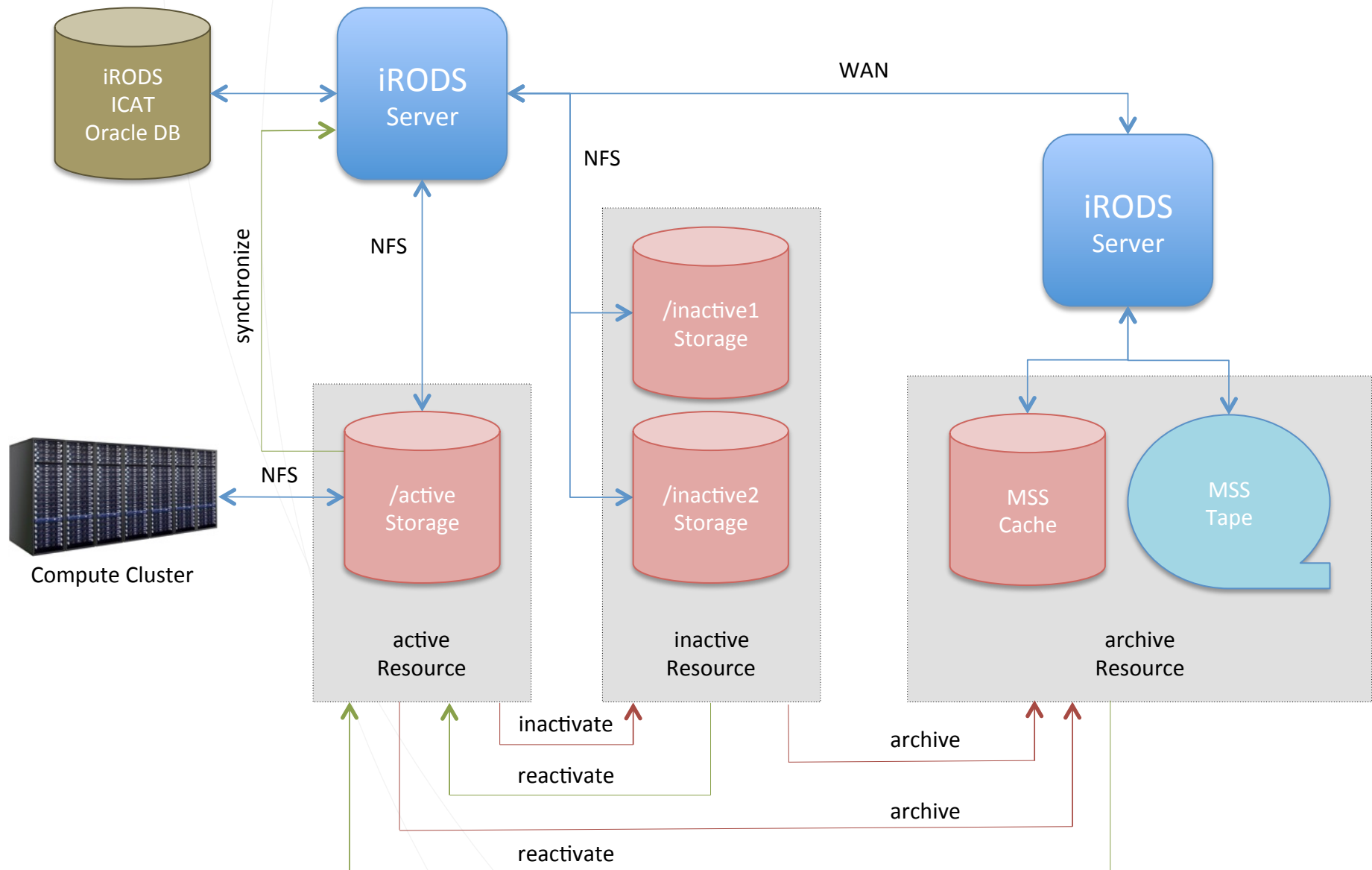
1. Put file or collection into archive and annotate with metadata

2. Results in random placement onto archive1\_2

4. Results in random placement onto archive2\_1



# Metadata Catalog and Lifecycle Management



# INCF Data Collaboration Using AWS

