Why this program?

- A large percentage of UCLA’s precious research data is currently backed up in “consumer-level” systems
- Increasing demand for multiple terabytes of “inexpensive,” “good enough” backup/archival research data storage.
- HPC storage is overkill, too expensive, and difficult to make available outside of our cluster environment.
- A need to satisfy various granting agency data plan requirements for retention and sharing
- Archival storage as a project had a 100% endorsement and the most funding from the recent IDRE Informatics grant process.
Cloud Archival Storage Program

- **Program Goal** - Provide archival and backup of research data with the following characteristics:
  - Professionally run, flexible, full-service offering
  - A highly competitive price which is
    - Lower than external Cloud Storage providers
    - Data is kept on campus; No extra data movement charges
  - “Good enough,” no extras that drive the cost up
  - A balance of ease of use, performance, reliability, maintainability and cost
  - Better than the usual DIY solutions which generally:
    - Have lower reliability, scalability, performance
    - Never account for labor costs – 2-4hrs/mo. ~$350-$700/yr. which is typically paid from research funds
  - Full integration with the campus rEcosystem
Cloud Archival Storage Program

- **Financial terms are key to the success of this program**
  - Requires buying a large amount of storage upfront
    - Competitive volume pricing from vendors
  - The system architecture is crucial
    - Low cost but with maximum flexibility
  - We picked SDSC as our representative market price top — can’t charge more — others are even higher
  - Our target price is close to what people are paying now
    - Competitive to the full cost to run a DIY system
  - Variable pricing model
    - Incentivize larger volume and longer time period purchases
Cloud Archival Storage Program

Cloud Storage Service Price Comparison Proposed Rates

1TB of Usable Storage per Year
Single Copy

<table>
<thead>
<tr>
<th>Service</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rackspace</td>
<td>$922</td>
</tr>
<tr>
<td>Amazon S3**</td>
<td>$776</td>
</tr>
<tr>
<td>Google**</td>
<td>$738</td>
</tr>
<tr>
<td>Dropbox***</td>
<td>$795</td>
</tr>
<tr>
<td>UCB*</td>
<td>$720</td>
</tr>
<tr>
<td>I2 Box.net***</td>
<td>$540</td>
</tr>
<tr>
<td>SDSC*</td>
<td>$390</td>
</tr>
<tr>
<td>IDRE Cloud</td>
<td>$276</td>
</tr>
<tr>
<td>DIY RAID6</td>
<td>$124</td>
</tr>
<tr>
<td>DIY RAID5</td>
<td>$93</td>
</tr>
</tbody>
</table>

* No single replica option. ** Without read traffic charges. ***Replica on user’s desktop.

Includes no labor costs.
Proposed IDRE Cloud Program versus IT Services Storage Program

The two programs differ in significant ways

- **Hardware** – We are optimizing on commodity, high-density, mass storage. Support hardware based on HPC systems.
- **Middleware** – We are using a combination of advanced VMware with our own custom-built automation overlay.
- **Expertise** – We are leveraging our HPC knowledge of high performance hardware, networks, interconnects and software development to engineer our storage system.
- **Network** – We are using a combination of 10Gb Ethernet and 40Gb Infiniband which we currently use for HPC.
- **Tight integration with our HPC resources, rEcosystem and external leadership class facilities.**
rEcosystem Conceptual Diagram

Medical Enterprises MEDNET

Gene sequencing data (Gonda)

Hoffman2 Shared Cluster (MSA, POD, CNSI)

SEM data (CNSI)

High Performance Storage

Cloud Storage

CTSI & Pathology Data

Dawson2 GPU Cluster (MSA)

CENIC, Esnet, StarLight, I2, OSG, XSEDE, National Supercomputing Centers

Data movement via
GlobusOnline using OpenFlow

10Gb Campus Backbone

Collaboration Tools

Workflow
Cloud Archival Storage Program

Next Steps

- **Establish a pilot** to substantiate demand, services, technologies and costs. *In process, estimate November 2012 for pilot start.*

- **In parallel, proceed with seeking financial model and pricing approvals from POSSSE. In process.**
  - Sales and service agreement is being submitted.

- **Secure a loan** based on a five-year payback period. *Need POSSSE approval first.*

- Assuming a technically and financially successful pilot, **officially roll the service out to the campus.**
  - Also assumes POSSSE approval and loan

- **Plan to be up and running by the end of Q1 2013**
Questions?