



## Data in the Research world



How to secure your scientific data and  
preserve your research findings with Atempo

*S3 "Managing Data Growth in Life Science" , January 26<sup>th</sup> 2011, Cambridge  
Speaker: Robert-Jan Overzee, Product Marketing Manager, Atempo*

# Today's Agenda



- Introduction
- Common issues and real world workflows to solve them
  - Workflow 1 : Secure raw data
  - Workflow 2 : Empower the end users
- Genetic Bio-molecular and cellular institute customer case study
- Conclusion

*Speaker: Robert-Jan Overzee, Product Marketing Manager, Atempo*



# Data in the Research world

## Introduction

*Speaker: Robert-Jan Overzee, Product Marketing Manager, Atempo*

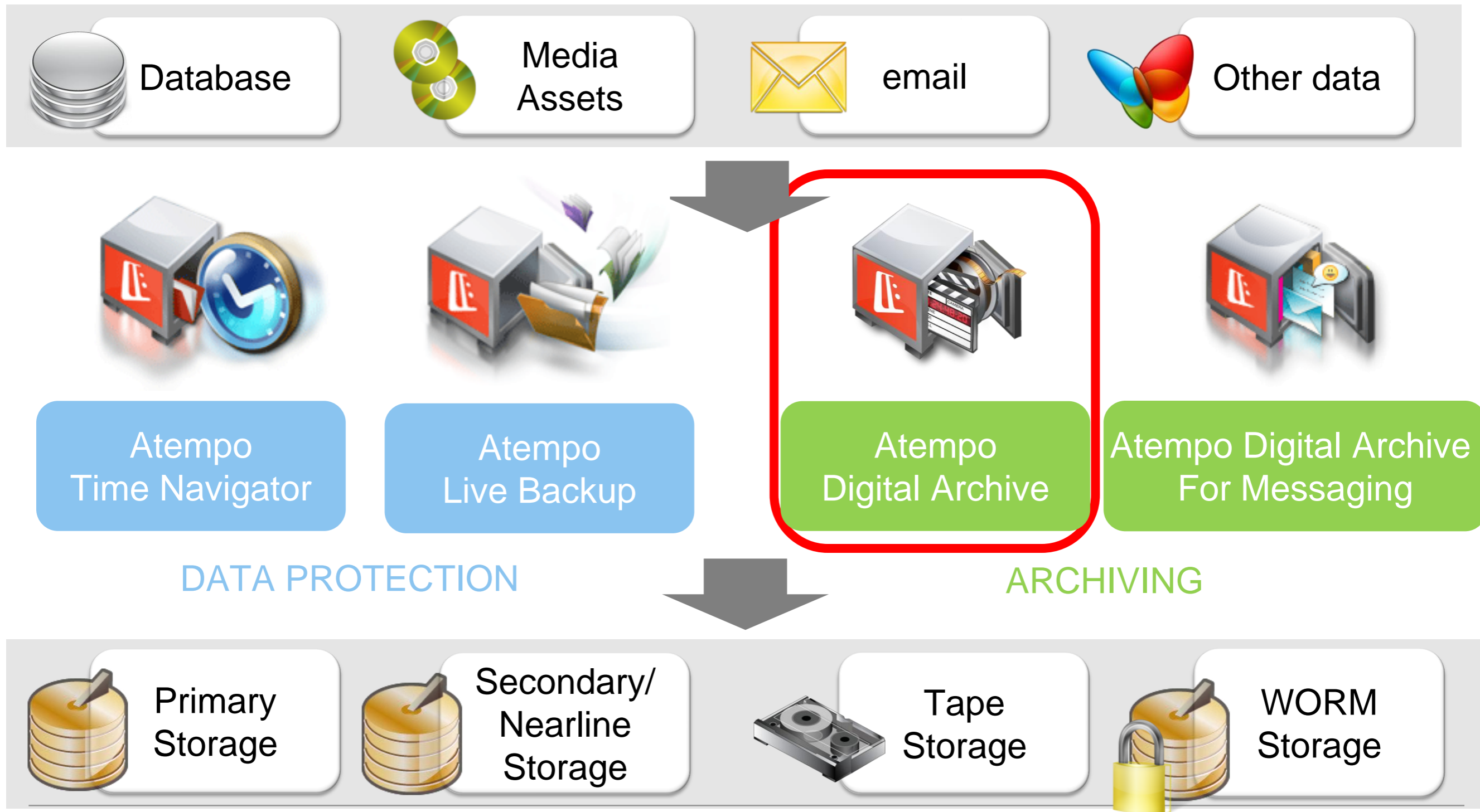
# About Atempo

---

- Headquartered in Palo Alto, California & Paris, France  
13 offices worldwide; N. America, Europe and Asia
- Established storage management solutions company with strong pedigree in data protection
- 4500+ corporate customers;
- 300 employees worldwide
- 200 + partners
- Focus: Protection & Retention of Digital Assets
  - Integrated data protection and digital archiving solutions
  - Broad system coverage from data centers to remote offices and laptop/desktops
  - Comprehensive application and content coverage
- Member of the Active Archive Alliance



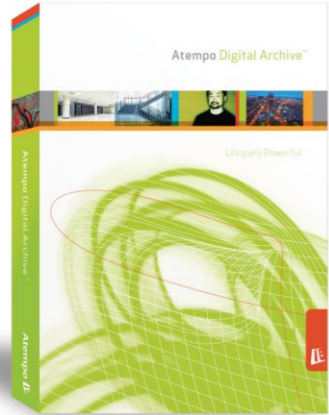
# Heterogeneity & Flexibility



# Atempo Digital Archive

---

## Archiving Solution for Files



- Economic archiving - Reduce footprint on expensive primary disk
- Long term retention
- Moves data from any of your storage to archiving storage
- Archiving performed, automatically or through policies in multiple copies
- Powerful Metadata retrieval.

## Technical description



Mac



- 18 years of all type of IT Storage support, Hardware vendor agnostic.
- Supported OS: Windows Mac, Linux, Solaris, HPUX, AIX
- Open format, Native file format or TAR v2.
- Open architecture with optimized usage of storage managers & CAS devices



# Data in the Research world

## Common issues and real world workflows to solve them

*Speaker: Robert-Jan Overzee, Product Marketing Manager, Atempo*

# The Supersizing of Storage

---

In 2009 the total storage in the world was 0.8 ZB

In 2010 this storage is now 1.2 ZB

or 1,200,000,000,000,000,000,000 bytes =  $1.2 \times 10^{21}$

- Cheap storage fuels the binge
- Rich Content
  - High-res images in reports
  - Animations in presentations
- Digital video & audio downloads
- **Science Raw data ...**



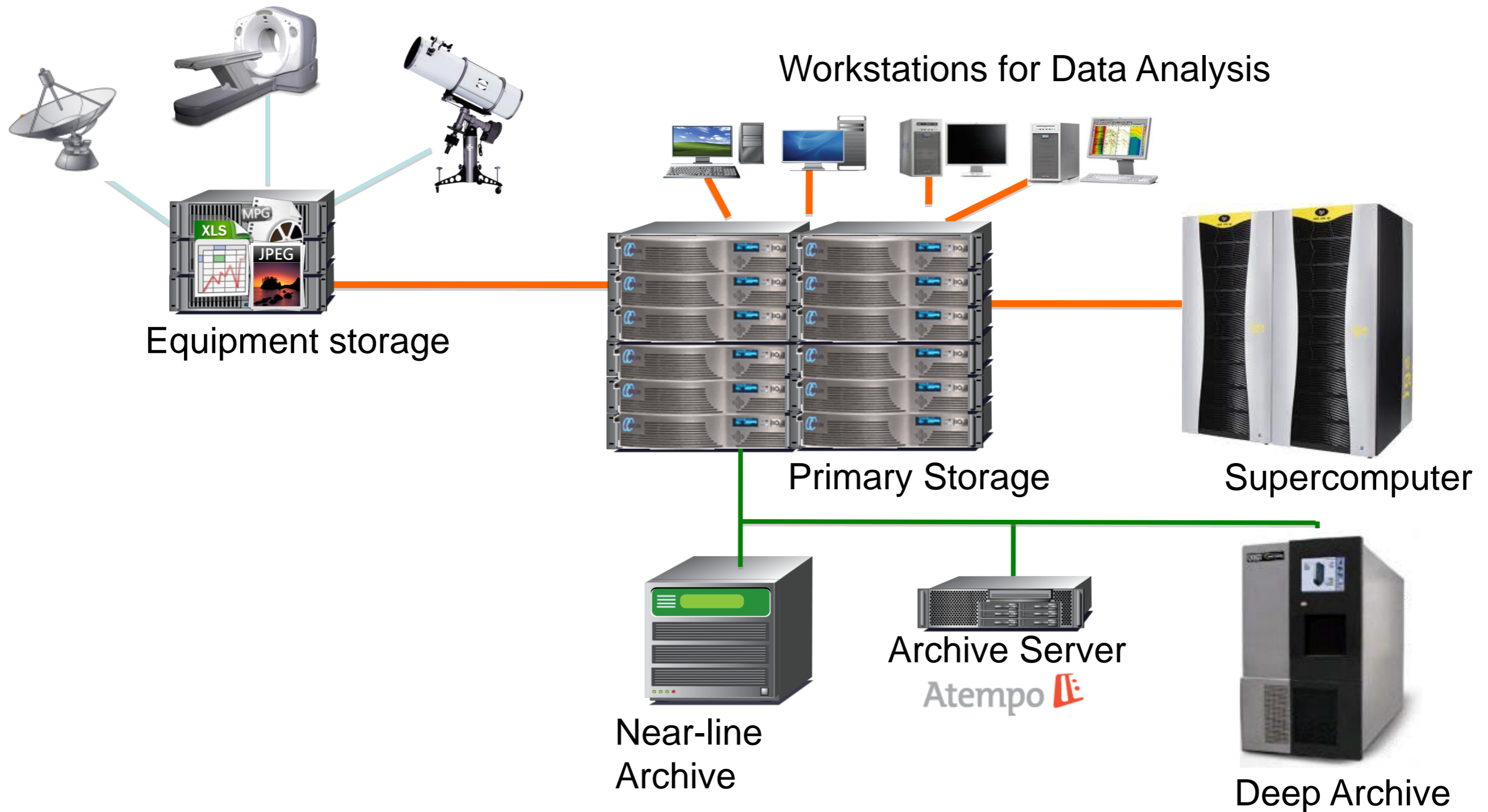
# Typical Science IT Storage symptoms

---

- Backup windows get longer and longer
- End users « solve » their storage needs with USB disks
- Backup misused as Archiving
- Several types of storage, from performant to deep archive
- Experiment raw data at risk
- Ever growing digital data
- Different retention times
- Finding data : a hassle
- Regulations impose Archiving



# Workflow 1-Step 1: secure raw data



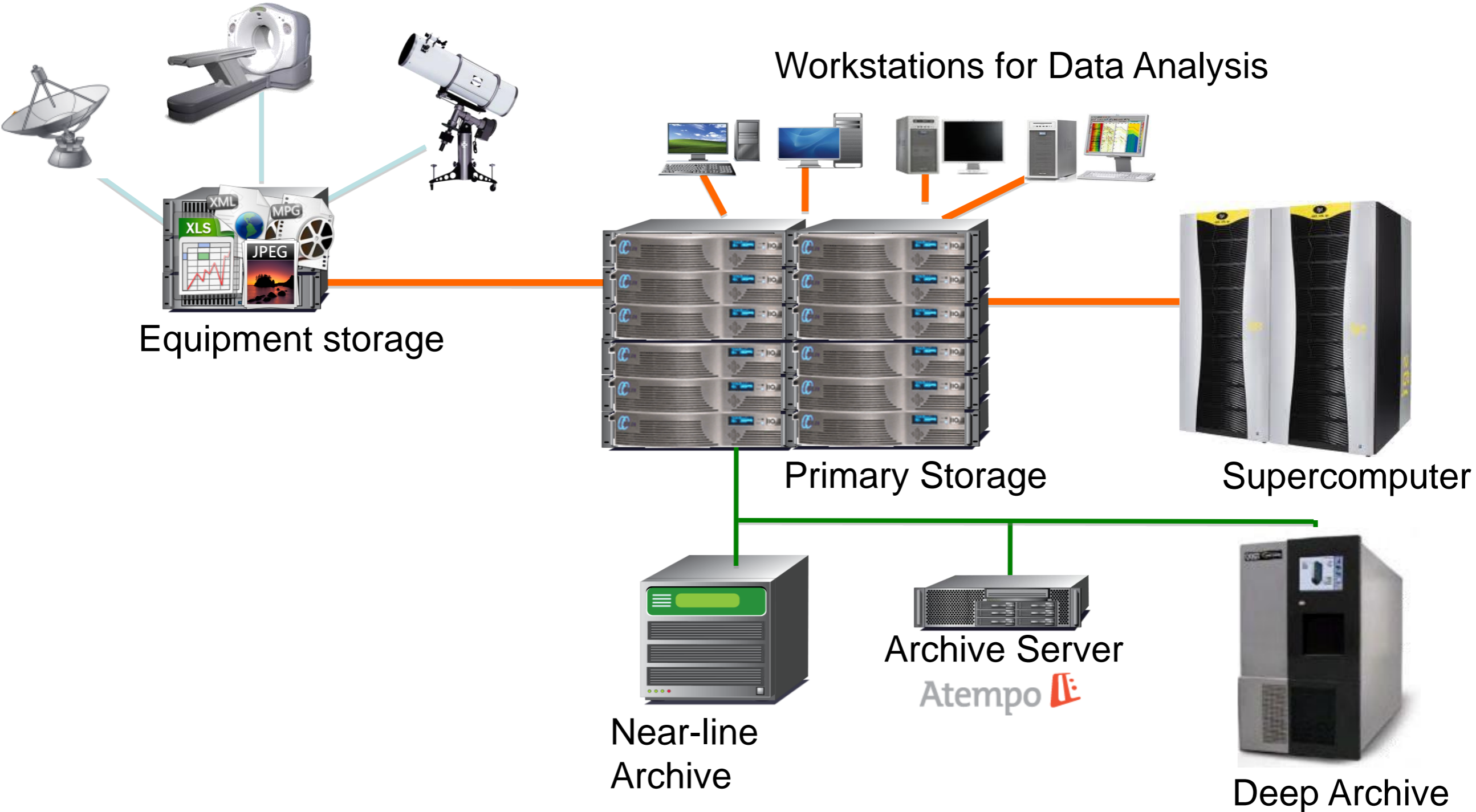
# Workflow 1-Step 1: secure raw data

---

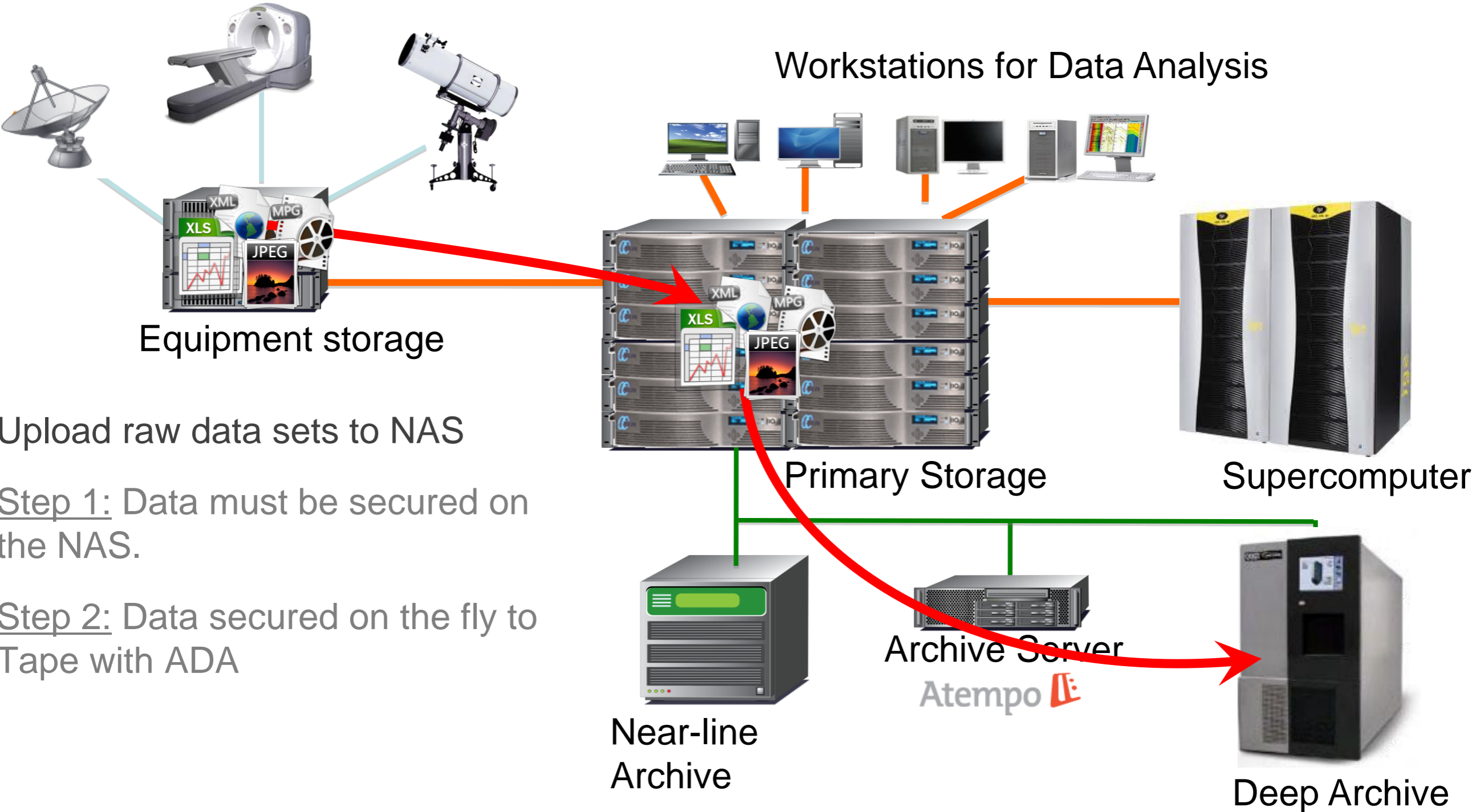
- Experiments and measurements are often
  - impossible to reproduce
  - expensive to reproduce
  - complicated to reproduce
- Its results must not be lost
- The data must be secure ASAP
- Access to data must remain easy
- Storage on the measurement tools must be freed



# Workflow 1-Step 1: secure raw data



# Workflow 1-Step 1: secure raw data

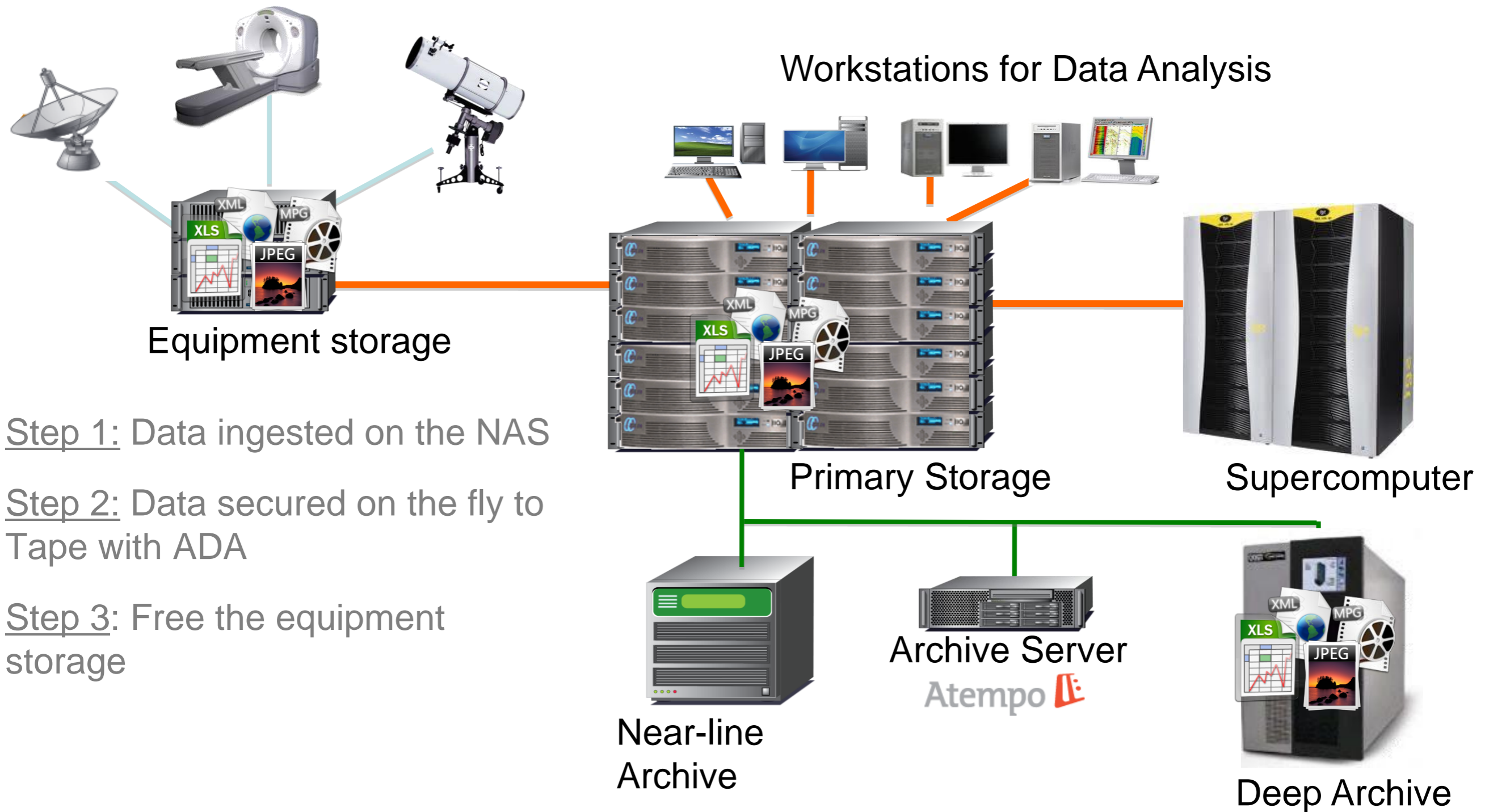


Upload raw data sets to NAS

Step 1: Data must be secured on the NAS.

Step 2: Data secured on the fly to Tape with ADA

# Workflow 1-Step 1: secure raw data

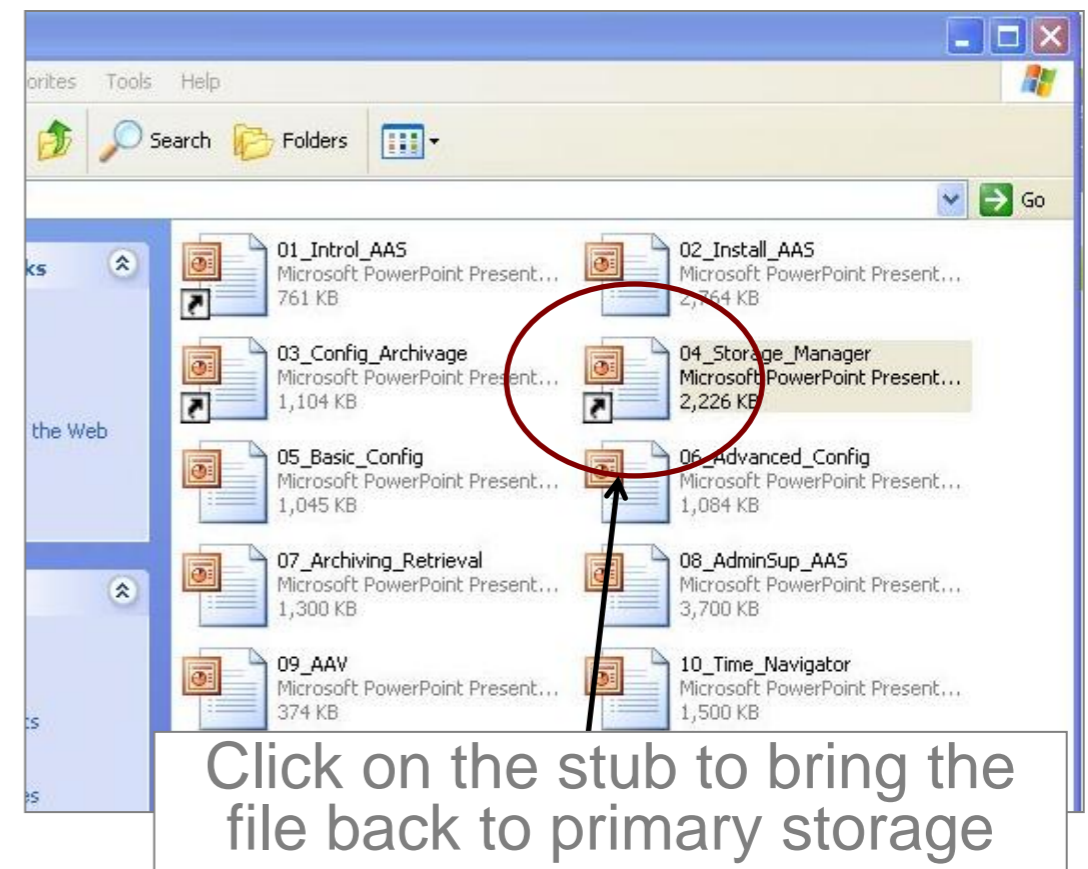


# Workflow 1-Step 2: free up space

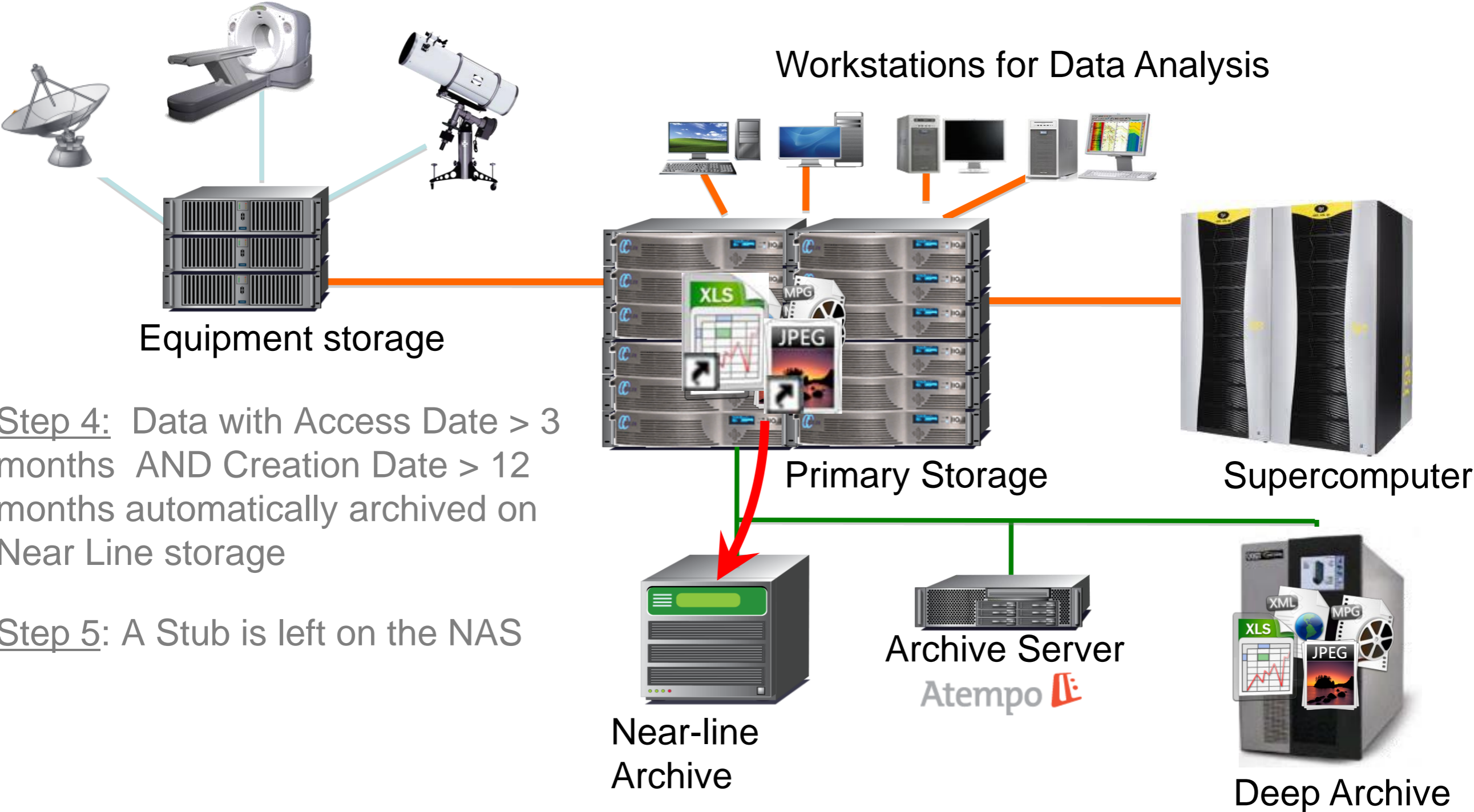
## *Create rules and schedules for automatic file archive*

This “archive” solution:

- Uses site-specific criteria like last access date and file type to automate archiving
- Leaves stubs on primary storage that point to the archived file
- Makes retrieving a file is as easy as clicking on the file in a file browser
- Native stubbing available on:  
Windows ; Mac OS; Netapp Filers ;  
EMC Celerra ; BlueArc Titan &  
Mercury



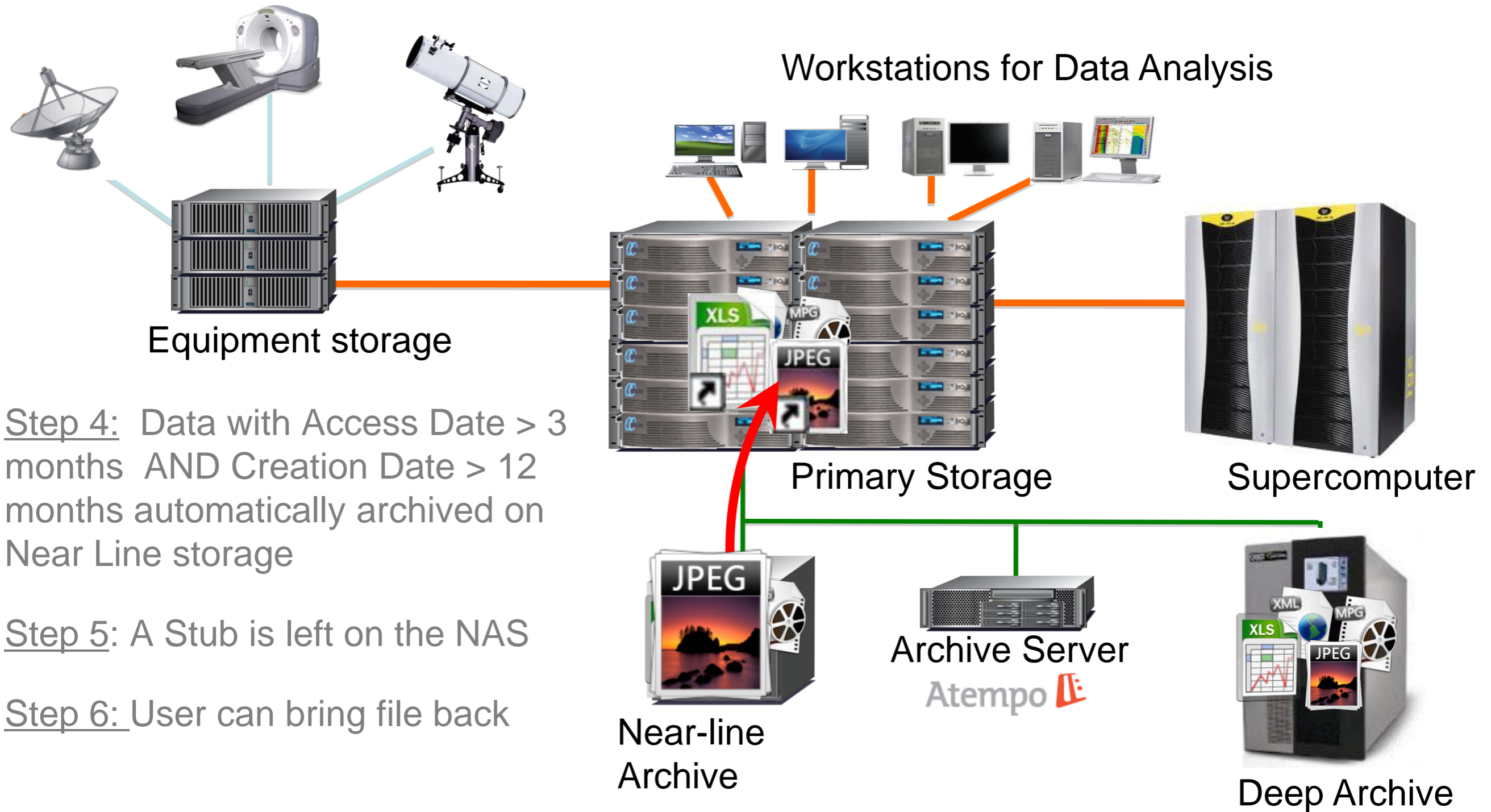
# Workflow 1-Step 2: free up space



Step 4: Data with Access Date > 3 months AND Creation Date > 12 months automatically archived on Near Line storage

Step 5: A Stub is left on the NAS

# Workflow 1-Step 2: free up space

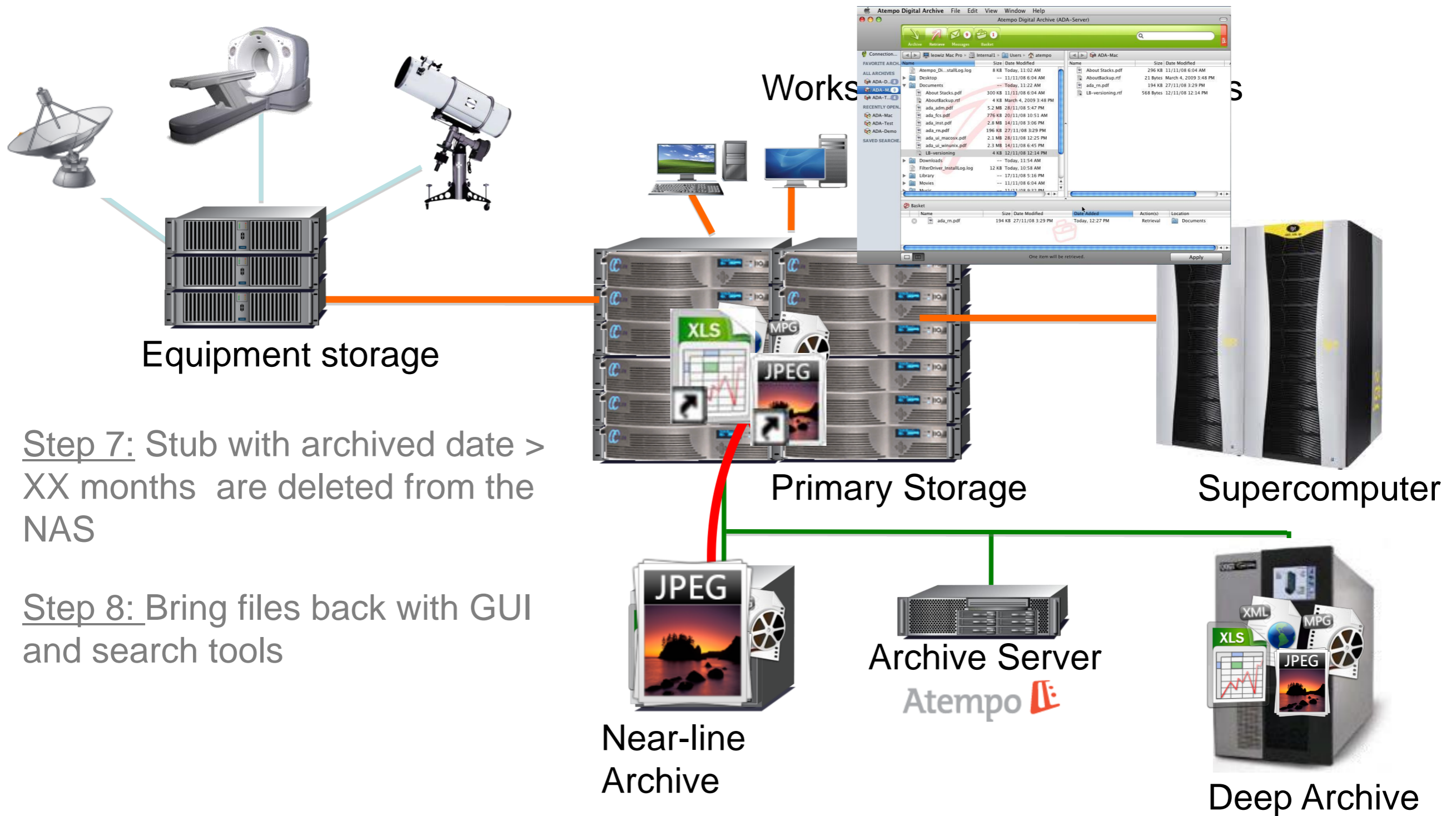


Step 4: Data with Access Date > 3 months AND Creation Date > 12 months automatically archived on Near Line storage

Step 5: A Stub is left on the NAS

Step 6: User can bring file back

# Workflow 1-Step 2: free up space



# Workflow 2: User-Initiated Archive

---

## Empower the Researchers

Data belongs to:

- A researcher
- A team

The IT archivist defines the classification system.

The end users do know:

- When they need their data
- When they are done with their data
- Where to find their data



# Workflow 2: User-Initiated Archive

Search by Metadata

Metadata List	Metadata Name	Metadata Value						
<ul style="list-style-type: none"> <li>⊕ ADA metadata</li> <li>⊕ Final Cut Pro</li> <li>⊖ petrochemical                             <ul style="list-style-type: none"> <li>📄 Benzene</li> <li>📄 Ethylene</li> <li style="background-color: #e0e0e0;">📄 Propylene</li> <li>📄 Toluene</li> <li>📄 Xylene</li> </ul> </li> </ul>	<div style="text-align: center; border: 1px solid black; padding: 5px; margin: 5px 0;"> <span style="color: green; font-size: 24px;">&gt;</span>  <span style="color: red; font-size: 24px;">&lt;</span>  <span style="font-size: 12px;">or</span>  <span style="color: red; font-size: 12px;">and</span>  <span style="font-size: 18px;">⌵</span> </div>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #e0e0e0;">/petrochemical/Benzene</td> <td>Yes</td> </tr> <tr> <td>/petrochemical/Ethylene</td> <td>No</td> </tr> <tr> <td style="background-color: #e0e0e0;">/petrochemical/Propylene</td> <td>No</td> </tr> </table>	/petrochemical/Benzene	Yes	/petrochemical/Ethylene	No	/petrochemical/Propylene	No
/petrochemical/Benzene	Yes							
/petrochemical/Ethylene	No							
/petrochemical/Propylene	No							

Results

- ⊖ Instances
  - /Chemical departments/Petrochemical studies@adhesives-classification.docx (01/14/2011 07:0...
  - /Chemical departments/Petrochemical studies@adhesives-reference.docx (01/14/2011 07:03:2...
  - /Chemical departments/Petrochemical studies@adhesives-tests.docx (01/14/2011 07:03:23 PM
  - /Chemical departments/Petrochemical studies@detergents-classification.docx (01/14/2011 07:(
  - /Chemical departments/Petrochemical studies@detergents-tests.docx (01/14/2011 07:05:04 PM
- ⊖ Jobs
  - 31
  - 32
  - 33

Run

i
-
⏻

Digital Archive™

- ives
- departments
- udies
- mental risks
- hemical studies
- epartments
- rchive
- ics archive
- dynamics archive
- physics departments
- rchive
- otron tests
- g Department

Archive

← Return



## Data in the Research world

# Genetic Bio-molecular and cellular institute customer case study

*Speaker: Robert-Jan Overzee, Product Marketing Manager, Atempo*

# Genetic institute

---



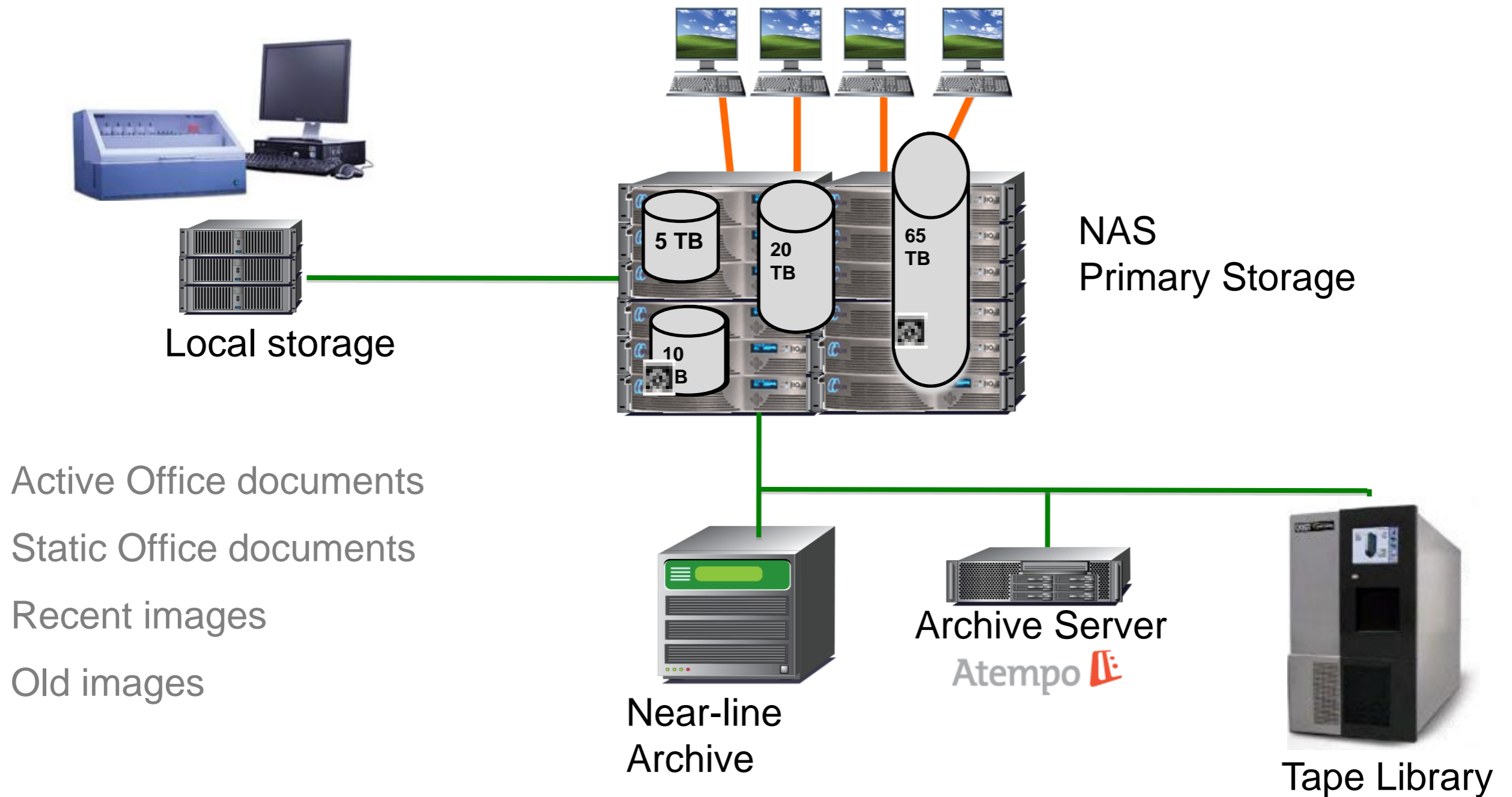
- One of the leading European centers of biomedical research
- Study of higher eukaryotic genomes; control of genetic expression; analysis of the function of genes and proteins.
- 100 000 mice hosted in the 'Mouse Clinic'
- 220 Publications each year
- 147 PhD and master students
- 250 engineers

# Existing Situation

---

- 100 TB of Scientist's Data is located on the Primary storage and backed up
- 2 weeks Retention (1 Full on Drive & 3 Full on tape) = 400 TB
- Data generated from the Microscope Platform are not backed-up (volume of 140 TB / Year)
- Full backup of the NAS every week takes around ~7 days with Legato Networker
- Manual Archiving to tape managed from the IT with xls

# Deployed architecture



# Results

---

- New workflows (from scanner to archive)
- All archived data is centralized
- There is always 2 copies of all data (ADA deduplication)
- Better use of existing storage
- Reduce and control the backup window to less than one day with a 12 months retention with Atempo Time Navigator
- Researchers can search their own archives thanks to Metadata, saving time to them and the IT team
- Open formats being used. Data is accessible for decades



# Data in the Research world

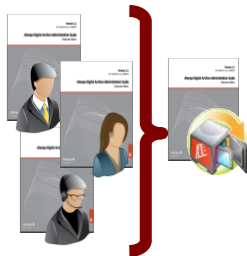
## Conclusion

*Speaker: Robert-Jan Overzee, Product Marketing Manager, Atempo*

# Summary

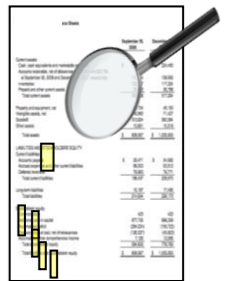
## Atempo Digital Archive ensures long-term retention of digital assets

- ❖ Reduces footprint on expensive primary disk to save money
- ❖ Reduces backup window & improves disaster recovery time



## ADA provides flexible archiving & retrieval methods

- ❖ Direct end user access including search of archived data
- ❖ Automated archiving with stub access to archived files
- ❖ Easy retrieval through GUI, stub files and via search
- ❖ Integrated in application workflows, APIs and CLI
- ❖ Broad archive media support: tape, disk and cloud storage
- ❖ Long term retention in Open Formats





“Atempo Digital Archive enables us to maintain our precious scientific assets in open formats while keeping these archives accessible and retrievable by our researchers.”

*Tien-Dung Le Van*

*IT manager Genethon, a non-profit center for biotherapy and DNA sequencing*

# Thank you

Visit our booth at the S3

“Managing Data growth in Life sciences Seminar”

*Speaker: Robert-Jan Overzee,*

*robert.overzee@atempo.com*