Cell Observatory Users Meeting

ESS Guidelines and Immediate Required Action Items

Aims

- Cell Observatory Data Storage Overview
- ESS Usage Guidelines – Active from tomorrow
- Required Immediate Action
  - 1st step By February 8th
  - 2nd step by February 26th

Cell Observatory Data Storage Overview

Automatically saved during Acquisition

Z1 Storage PC
aka SWAP

Inspect and Manual Move To *Raw_Data*

ESS: ~430Tb Active Storage

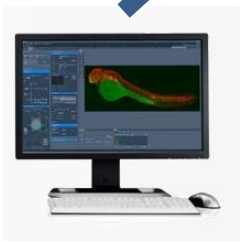
Almost Full
Very Messy

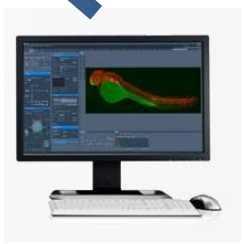
Automatically Archive All Data from *Raw_Data* & *Analyzed_Data*

Recovery of Archived Data

Long Term Data Archive

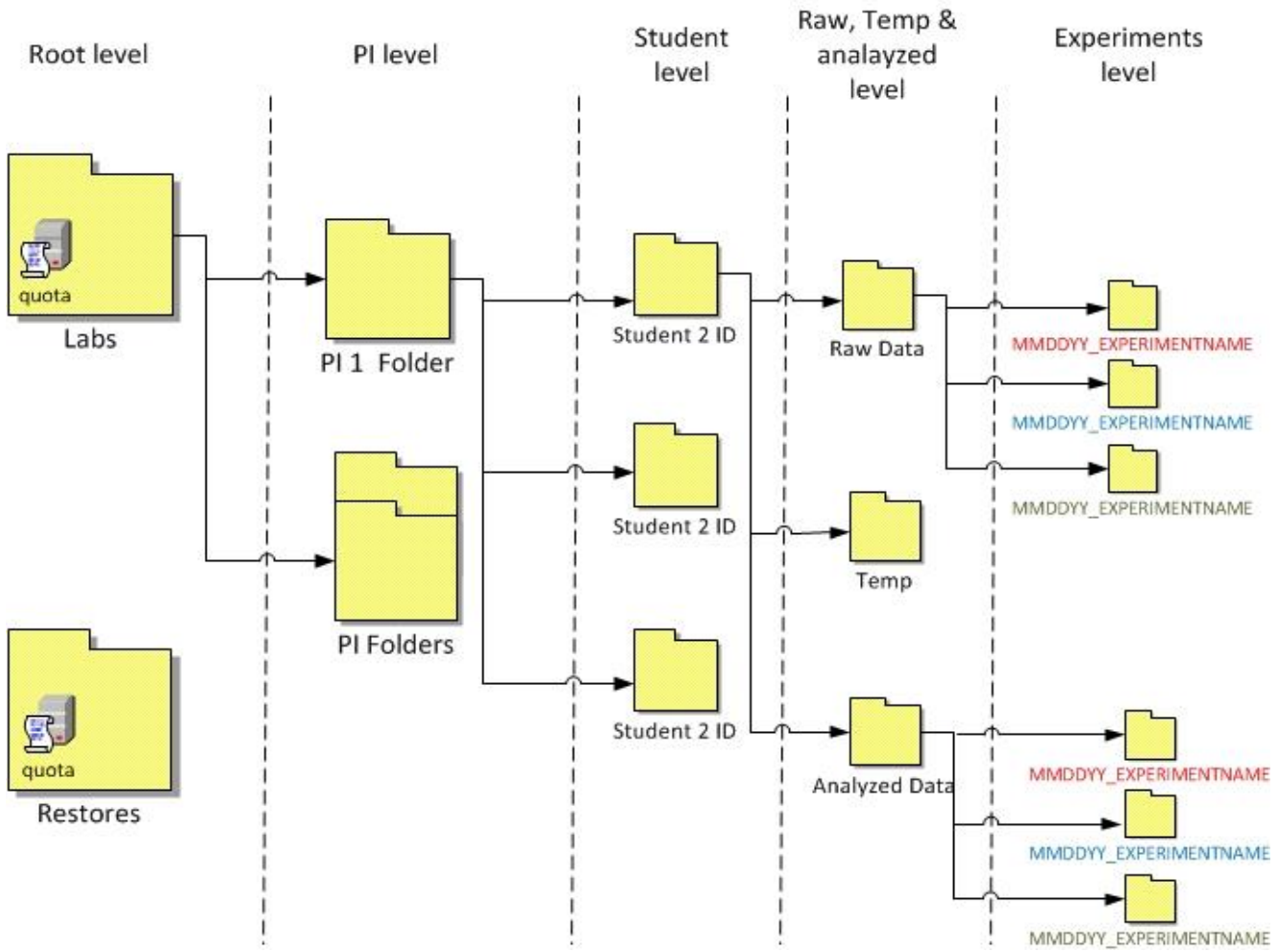
Currently ~800TB * 2 !
Will be Charged Soon





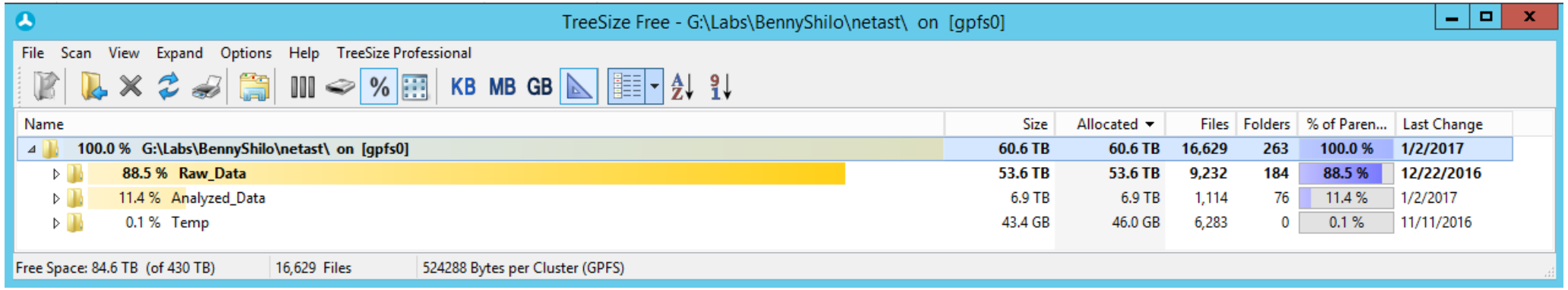
Visualize / Analyze directly from ESS (GPFS protocol over fast network)
Create new Data on *Temp* & *Analyzed_Data*

ESS Folder Structure



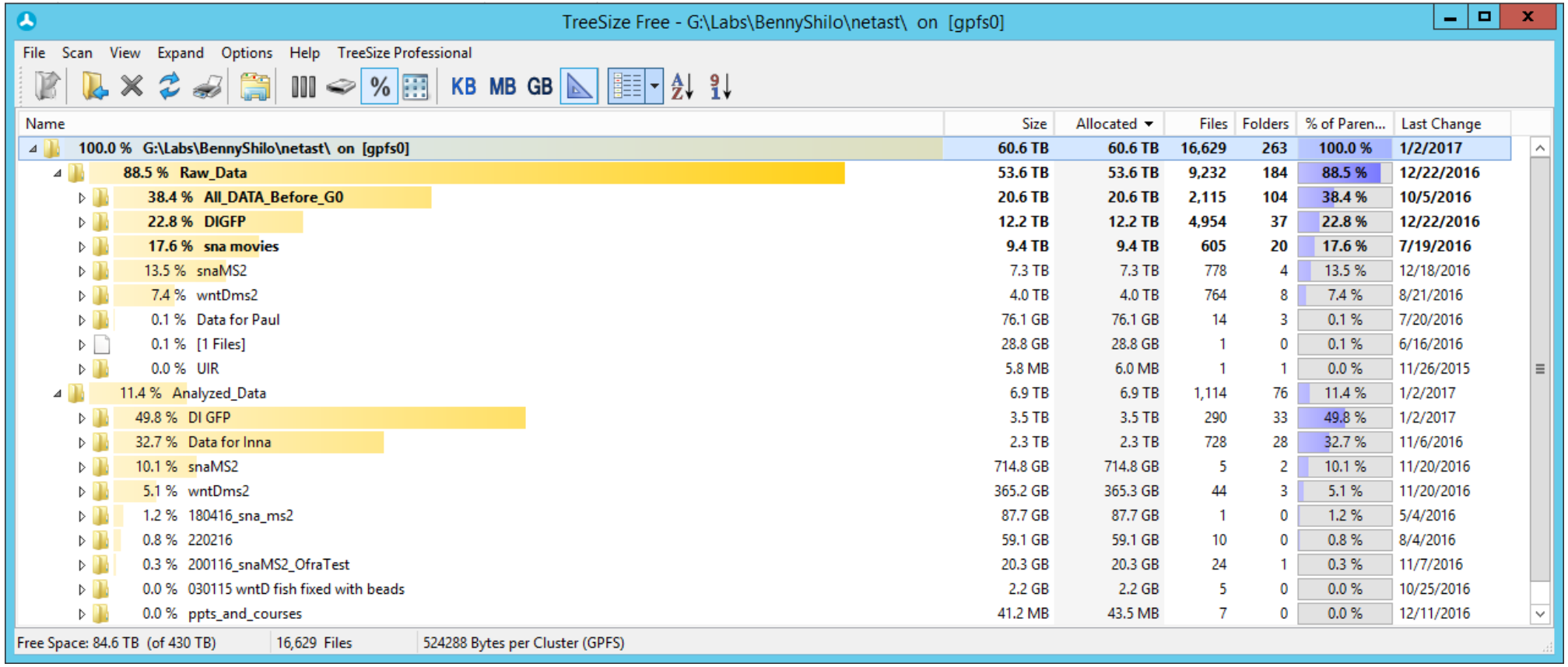
My ESS Disk Usage

Use TreeSize Free to inspect your ESS disk Usage



My ESS Disk Usage

Use TreeSize Free to inspect your ESS disk Usage



Typical Workflow: Data Duplication Problem

Data Acquisition	ZEN	Original File .czi	Delete after Fusion
Dual Side Fusion	ZEN / Fiji SPIM Registration	Fused file .czi	
Multi-View Fusion	ZEN / Fiji SPIM Registration	MV-Fused file .czi	
Load into Arivis	Arivis	Non-stitched Arivis file .sis	Delete After Stitching
Stitching	Arivis / Fiji / TeraStitcher	Stitched Arivis File .sis	
Export	Arivis	.tif files	Delete After Import to Imaris
Analyze	Imaris	Imaris file .ims	

ESS Usage Guidelines

- Follow File and folder naming guidelines (meaningful name, no spaces, no special characters, don't save files directly under Raw/Analyzed/Temp)
- Raw Data should be always archived
- Never Duplicate Raw Data files on ESS (Under *Raw_Data* folder, or different user)
- Minimize (Saving of) Data Duplicates
- Dual Side Fusion and Multi-View Fusion (using ZEN) will be saved under *Raw_Data* All Further Data Visualization & Analysis should be done in the user's *Temp* folder
- Carefully Select which Analyzed data to archive, **move** it to *Analyzed_Data* (**Don't copy**) – Charging will be based on size of archived data
- Use Matched Folders Names in *Raw_Data* / *Analyzed_Data* / *Temp*
- Quota / Time policy will be imposed

Active for new data from Jan 24th

Data Archive & Retrieval

- Data will not be saved on ESS forever
- All Data is stored on 2 copies
- Raw and Processed Archived Data can be recovered from the Tapes
- Note that any change to a file create another archived copy (analyzed data) Archived data will be charged by TB
- In order to retrieve files from the Tapes a full path should be provided
- To enable data retrieval a database application will be created
- Retrieval is into *Restore* folder

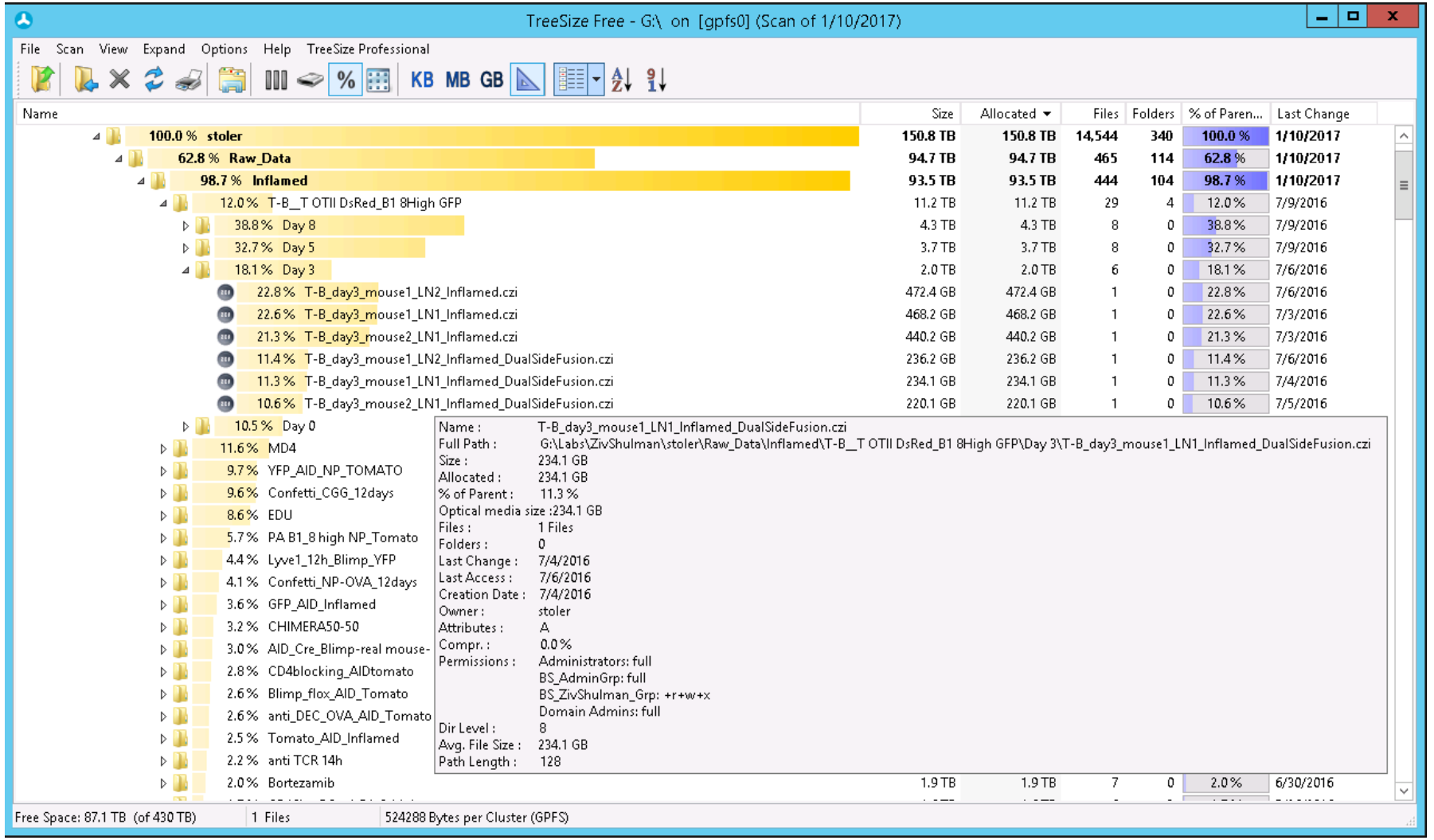
Full Data Retrieval Instructions will be posted soon

Data Annotation

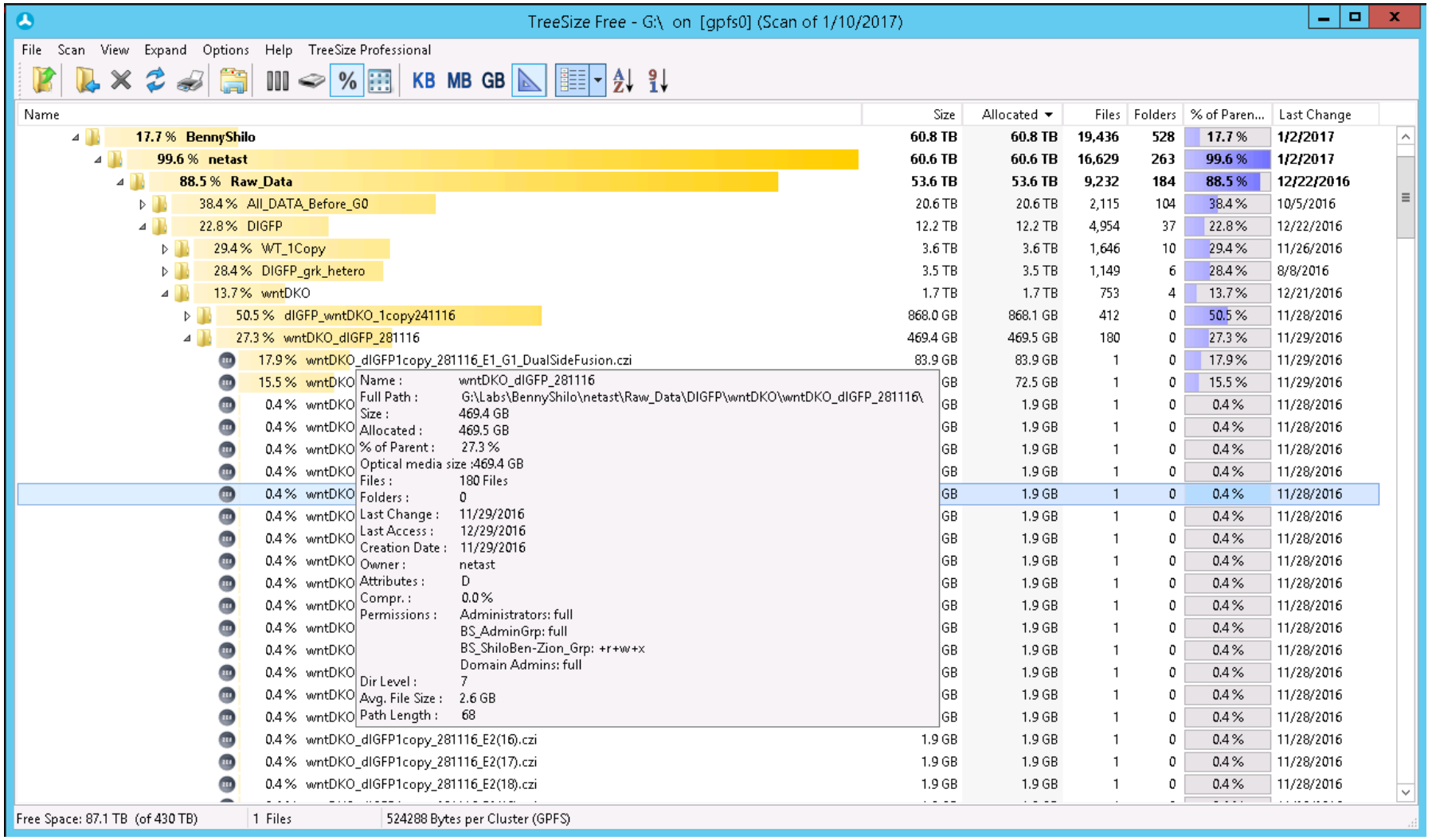
- Storing the information about the datasets names and content is required to enable data retrieval
- A database application will help you keep track of your files and experiments.
- Impossible to save information at the single file level (optionally multiple files for experiment)
- **Path and description will be saved at the folder level only and not at the single file level**
- A new item will be added for each subfolder that contain data files
- Automatically saved info: Folder name, Full path, Folder creation date, Number of files, size, Data ownership
- The user is required to fill in additional information:
  - Experiment information (textual): conditions, date , number of files of each experiment
  - Anything that you think you or your PI will need 2 years from now to identify the experiment
  - Base File names, multiple files separated by comma

Full Data Annotation Instructions will be posted soon

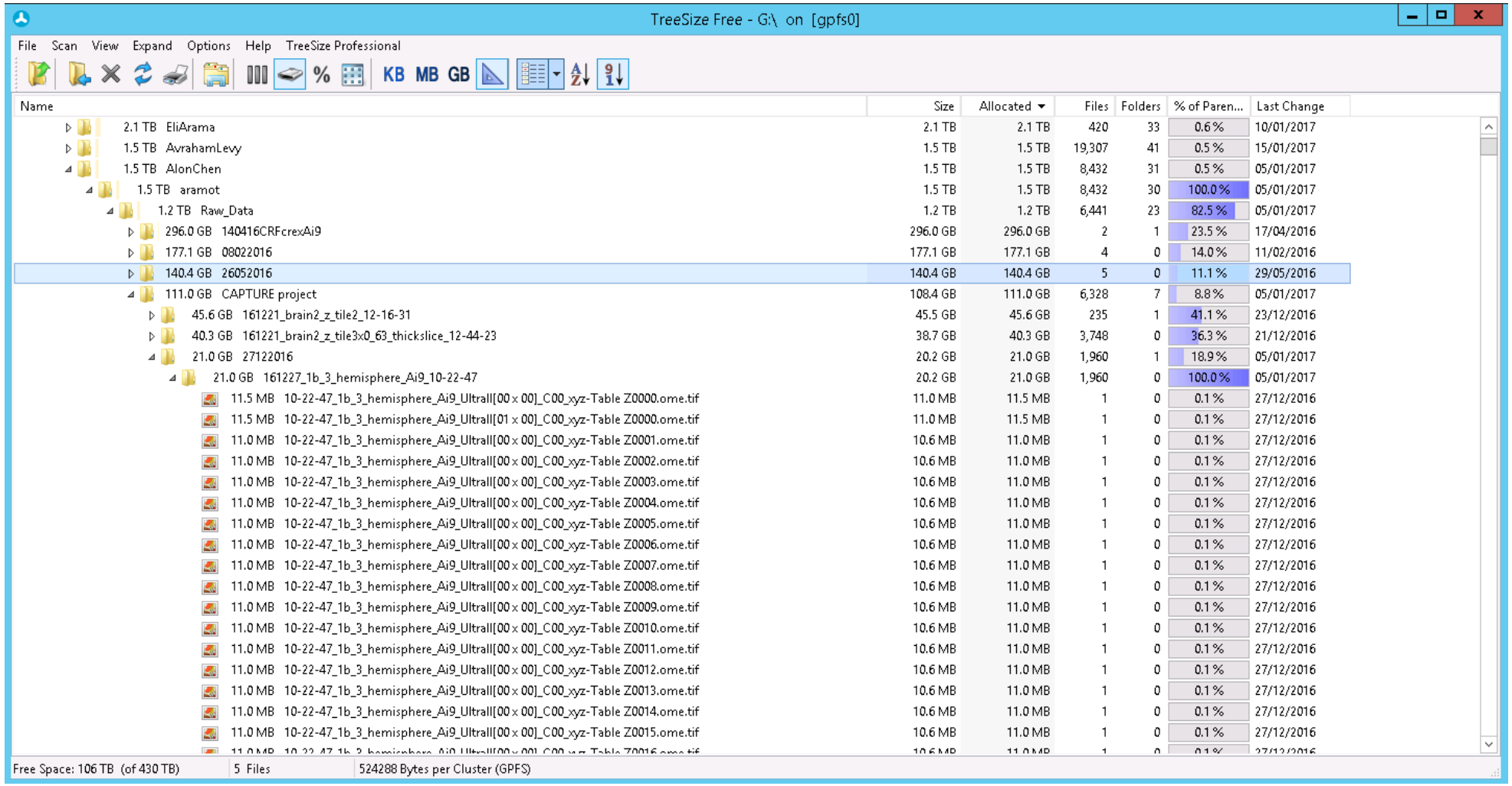
Nested Folders, Single File per Experiment



Nested Folders, Multiple Files per Experiment



Nested Folders, Multiple Files per Experiment



ESS Reorganization & Cleanup

Required Action Step 1: **by February 8**

- Delete all data that is
    1. Not needed (bad experiment, temporary)
    2. No need to archive it
- Reorganize the data

Create updated fresh archive

Required Action Step 2: **TBA, tentatively by February 26**

- Annotate data to enable retrieval
- Delete data that is not needed currently (once archived)

Required Action Step 1: By February 8

Delete unsuccessful experiments

- From all folders

Remove Duplicated Data Step1

- For all files stitched in Arivis – Delete the non-stitched Arivis files
- Delete all temporary tif files (exported from Arivis for import in Imaris)
- Remove all other un-needed data (eg. Duplicated files) from ESS and from WS

Reorganize Data

- Match folders names under *Raw_Data* / *Analyzed_Data* / *Temp*
- Move analyzed data that you do not want to archive into *Temp* folder

Required Action Step 2: TBA By February 26

Annotate Data to enable Data Retrieval

- We are currently building an application for this, more instructions will be posted soon

Remove Duplicated Data step 2

- For all files for which Dual-Side Fusion was done successfully – Delete the original files (before fusion)
- Remove all other un-needed data (eg. Duplicated files)

Remote Data Access and Further Info

Connecting to the ESS from your lab

- Open my computer → Map network drive → Folder → \\analyis80t.wismain.weizmann.ac.il\Labs\lab_name
- At security prompt enter:
  - username - wismain\userID
  - password - your WIS regular password

Further info can be found on the **Unit WIKI page**:

https://bbcunit.atlassian.net/wiki/display/BIMGP/Advanced+Light+Microscopy+Unit