An Open-Source Data Management and Processing Framework on a Central Server for Scientific Experimental Data

EuclidLIMS: A habit-changing way of saving and using data

Ao Liu,
Deputy Director of Research, Euclid Techlabs
On behalf of the EuclidLIMS team: Dr. Weinan Si, John Callahan, Dr. Shashi Poddar, Dr. Jie Gao

First things first – our umbrella

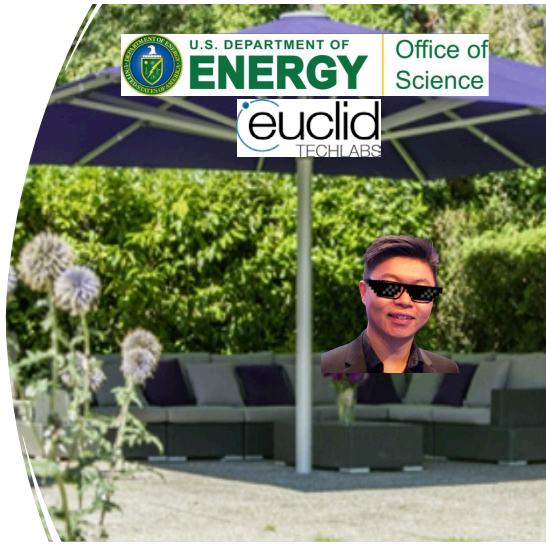
- Euclid Techlabs, LLC: a R&D company specialized in:
  - Linear particle accelerators
  - Ultrafast electron microscopy
  - Advanced material technologies.
- Formed in 2003; Spun-off: Euclid Beamlabs in 2014. Same research interests but different projects.
- Products and services (www.euclidtechlabs.com):
  - Time-resolved ultra-fast electron microscopy;
  - Ultra-compact linear accelerators and electron guns with various types of cathodes;
  - Fast tuners for SRF cavities;
  - Advanced dielectric materials;
  - HPHT and CVD diamond growth and applications;
  - Thin-film for accelerator technologies;
  - Robotic automation and its applications;
  - More!
- Presently: 27 research staff (16 Ph.D.) and 5 administrative. 2 labs: Bolingbrook, IL (accelerator R&D lab) and Beltsville, MD (material science lab). Collaborations with many national labs and universities.
- Eucliders are at NAPAC 2022! Find schedules for our talks/posters in the book of abstracts.
- Project funded by the DOE Office of Science under contract DE-SC0021512.

First things first – our umbrella

- Euclid Techlabs, LLC: a R&D company specialized in:
  - Linear particle accelerators
  - Ultrafast electron microscopy
  - Advanced material technologies.
- Formed in 2003; Spun-off: Euclid Beamlabs in 2014. Same research interests but different projects.
- Products and services (www.euclidtechlabs.com):
  - Time-resolved ultra-fast electron microscopy;
  - Ultra-compact linear accelerators and electron guns with various types of cathodes;
  - Fast tuners for SRF cavities;
  - Advanced dielectric materials;
  - HPHT and CVD diamond growth and applications;
  - Thin-film for accelerator technologies;
  - Robotic automation and its applications;
  - More!
- Presently: 27 research staff (16 Ph.D.) and 5 administrative. 2 labs: Bolingbrook, IL (accelerator R&D lab) and Beltsville, MD (material science lab). Collaborations with many national labs and universities.
- Eucliders are at NAPAC 2022! Find schedules for our talks/posters in the book of abstracts.
- Project funded by the DOE Office of Science under contract DE-SC0021512.

First things first – our umbrella

- Euclid Techlabs, LLC: a R&D company specialized in:
  - Linear particle accelerators
  - Ultrafast electron microscopy
  - Advanced material technologies.
- Formed in 2003; Spun-off: Euclid Beamlabs in 2014. Same research interests but different projects.
- Products and services (www.euclidtechlabs.com):
  - Time-resolved ultra-fast electron microscopy;
  - Ultra-compact linear accelerators and electron guns with various types of cathodes;
  - Fast tuners for SRF cavities;
  - Advanced dielectric materials;
  - HPHT and CVD diamond growth and applications;
  - Thin-film for accelerator technologies;
  - Robotic automation and its applications;
  - More!
- Presently: 27 research staff (16 Ph.D.) and 5 administrative. 2 labs: Bolingbrook, IL (accelerator R&D lab) and Beltsville, MD (material science lab). Collaborations with many national labs and universities.
- Eucliders are at NAPAC 2022! Find schedules for our talks/posters in the book of abstracts.
- Project funded by the DOE Office of Science under contract DE-SC0021512.

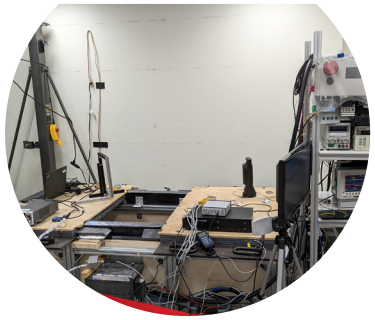


Why are we doing this

Considering the following scenarios:

- You did a **long** experiment and saved your files on a flash drive, went to a library.

  Then lost the drive. Files are not backed up. A report is due in 2 days.

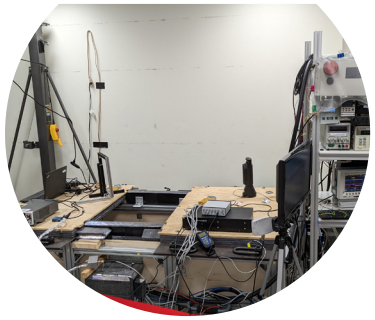


Why are we doing this

Considering the following scenarios:

- You did a **long** experiment and saved your files on a flash drive, went to a library.

  Then lost the drive. Files are not backed up. A report is due in 2 days.



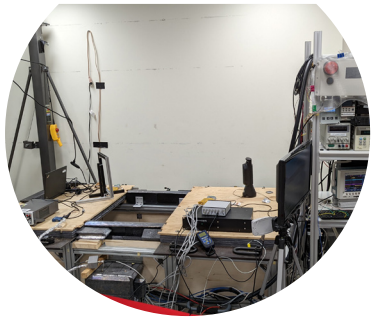
- The second time, you saved the files on a network file system (NFS). Gave the report, went on vacation

Why are we doing this

Considering the following scenarios:

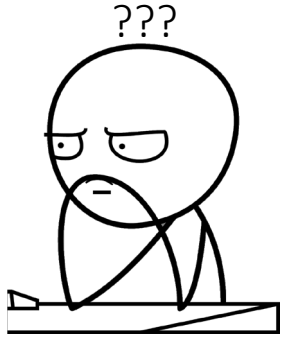
- You did a **long** experiment and saved your files on a flash drive, went to a library.

  Then lost the drive. Files are not backed up. A report is due in 2 days.



- The second time, you saved the files on a network file system (NFS). Gave the report, went on vacation

  Then your colleague wants to replot the data because an error was found. He/She has no -r permission



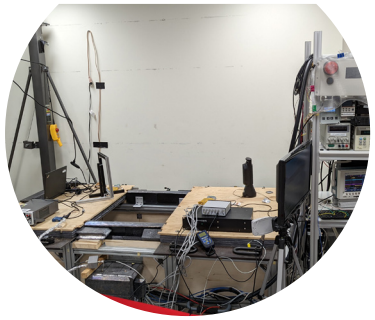
A revision is due 1 day after your vacation ends

Why are we doing this

Considering the following scenarios:

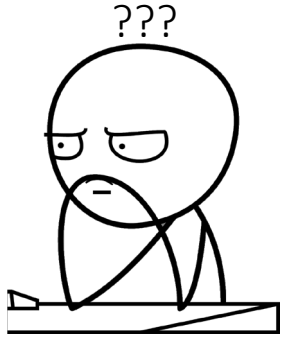
- You did a **long** experiment and saved your files on a flash drive, went to a library.

  Then lost the drive. Files are not backed up. A report is due in 2 days.



- The second time, you saved the files on a network file system (NFS). Gave the report, went on vacation

  Then your colleague wants to replot the data because an error was found. He/She has no -r permission



A revision is due 1 day after your vacation ends

Why are we doing this

You need a data manager

- Not this type of manager:

You need a data manager

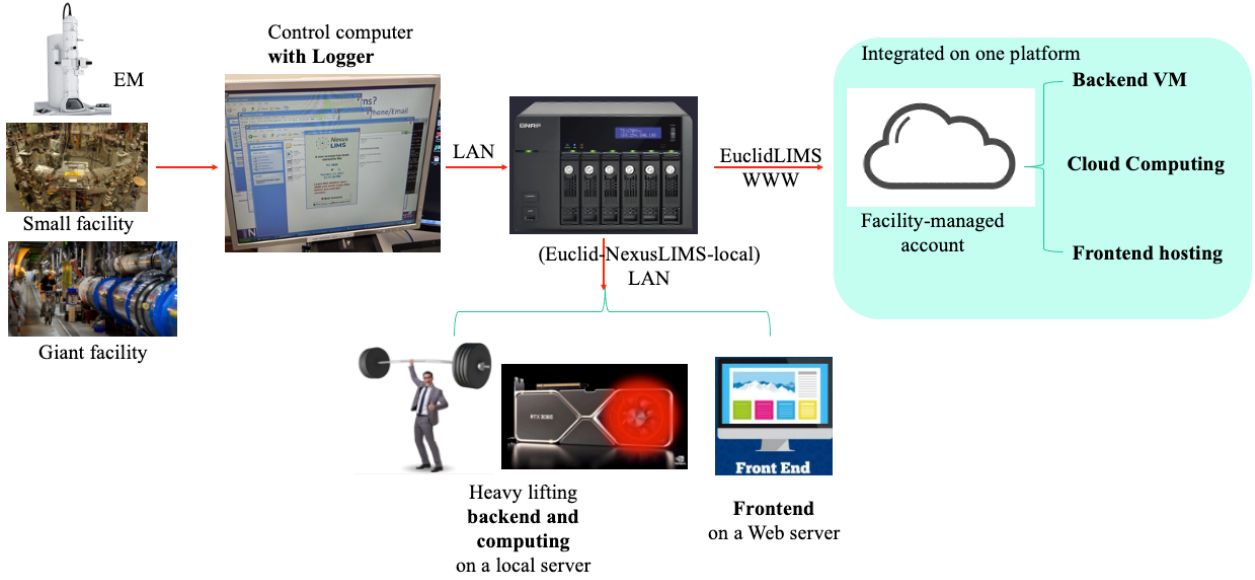
- Not this type of manager:

You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts

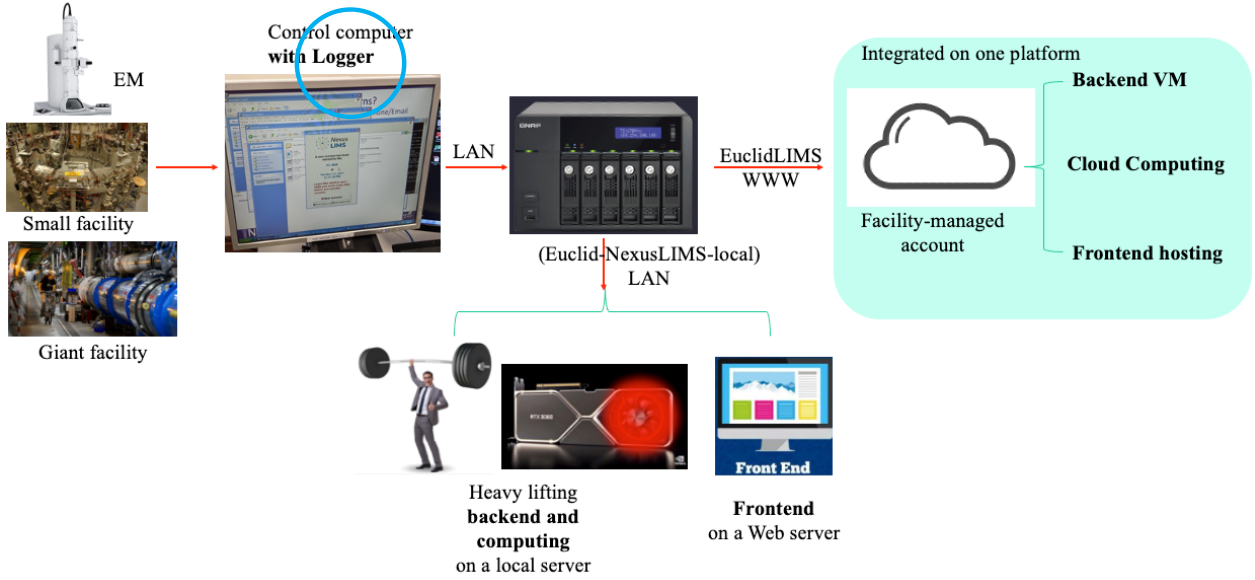
You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:



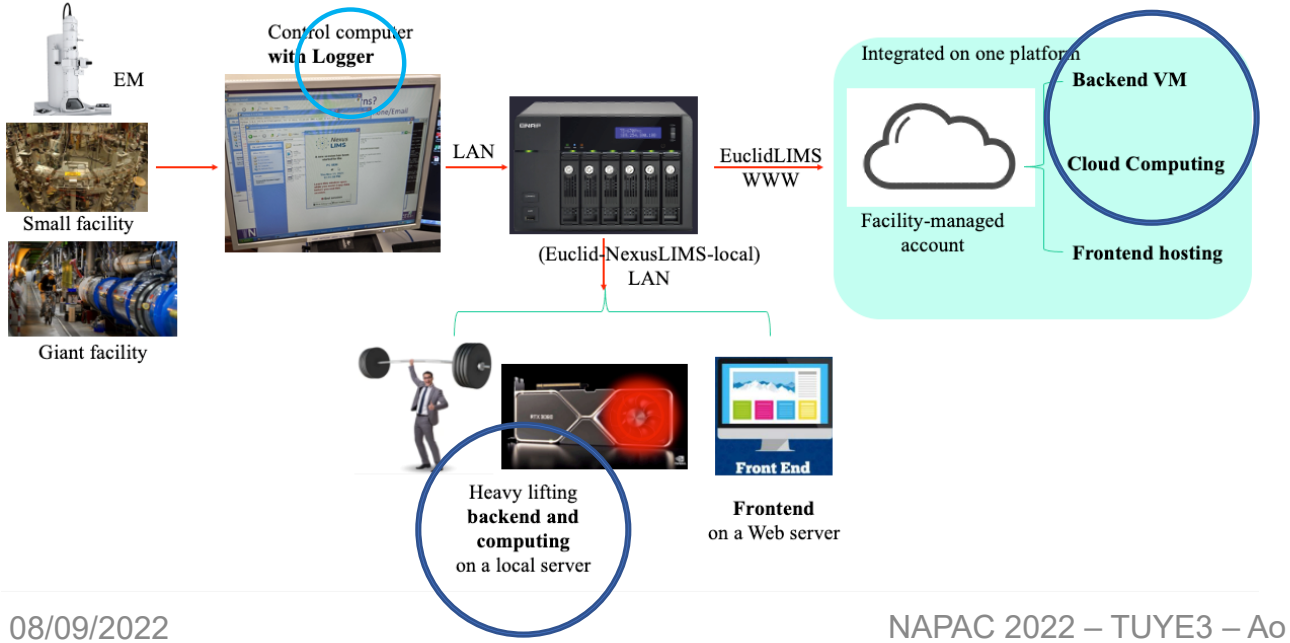
You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:



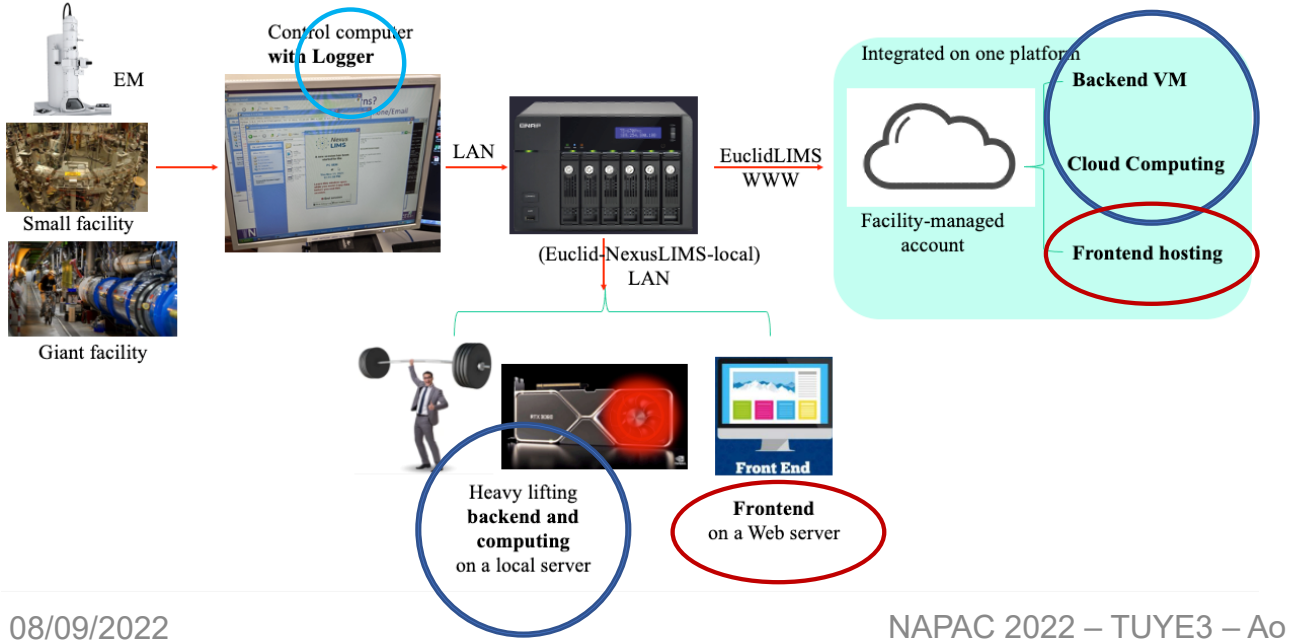
You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:



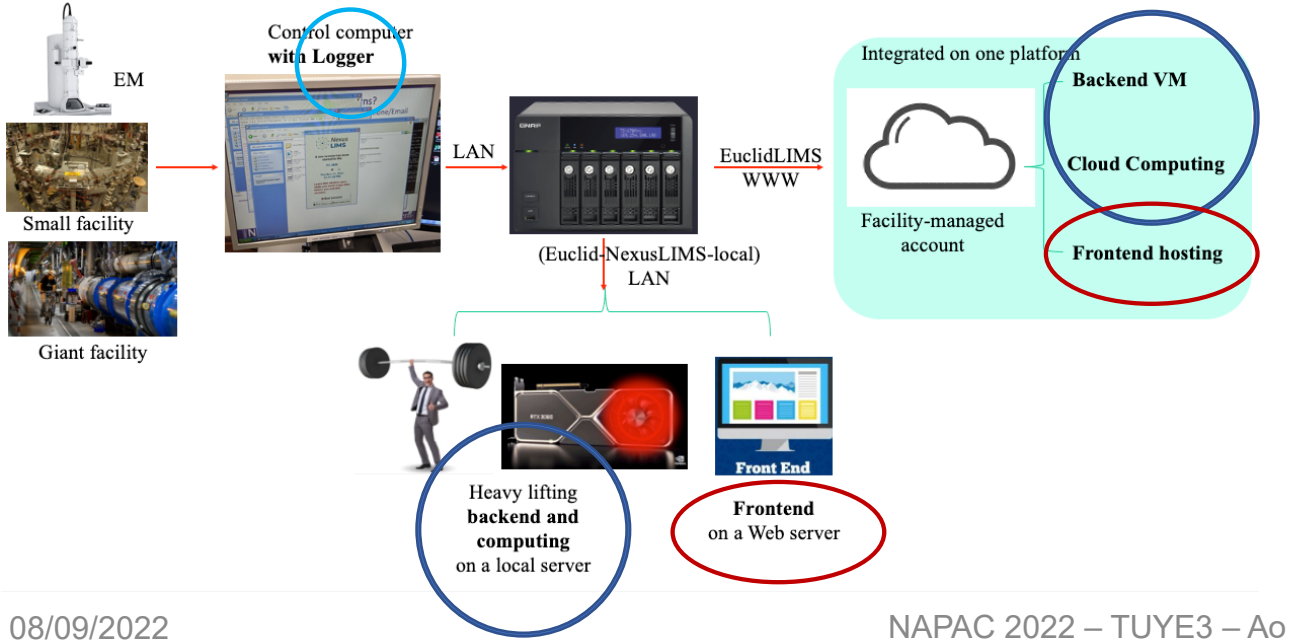
You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:



You need a data manager

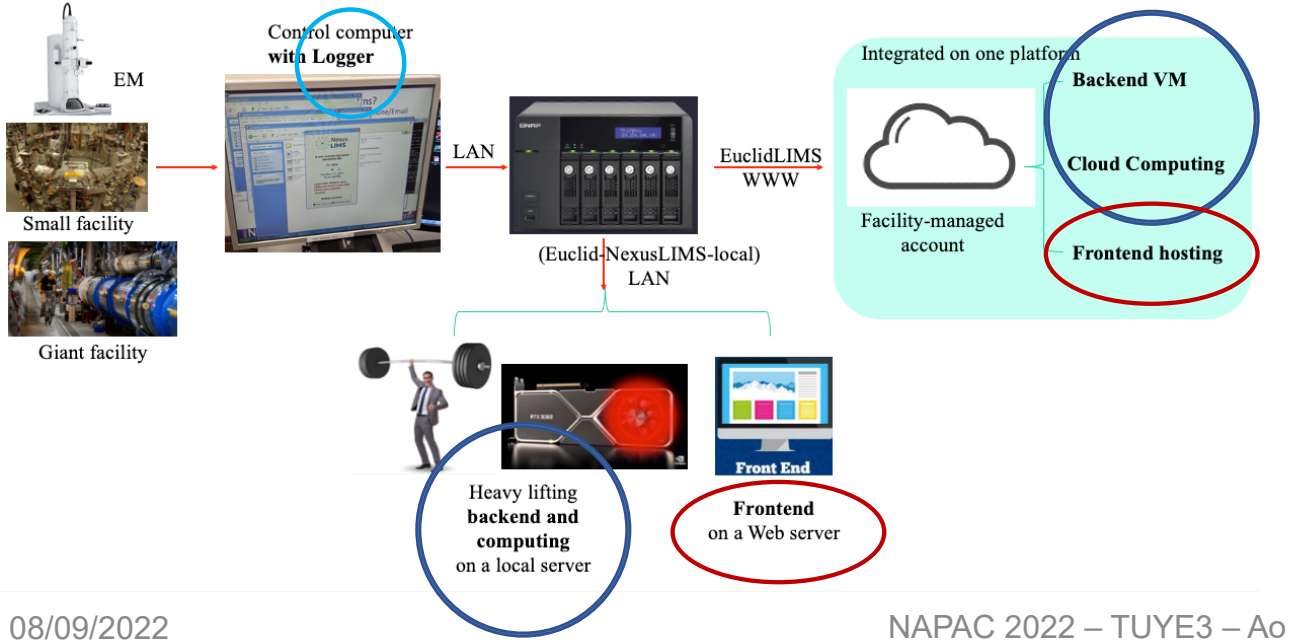
- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:



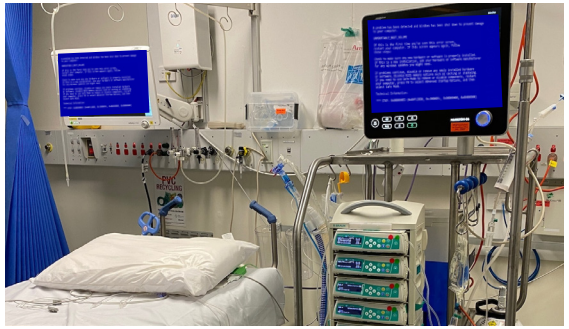
Yes, we support XP.
No certainly no 2000

You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:

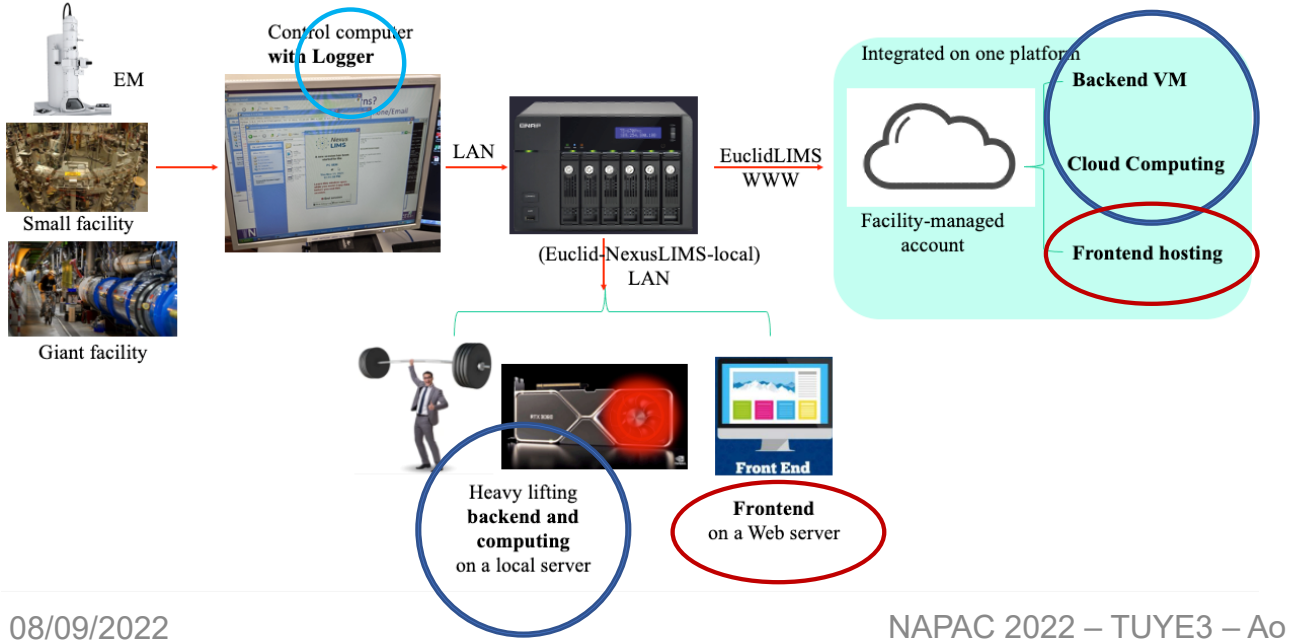


Yes, we support XP.
No certainly no 2000

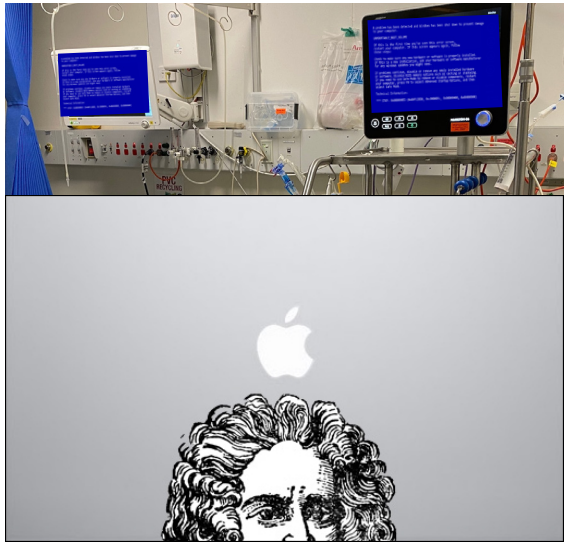


You need a data manager

- Not this type of manager:
- But actually a LIMS:
  - LIMS = laboratory information management system
  - Adopted by many labs, universities, and industry
  - Many variations and concepts
- In our case, EuclidLIMS works in the following way:

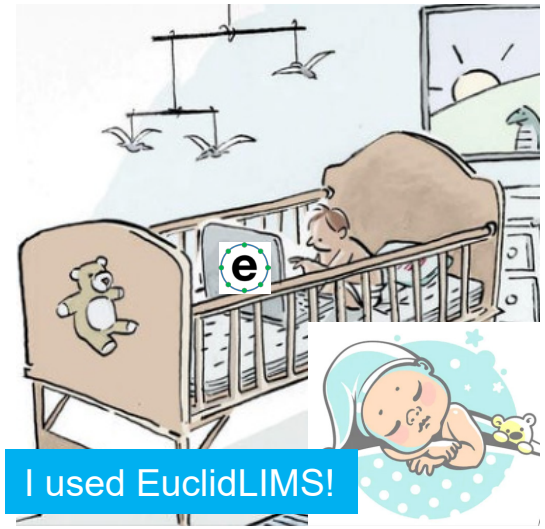


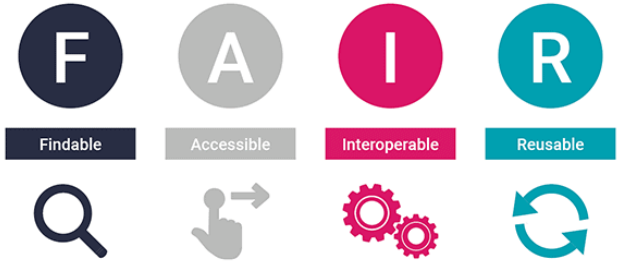
Yes, we support XP.
No certainly no 2000

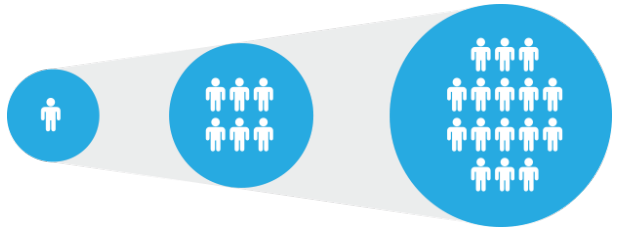


Our vision for euclidLIMS

- Findability, Accessibility, Interoperability, and Reusability (FAIR) principles of digital assets;
- Reliability, reproducibility, scalability;







- "Least invasive" (simplicity), adaptability to various data types;
- Infrastructure independency.

Our vision for euclidLIMS

euclid TECHLABS

Otherwise:

Standing on the giants' shoulders

- Giant: NIST's NexusLIMS, open-source – a collaboration
- Same concepts, same principles, similar goals
- Similar framework but infrastructure-dependent

Standing on the giants' shoulders

- Giant: NIST's NexusLIMS, open-source – a collaboration
- Same concepts, same principles, similar goals
- Similar framework but infrastructure-dependent
- Don't reinvent the wheel

Standing on the giants' shoulders

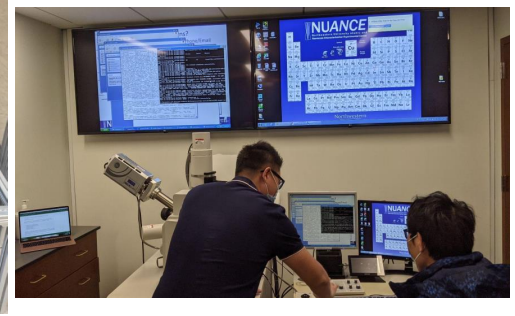
- Giant: NIST's NexusLIMS, open-source – a collaboration
- Same concepts, same principles, similar goals
- Similar framework but infrastructure-dependent
- Don't reinvent the wheel

  Optimize it!
- Birth of Euclid-NexusLIMS-local:

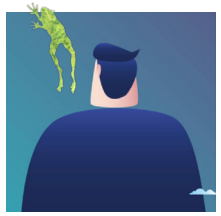
Standing on the giants' shoulders

- Giant: NIST's NexusLIMS, open-source – a collaboration
- Same concepts, same principles, similar goals
- Similar framework but infrastructure-dependent
- Don't reinvent the wheel

  Optimize it!
- Birth of Euclid-NexusLIMS-local:
  - Infrastructure-independent: deploy on a LAN server!
  - "Local": completely "off the WWW"
  - Enhanced functionalities
  - Added computing support (via *JupyterHub* + *Slurm*)



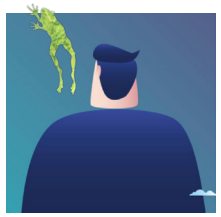
Leaping from the giants' shoulders



- Our mind is "cloudy"

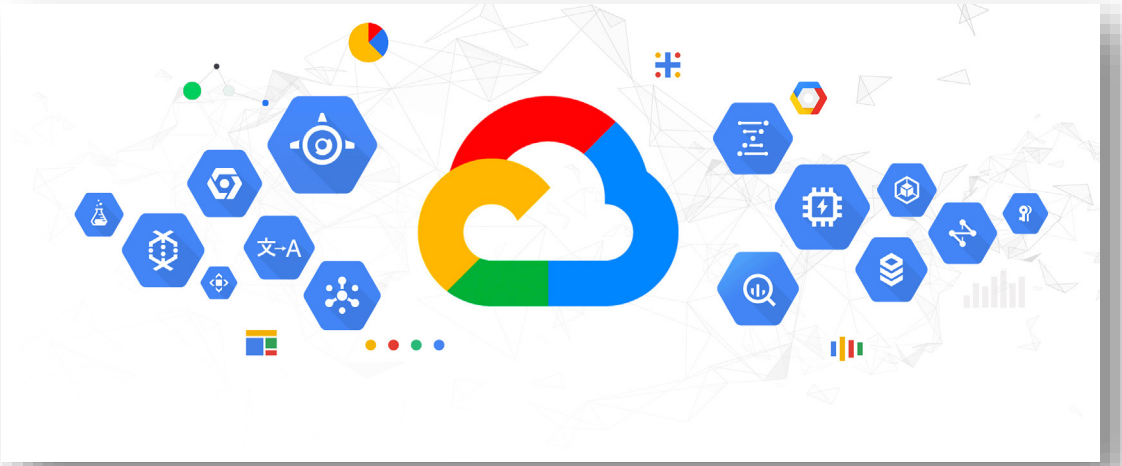
- Cloud computing provides: reliability, scalability, accessibility, plasticity
- No need to maintain a server: power failure, network issues, RAID, or even more challenging: "I can't log in please reset my password".
- Google Cloud Platform (GCP)

Leaping from the giants' shoulders

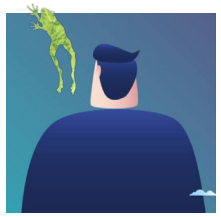


- Our mind is "cloudy"

- Cloud computing provides: reliability, scalability, accessibility, plasticity
- No need to maintain a server: power failure, network issues, RAID, or even more challenging: "I can't log in please reset my password".
- Google Cloud Platform (GCP)
  - An integration of "you name it" services
  - Convenient for Eucliders (G Suite users)
  - Great Identity and Access Mgmt (IAM)

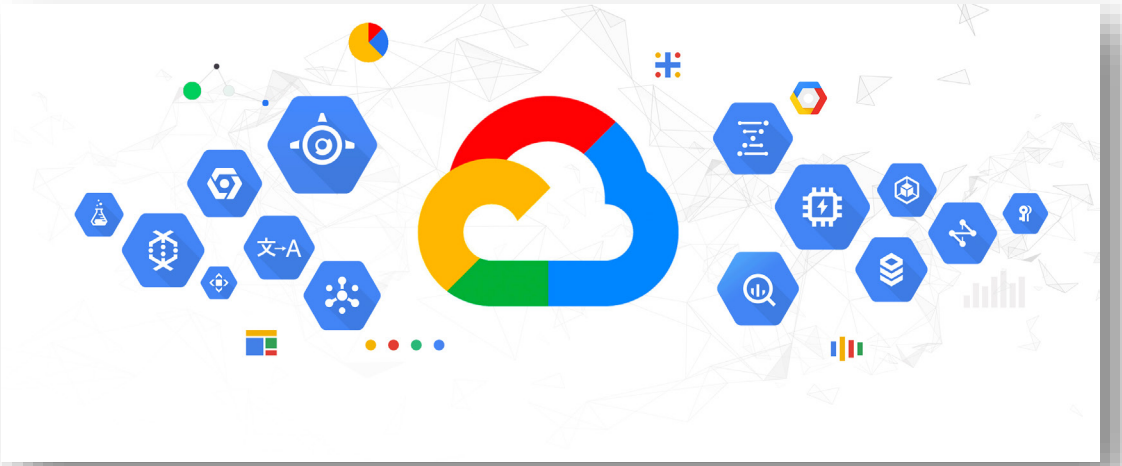


Leaping from the giants' shoulders

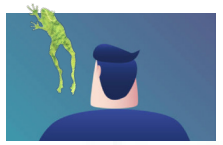


- Our mind is "cloudy"

- Cloud computing provides: reliability, scalability, accessibility, plasticity
- No need to maintain a server: power failure, network issues, RAID, or even more challenging: "I can't log in please reset my password".
- Google Cloud Platform (GCP)
  - An integration of "you name it" services
  - Convenient for Eucliders (G Suite users)
  - Great Identity and Access Mgmt (IAM)
  - Everyone likes to Google



Leaping from the giants' shoulders



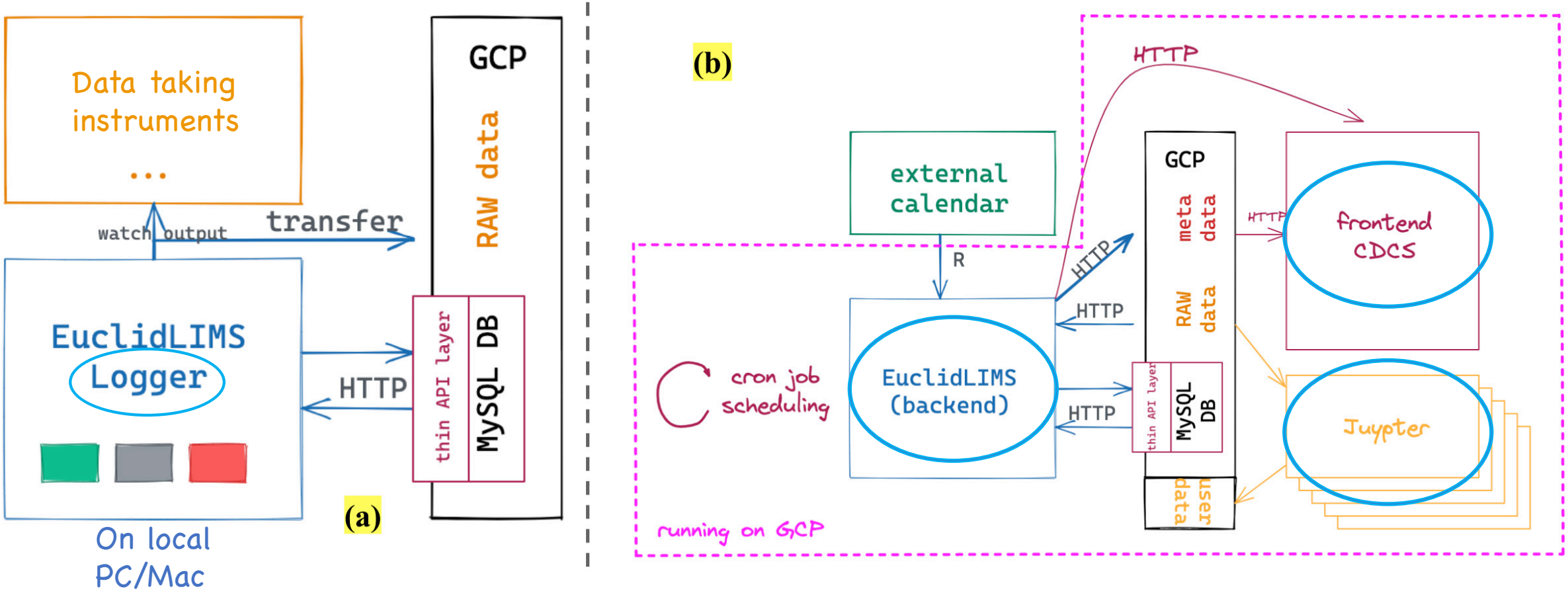
euclid LABS

- (
- (
- |
- (

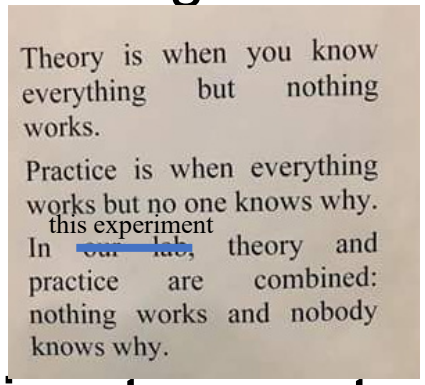
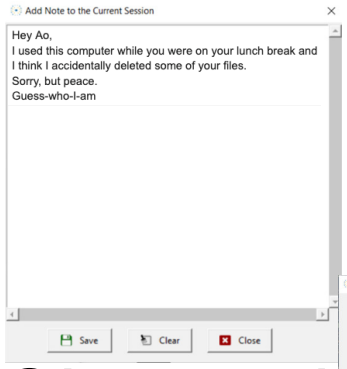
– Everyone likes to Google

The cloud-oriented EuclidLIMS

- Flowchart:

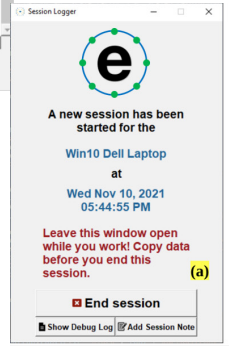


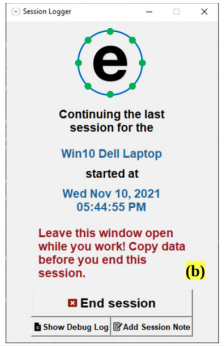
First component: *logger*

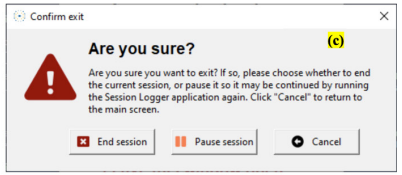
- It's like a punch card machine!
- Well… Similar:
  - It records session time;
  - PC and person names that generated the data;
  - Session notes: not  but 
  - Uploads files to GCP
- **Start** and **end** a session to create an SQL record
- Pause if on a break
- OK if PC has no WWW access:
  - Logger-instrument and Logger-hub

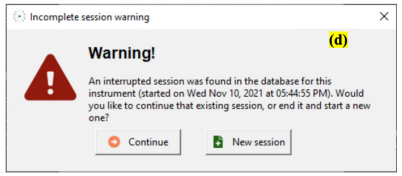
Python 3.4; Tkinter; XML parser:

Miniature to fit any computer!

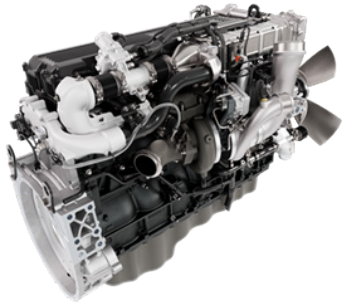


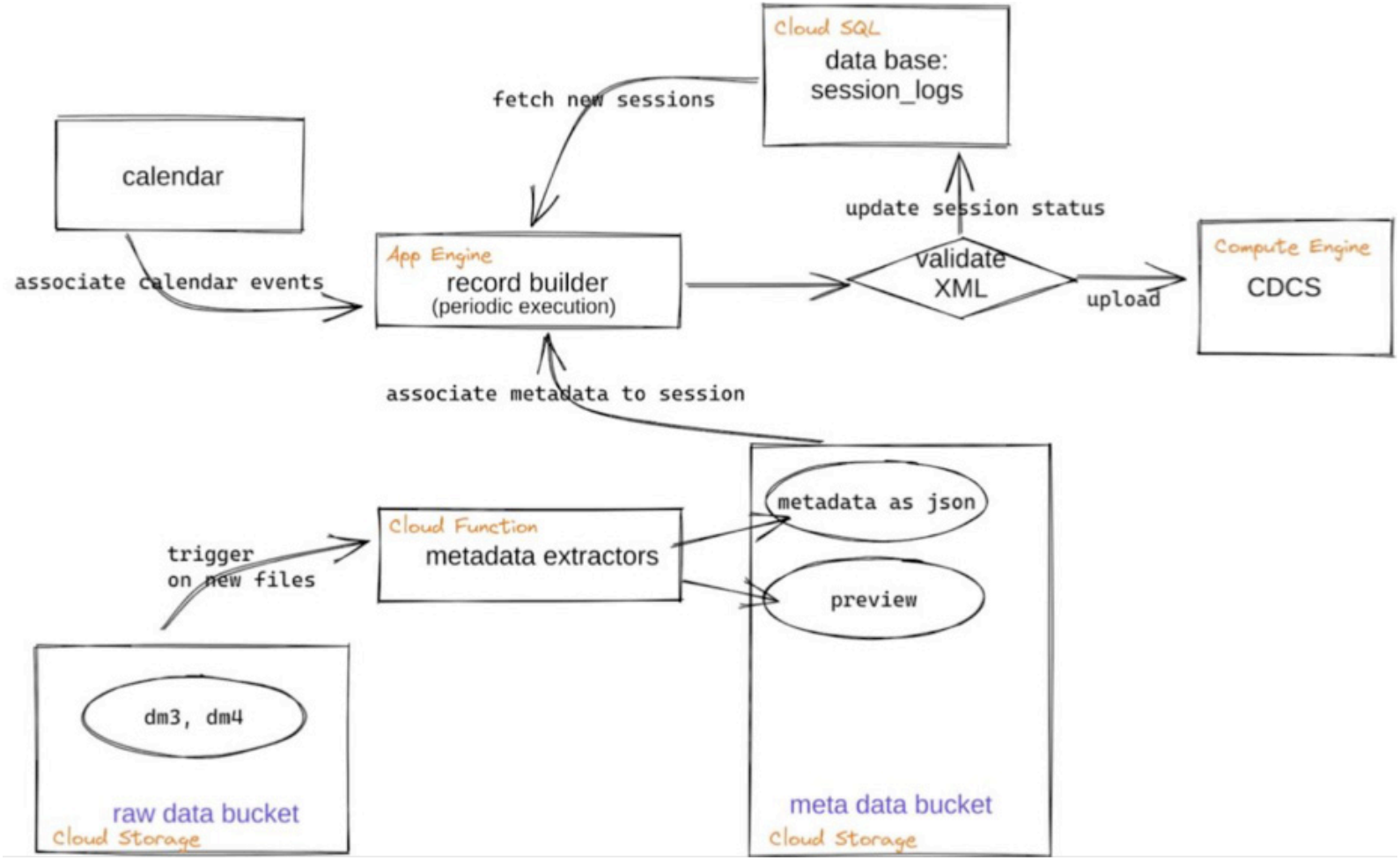




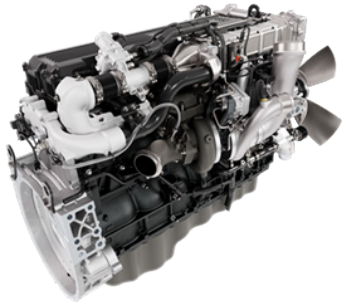


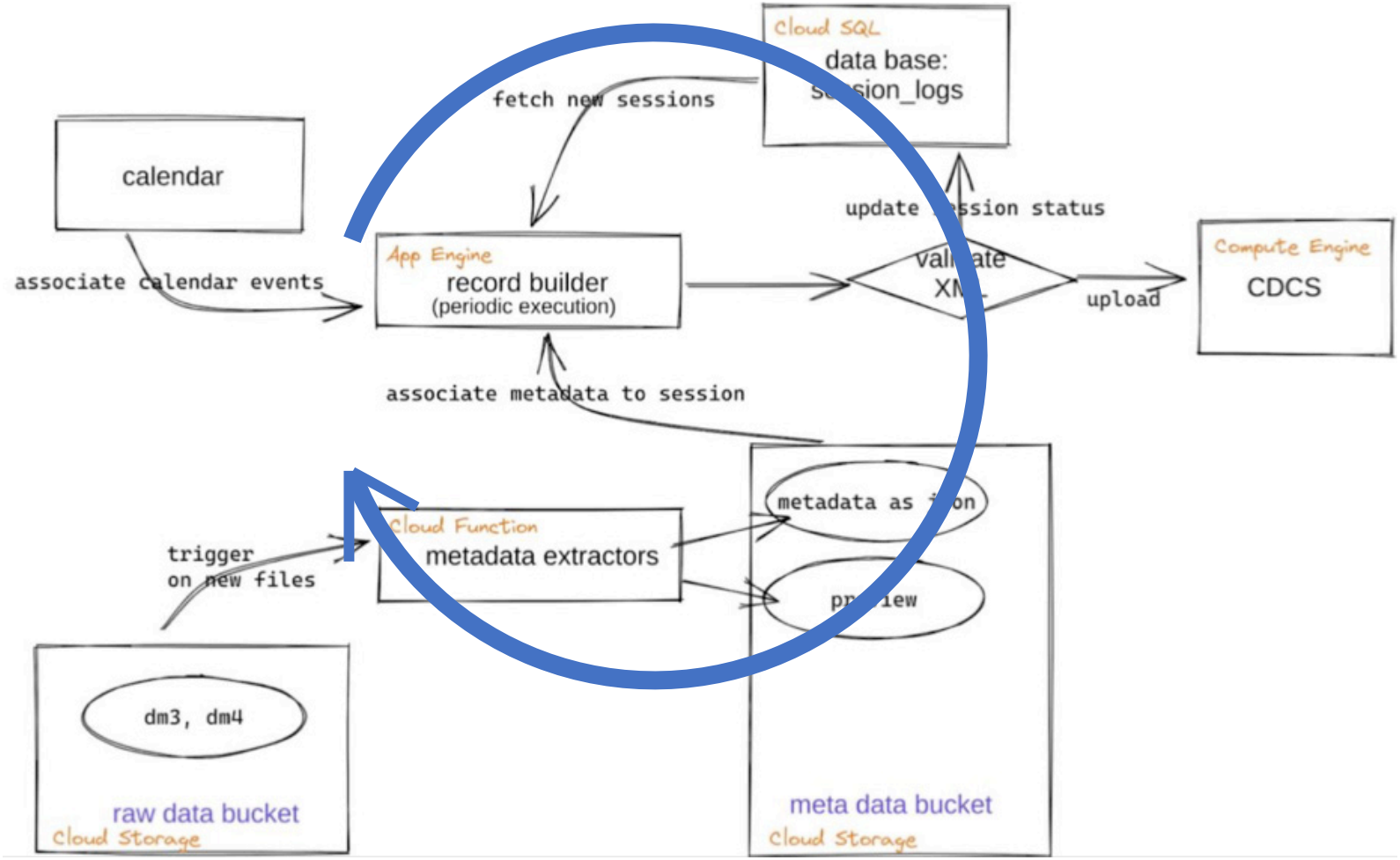
Second (core) component – backend





Second (core) component – backend



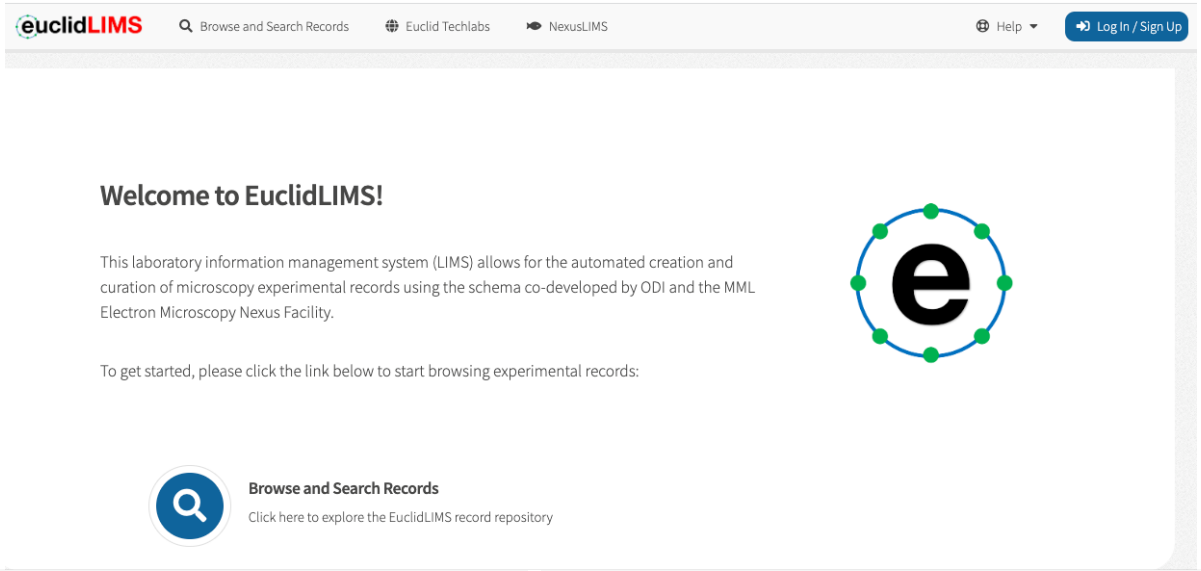


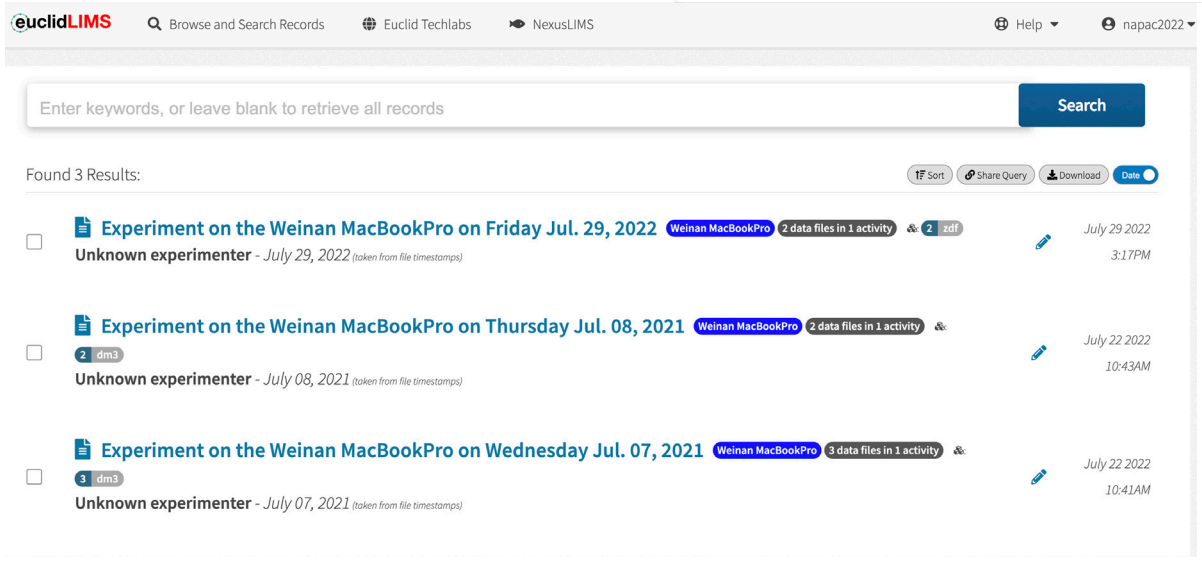
The human part – frontend

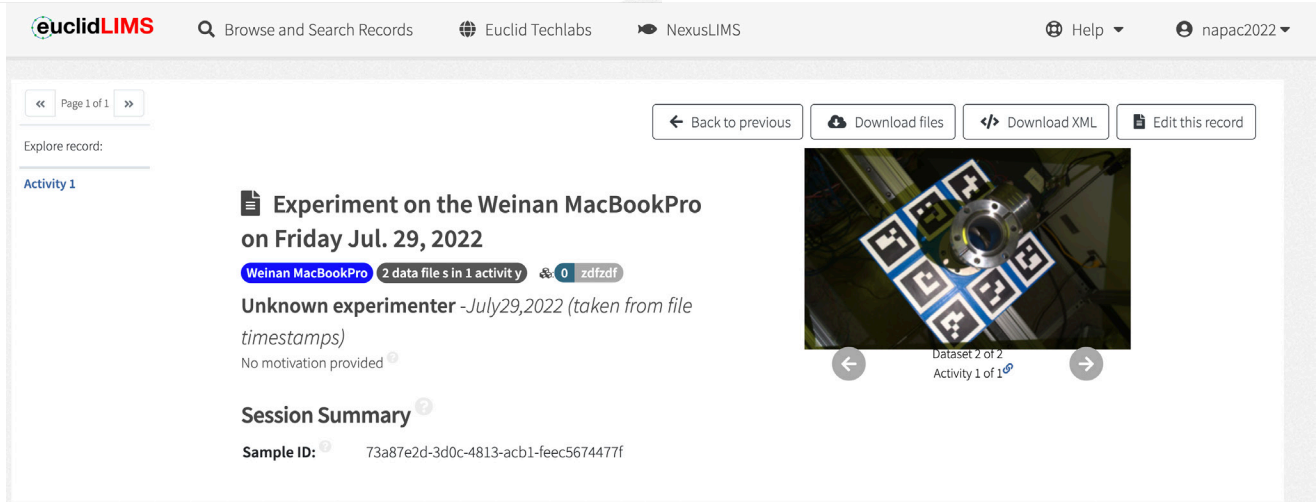
Powered by CDCS, which is developed by the NIST team

euclidlims.beamphysics.com
Username: napac2022
Password: Test#test2022

(Temporary site, available till 08/12/2022)







The Swiss army knife – GCP computing

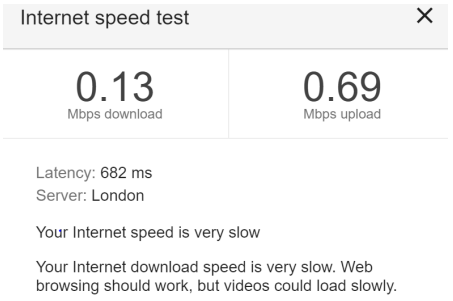
- All data in the cloud storage buckets and accessible, now what?

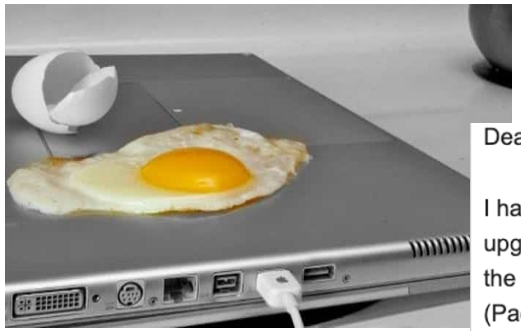
The Swiss army knife – GCP computing

- All data in the cloud storage buckets and accessible, now what?
- Old way: download to the local storage (laptop or server), then analyze.

The Swiss army knife – GCP computing

- All data in the cloud storage buckets and accessible, now what?
- Old way: download to the local storage (laptop or server), then analyze.



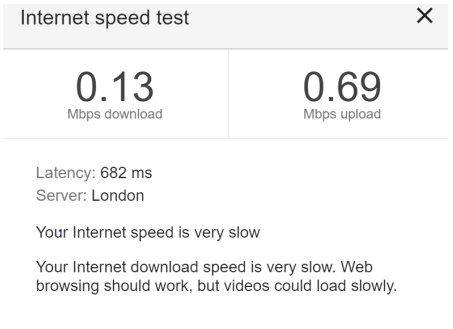


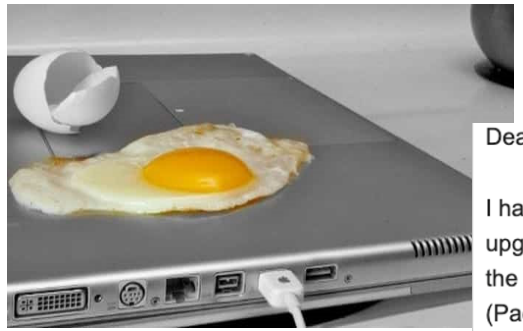
Dear [redacted] Users:

I have just learned that due to new safety procedures required during this period, the electrical system upgrade scheduled for the weekend of June 27-28 will require an additional day in order to complete the work. Therefore we will begin shutting down [redacted] resources on **Thursday, June 25** at 6 am (Pacific time) rather than Friday as originally planned. [redacted] systems will be brought back online on Monday, June 29 by the end of the day as originally scheduled.

The Swiss army knife – GCP computing

- All data in the cloud storage buckets and accessible, now what?
- Old way: download to the local storage (laptop or server), then analyze.





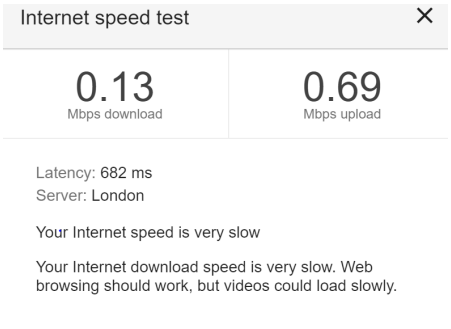
Dear [redacted] Users:

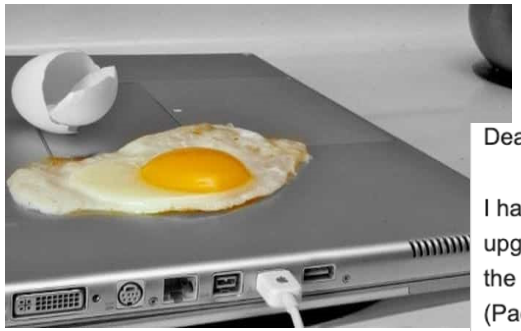
I have just learned that due to new safety procedures required during this period, the electrical system upgrade scheduled for the weekend of June 27-28 will require an additional day in order to complete the work. Therefore we will begin shutting down [redacted] resources on **Thursday, June 25** at 6 am (Pacific time) rather than Friday as originally planned. [redacted] systems will be brought back online on Monday, June 29 by the end of the day as originally scheduled.

- New and smarter way: do it on the cloud
  - Google Kubernetes Engine (GKE): auto-scaling, auto-load-balancing, node pools
  - JupyterHub on GKE: spawned pods for users, convenient scripting, IAM incorporated.

The Swiss army knife – GCP computing

- All data in the cloud storage buckets and accessible, now what?
- Old way: download to the local storage (laptop or server), then analyze.



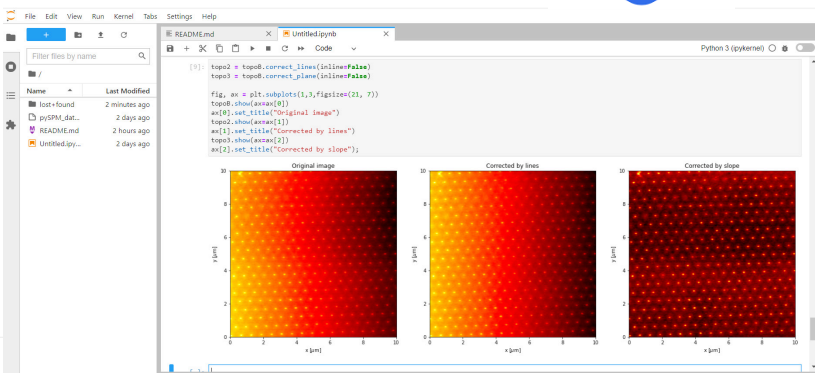


Dear [redacted] Users:

I have just learned that due to new safety procedures required during this period, the electrical system upgrade scheduled for the weekend of June 27-28 will require an additional day in order to complete the work. Therefore we will begin shutting down [redacted] resources on **Thursday, June 25** at 6 am (Pacific time) rather than Friday as originally planned. [redacted] systems will be brought back online on Monday, June 29 by the end of the day as originally scheduled.

- New and smarter way: do it on the cloud
  - Google Kubernetes Engine (GKE): auto-scaling, auto-load-balancing, node pools
  - JupyterHub on GKE: spawned pods for users, convenient scripting, IAM incorporated.

```
jovyan@jupyter-admin:~$ gsutil ls gs://nexuslims_example-datafiles/Weinan-MBP16
gs://nexuslims_example-datafiles/Weinan-MBP16/image_3d_220106165017.zdf
gs://nexuslims_example-datafiles/Weinan-MBP16/image_3d_220106170220.zdf
jovyan@jupyter-admin:~$ gsutil cp gs://nexuslims_example-datafiles/Weinan-MBP16/image_3d_*.zdf ./
Copying gs://nexuslims_example-datafiles/Weinan-MBP16/image_3d_220106165017.zdf...
Copying gs://nexuslims_example-datafiles/Weinan-MBP16/image_3d_220106170220.zdf...
- [2 files][ 42.1 MiB/ 42.1 MiB]
Operation completed over 2 objects/42.1 MiB.
```



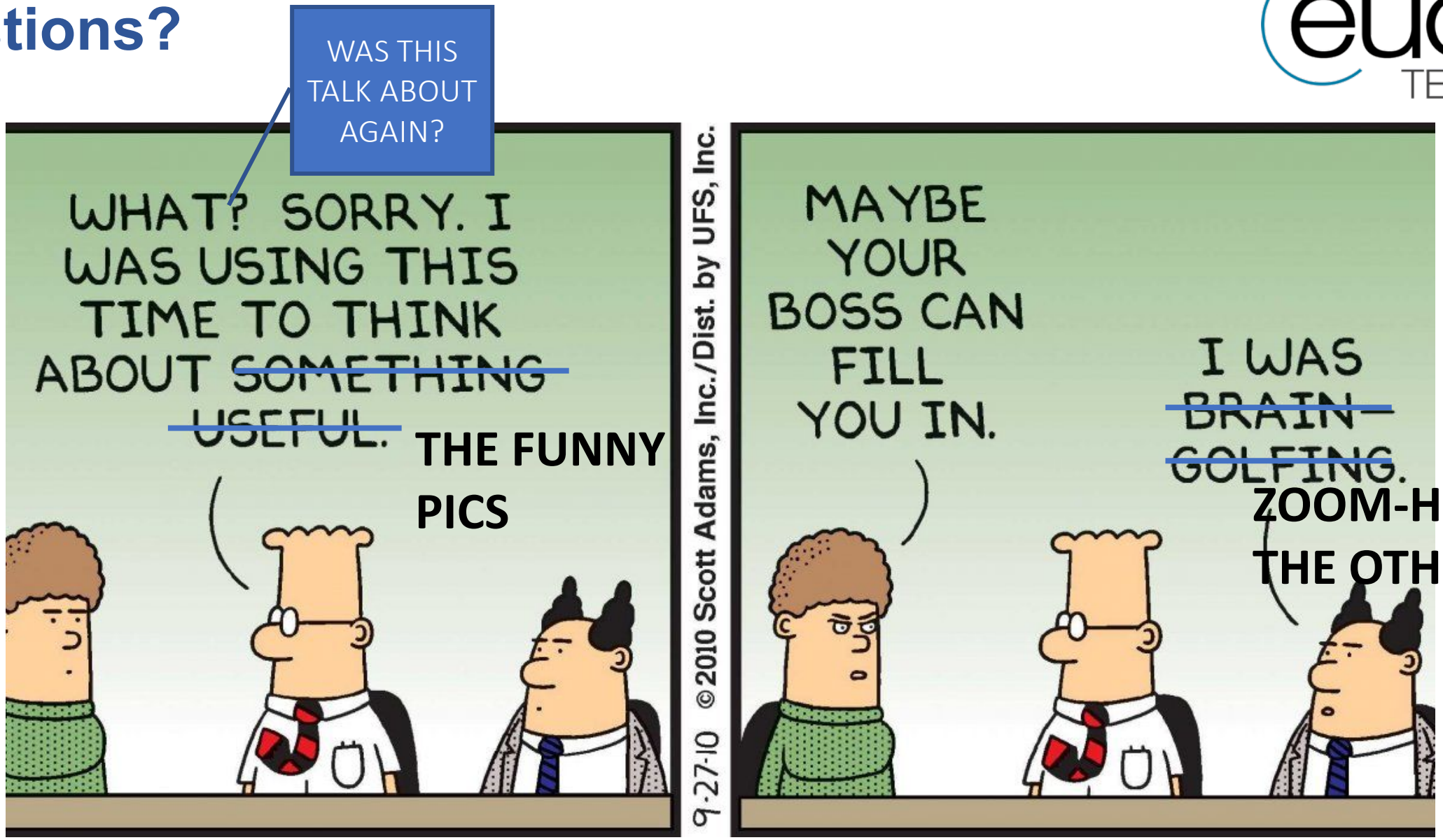
Conclusion

- EuclidLIMS, a cloud platform-based data management and processing software framework, is here.
  - A local version, Euclid-NexusLIMS-local, which has similar principles is available
- EuclidLIMS is easy to maintain and deploy.
- EuclidLIMS does not depend on a local data/computing server.
- EuclidLIMS allows data operations to be done virtually anywhere.
- EuclidLIMS frontend can be customized to support all file formats.
- EuclidLIMS is “non-invasive”: after deployment, users hardly feel any difference other than using the logger.
- **EuclidLIMS brings you joy (at least during this talk). If not, the conclusion is that we (at least I) failed.**

Acknowledgement and Appreciation

- All Eucliders!
  - Particularly Weinan for hard coding & making the EuclidLIMS possible
- NIST collaborators: J. Lau, J. Taillon
- Northwestern customers: R. Reis, L. Bartolo
- Sincere thanks to:
  - NAPAC organizing committee
  - DOE Office of Science
- **Most importantly thanks to the audience today!**

Questions?


WAS THIS TALK ABOUT AGAIN?
WHAT? SORRY. I WAS USING THIS TIME TO THINK ABOUT ~~SOMETHING USEFUL.~~
THE FUNNY PICS
9-27-10 ©2010 Scott Adams, Inc./Dist. by UFS, Inc.
MAYBE YOUR BOSS CAN FILL YOU IN.
I WAS ~~BRAIN-GOLFING.~~
ZOOM-HEARING THE OTHER TALK