ET site repository

Massimiliano Razzano

(University of Pisa & INFN-Pisa)

On behalf of the WG1

- **Data repository**
  - Have a place to store the data
  - Accessible from the site and other places
  - Doing some interactive analysis

- **Software repository**
  - Standard development tools online
  - Software versioning
  - Developers groups etc

- **Developed on Github**
  - Git-based, (almost) easy to use
  - Interface with CI tools (e.g. CircleCI)
  - Public and (special conditions) private repositories
  - Offers great interface with other services (notebooks, overleaf, etc..)

- **Our sw repo**
  - Created an ET organization (in Github jargon, shared account for group of projects/developer)
  - Everyone in our group can get access (see later)
  - Special bonus: private repositories (under unipi license)
  - Created some test projects, and group of projects

Software repository

- https://github.com/et-sw

Data repository

- **Virtual machine dedicated to our analysis**
- Hosted @ Computing center of the University of Pisa computing center (request by F. Fidecaro & M. Razzano)
- 8 cores ($\rightarrow$ upgraded to 16)
- RAM: 16 Gb $\rightarrow$ upgrated to 32
- Large disk space: 4.6 Tb
- OS: Centos7 64 bit
- Performance can be increased in need, so far so good
- Can be accessed remotely

- **Data space (/etrepo)**
  - Temporary data-sandbox (writable by users), used to temp store input data
  - data-sites: permanent data storage (read only)
    - Automatic script to move from sandbox → data-sites
  - Under sandbox and data-sites contains subdir named as sites
    - Names so far present and agreed with Jan: sosenattos, terziet
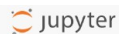   - Work to re-organize datasets

- **ET software (/etrepo/et-software)**
  - Meant to host general-purpose sw and ET-specific software

- **User work directories** (now under /etrepo/etuser-work/$USER)
  - Workspace for each user.
  - Use this space for your analysis, not your /home (here you have more space!)

Interactive notebooks

- **What about looking at the data?**
  - We have set up a JupyterLab server to allow analysis
  - Each user can access the interface
  - Can run Linux terminal Python IDE, and of course Python Notebooks
  - Using SSL encryption-based protocol

Accessible at https://etrepo.df.unipi.it:8000
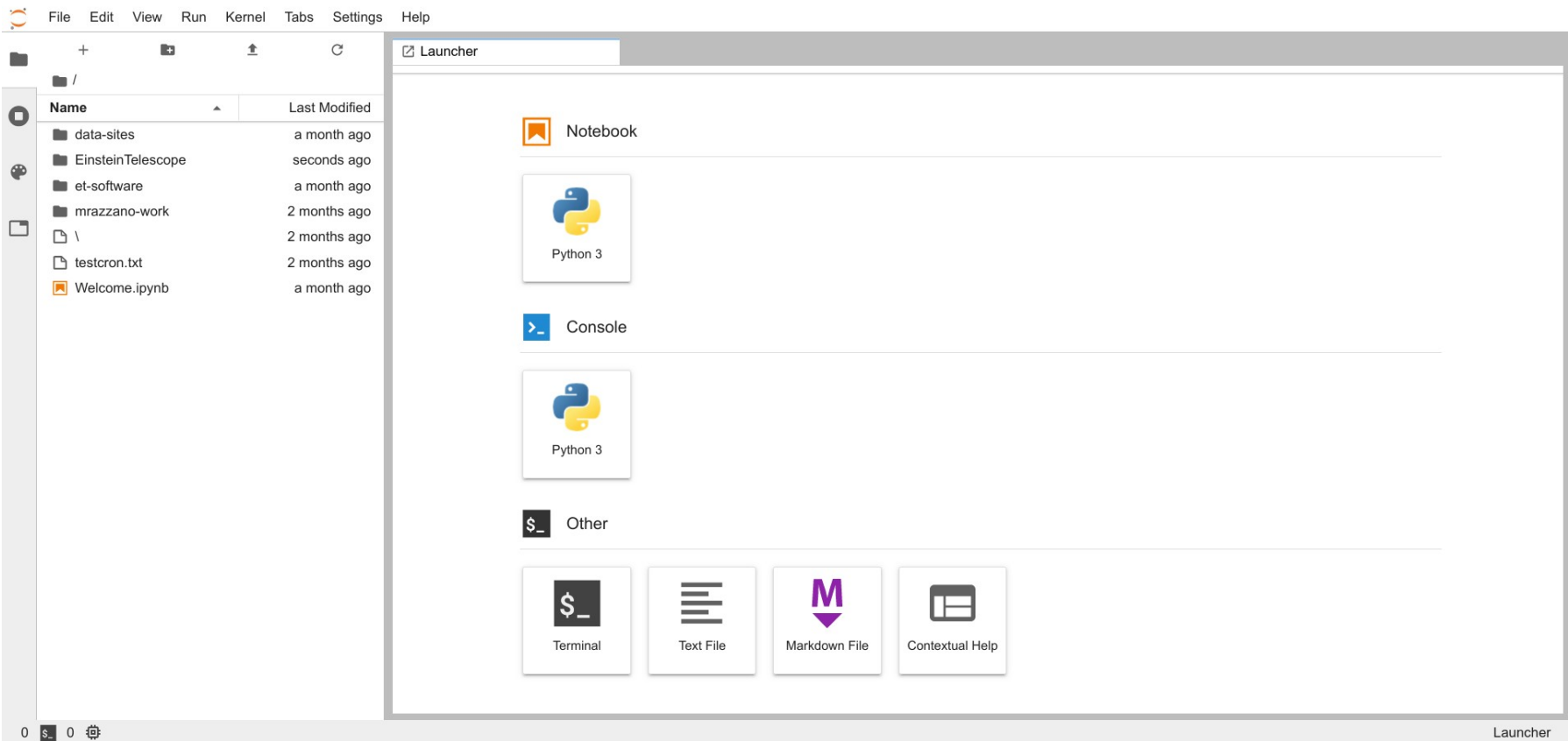
Interactive notebooks

File Manager

Workspace

Interactive notebooks

Welcome (sample) notebook

- **Automatic accounting system**
- Fill the form at https://forms.gle/n2MpK1cg2Mxfdz1o8 (sent around by email, will send again and put in a more convenient place)
  - Scripts will take your requests, make an account for you, set up directories and send an email to you with username and temp pwd
  - Usernames as name+surname → nsurname
  - Latency half hour, can be longer in some cases
  - Will also add github account to et-sw (I am doing it, so it will take some extra time)
  - In case of problems, let me know!

Documentation at https://tinyurl.com/y4ukh98d

Howto: access and upload data

- **Access**
  - Now via ssh and scp
  - Implement (if useful) web-based

- **Upload data**
  - Scp to the data-sandbox/sitename
  - The system will copy the files to the final storage space (data-sites)

- **Use JupyterLab interface**
  - Go to https://etrepo.df.unipi.it
  - Login with your username and pwd
  - Write code and enjoy

Conclusions

- **Repositories**
  - Software repository ready and working in Github
  - Data repository upgraded and running
  - Automatic accounting system
- **Status**
  - Repositories are up and running
  - Data already copied (Luca, Jan, Carlo)
- **Future perspectives**
  - Update the organization of material
  - Upgrade the packages on the Jupyter nb (e.g. obspy)