



EURO

Source control management with Git at Gitlab, Github and Bitbucket

Leon Kos, University of Ljubljana

12 April 2022, Zoom meeting



Source code management (SCM) with Git provides support for versioning with branching and merging in a collaborative work. Git as distributed SCM is usually connected to a central Git server that provides web functionality for source review, pull requests and integration with other services such as continuous integration (CI) and code documentation. In this seminar we will take a look into Git development process and popular servers and integrated services (CI, Read the docs). How to Git version large files in HPC environment along with code and documentation will be discussed from practical and data provenance viewpoint.

Git – daily work from command line

Setup

1. Add ~/.ssh/id_rsa.pub to the list of SSH keys on the portal
2. `git clone ssh://git@git.../...git`
3. `git branch -r`
4. `git checkout develop`
5. `git add .gitignore`

Develop

1. `git status`
2. `git add file ...`
3. `git pull`
4. `git status`
5. `git commit -m "Short description of changes"`
6. `git push`

Gitlab



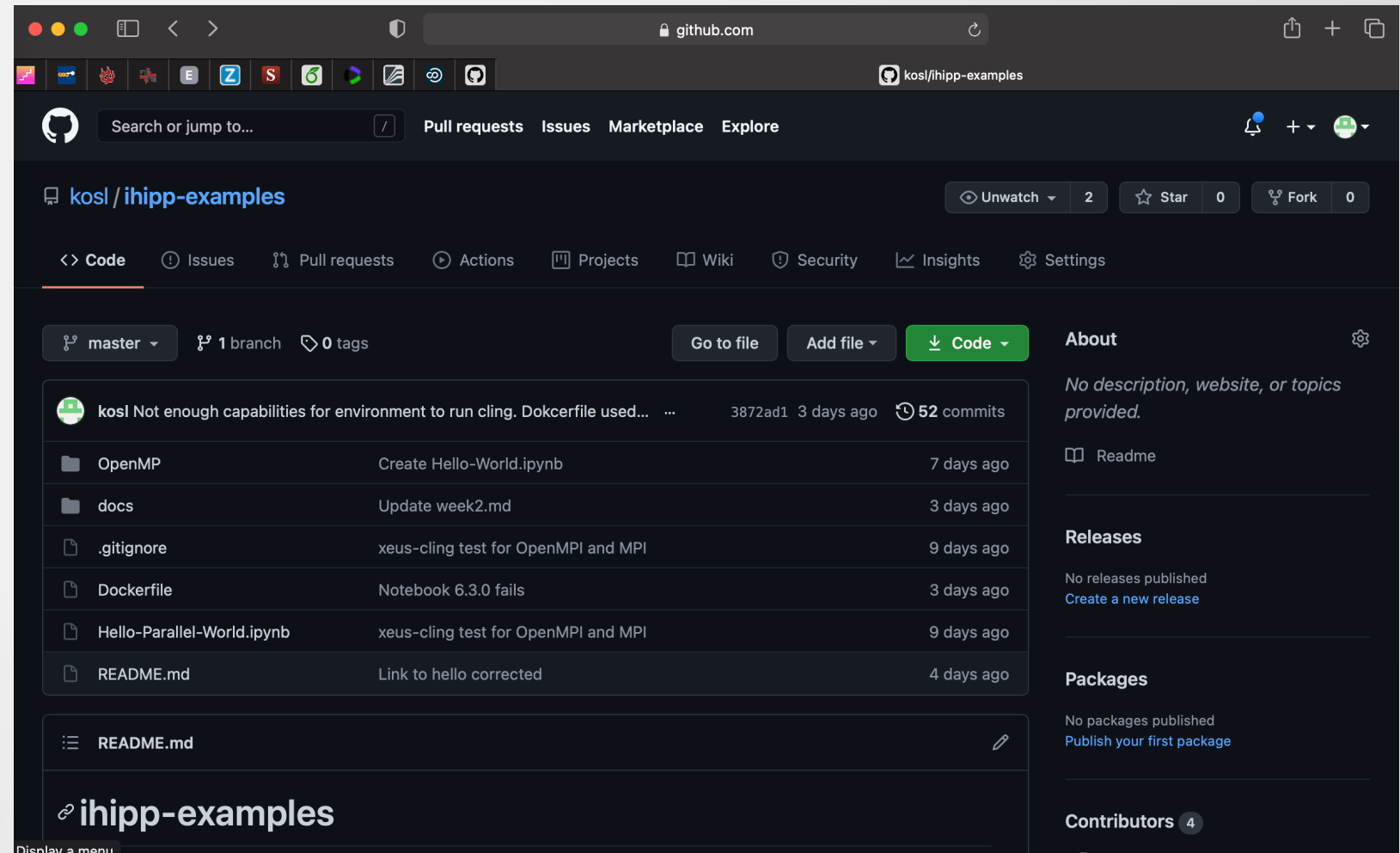
- Good for institutional hosting by deploying Community Edition
gitlab.com/gitlab-org/gitlab
- Not that open anymore (Free trials)
<https://gitlab.com/explore>
- Integrated Wiki, Snippets, Analytics
- Continuous Integration with Kubernetes
- Demo of projects at
<https://gitlab.eudat.eu/eudat-prace-2019>

The screenshot shows the GitLab web interface for a project named 'computing-architecture'. The left sidebar contains navigation links: Project overview, Details, Activity, Releases, Repository, Issues (0), Merge Requests (0), CI / CD, Operations, Analytics, Wiki, Snippets, Members, and a Collapse sidebar button. The main content area displays project statistics (4 Commits, 1 Branch, 0 Tags, 143 KB Files, 143 KB Storage) and a list of recent commits. A commit by Leon Kos is highlighted, showing the addition of 'miniconda.yaml'. Below the commit list, there are buttons for adding various files and setting up CI/CD.

Name	Last commit	Last update
.gitignore	MPI in Python	1 year ago
README.md	Add miniconda.yaml	1 year ago
conda-data_analysis-env.yaml	Add miniconda.yaml	1 year ago
conda-rtd-env.yaml	Add miniconda.yaml	1 year ago
hello-mpi.py	MPI in Python	1 year ago
hello-mpi.sbatch	MPI in Python	1 year ago

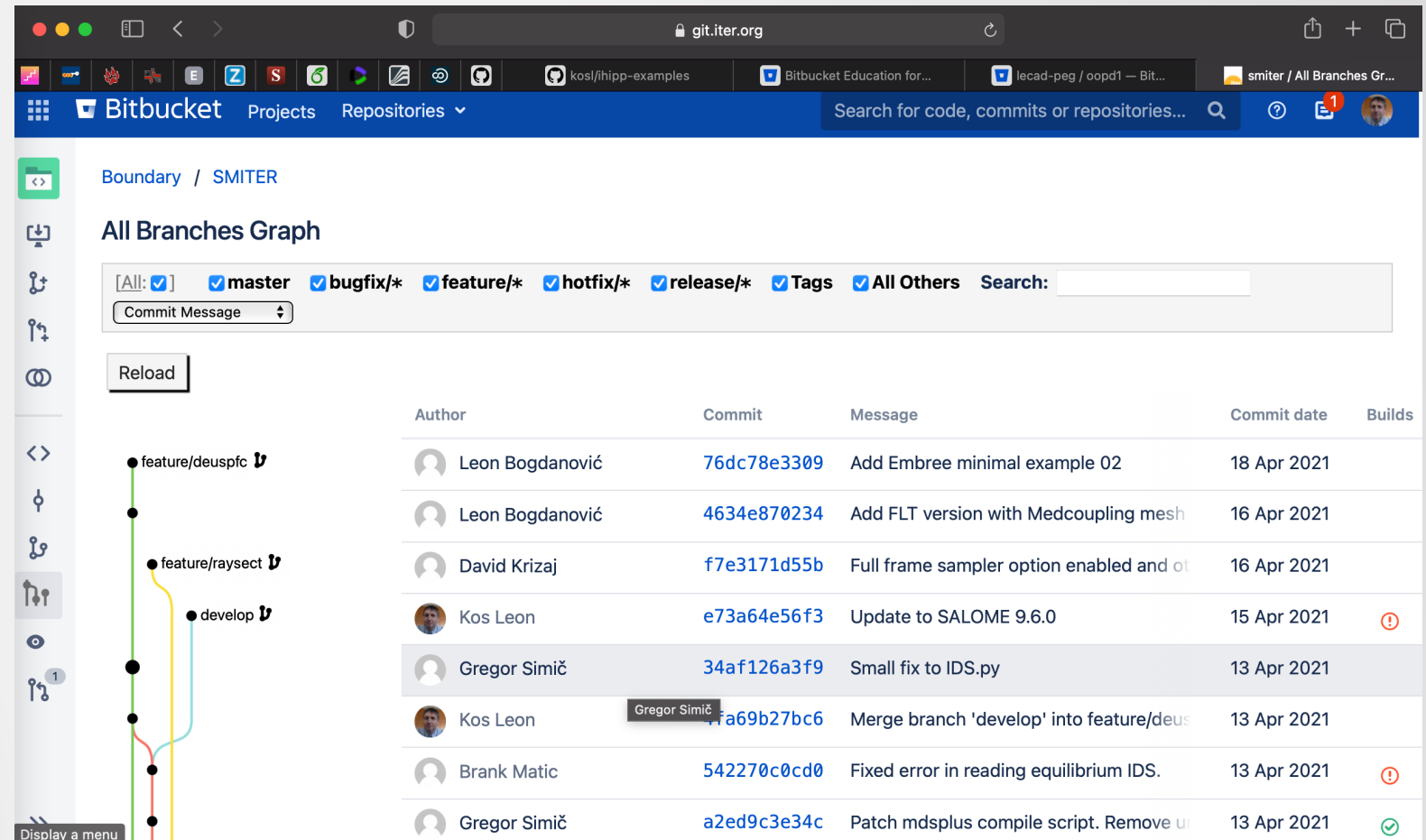
Github

- Open source hosting friendly (e.g. <https://github.com/torvalds/linux>)
- Integration to many other websites (Overleaf, Binder, CircleCI, ...)
- Private repositories, Gists
- Not that great interface
- Many source code rendering and integrated editors
- Github desktop for beginners
- *.github.io websites
- Actions
- Demo of project at <https://github.com/kosl/ihipp-examples>



Bitbucket

- Commercial hosting only
- Form 2024 only as cloud offering
- Great Web based integration (Confluence tools, Stash, Jira, Bamboo, Servicedesk, ...)
- Free for academic hosting
<https://bitbucket.org/product/education> Private repositories
- Limited source code rendering and integrated editors
- Demo of the project at
<https://bitbucket.org/lecad-peg/oopd1/>
- Demo of Bamboo CI agent



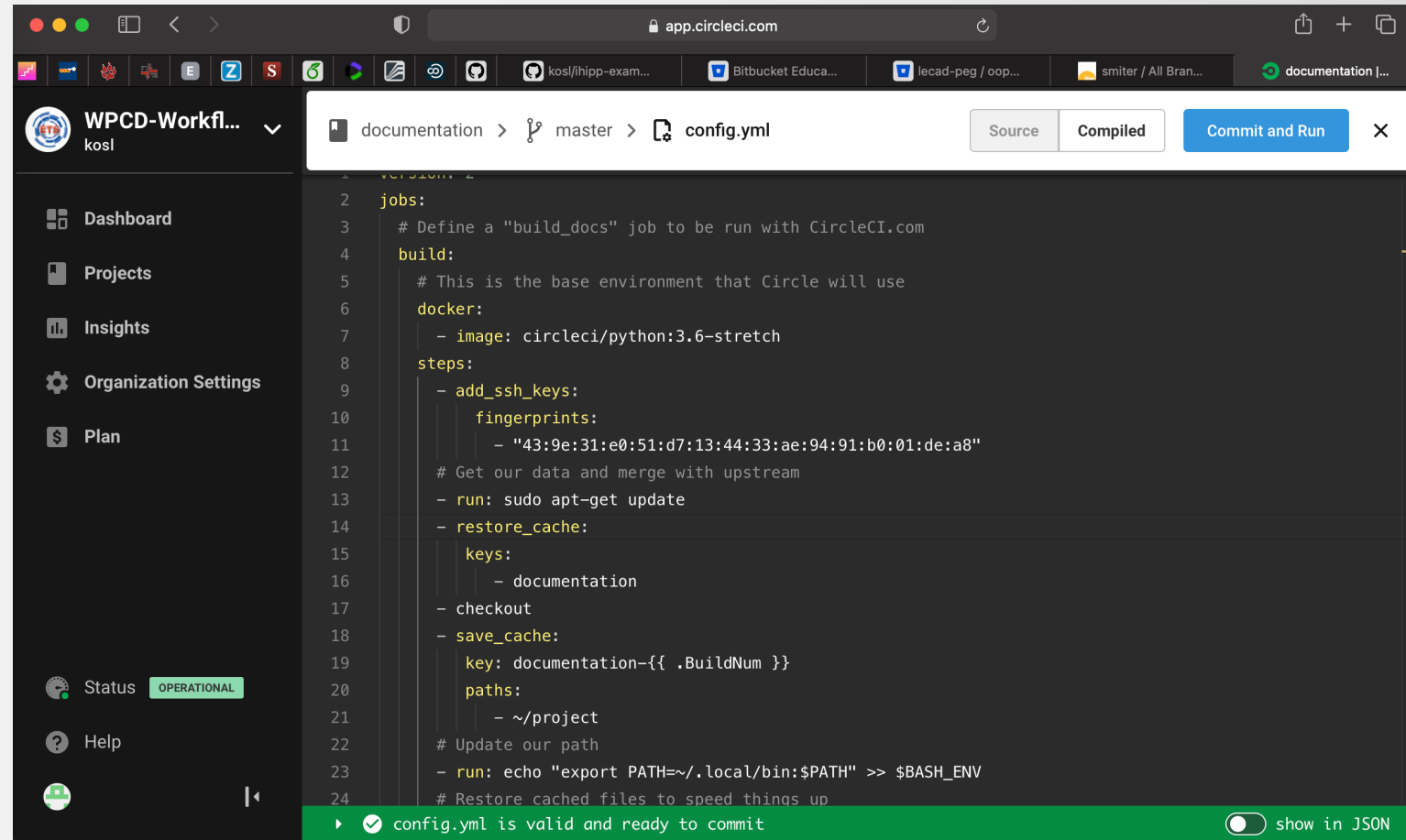
Author	Commit	Message	Commit date	Builds
Leon Bogdanović	76dc78e3309	Add Embree minimal example 02	18 Apr 2021	
Leon Bogdanović	4634e870234	Add FLT version with Medcoupling mesh	16 Apr 2021	
David Krizaj	f7e3171d55b	Full frame sampler option enabled and ot	16 Apr 2021	
Kos Leon	e73a64e56f3	Update to SALOME 9.6.0	15 Apr 2021	!
Gregor Simič	34af126a3f9	Small fix to IDS.py	13 Apr 2021	
Kos Leon	fa69b27bc6	Merge branch 'develop' into feature/deus	13 Apr 2021	
Brank Matic	542270c0cd0	Fixed error in reading equilibrium IDS.	13 Apr 2021	!
Gregor Simič	a2ed9c3e34c	Patch mdsplus compile script. Remove u	13 Apr 2021	✓

Continuous Integration (Circle CI)

- Testing and building automatism
 - Testing on different compilers and platforms
 - Build Releases
 - Documentation
- Free minutes can be exhausted unless caches are used for rebuild

Demo of CircleCI

<https://app.circleci.com/projects/github/WPCD-Workflows/documentation/config/>



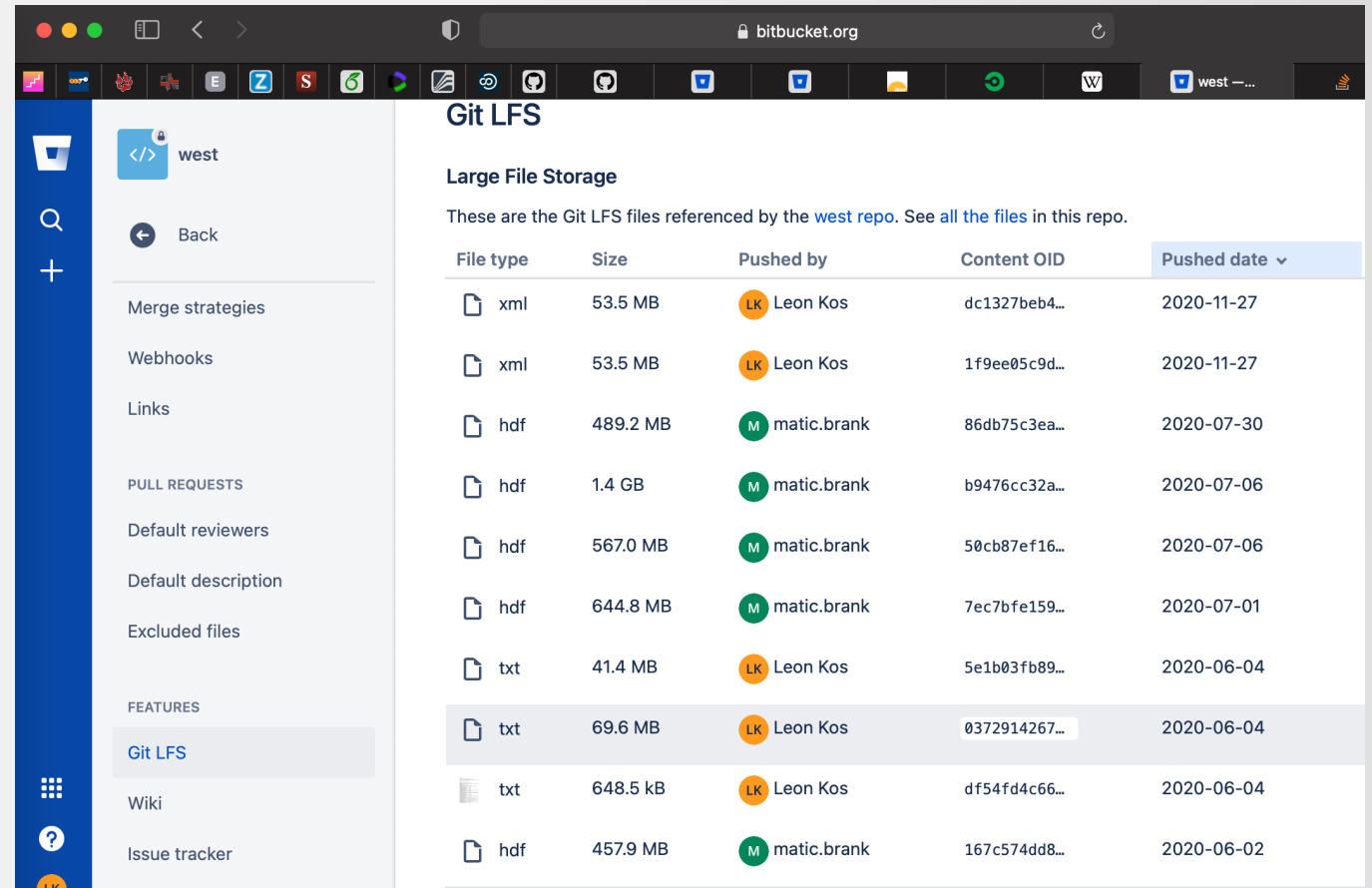
The screenshot displays the CircleCI web interface in a browser. The address bar shows 'app.circleci.com'. The left sidebar contains navigation links: Dashboard, Projects, Insights, Organization Settings, and Plan. The main content area shows the configuration file 'config.yml' for the 'documentation' project. The file is edited in a dark-themed editor. The configuration defines a 'build_docs' job using a Docker image 'circleci/python:3.6-stretch'. It includes steps for adding SSH keys, updating packages, restoring caches, and setting up the environment. A green status bar at the bottom indicates 'config.yml is valid and ready to commit'.

```
1 version: 2
2 jobs:
3   # Define a "build_docs" job to be run with CircleCI.com
4   build:
5     # This is the base environment that Circle will use
6     docker:
7       - image: circleci/python:3.6-stretch
8     steps:
9       - add_ssh_keys:
10         fingerprints:
11           - "43:9e:31:e0:51:d7:13:44:33:ae:94:91:b0:01:de:a8"
12       # Get our data and merge with upstream
13       - run: sudo apt-get update
14       - restore_cache:
15         keys:
16           - documentation
17       - checkout
18       - save_cache:
19         key: documentation-{{ .BuildNum }}
20         paths:
21           - ~/project
22       # Update our path
23       - run: echo "export PATH=~/.local/bin:$PATH" >> $BASH_ENV
24       # Restore cached files to speed things up
```

config.yml is valid and ready to commit

Git Large File Support (LFS)

- Install from <https://git-lfs.github.com>
 - git lfs install
 - git lfs track "*.stp"
 - git add .gitattributes
 - git lfs pull
- Limited storage size (5GB) on public repositories. Use [git filter-repo](#) to remove the objects from the repository.
- No site-wide cache!
 - Increased local storage
 - bandwidth



The screenshot shows the Bitbucket web interface for a repository named 'west'. The left sidebar contains navigation links for 'Merge strategies', 'Webhooks', 'Links', 'PULL REQUESTS', 'Default reviewers', 'Default description', 'Excluded files', 'FEATURES', 'Git LFS', 'Wiki', and 'Issue tracker'. The main content area is titled 'Git LFS' and 'Large File Storage'. It states: 'These are the Git LFS files referenced by the west repo. See [all the files](#) in this repo.' Below this is a table listing the files.

File type	Size	Pushed by	Content OID	Pushed date
xml	53.5 MB	LK Leon Kos	dc1327beb4...	2020-11-27
xml	53.5 MB	LK Leon Kos	1f9ee05c9d...	2020-11-27
hdf	489.2 MB	M matic.branc	86db75c3ea...	2020-07-30
hdf	1.4 GB	M matic.branc	b9476cc32a...	2020-07-06
hdf	567.0 MB	M matic.branc	50cb87ef16...	2020-07-06
hdf	644.8 MB	M matic.branc	7ec7bfe159...	2020-07-01
txt	41.4 MB	LK Leon Kos	5e1b03fb89...	2020-06-04
txt	69.6 MB	LK Leon Kos	0372914267...	2020-06-04
txt	648.5 kB	LK Leon Kos	df54fd4c66...	2020-06-04
hdf	457.9 MB	M matic.branc	167c574dd8...	2020-06-02

„Externaldata“ Git LFS support

- External repository to push assets
 - http for public hashes
 - ssh for private repos
- Site-wide cache possible
 - One copy only with symlinks
- Github example <https://github.com/WPCD-Workflows/documentation/blob/master/Makefile>
- GForge SSH Kerberos example <https://gforge-next.eufus.eu/#/project/publicdb/scm/file/Makefile>
- SMITER demo

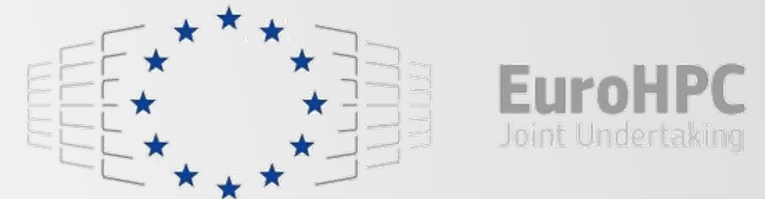
```
258 # If the asset already exist the we don't want to change its timestamp as
259 # this may impact other's (re)builds. Therefore MD5 and timestamps needs
260 # to be backdated in this case.
261 % : %.md5-stamp
262     if test -f $(ASSETDIR)/$(shell cat $^); \
263     then touch -r $(ASSETDIR)/$(shell cat $^) -d "-1 minute" $^; \
264     touch -r $^ -d "-1 minute" $(patsubst %.md5-stamp, %.md5, $^); \
265     else make $(ASSETDIR)/$(shell cat $^); fi
266     $(ASSET_LINK) $(ASSETDIR)/$(shell cat $^) $@
267
268 %.md5-stamp : %.md5
269     cat $^ > $@
270
271 % : %.zip
272     unzip -q -d $@ $^
273
274 # Downloads asset and makes it available to the group for linking to it
275 $(ASSETDIR)/% :
276     if ! test -d $(ASSETDIR); \
277     then mkdir -m 3775 -p $(ASSETDIR); \
278     if test "x$(ASSETDIR_GROUP)" != "x"; \
279     then chgrp $(ASSETDIR_GROUP) $(ASSETDIR); \
280     fi; \
281     fi
282     curl --location --fail --insecure \
283     --output $(ASSETDIR)/$(@F) $(EXTERNALDATA)/$(@F)
284     if test $$$(md5sum $(ASSETDIR)/$(@F)|cut -c 1-32) = $(@F) ; then \
285     chmod g+rw $(ASSETDIR)/$(@F); \
286     else \
287     rm -f $(ASSETDIR)/$(@F); \
288     echo "Error: Checksum of $(ASSETDIR)/$(@F) incorrect! Removing it."; \
289     exit 1; \
290     fi
```

Restructured text with Sphinx

- Repository should version
 - Source Code
 - Large data
 - Documentation
- Data provenance
- Containers?
- Binder?



Thanks!



This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 951732. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, United Kingdom, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Switzerland, Turkey, Republic of North Macedonia, Iceland, Montenegro