WWU
MÜNSTER

Tour of Jupyter- and BinderHub

Adrian Brömmelsiek

08.10.2019

living.knowledge

ZENTRUM FÜR
INFORMATIONS
VERARBEITUNG

Jupyter notebooks

- may contain live code, equations, visualizations and narrative text
- use an open format
  - shareable
  - integrateable with third party products, e.g. previews in GitHub / GitLab

Jupyter Notebook

- web app for working with Jupyter notebooks
- makes notebooks interactive; great for data analysis and learning
- easily started once installed: `jupyter notebook`
- you setup your own workspace (libraries, data, ...) and work on your own

What if you wanted to share your (non-trivial) workspace?

- notebooks and data: easy
- installed libraries, interpreters/compilers/tools: not that easy
  - you can only assume a python interpreter in an unknown version (probably 3.x)
  - include a list of things you'd have to install to get your notebook running? cumbersome ...
  - install script? shifts complexity from end-user back to you, also difficult to maintain for many platforms
  - docker image? works nicely, but needs power tools
- How about offering your workspace "as a service"?

JupyterHub

- “A multi-user version of the notebook designed for companies, classrooms and research labs”
- under the hood: actually just support structures for the Notebook app
  - dynamically choose which notebook image to run
  - (external) authentication
  - persistent storage (= homes)
  - CPU time, RAM, storage quotas

Demo: WWU JupyterHub

Benefits of running a JupyterHub

- instantly available fine-tuned environments
  - good for hosting lectures and seminars, zero prep time for students
  - no need to debug individual setups
- nice side effect: access to compute resources other than your local machine

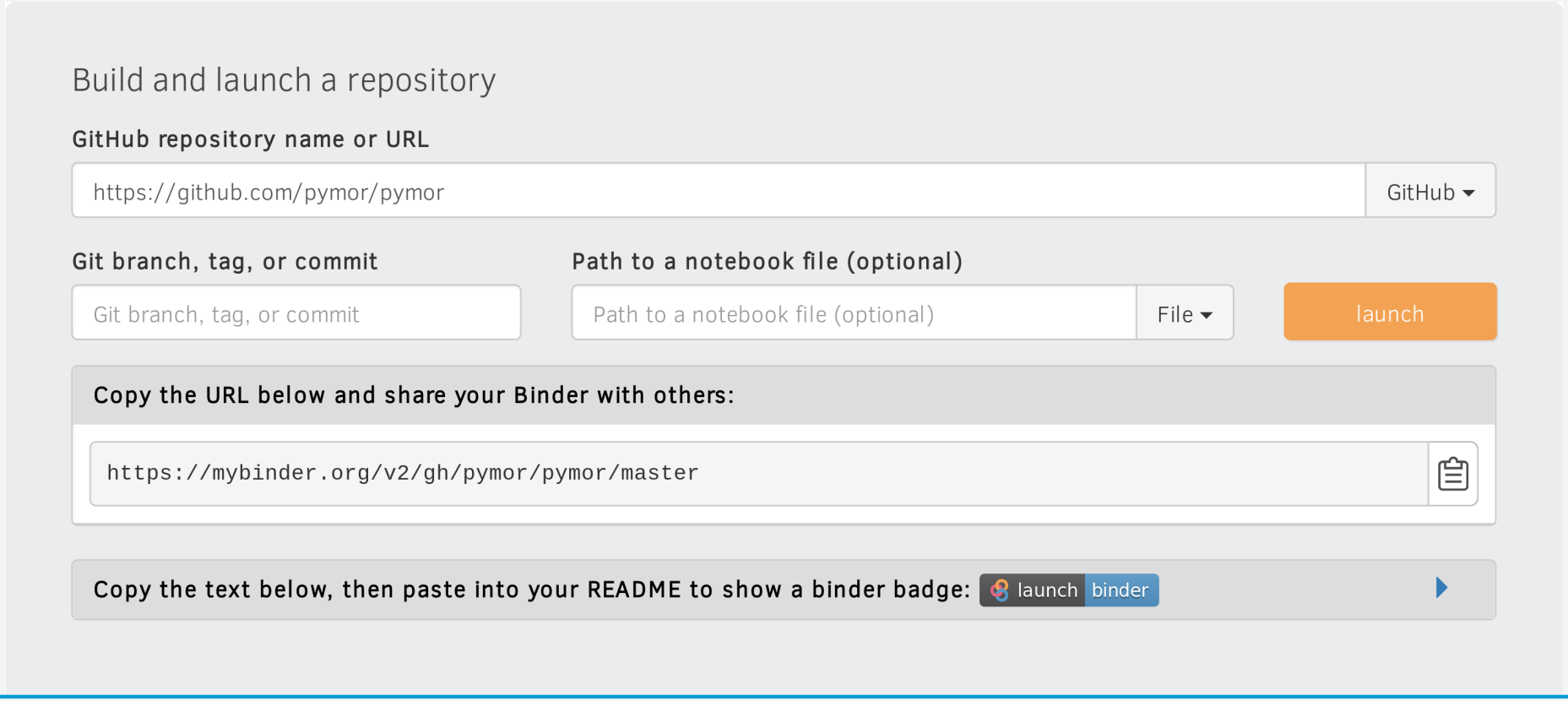
Sharing with strangers

What if you wanted to share your workspace *with people outside your organization*?

- create individual accounts? only possible if you know who you share with, not always possible
- go back to distributing zips and README files? come on!
- JupyterHub for guests?

BinderHub

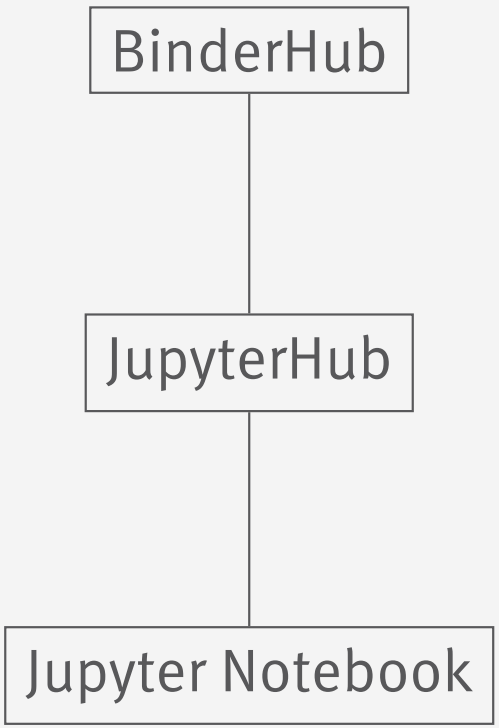
- JupyterHub for guests (authentication is very optional)
- also: on demand image generation!



BinderHub image generation

- relies on `repo2docker`
- tries to guess what could be needed to make your repository contents usable in a JupyterHub
  - scans standard files such as `requirements.txt`, `Pipfile` or `DESCRIPTION`
  - e.g.: if it finds `requirements.txt` file, you're getting a python notebook with these requirements installed
  - some BinderHub-specific files are also looked for
- if you don't want to rely on magic, you may use a `Dockerfile` instead

Overview



"Generalized JupyterHub", provide anybody with arbitrary Notebook environments

"Managed Jupyter Notebook", feature-rich platform for select users. May require beefy machines and an admin(-y person)

Great for working on your own, sharing is possible if your setup is simple enough (YMMV)

How do I get access to any of that?

- Jupyter Notebook: install it yourself
- JupyterHub: make somebody host it for you[1]
- BinderHub: use *mybinder.org*, or make somebody host it

[1] or be ready to invest a good share of your time into doing it yourself

What we have in mind for BinderHub

- security enhancements (subuid mapping, rootless operation)
- non-Docker build systems and execution environments
- *optional* attachment of persistent storage (own data, foreign code or vice versa)
- better performance than mybinder.org (more cores, RAM)
- HPC integration, GPU cores (?)
- configurable lifetimes
- better support for non-notebook build output

Fin

Questions?