PHENIX DAP Update: The site, Zenodo Communities and GitHub Integration

Maxim Potekhin
(BNL, NPPS)
04/23/2020

The DAP website

- Deepali made a contribution to the HBD page
  - Thank you. The pull request process worked fine.
- Improvements in the site mechanics are in the works:
  - Better organization of menus, making it easier to add material - just plug in entries in the menu descriptor instead of tweaking weights in each file
  - Will be ready in a day or two
- A template for the analysis pages - this still requires thought
- Referencing analysis code on the DAP site:
  - See the discussion of **GitHub and Zenodo** in the following slides
  - But also a matter of policy: how much of the code do we publish on the site?
    - cf. the site is public
    - A small fragment will not achieve anything since the goal is reproducibility

Zenodo

- Recap of previous presentations:
  - Low entry cost
  - Metadata support
  - Can store datasets, code or any other digital products
  - Generates and supports official DOIs (i.e. doi.org)
  - Prime candidate for use in PHENIX
- Today’s Zenodo items:
  - Communities
  - Communication with the developers
  - GitHub integration

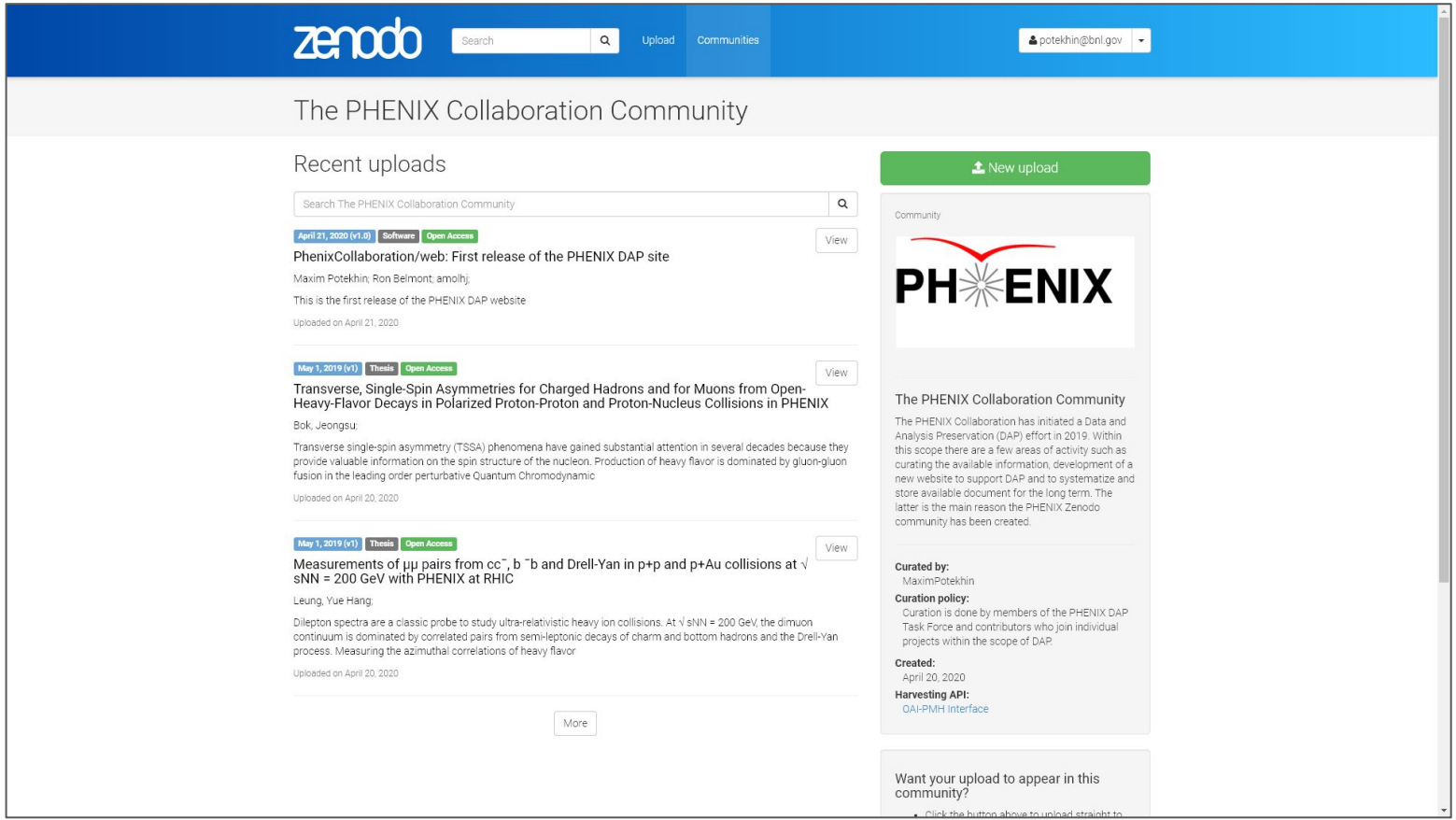
Zenodo Community

- A way to organize material, and to consistently attribute materials to a collaboration/project/experiment - keeping a consistent brand
- An improvement in visibility/discoverability/PR
  - An addition to the already existing metadata query aids in discovery of materials and in practice works better
- Anyone can upload a material to the community which is subject to **curation**
  - The curator gets notified and inspects the submission
  - If accepted, it becomes posted under the community umbrella
  - If rejected, it still remains on Zenodo site but is not officially owned/acknowledged by the community, this is an accordance to the “open access” platform
  - *There is currently one curator per community and there is no easy way to transfer this duty to a different account* (something few people expected) but a fix is on the way according to the lead developer and other team members. Unofficial ETA is late 2020.

Zenodo “PHENIX community” - initial test

- As an initial test (which can be extended) I created a community for the PHENIX Collaboration under the tag “**phenix-collaboration**” similar to our organization name on GitHub
  - Also, “phenix” was already taken by a French bio lab
- As with most other features of Zenodo, the setup was fairly easy

Zenodo Community Example: the PHENIX proto



Zenodo issues and next steps

- I believe it was agreed that we need to start uploading and tagging theses
  - Already started doing this as a part of the test
- Having these materials posted under the PHENIX Community umbrella seems like precisely the right solution
- Are we OK with having me (Maxim) as the curator of the community?
  - I'd rather not distribute my own credentials
- Alternatively, the prototype community can be deleted and a new production account created on Zenodo with multiple people having access to it
  - this creates a separate set of (minor) problems TBD e.g. with the GitHub link

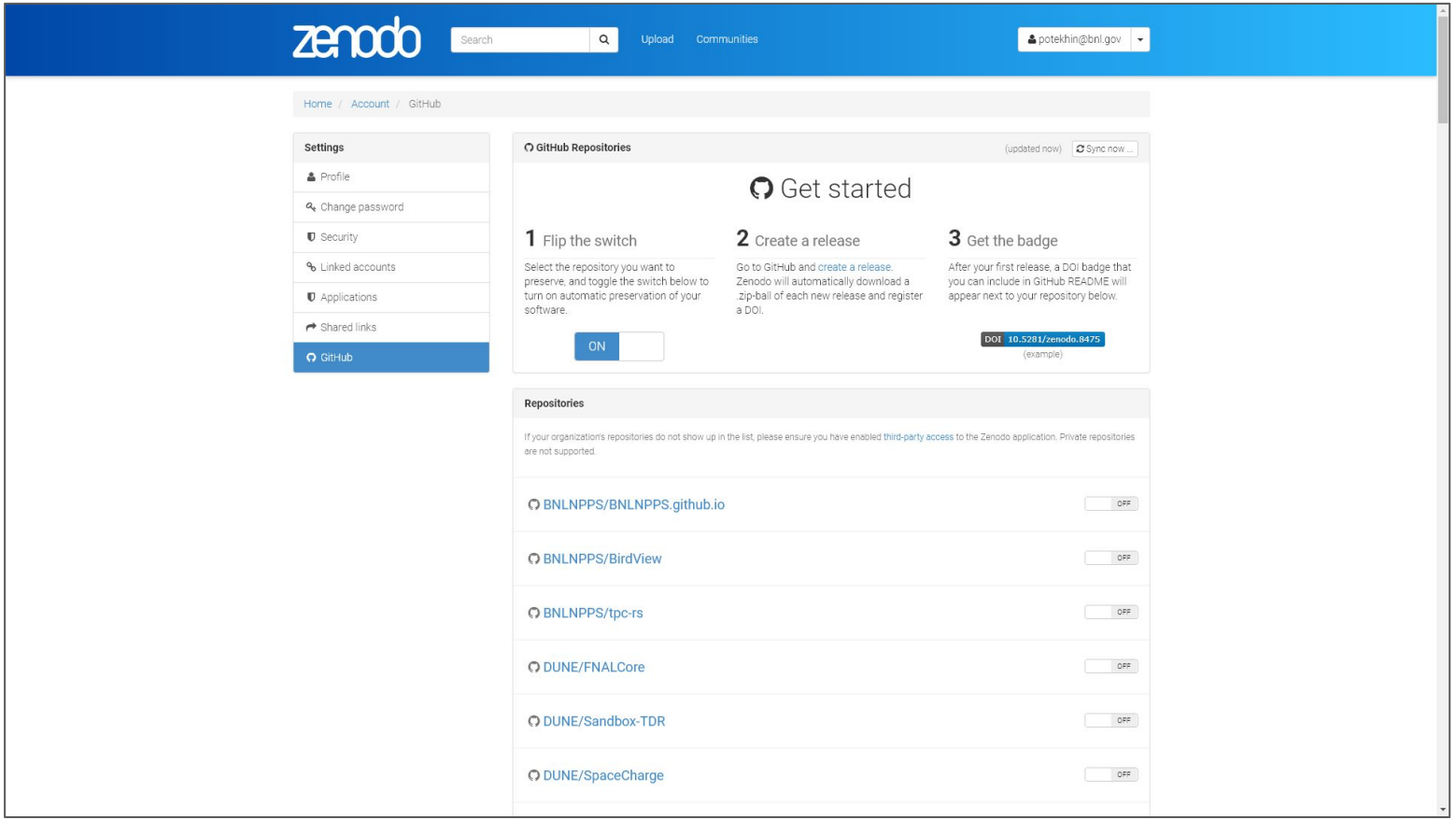
Zenodo uploads, going forward

- I need help with uploads of theses and other materials
  - Assigning metadata (keyword) to documents is best done by domain experts
  - Students/Grad Students/Other community members can easily contribute
    - Low-hanging fruit, real results visible immediately
  - Curation makes it easy to accept uploads from anyone - any uploads, of any kind
  - If theses authors do this themselves this is even better
  - Should also benefit from the ORCID reference mechanism
- What other document (in addition to theses) should be considered public?
  - We need to reduce storage requirements for the GitHub repo of the site, and there are some large documents there already which really belong in the long-term document database
  - Of course can also create a separate repo and will probably do so
  - But for example should we just move the decadal plans and detector writeups to Zenodo?

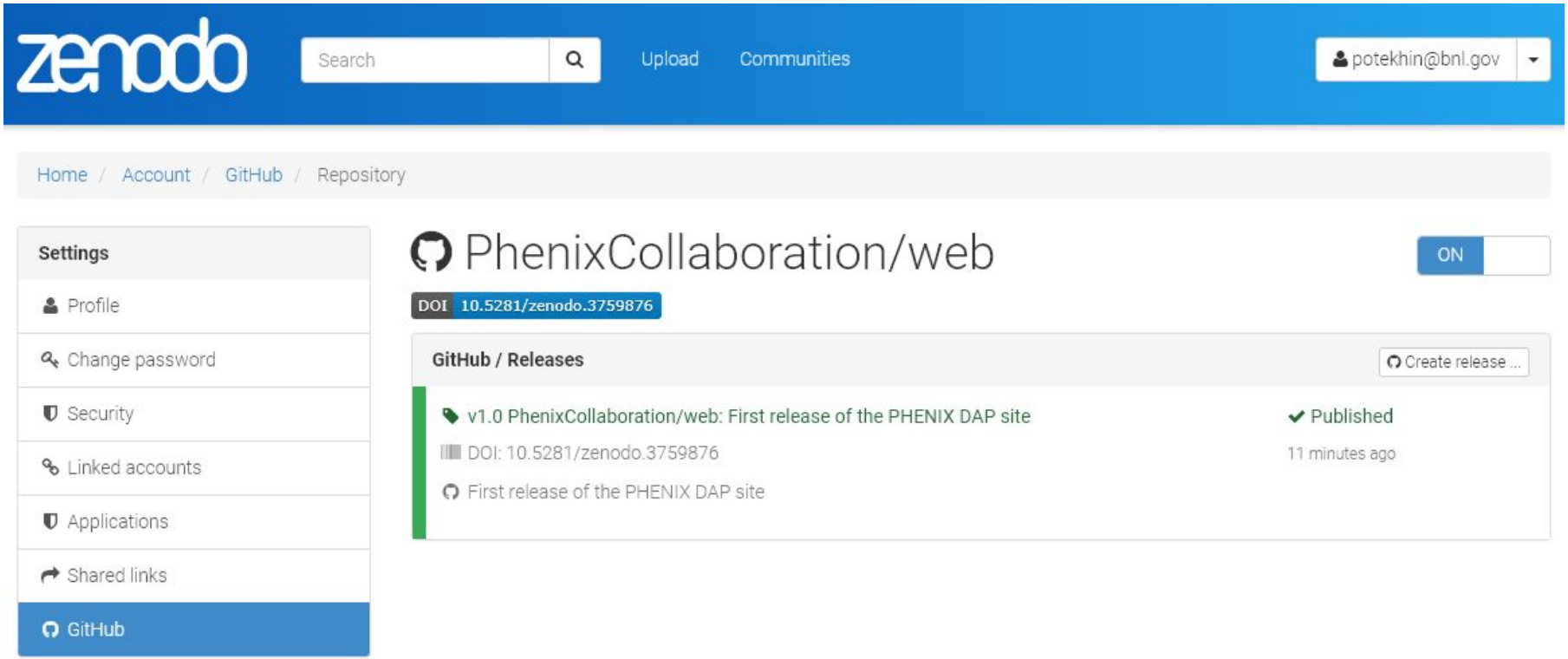
GitHub/Zenodo mechanics

- A snapshot of a GitHub repository can be included in Zenodo organically, and a DOI generated
  - Prepares and preserves tarballs of your releases
  - Makes your code easy to find (using the metadata) and to reference by a unique ID
  - Nice GUI
  - DOI reference to analysis code
- Easy to use
  - I tested this functionality and it was quite simple
  - DOIs take some time O(10min) to propagate to the DOI.org system

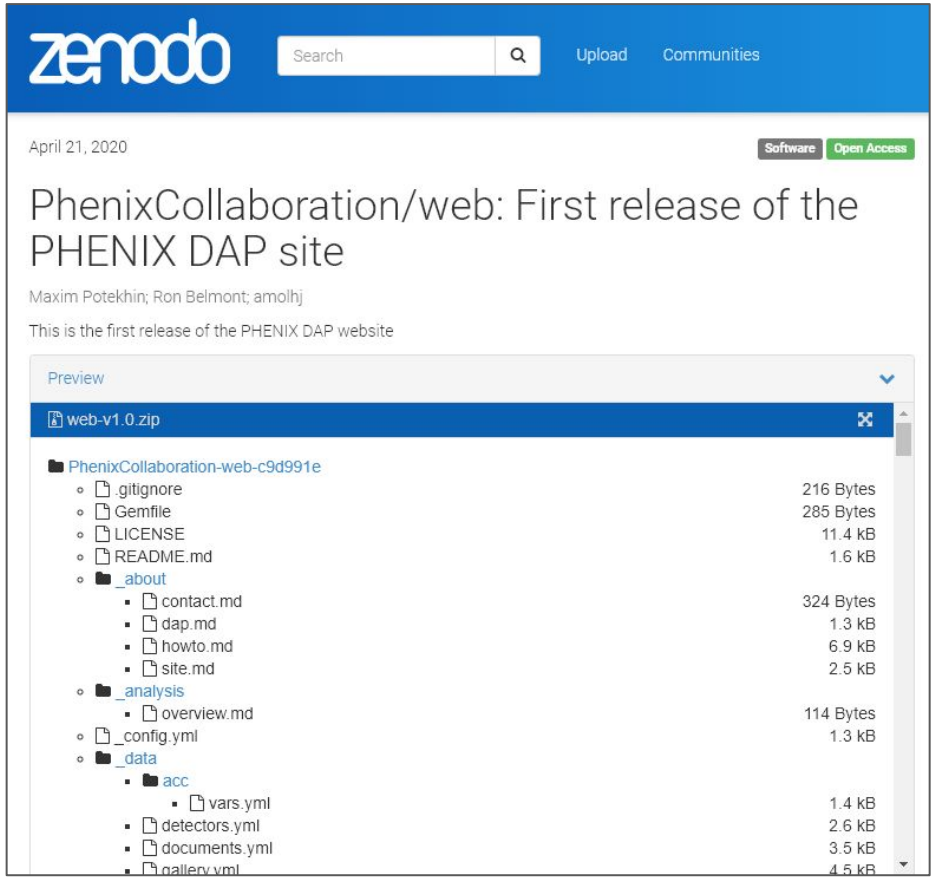
Zenodo - GitHub panel - repo selection



Zenodo - GitHub panel - published release

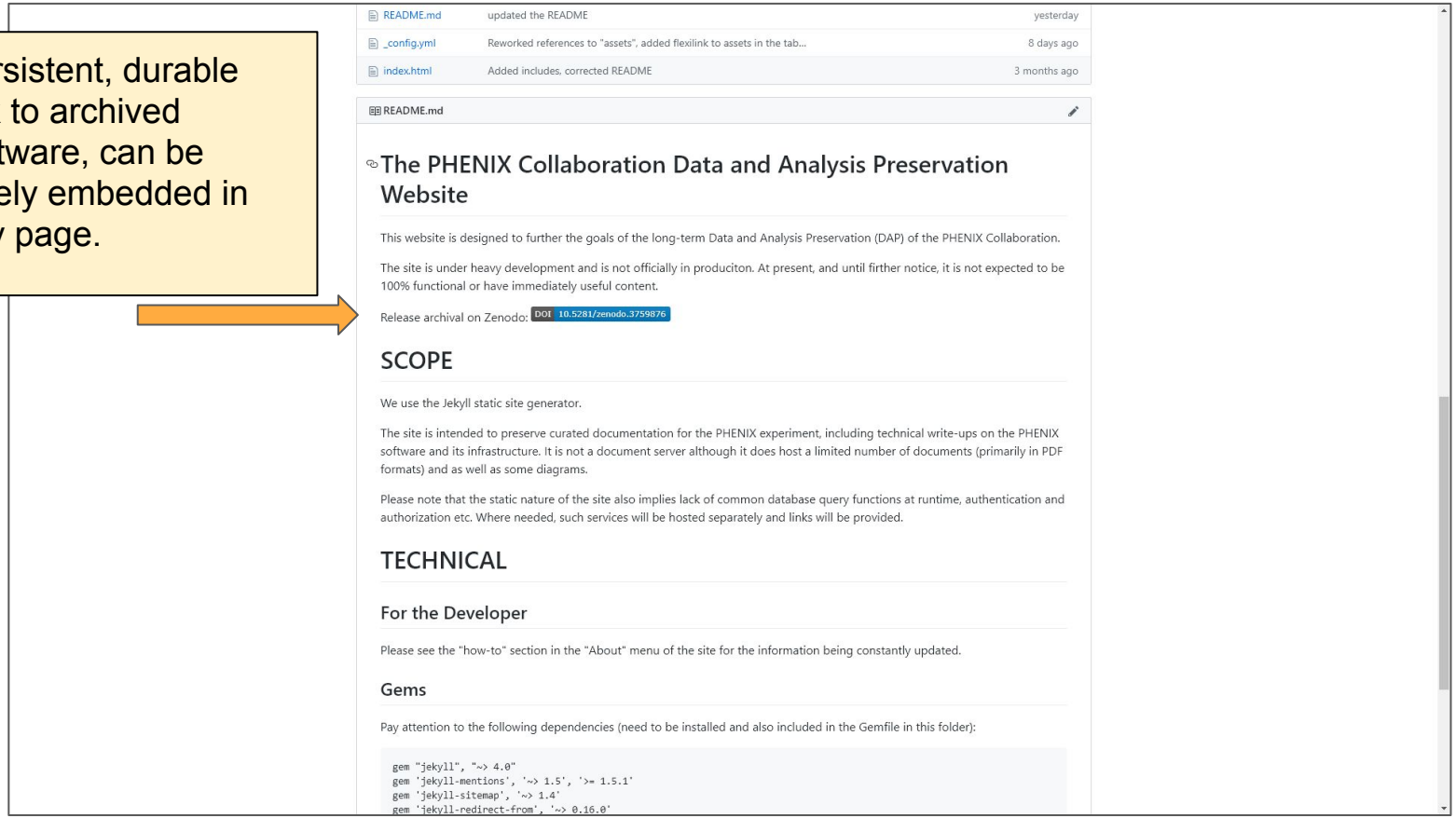


Zenodo - GitHub panel - published release browser



Zenodo - an (elegant) way to reference software

Persistent, durable link to archived software, can be nicely embedded in any page.



GitHub/Zenodo benefits for DAP

- Analysis note process: current policy as per the template is to create a tarball and indicate its location in HPSS
  - Very low-tech method of preserving the code, may be hard to find in a context other than the analysis note itself
  - Current solution - “leave for later”
- With GitHub (with optional Zenodo integration):
  - Analysis code/macros tagged with metadata, traceable and public (?)
  - Very easy to aggregate all information on one web page that works
  - Permanent/durable links in DAP documentation
- ...which brings us back to the policy issue

Policy

- Should we declare the analysis code public?
  - On the technical side, this would simplify things quite a bit
- If not, what are implications for DAP, especially for the DAP site which is public?
- I would argue that there is “security by obscurity” i.e. it would be impossible for an arbitrary person to make an arbitrary use of PHENIX code since it’s tied to the BNL infrastructure, and all access will need to be properly authorized