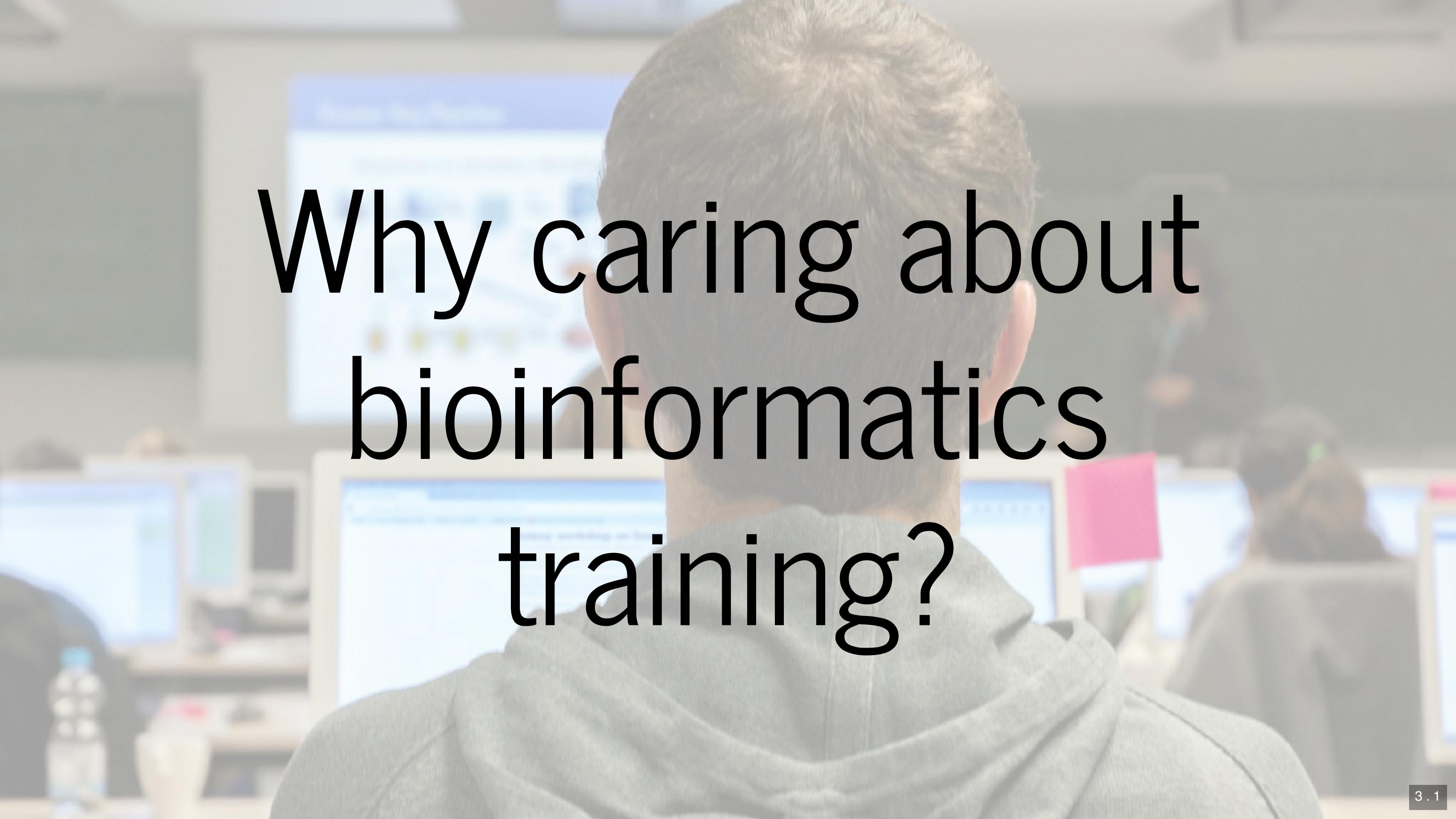


# Building an open, collaborative, online infrastructure for bioinformatics training



Bérénice Batut, Galaxy Training Network,  
Dave Clements, Björn Grüning

BOSC  
July 2017

A person with light brown hair, wearing a grey hoodie, is shown from the back and slightly to the side. They are looking towards a computer monitor which displays a presentation slide. The slide features a large, bold, black font question: "Why caring about bioinformatics training?".

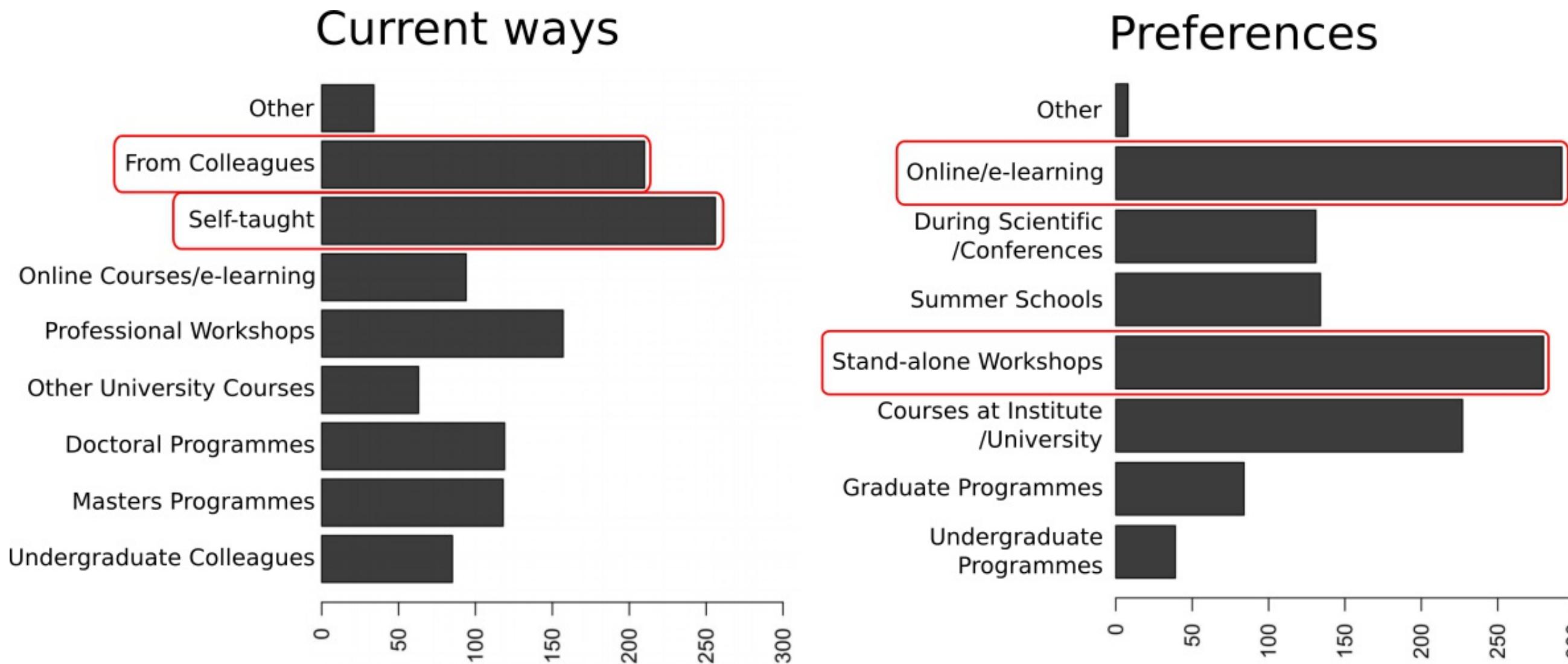
Why caring about  
bioinformatics  
training?

# Need for bioinformatic training

*Bioinformatics has become too central to biology  
to be left to specialist bioinformaticians*

- Explosion of data to analyze
- Access to computational power
- Thousand of possible tools for specialized analyses

# An increasing demand for learning bioinformatics



Graphs of **Brazas et al, 2017**

## Tools

[plotHeatmap](#)

## deepTools

[plotProfile](#) creates a profile plot for score distributions across genomic regions.

[computeMatrix](#) prepares data for plotting a heatmap or a profile of given regions.

[plotHeatmap](#) creates a heatmap for score distributions across genomic regions.

## Workflows

[All workflows](#)

2: patient4\_input\_poor\_outcome.bam (as bigwig)  
1: patient4\_ChIP\_ER\_poor\_outcome.bam (as bigwig)

You can generate a bigWig file from a BAM file using the bamCoverage tool. (-scoreFileName)

## computeMatrix has two main output options

## scale-regions

In the scale-regions mode, all regions in the BED file are stretched or shrunken to the same length (in bases) that is indicated by the user. Reference-point refers to a position within the BED regions (start or end of each region). In the reference-point mode only those genomic positions before (upstream) and/or after (downstream) the reference point will be considered.

## Distance in bases to which all regions are going to be fit

500

(-regionBodyLength)

## Set distance up- and downstream of the given regions

no

## Show advanced output settings

no

## Show advanced options

no

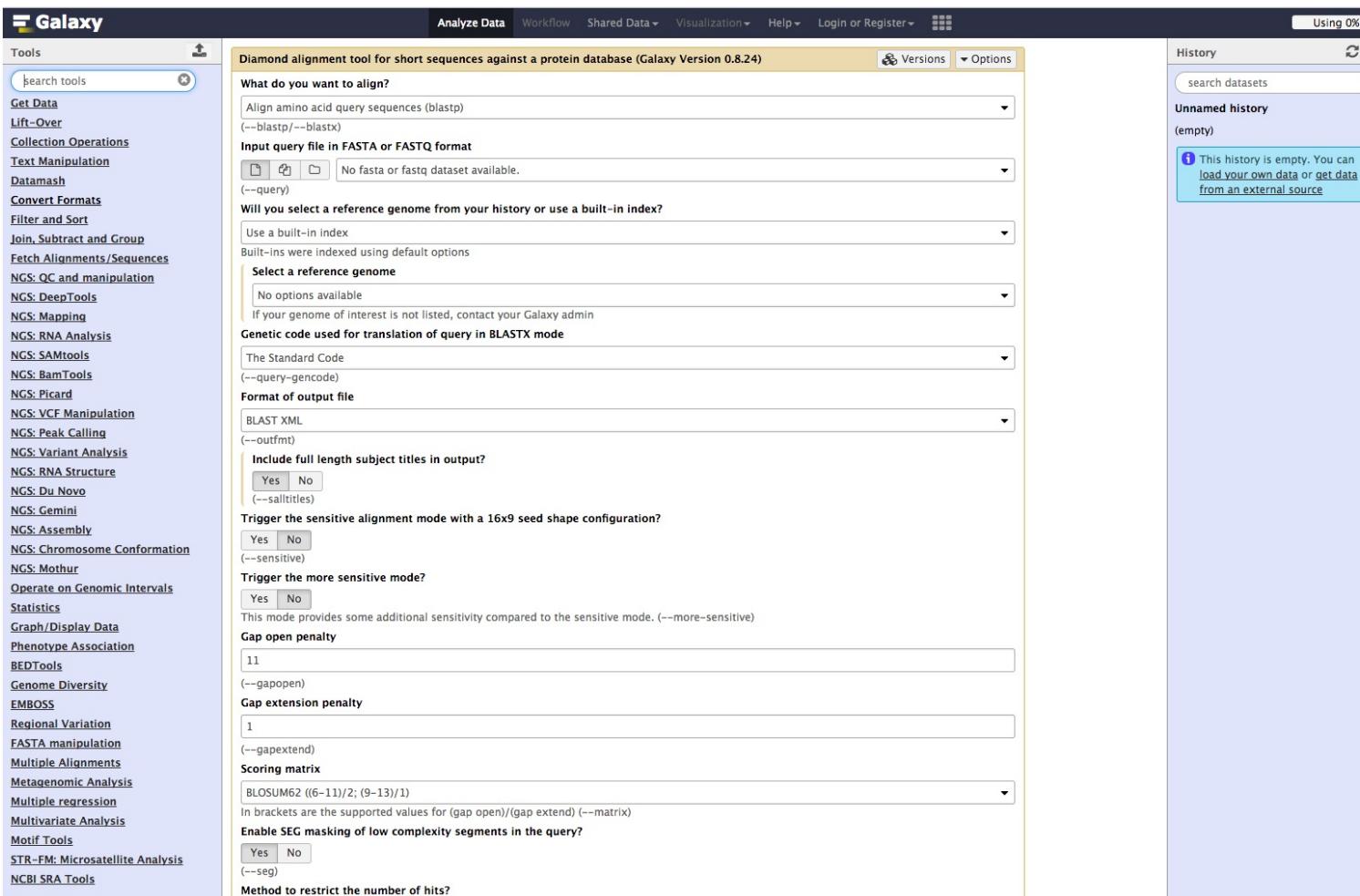
## What it does

This tool generates an input file in bigWig format containing scores associated with genomic regions. This can be used for example to compare ChIP-seq and control samples. The generated bigWig files can be visualized in Galaxy's visualization tools.

Firefox sendet automatisch einige Daten an Mozilla, damit die Benutzerzufriedenheit verbessert werden kann.

# Galaxy a great solution !

# Computational knowledge: Not required!



- Web interface for numerous bioinformatics tools
- Scalable
- No issue with computer configuration during training

Platform  
recommendations

Best  
Practices

Trainer  
Directory

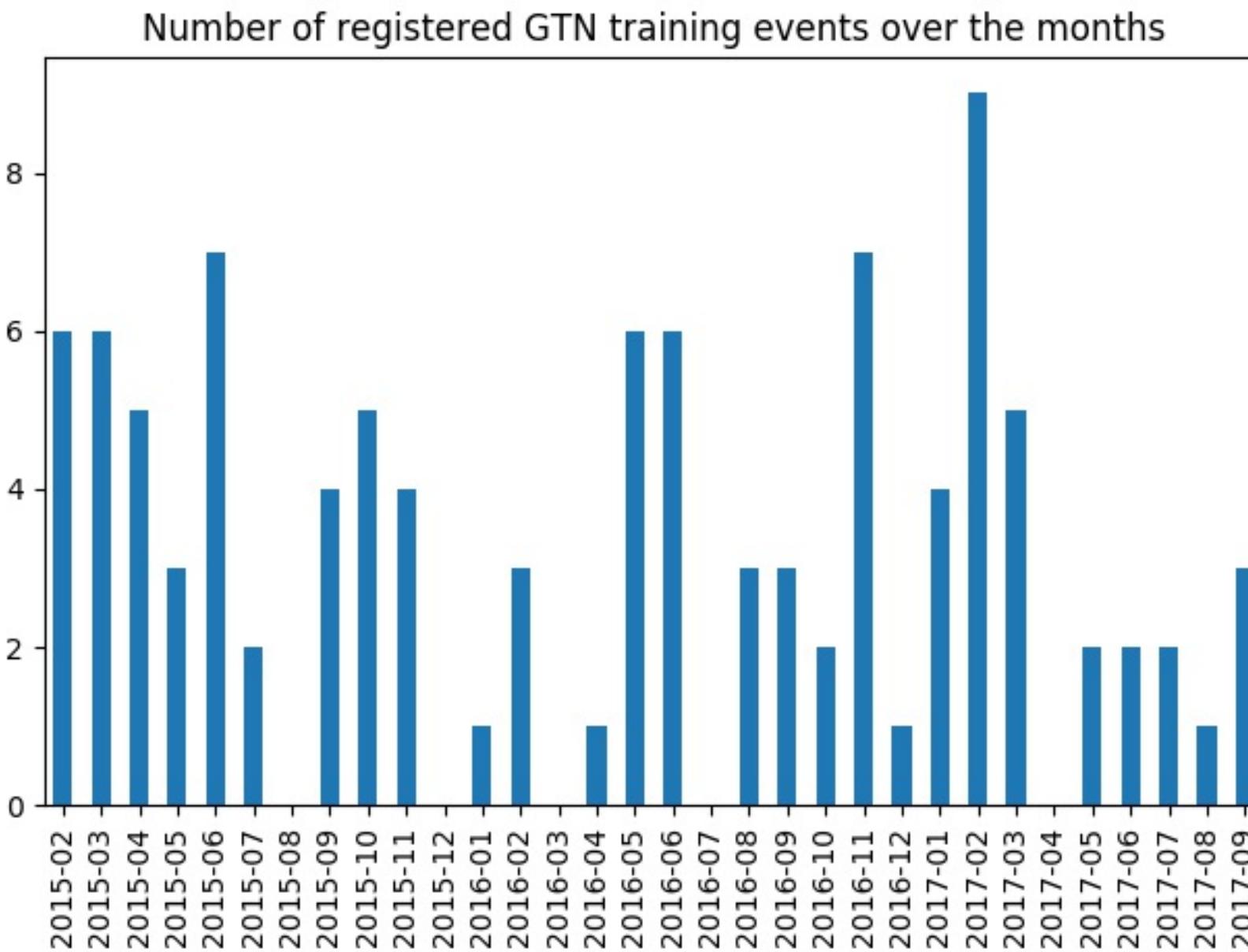
Training  
ressources

Galaxy

Training  
Network



# Quite a lot of events...



# Worldwide!

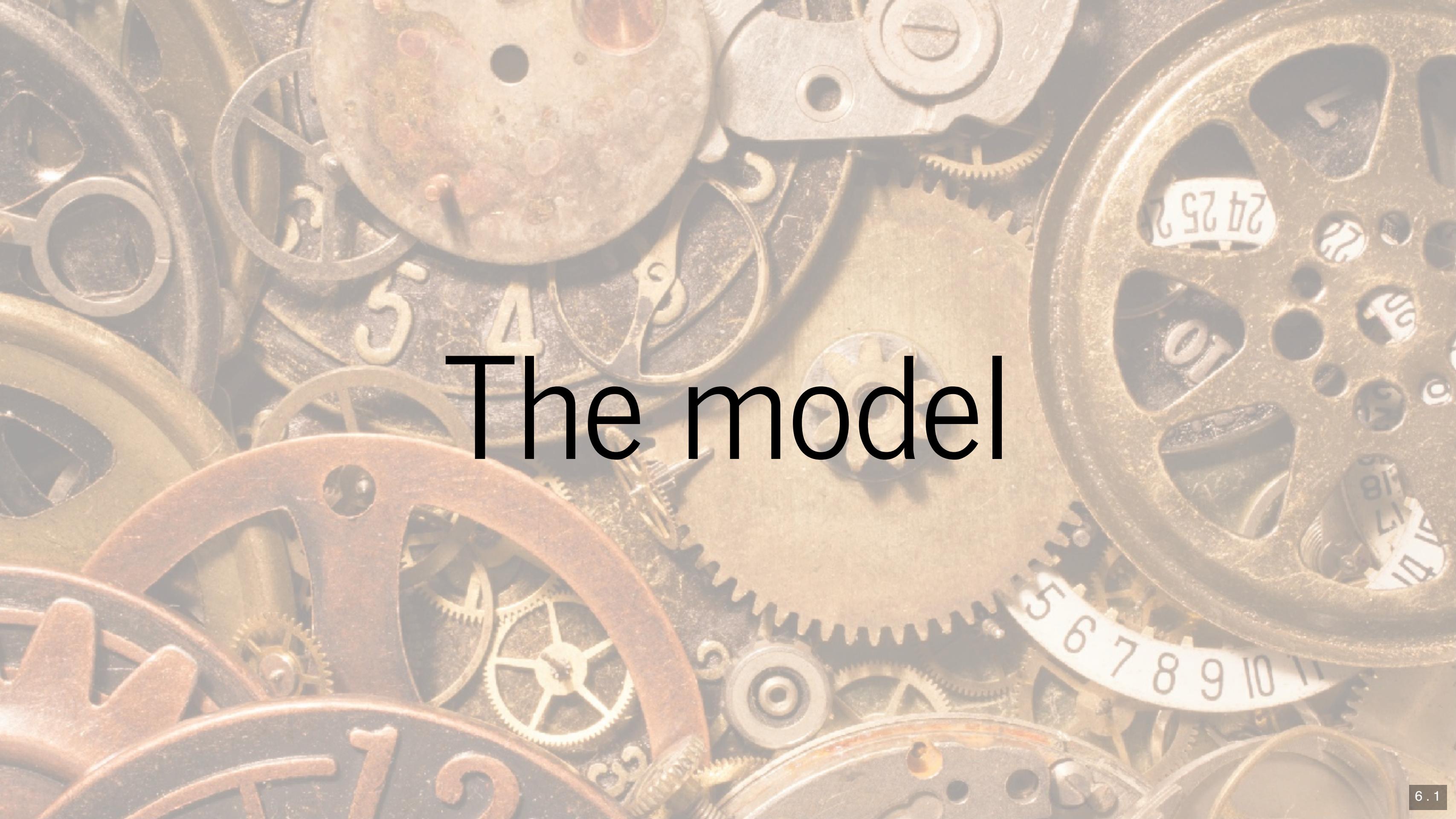




**Building a new open, collaborative and FAIR  
model for bioinformatics training**

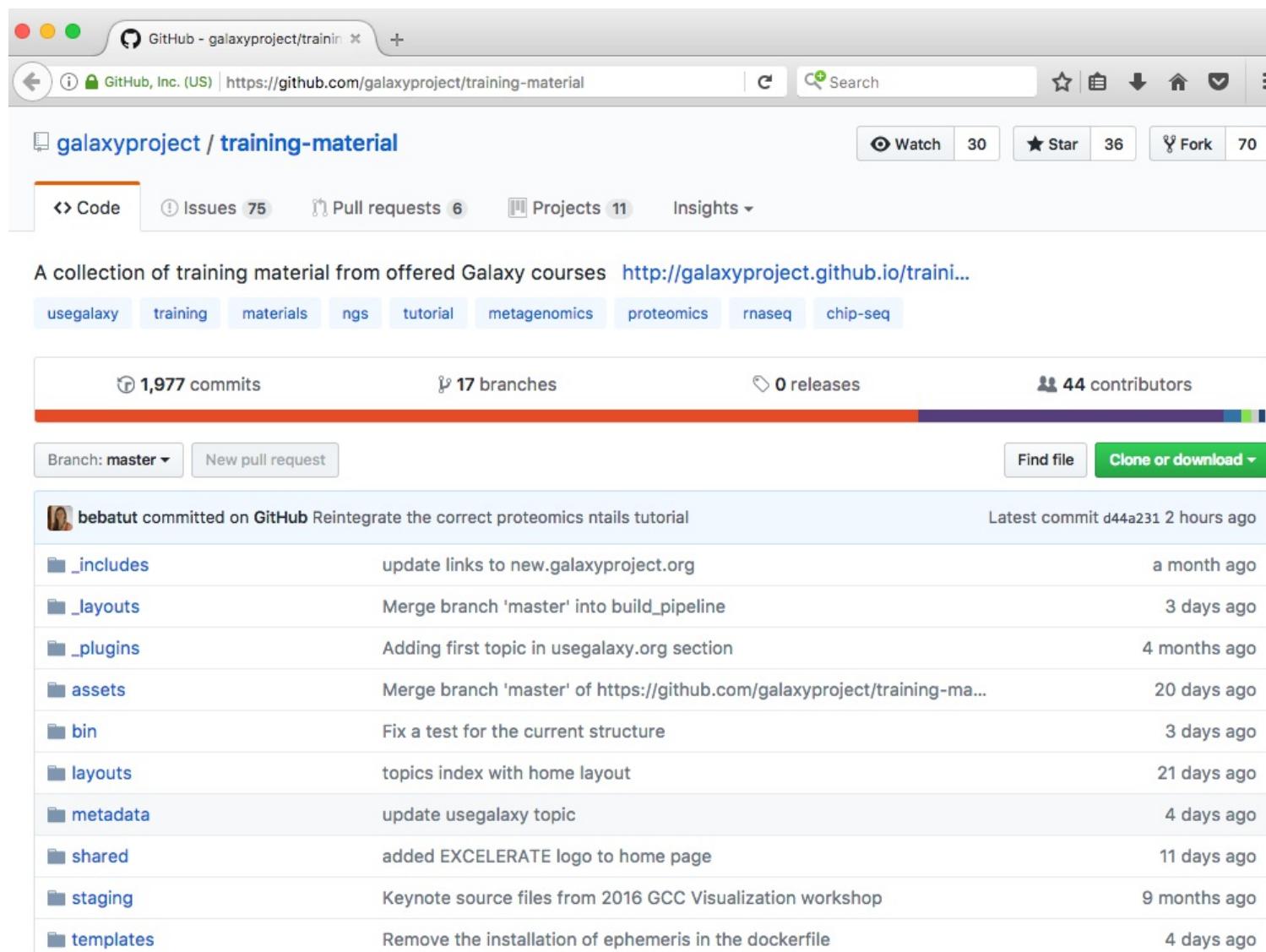
# Requirements

- Easy to use
- Support for effective training for
  - Individual users
  - Instructors
- Definition of technological infrastructure
- Limited redundancy



# The model

# One repository to collect everything

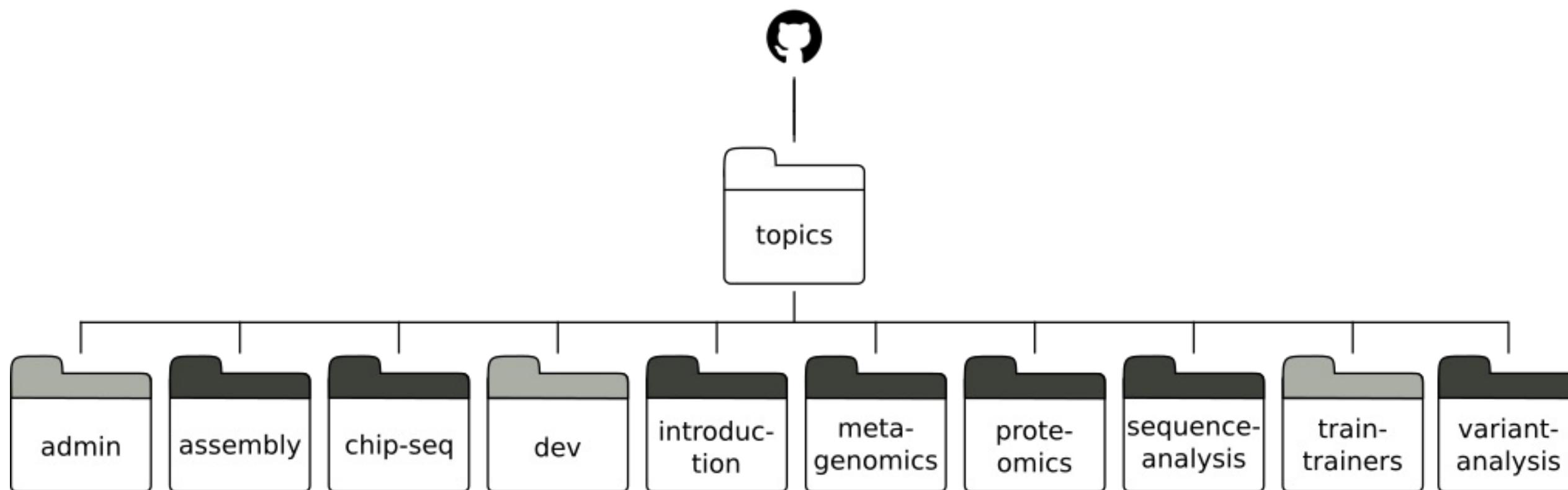


A screenshot of a GitHub repository page. The repository is named "galaxyproject / training-material". The page shows a collection of training material from offered Galaxy courses. It features a header with navigation links like "Code", "Issues 75", "Pull requests 6", "Projects 11", and "Insights". Below the header, there's a summary bar with metrics: 1,977 commits, 17 branches, 0 releases, and 44 contributors. A list of recent commits is displayed, each with a small icon, the author's name, the commit message, and the time it was made. The commits are as follows:

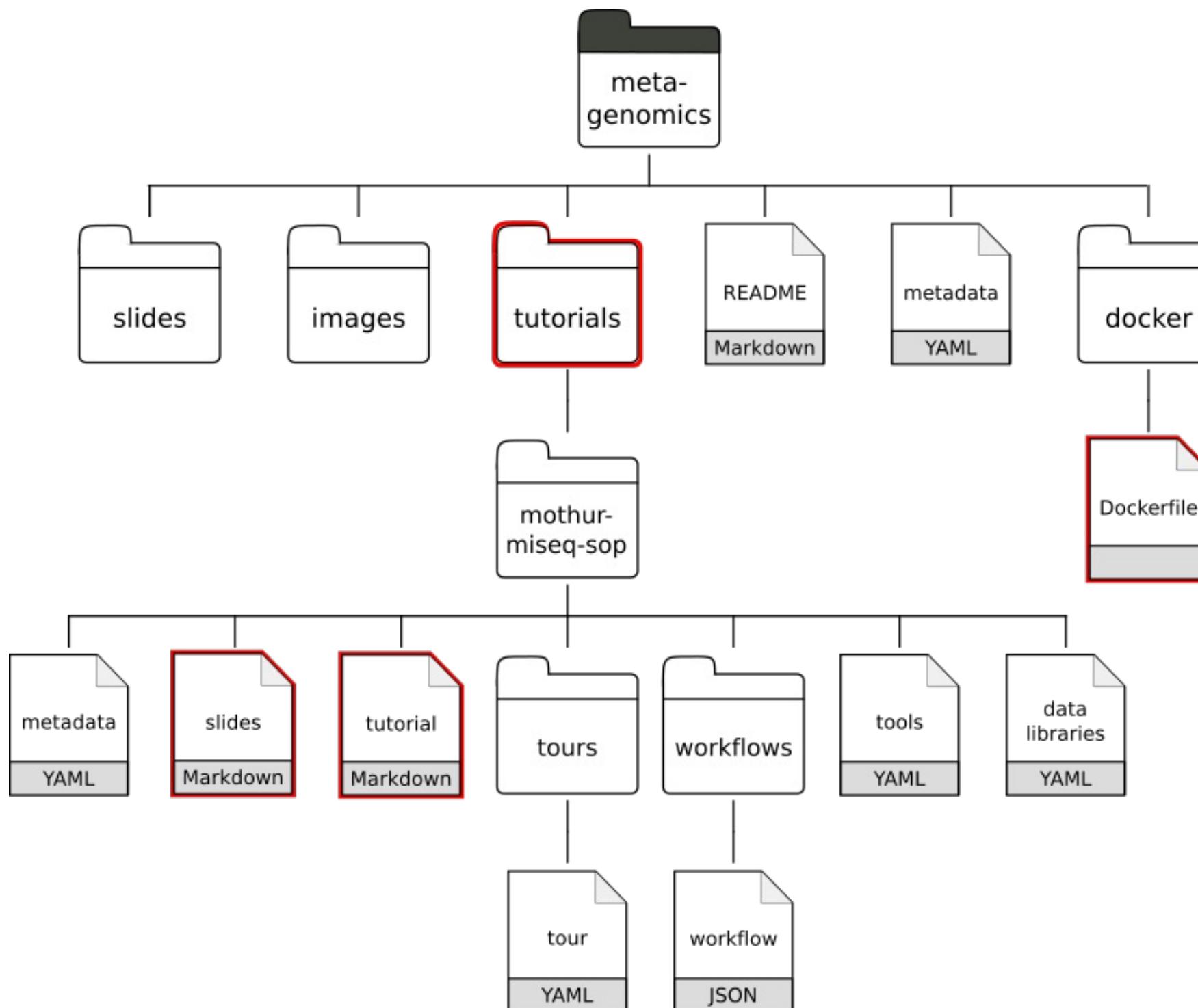
Author	Commit Message	Time Ago
bebatut	Reintegrate the correct proteomics ntails tutorial	Latest commit d44a231 2 hours ago
	update links to new.galaxyproject.org	a month ago
	Merge branch 'master' into build_pipeline	3 days ago
	Adding first topic in usegalaxy.org section	4 months ago
	Merge branch 'master' of https://github.com/galaxyproject/training-ma...	20 days ago
	Fix a test for the current structure	3 days ago
	topics index with home layout	21 days ago
	update usegalaxy topic	4 days ago
	added EXCELERATE logo to home page	11 days ago
	Keynote source files from 2016 GCC Visualization workshop	9 months ago
	Remove the installation of ephemeris in the dockerfile	4 days ago

GitHub: [galaxyproject/training-material](https://github.com/galaxyproject/training-material)

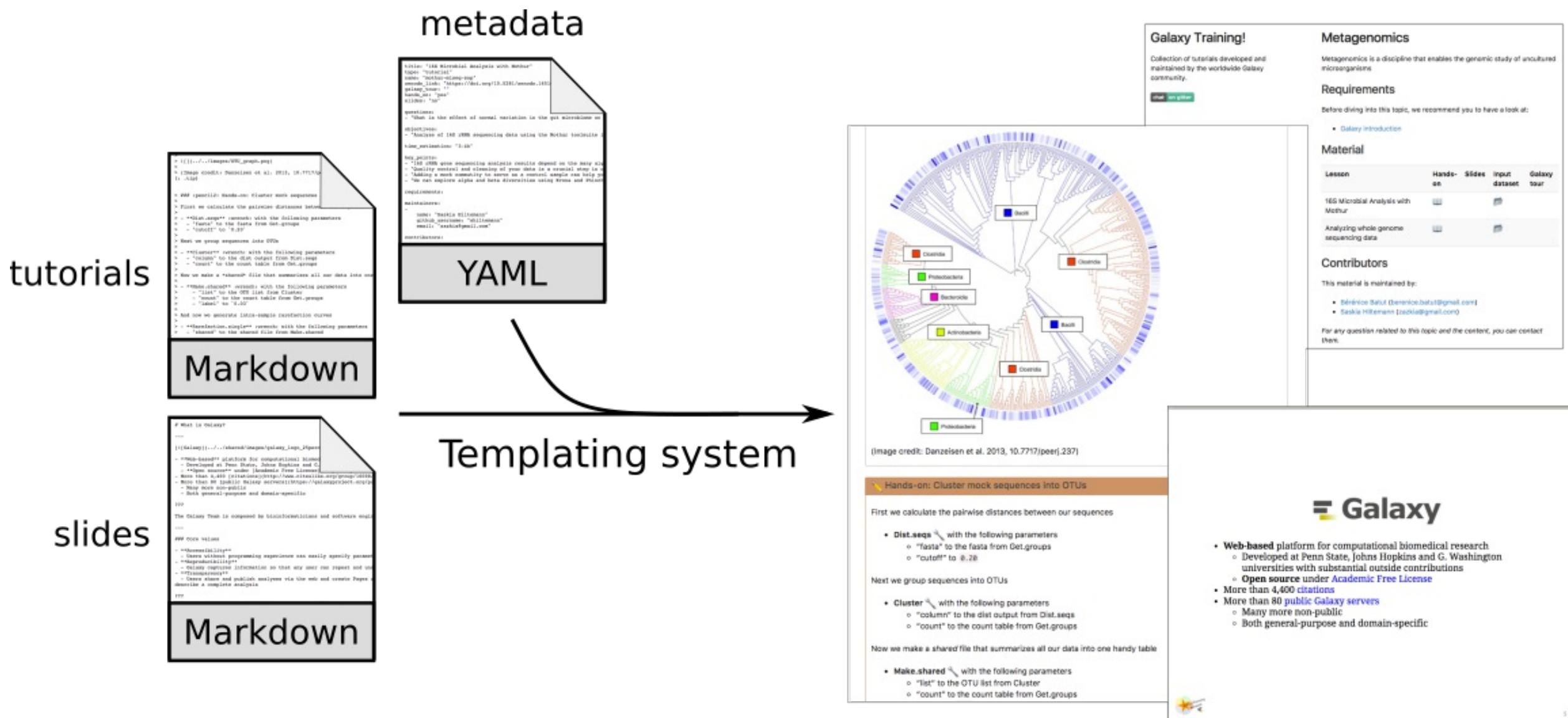
# Topics for different targeted users



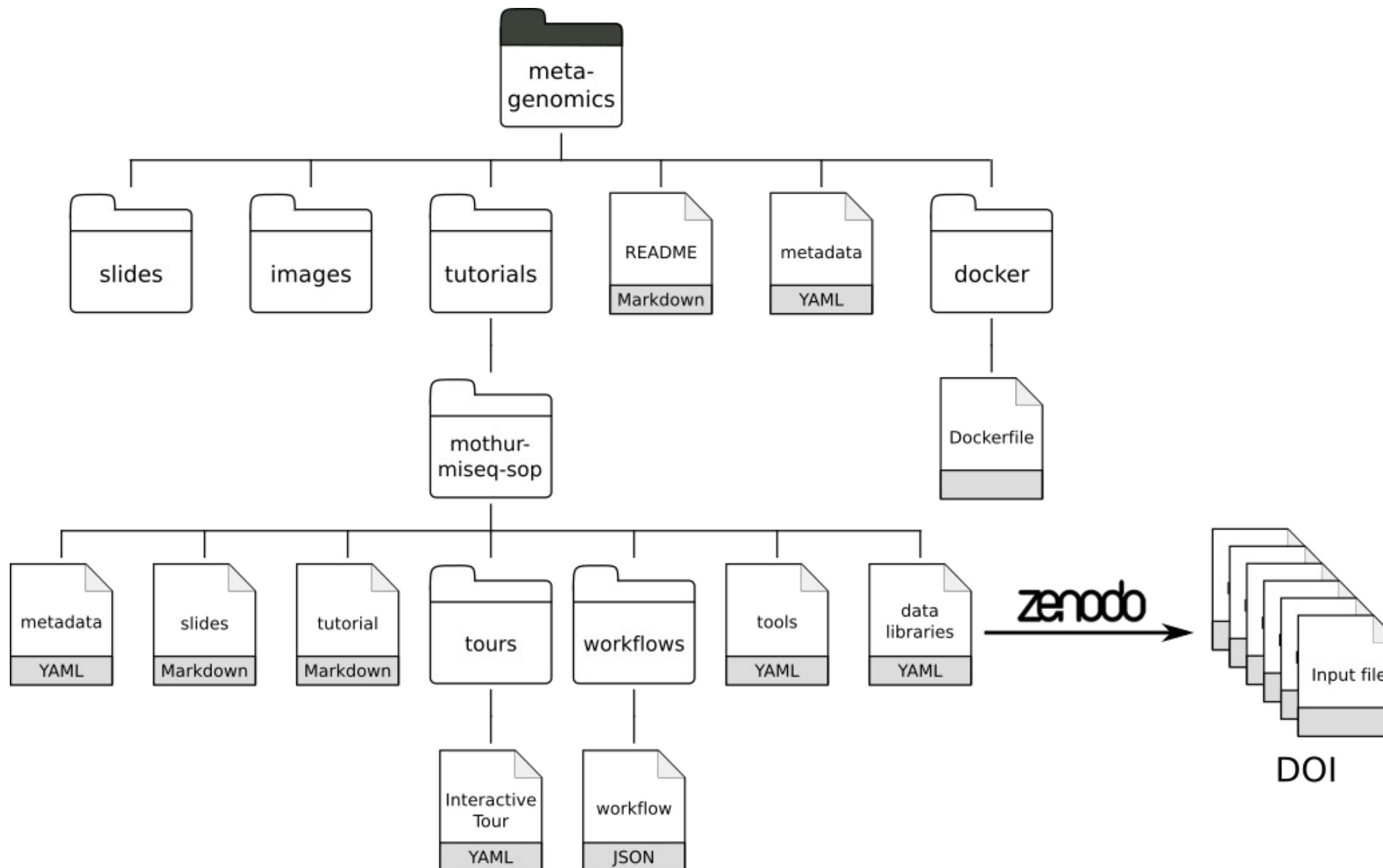
# Similar structure, content and formats



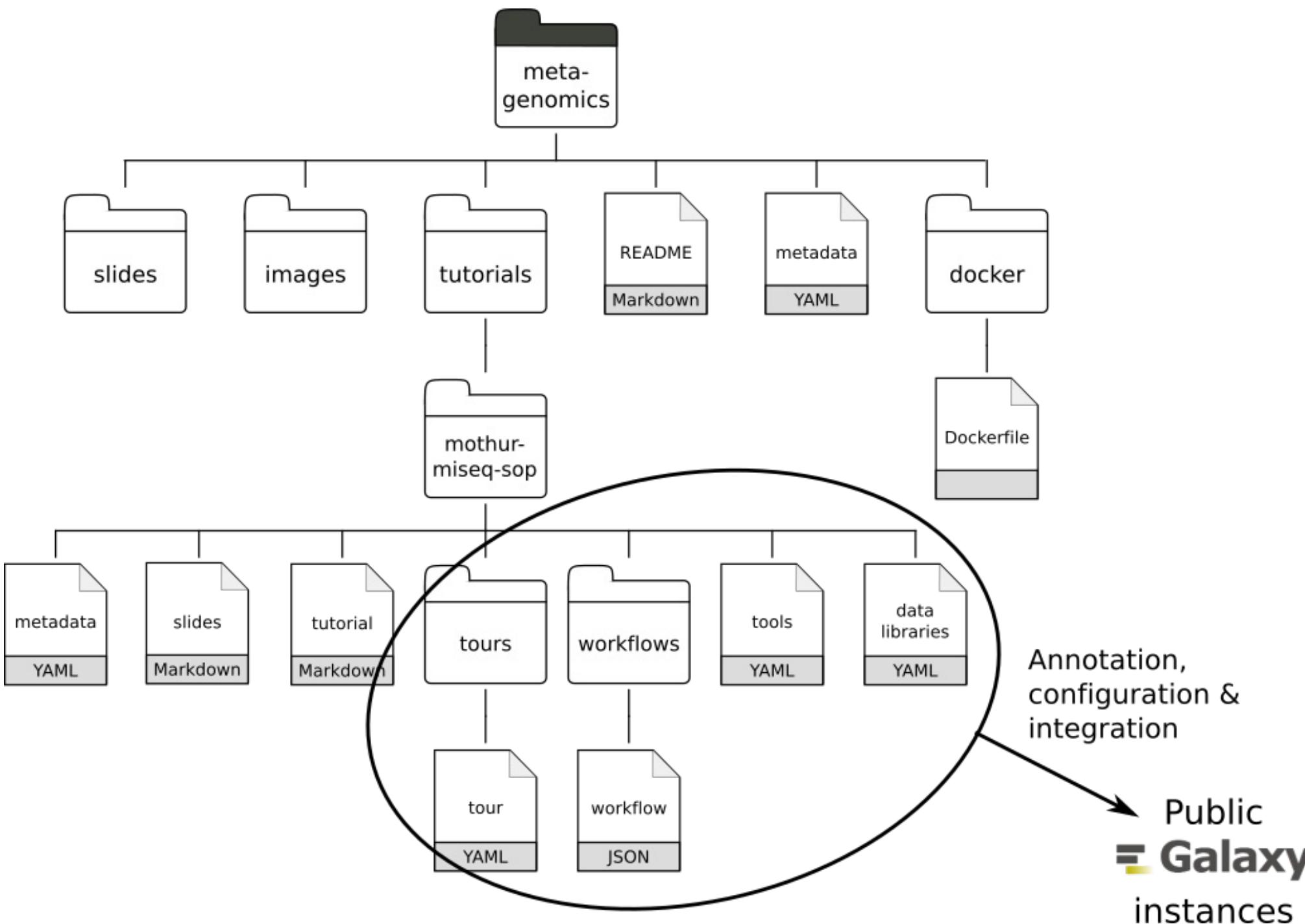
# Separation between content and formatting



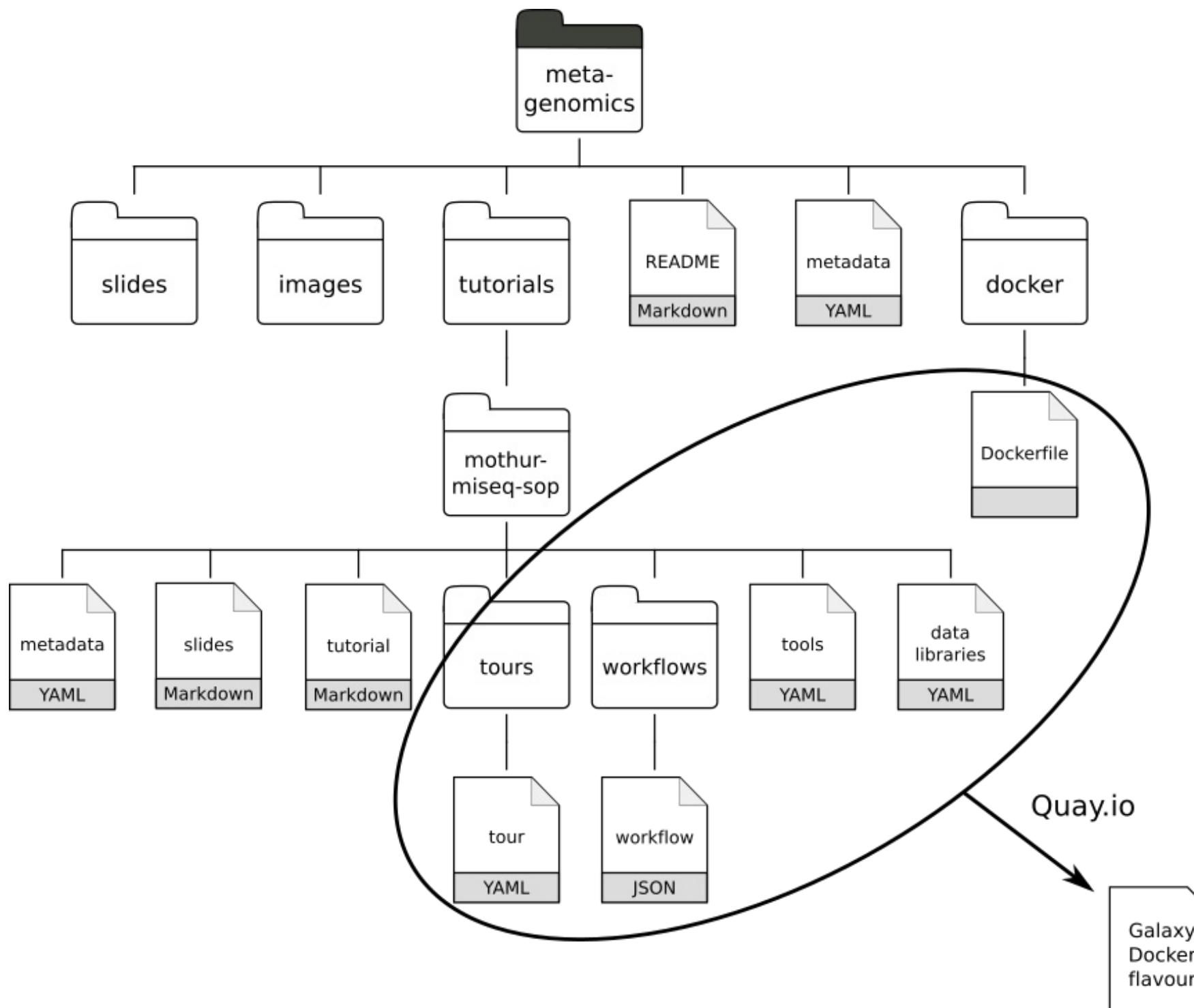
# Citable data & Credit



# Definition of the technical infrastructure



# Definition of the technical infrastructure



Quay.io

Galaxy  
Docker  
flavour

**F**indable **A**ccessible **I**nteroperable **R**Reusable



model



# Findable

```
YAML  
metadata
```

Automatically  
added →

The screenshot shows a web interface for the elixir TeSS Training Portal. At the top, there's a navigation bar with links for Events, Materials (which is currently selected), Packages, Workflows, Providers, Nodes, About, and a search bar. Below the navigation, the URL indicates a specific training material: Home / Materials / Metagenomics - 16S Microbial Analysis with Mothur. The main content area is divided into two columns. The left column, titled 'Content provider', features a logo for 'Galaxy Training Network' and 'Galaxy Training'. It describes the provider as a GitLab repository and its Wiki companion containing training materials for teaching next-generation sequencing data analysis. The right column contains the title 'Metagenomics - 16S Microbial Analysis with Mothur', a link to the original source (<http://galaxyproject.github.io/training-material/Metagenomics/tutorials/mothur-miseq-sop>), author information (Bérénice Batut, Saskia Hiltemann), a DOI (<https://doi.org/10.5281/zenodo.165147>), and a section for 'External resources'.

TeSS:  elixir Training Portal

<http://tess.elixir-europe.org/>



## Accessible

- Online

<http://training.galaxyproject.org>

- Technical support

- Self-training boxes with Galaxy Docker flavor
- Annotated public Galaxy instances



## Interoperable

- Metadata description in YAML and integrating EDAM ontology
- Content for different targets (workshops and self-training)
- Technical support for different platforms



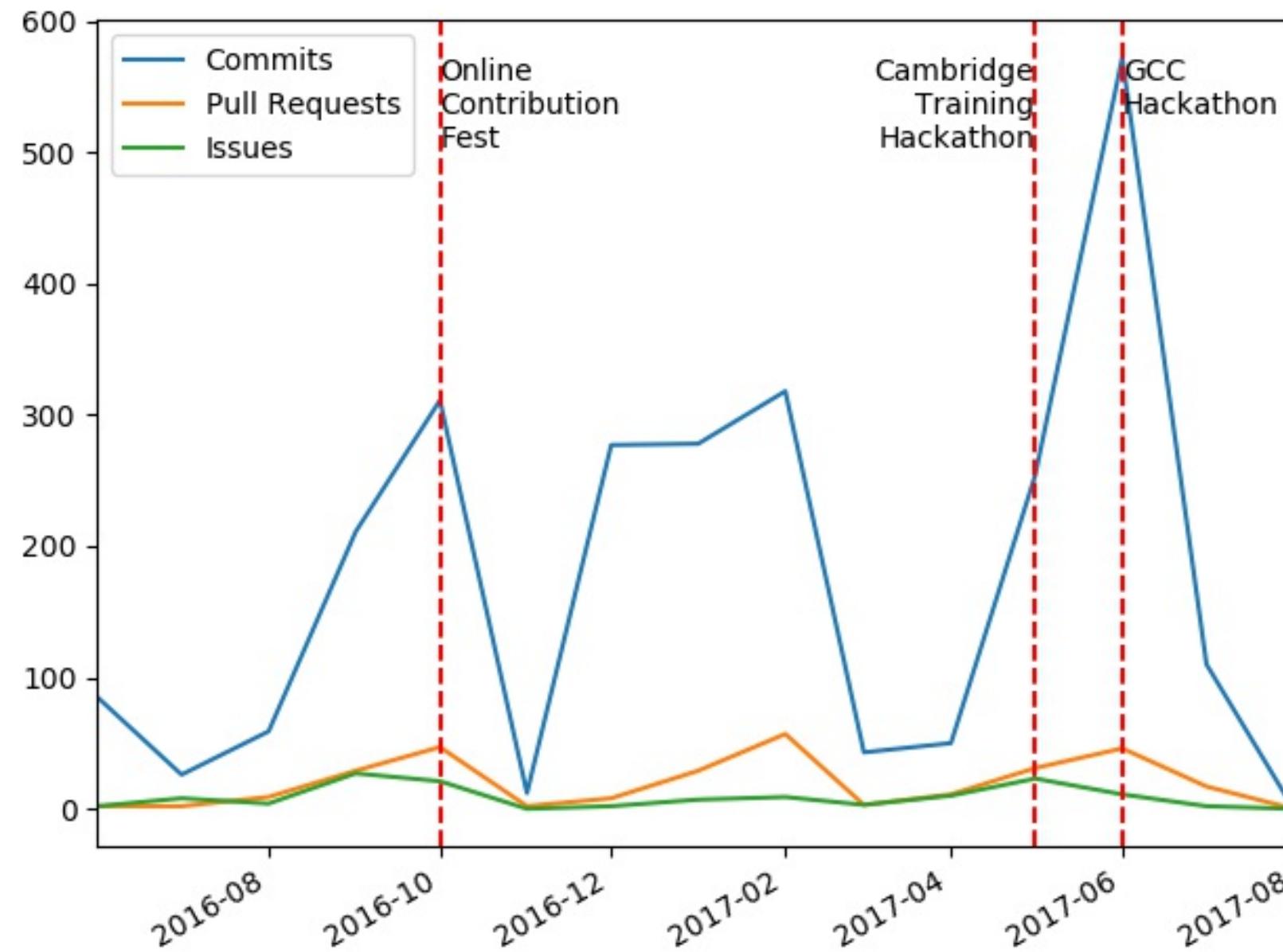
## Reusable & Open

- CC BY 4.0 license
- DOI for the input datasets
- Open development process on GitHub & via Gitter
- Open education movement

A hand is shown on the right side of the frame, holding a piece of chalk and drawing several white outlines of lit lightbulbs on a dark green chalkboard. The lightbulbs vary in size and are scattered across the board. One large lightbulb on the right has a thought bubble drawn around it, containing a smaller, faint outline of a lightbulb.

# Community effort

# Numerous contributions



And 3 successful hackathons!

# Numerous discussions

Galaxy Training Network/Lobby <https://new.galaxyproject.org/MailingLists/>

Gildas Le Corguillé @lecorguille May 31 16:01  
thumb up

John Chilton @jmchilton Jun 01 21:11  
I tried checking out build\_pipeline and now only the usegalaxy tutorials show up in my training site  
Not admin or dev tutorials at all - any ideas?

Björn Grüning @bgruening Jun 01 21:12  
This branch is not yet working ... @dannon needs to finish up his methal smith build magic

John Chilton @jmchilton Jun 01 21:18  
Are there slide content changes or just restructuring and metadata reorganization in that branch?

Dannon Baker @dannon Jun 01 21:18  
Yep, the build [Blocked Plug-in](#) contains a full reorganization of the content, which doesn't build quite yet. (but that's where we'll want new stuff for now)  
I wouldn't guess there are likely conflicting slide content changes

Björn Grüning @bgruening Jun 01 21:22  
Afaik there are no content changes, just the build procedure and the organisation.

Victoria Dominguez del Angel @vdda Jun 02 16:15  
I'm with Berenice, please take care 🙏

Slugger70 @Slugger70 Jun 05 12:29  
Hi all, Torsten Seemann came up with an idea after looking at the GTN website. He would like to see tags on the various tutorials for things like Eukaryotic vs prokaryotic specific, or virus etc etc... I reckon it's a good idea.

Björn Grüning @bgruening Jun 05 12:29  
thumb up  
We need more tags, also for supported Galaxy instance etc ...

Slugger70 @Slugger70 Jun 05 12:32  
I agree. Minimum Galaxy version at least.

Yvan Le Bras @yvanlebras Jun 05 12:50  
+1

Mallory Freeberg @malloryfreeberg Jun 05 16:04  
+1  
Also willing to help with this 😊

Slugger70 @Slugger70 Jun 05 16:07  
Could be something to add at the hackathon?

Mallory Freeberg @malloryfreeberg Jun 05 16:08  
Absolutely. I'll add it to suggested data hack topics

Mallory Freeberg @malloryfreeberg Jun 05 16:16  
Click here to type a chat message. Supports GitHub flavoured markdown.

PEOPLE: [ADD](#) [SEE ALL \(50 PEOPLE\)](#)

ACTIVITY

shiltemann on general\_metagenomics\_tutorial update tutorial (compare) Jun 01 21:11

nsoranzo commented #358 Jun 15

shiltemann on master Jun 15  
Add authors of Introduction sli...  
Merge pull request #359 from ns... (compare)

shiltemann closed #359 Jun 15

nsoranzo opened #359 Jun 15

shiltemann on fix-slides Jun 15  
change slide deck type (compare)

shiltemann synchronize #358 Jun 15

shiltemann opened #358 Jun 15

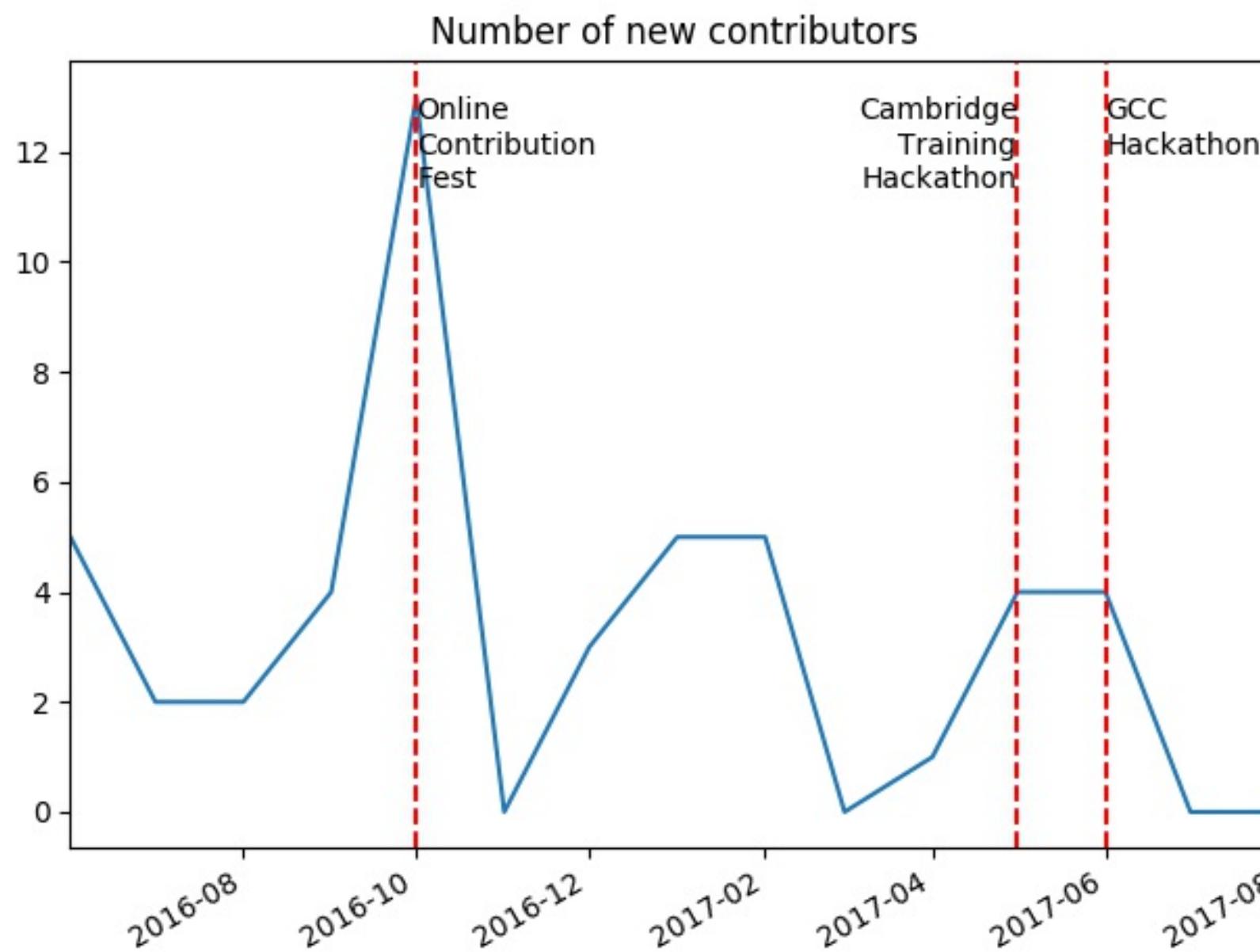
shiltemann on fix-slides Jun 15  
change slide deck type (compare)

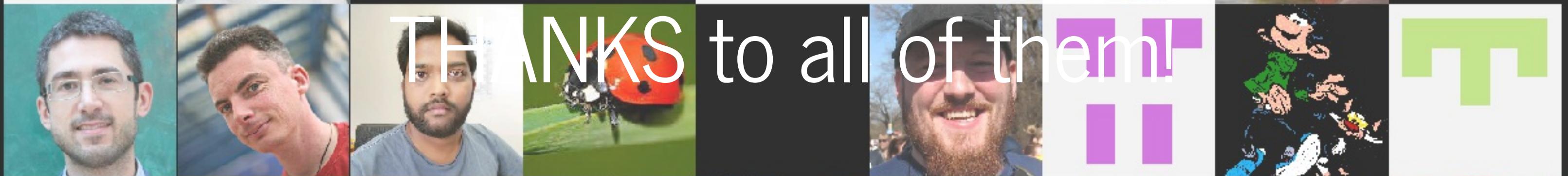
nsoranzo commented #354 Jun 15

shiltemann on general\_metagenomics\_tutorial start updating amplicon part (compare)

Gitter: Galaxy-Training-Network/Lobby

# A constantly growing community





# Thank you!

## Poster A166

### Scaling Bioinformatics Training



Bérénice Batut<sup>1</sup> (berenice.batut@gmail.com)

Björn Grüning<sup>1</sup>, Frederik Coppens<sup>2</sup>, Hans-Rudolf Hotz<sup>3</sup>, Gabriella Rustici<sup>4</sup>, Celia W G van Gelder<sup>5</sup>, Dave Clements<sup>6</sup>  
On behalf of : Galaxy Training Network Members, May 2017 Hackathon Participants, ELIXIR Galaxy WG

**Reusable and community extensible bioinformatics training materials,  
available online for trainers and individual life science researchers**

**A newly developed and already used Galaxy Training infrastructure**

