



HAL
open science

Galaxy CoDex for finding tools, workflows, and training

Bérénice Batut, Wendi Bacon, Paul Zierrep, Matthias Bernt, Nicola Soranzo,
Ove Johan Ragnar Gustafsson

► **To cite this version:**

Bérénice Batut, Wendi Bacon, Paul Zierrep, Matthias Bernt, Nicola Soranzo, et al.. Galaxy CoDex for finding tools, workflows, and training. GCC 2024 - Galaxy Community Conference, Jun 2024, Brno, Czech Republic. 10.7490/f1000research.1119764.1 . hal-04644966

HAL Id: hal-04644966

<https://hal.science/hal-04644966>

Submitted on 11 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Galaxy CoDex for finding tools, workflows, and training

B er enice Batut, Wendi Bacon, Paul Zierrep, Matthias Bernt,
Nicola Soranzo, Johan Gustafsson

Galaxy Community Conference 2024

The Problem of the [microGalaxy] community

Giving visibility to the resources for the [microGalaxy] community

- What is available in the Galaxy ecosystem for [microbiological] data analyses?
 - Tools
 - Workflows
 - Training
- How can we make a catalog that can be up-to-date?
- How can we get usage of these resources?

A black and white photograph of various tools including pliers, a hammer, and a screwdriver on a wooden workbench. The tools are arranged on a wooden surface, with a hammer and a pair of pliers in the foreground and a screwdriver and another pair of pliers in the background. The text "First step" is overlaid in the center of the image.

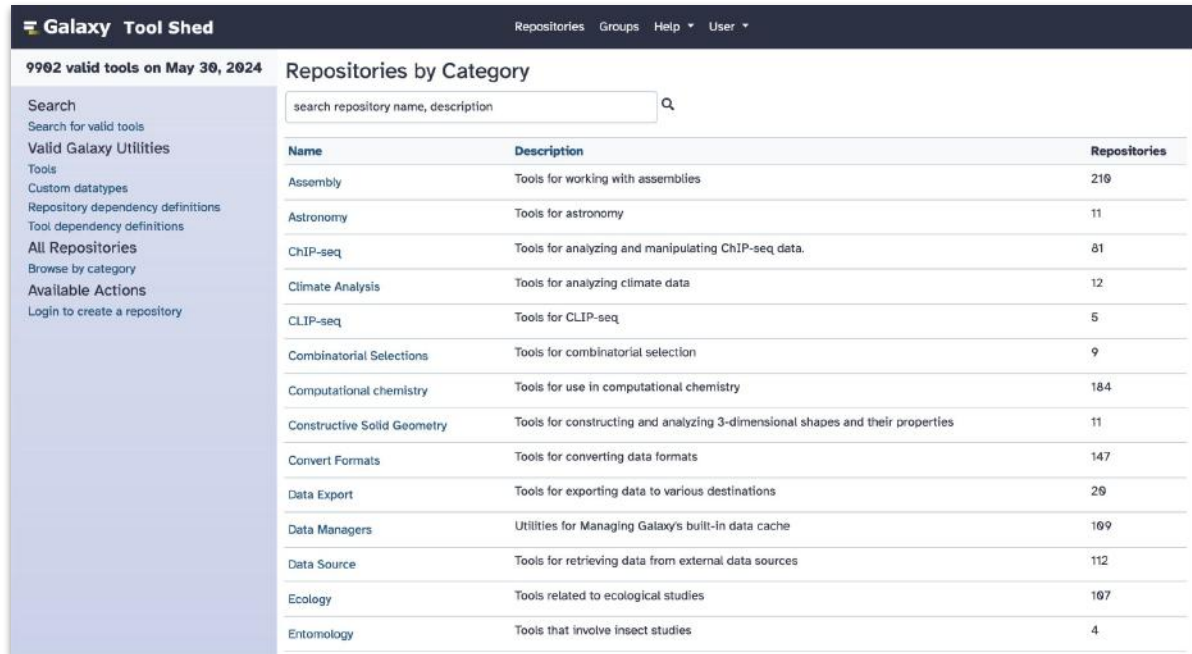
First step

Extract Tools

Extract Tools from the Galaxy ToolShed

BUT

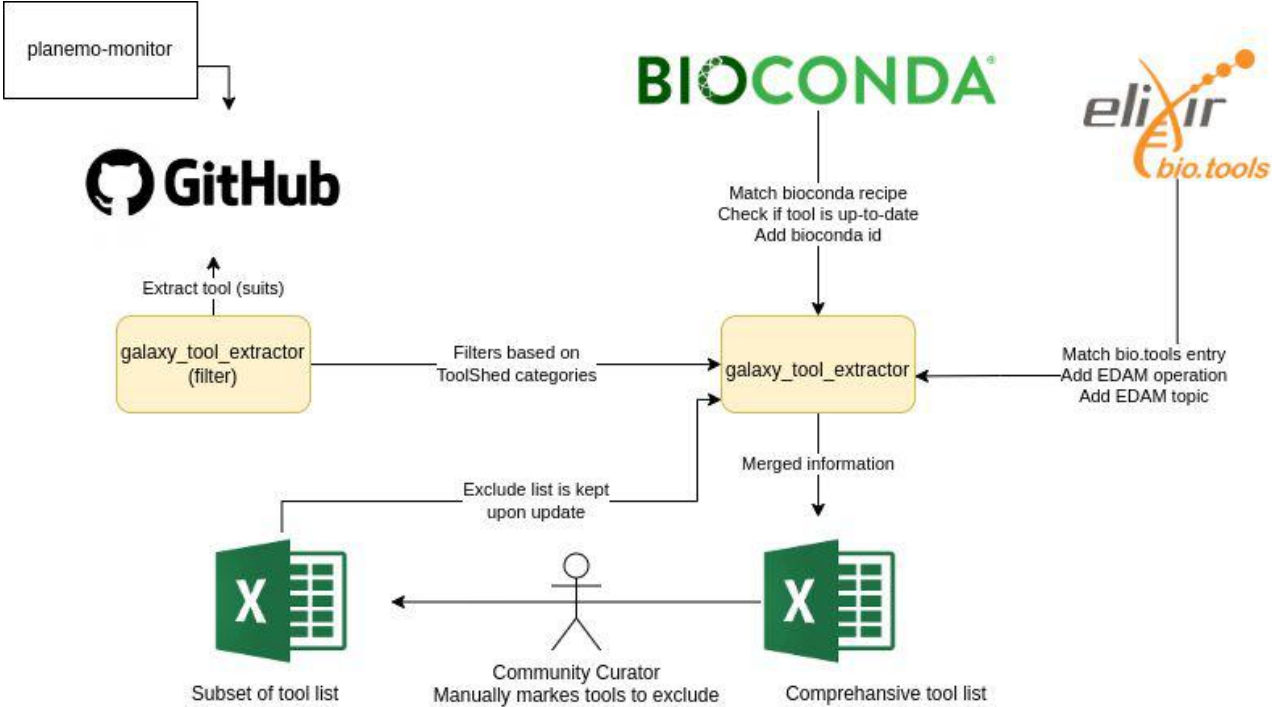
- Not interested all individual tools/wrappers but suites
- Hard to query programmatically



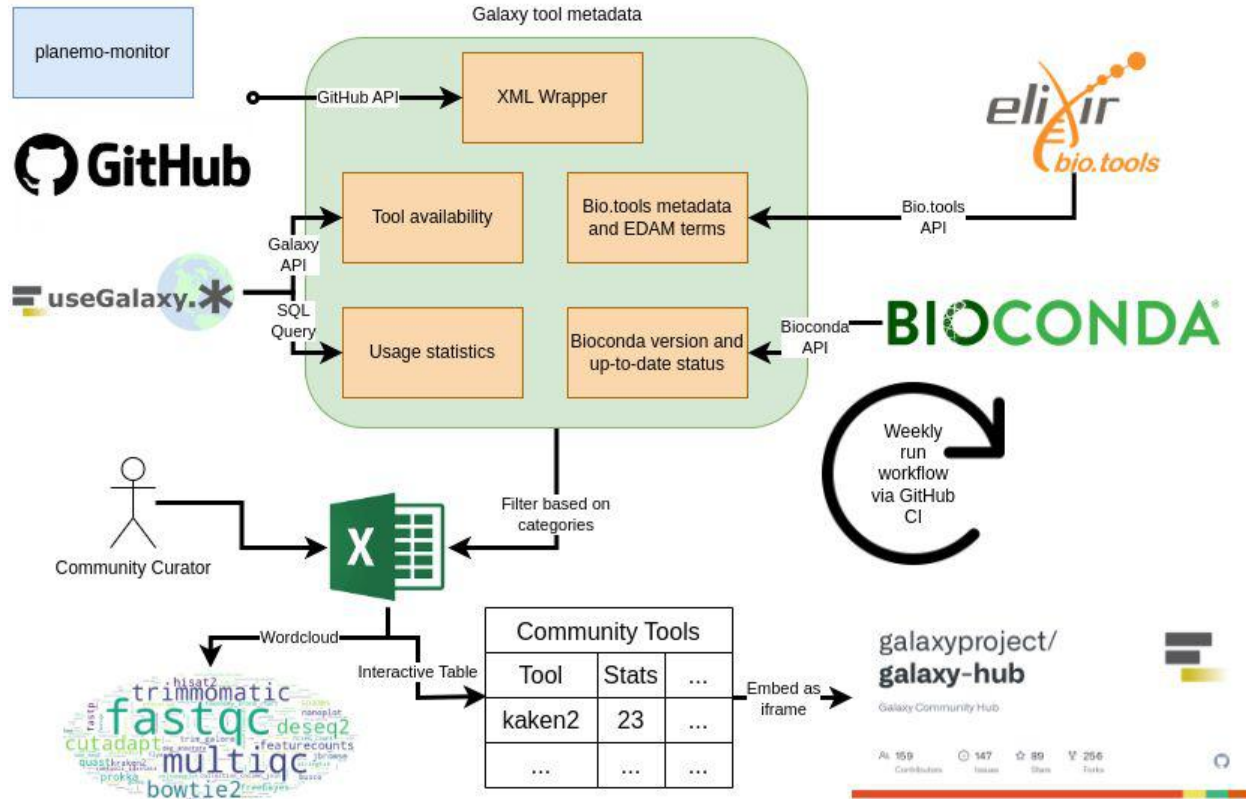
The screenshot shows the Galaxy Tool Shed interface. At the top, it says "Galaxy Tool Shed" and "9902 valid tools on May 30, 2024". The main heading is "Repositories by Category". There is a search bar for repository names and descriptions. Below the search bar is a table with three columns: Name, Description, and Repositories. The table lists various categories and their corresponding tool counts.

Name	Description	Repositories
Assembly	Tools for working with assemblies	210
Astronomy	Tools for astronomy	11
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	81
Climate Analysis	Tools for analyzing climate data	12
CLIP-seq	Tools for CLIP-seq	5
Combinatorial Selections	Tools for combinatorial selection	9
Computational chemistry	Tools for use in computational chemistry	184
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11
Convert Formats	Tools for converting data formats	147
Data Export	Tools for exporting data to various destinations	20
Data Managers	Utilities for Managing Galaxy's built-in data cache	109
Data Source	Tools for retrieving data from external data sources	112
Ecology	Tools related to ecological studies	107
Entomology	Tools that involve insect studies	4

Extract Tools: Galaxy Tool extractor



Galaxy Tool Extractor → Galaxy Tool Metadata Extractor



Galaxy Tool Metadata Extractor output

Output

1. JSON
2. TSV Table
3. Online, interactive, and queryable table

https://galaxyproject.github.io/galaxy_codex/

Including

- Metadata (EDAM, etc)
- Tool availability on usegalaxy.*
- Usage on usegalaxy.eu

Custom Search Builder (1)

Filter: Data | Condition | Value

Show 10 entries

Search: kraken

Expand	Galaxy wrapper id	Galaxy wrapper version	Conda version	Conda id	Status	bio.tool id	bio.tool name
	kraken2tax			1.2+galaxy0			
	bracken		2.9				
	kraken_biom		1.2.0				
	kraken_taxonomy_report		0.0.3				
	krakentools		1.2				
	pureclip		1.0.4				
	kraken						
	kraken2		2.1.1				

EDAM operation Taxonomic classification

EDAM topic Taxonomy, Metagenomics

Description Kraken2 for taxonomic designation.

bio.tool description Kraken 2 is the newest version of Kraken, a taxonomic classification system using exact k-mer matches to achieve high accuracy and fast classification.

bio

Status To update

Source <http://ccb.jhu.edu/software/kraken/>

ToolShed categories Metagenomics

ToolShed id kraken2

Galaxy wrapper owner iuc

Galaxy wrapper source https://github.com/galaxyproject/tools-iuc/blob/master/tool_collections/kraken2/kraken2/

Showing 1 to 8 of 8 entries (filtered from 1,336 total entries)

Previous 1 Next

Getting Tools for a community

1. Add your community to the Galaxy Tool Metadata Extractor pipeline by adding new folder in the data/community folder
2. Review the generated table to add status (keep/exclude/deprecated) to tools
3. Embed the interactive table in your community page on the Hub

Tutorial : <https://gxy.io/GTN:T00426>

The screenshot shows a Galaxy Training page with the following content:

- Galaxy Training!** header with navigation links: Development in Galaxy, Learning Pathways, Help, Settings, and Search Tutorials.
- Title:** Creation of an interactive Galaxy tools table for your community
- Author(s):** Bérénice Batut
- Overview:**
 - Questions:**
 - Is it possible to have an overview of all Galaxy tools for a specific scientific domain?
 - How can I create a new overview for a specific Galaxy community or domain?
 - Objectives:**
 - Create a community reviewed table for Galaxy tools within a specific scientific domain
 - Embed an interactive table in a community page
 - Time estimation:** 1 hour
 - Level:** Introductory
 - Supporting Materials:**
 - Published:** Mar 13, 2024
 - Last modification:** Feb 27, 2024
 - License:** Tutorial Content is licensed under [Creative Commons Attribution 4.0 International License](#). The GTN Framework is licensed under [MIT](#)
 - PURL:** <https://gxy.io/GTN:T00426>
 - Revision:** 2
- Table of Contents:**
 - Add your community to the Galaxy Tool Metadata Extractor pipeline
 - Review the generated interactive table
 - Embed the interactive table in your community page on the Hub
 - Conclusion
 - Frequently Asked Questions
 - Feedback
 - Citing this Tutorial
- Main Text:**

Galaxy offers thousands of tools. They are developed across various GitHub repositories. Furthermore, Galaxy also embraces granular implementation of software tools as sub-modules. In practice, this means that tool suites are separated into Galaxy tools, also known as wrappers, that capture their component operations. Some key examples of suites include [Methur](#) and [OpenMS](#), which translate to tens and even hundreds of Galaxy tools.

While granularity supports the composability of tools into rich domain-specific workflows, this decentralized development and sub-module architecture makes it **difficult for Galaxy users to find and reuse tools**. It may also result in Galaxy tool developers **duplicating efforts** by simultaneously wrapping the same software. This is further complicated by a lack of tool metadata, which prevents filtering for all tools in a specific research community or domain, and makes it all but impossible to employ advanced filtering with ontology terms and operations like [EDAM ontology](#).

The final challenge is also an opportunity: the global nature of Galaxy means that it is a big community. Solving the visibility of tools across this ecosystem and the potential benefits are far-reaching for global collaboration on tool and workflow development.

To provide the research community with a comprehensive list of available Galaxy tools, [Galaxy Tool Metadata Extractor](#) was developed to collect Galaxy wrappers from a list of Git repositories and automatically extract their metadata (including Conda version, [bio.tools](#) Identifiers, and EDAM annotations). The workflow also queries the availability of the tools and usage statistics from the three main Galaxy servers (usegalaxy.*).

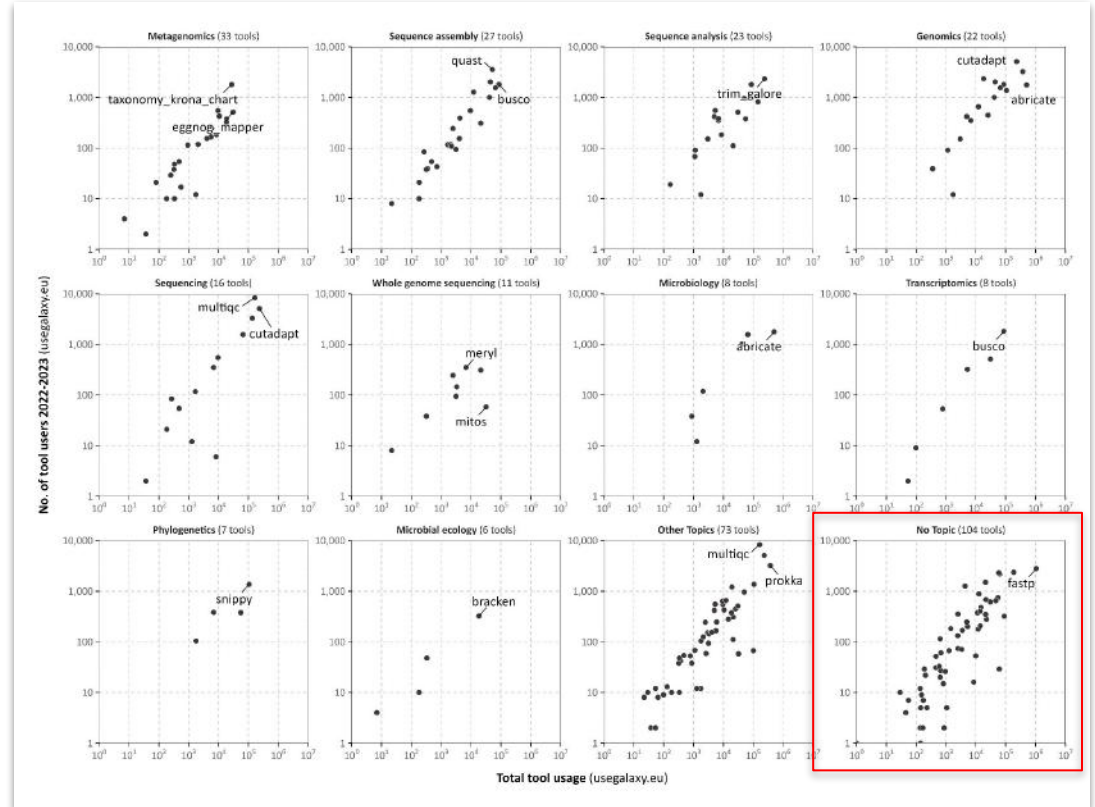
Example of the microGalaxy community

300+ tools



BUT

100+ tools without EDAM annotations



Community effort to improve tool annotation

[← Back to Events](#)

Hackathon - Improving the annotation of Galaxy resources for microbial data analysis and beyond

This hackathon aims to improve the annotation of Galaxy resources for microbial data analysis and beyond

Date: March 11 - 15, 2024

Contact: Bérénice Batut, Paul Zierep

Objectives

The objective of this hackathon is to improve the annotation of the Galaxy resources (tools, workflows, tutorials) for data analysis by:

- Linking microbial Galaxy tools to bio.tools to obtain EDAM ontology annotation
- Improving bio.tools annotations
- Annotating existing microbial-related tutorials with EDAM terms
- Reflecting on the addition of EDAM terms to workflows
- Reflecting on missing terms in the EDAM ontology for microbial data analyses
- Brainstorming about a way to connect tool annotations to improve training and workflow

This hackathon is open to other communities than microGalaxy

Adding and updating best practice metadata for Galaxy tools using the bio.tools registry

Author(s) Bérénice Batut Johan Gustafsson Paul Zierep

Overview

Questions:

- How are Galaxy tools linked to EDAM ontology?
- How to connect Galaxy tools to bio.tools?
- What is bio.tools?
- How to add, and update, entries in bio.tools?

Objectives:

- Identify Galaxy tools without bio.tools entry
- Create a bio.tools entry
- Update a bio.tools entry
- Add EDAM ontology terms to a bio.tools entry
- Link a Galaxy tool to its corresponding bio.tools entry

Time estimation: 1 hour

Level: Introductory

Supporting Materials:

Published: Mar 6, 2024

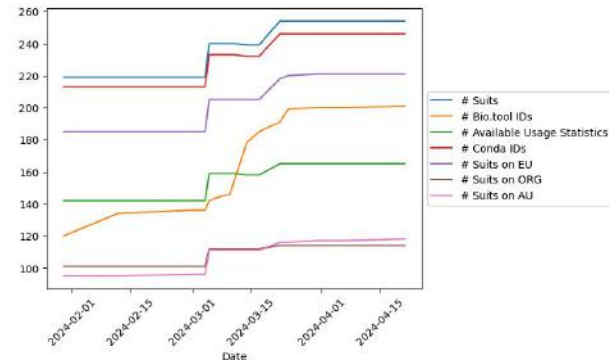
Last modification: Mar 5, 2024

License: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License. The GTN Framework is licensed under MIT

PURL: <https://gxy.io/GTN:T09421>

Revision: 8

Progress of the Galaxy Tool Metadata Extractor

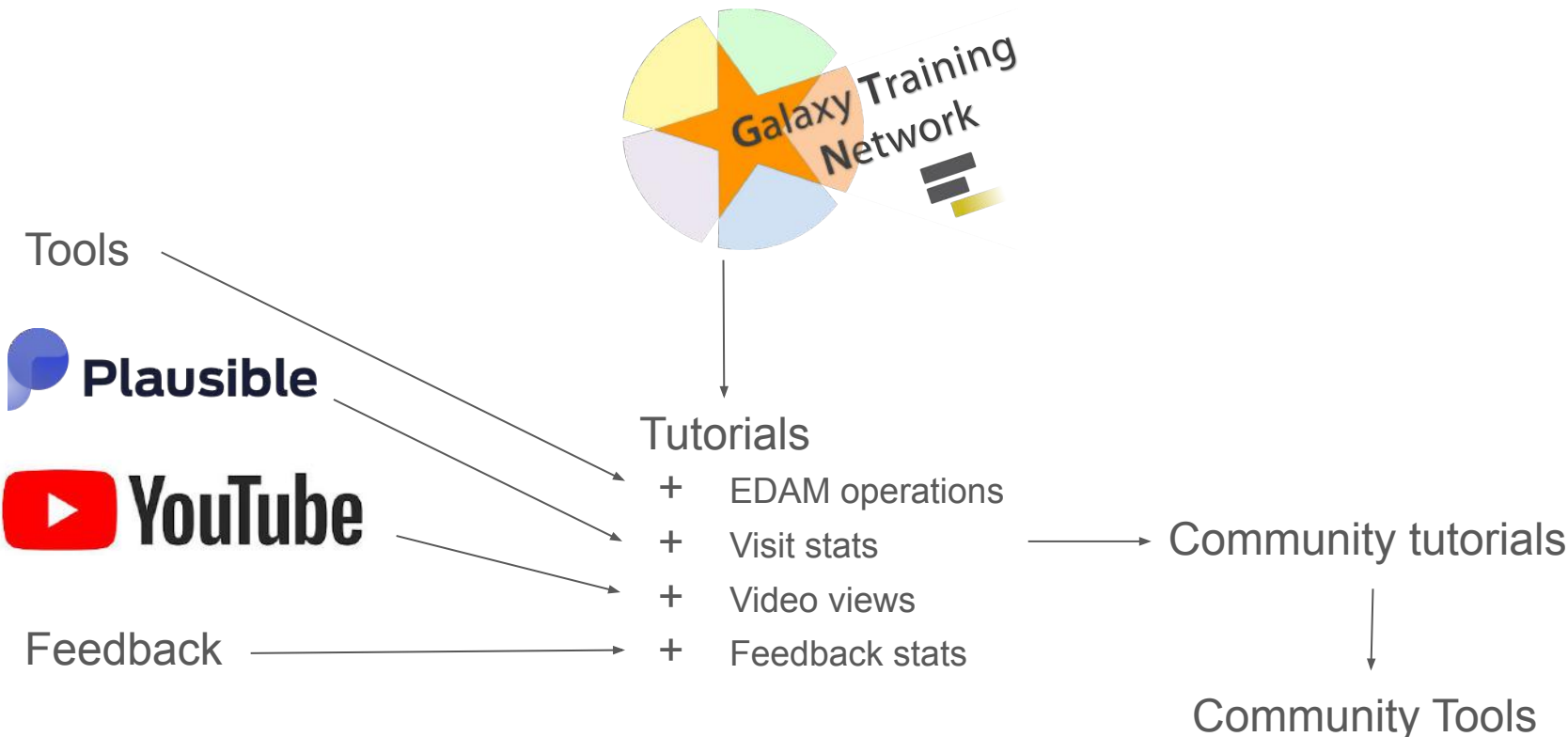




Beyond Tools

From Galaxy Tool Metadata Extractor to
Galaxy Communities Dock or Galaxy CoDex

Extract Training





WHAT
NOW

Next steps

1. Extract Workflows

Next steps

1. Extract Workflows
2. Create a packaging for community resources
 - a. Home page template
 - b. Deployment configuration files

Next steps

1. Extract Workflows
2. Create a packaging for community resources
 - a. Home page template
 - b. Deployment configuration files
3. Develop best practices

Next steps

1. Extract Workflows
2. Create a packaging for community resources
 - a. Home page template
 - b. Deployment configuration files
3. Develop best practices
4. Integrate more communities
 - a. Current communities
 - i. microGalaxy
 - ii. Imaging
 - iii. Biodiversity
 - b. Plan
 - i. Single-cell
 - ii. Yours?

5. Involve the whole Galaxy community

Who?

- Community leads
- Community members
- Tool developers
- Web developers
- Admin
- PIs
- Everyone

To do what?

- Test
- Give Feedbacks
- Suggest Ideas / Needs
- Help with interface
- & more

Goal: Make the Galaxy CoDex for the community by the community

Thank You!

github.com/galaxyproject/galaxy_codex

