Biodiversity
Genomics
Europe

WorkflowHub

The BGE guide to using WorkflowHub

The Biodiversity Genomics Europe project of 33 partners aims to align the resources and research agendas of two networks working in DNA barcoding (BIOSCAN Europe) and reference genome generation (European Reference Genome Atlas). A driver is to build capacity and share knowledge between the networks, and democratise participation within the networks. A major pillar of BGE is the development of large-scale biodiversity genomic data generation pipelines for Europe to accelerate the production and accessibility of genomic data for biodiversity characterisation, conservation and biomonitoring.

The BGE project is producing pipelines (aka workflows). We need to register those workflows in WorkflowHub:

- To give visibility to the workflows created by the project and by the different networks within the project
- To give visibility to the workflows used by project that were created by the different networks
- To share workflows across the project, within the networks and externally
- To credit and cite the people making the workflows and the networks to which they belong
- To track the versions of workflows as they are produced, and
- To retain the workflows for post-project use by the networks

BGE workflows organisation in WorkflowHub

The Biodiversity Genomics Europe project is complex. The networks that have come together need to be visible as workflow developer communities in their own right. At the same time they need to have both the workflows they bring to the project, and develop for the project, visible as BGE workflows.

The WorkflowHub has 3 primary organisational structures that we will use in BGE.

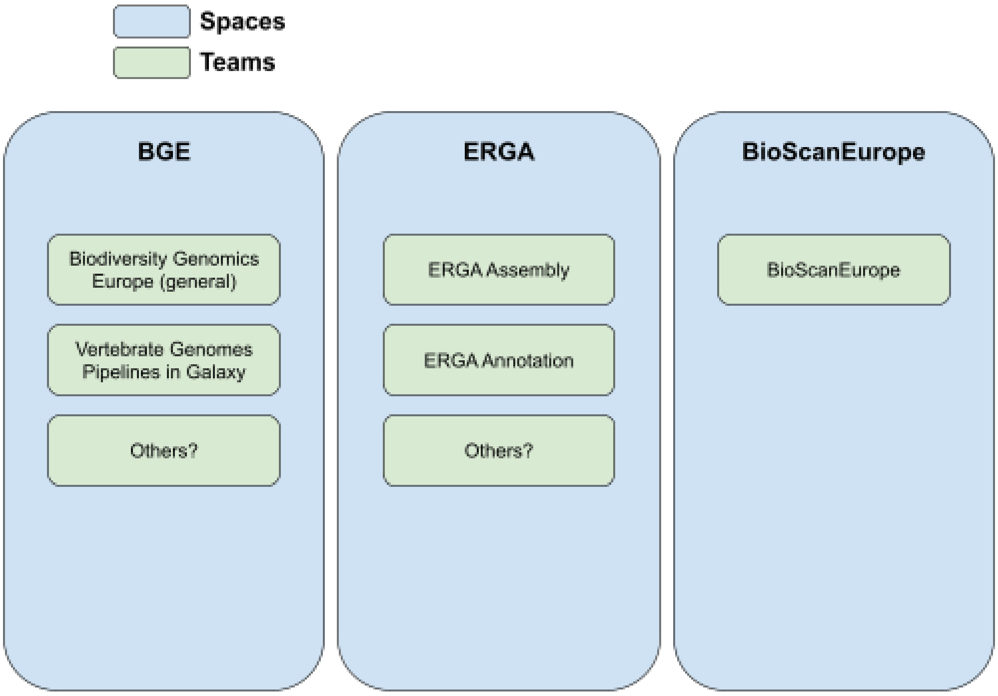
Space	Describes a broad activity, such as a project consortium or a long-running activity. Within the Space are Teams. A Space is self-administered. The administrator has the power to create Teams. A Space can have any number of Teams. A Team can only be associated with one Space.
Team	A group of one or more people collaborating together on a particular activity. The Team provides the ability to describe and promote the activity and give credit back to the team. A Team is self-administered. The Team administrator has the power to control the Team membership, and set sharing permissions and licence across all members. **Before you can register anything you need to be a member of the Team.** You can be a member of any number of Teams. All workflows have to be registered with at least one Team, and can be registered with any number of Teams, which could belong to different Spaces.

Collection	A group of workflows explicitly curated into a collection. There are ways of gathering workflows in one collection from across different Teams and different Spaces. A collection is self-administered and managed by one or more nominated maintainers. The workflows that are being put into the collection must already be registered with at least one Team. A workflow can be registered with any number of collections.

BGE Spaces and Teams

In BGE we have the following Spaces, Team and Collections.

Space	**Teams**
BGE Admin: Stian Soiland-Reyes	Biodiversity Genomics Europe (general) This Team will register the workflows that can be organised into collection(s) of workflows used in BGE. Vertebrate Genomes Pipelines in Galaxy Admin: Stian <what other teams should be here?>
ERGA Admin:Tom Brown, Diego De Panis	ERGA Annotation (1) Admin: Tom Brown ERGA Assembly (7) Admin: Tom Brown <what other teams should be here?>
BioScanEurope Admin:	BioScanEurope Admin: <what other teams should be here>



BGE Collections

Collection	Administering Team
Vertebrate Genomes Pipelines (VGP) workflows	Vertebrate Genomes Pipelines in Galaxy
ERGA Assembly Galaxy HiFi & HiC Pipelines (Hifiasm-HiC + Purge_Dups + YaHS)	ERGA Assembly
Workflows used by BGE project (We may on request add these or others: BGE pillar: Node development, BGE pillar: Sampling, BGE pillar: Sequencing, BGE pillar: Data, PGE pillar: Applications) <what other teams should be here?> By default any member of the Biodiversity Genomics Europe (general) team has access to add to the Workflows used by BGE project collection.	Biodiversity Genomics Europe (general)

Collections are administered by a set of maintainers. By default any member of the Biodiversity Genomics Europe (general) team can manage the Workflows used by BGE project collection. Members of other teams can be added as “Maintainers” if required, which allows them to add and remove workflows from the collection.

BGE workflow developers will be invited to help curate the BGE Teams and Collections.

Registering BGE workflows in WorkflowHub

Joining WorkflowHub

1. Login into WorkflowHub
    a. Register an account https://workflowhub.eu/signup, or
    b. Log-in using GitHub or institutional credentials.

2. Make sure your WorkflowHub institution matches one of the listed BGE partners
   https://cordis.europa.eu/project/id/101059492
   https://biodiversitygenomics.eu/the-project/network/

3. Request to join the team Biodiversity Genomics Europe (general)

4. Request to join any additional teams that may be relevant to you from the ERGA space or BioScanEurope Space.

Registering a BGE workflow in WorkflowHub

To register a workflow:

1. Log into WorkflowHub
2. Pick “Workflow” from the “+” menu in the top navigation bar.
3. You will be asked for the Team(s) it should be registered with. A workflow may be registered with many Teams. **You must be a member of the Team to be able to register it there.** The choices are:
    - The Biodiversity Genomics Europe (general) team
    - One of the more specific teams, such as ERGA Assembly.

4. Any BGE workflow should be added to the Workflows used by BGE project collection.
    - Remember to make sure you are a member of the Biodiversity Genomics Europe (general) team to be able to do this.

5. Add description metadata to the workflow including tags. Please tag workflows with ”BGE” and if appropriate “ERGA” and/or “BioScanEurope”. Other tags you might consider: “Biodiversity", "Genomics", “Genome Assembly”, “Assembly”, “Annotation”, “Sampling”, “Sequencing”. New tags can be added if required. <what other tags do we recommend?>

Adding and removing a Workflow to one of the BGE collections

To add a Workflow to a Collection

1. Log into WorkflowHub
2. Select the workflow you want to add by selecting the magnifying glass icon on the top bar next to the WorkflowHub logo
3. Select the button “Add Collection”in the top right (next to all the other buttons) for the Collections that you have access to. Select the one you want to add to
4. If the Collection you want to add to isn’t there then you are not in a Team associated with that Collection. You will need to go look at the Collection to see the Teams associated with it and request to join one of them.

To remove a Workflow from a Collection or reorder it within the Collection

1. Log into WorkflowHub
2. Go to the Collection
3. Select the “Actions” button top right
4. To remove a workflow, click the “Delete?” checkbox.
5. To re-order a workflow in the Collection click and drag the item using the blue numbered button on the left side above or below another item.
6. To modify the comment on an item, type into the text box.
7. Click the “Update” button at the bottom of the page

Adding a workflow to another BGE team

You uploaded a Workflow to a Team and now want to add it to another Team, maybe in a different Space:

1. Log into WorkflowHub
2. Select “Manage Workflow” from the “Actions” drop down menu.
3. In the “Manage Workflow” pop-up, select the team that you would like to add to
4. Click the “Update” button at the bottom of the screen.

Moving a workflow from one BGE team to another

You realised you put the workflow in the wrong Team?

1. Log into WorkflowHub
2. Select “Manage Workflow” from the “Actions” drop down menu.
3. In the “Manage Workflow” pop-up:
    - Add the workflow to the Team you want it added to.
    - Click the red cross next to the Team name you want it removed from.
4. Click the “Update” button at the bottom of the screen.

Linking your workflow GitHub repository with WorkflowHub

During registration of the workflow

1. Log into WorkflowHub

2. Pick “Workflow” from the “+” menu in the top navigation bar.
3. Make sure to choose the “Import Git Repository” option in the workflow registration wizard
4. Add your GitHub repository URL and click “Register”.

Using the LifeMonitor App for automatic integration and versioning of workflows

1. Follow the instructions available in the LifeMonitor documentation to install the app in your repository or organisation.
2. Make sure to configure the “GitHub integration” in the “Settings” menu to:
    - Version based on a specific branch, and
    - Automatically check for issues and suggest pull requests
3. To integrate with WorkflowHub you need to activate this feature in the “Registry Integration” tab in the “Settings” menu. You also need to give the LifeMonitor app access permissions for your WorkflowHub account.

Keeping workflows up-to-date

Workflows will keep being updated and those new versions need to appear on the WorkflowHub. Before you update a registered workflow, you need to ensure that you

- are logged into WorkflowHub.
- have the rights to change the workflow.
- are on the workflow page of the workflow.

To update the metadata of a workflow

1. Go to the “Actions” dropdown menu
2. Click “Edit Workflow” to change the metadata of the registered workflow.

Note that updates made in the metadata will not affect the metadata in the RO-crate, yet (upcoming feature).

To update the workflow version manually

If the workflow was imported via Git

Files can be added/removed/modified via the “Files” tab if the current version is “Open”.

A version must be “Frozen” before a new version can be created, or before a DOI can be minted.

To freeze a version

- Go to the “Actions” dropdown menu
- Click “Freeze Version” to lock the current version to changes

To create a new version

- Go to the “Actions” dropdown menu
- Click “New Version”
- Decide either to continue managing the files on WorkflowHub (“Upload/Import files”), or switch to using a Git repository (“Import Git Version”).

If the workflow was uploaded manually, or via an RO-Crate

- Go to the “Actions” dropdown menu
- Click “New Version”
- Either confirm the repository URL, or enter a different one (in the event that the workflow has moved to a different repository)
- Select the branch/tag to import
- You may be prompted to reselect the main workflow if it can no longer be found
- Check/complete the metadata

To update the workflow version automatically from GitHub

Updating workflow versions automatically from GitHub or local Git is a planned feature.

To change the main workflow/abstract CWL/diagram

- Click on the “Files” tab
- Click “Annotate Files”
- Select a new main workflow/abstract CWL/diagram by clicking the respective field and navigating to the appropriate file in the popup that appears
- Optionally tick the “Re-extract metadata?” checkbox if you would like WorkflowHub to attempt to extract metadata from the newly selected main workflow/abstract CWL
- Optionally check and confirm the metadata if the checkbox was ticked

Linking workflows with other workflows

Workflows can be linked to other workflows in a number of ways

1. **By being in the same collection** (see adding workflows to collections)
2. **By attribution**, linking a workflow to one that it is derived from, as workflows are often variants of others. To attribute a workflow:
    - Go to the *Actions* dropdown menu
    - click **Edit Workflow** and scroll down to the *Attributions* section, and click to expand it if it is not yet open.
    - Begin typing the title of a workflow to search, and select one of the results that appear.
    - Scroll down and click **Update** to confirm the changes.

Publishing workflows

Workflows registered in https://workflowhub.eu/ that are *public* can be assigned a DOI (Digital Object Identifier) so that they can more easily be cited in academic work and enter the scholarly knowledge graphs.

WorkflowHub DOIs are registered with DataCite using the DataCite Metadata Schema, meaning that the metadata joins the scholarly knowledge graphs DataCite Commons PID Graph and OpenAIRE Research Graph (example).

You can assign DOIs for workflows you have registered and that you have been given write access to. You are not able to make DOIs for other people's workflows.

The citation will appear on the workflow's overview page.

Assigning a DOI has two steps

1. **Check metadata**

   The metadata for the workflow will be part of the DataCite registration.

   Before you mint a DOI, make sure you check the correctness of the metadata of the workflow entry, and that it makes sense globally,.

   Use *Actions -> Edit Workflow* and *Actions -> Manage workflow* to fill in:

   - Title (is it specific enough)
   - Creators (you can include registered WorkflowHub users or others by name)
   - Status of workflow
   - Description
   - License

   For each of the registered users, check that their ORCID identifier has been registered. This ensures that the people will be uniquely identified. Users can edit *"My Profile"* to link to ORCID.
2. **Freeze the workflow version**

   Once a DOI is assigned that workflow version is fixed as a snapshot. So before minting a DOI, we need to *freeze* the version of the workflow in WorkflowHub, so that its metadata stays consistent with the DOI registration.

   Each version of the workflow must be registered separately.

   You can create a new version of a workflow that has been assigned a DOI, without necessarily giving the new version a DOI yet.

Adding new teams and collections

Adding new BGE teams

Adding a Team:

- Log into WorkflowHub
- Pick “Team” from the “+” menu on the top bar.
- Untick the “WorkflowHub managed Space? ”
- Enter “BGE” into the “Specify a title for a new Space” field.
- Add the Team metadata - the more the better. Remember this will be displayed and no description looks bad.
- Add the name of the affiliation that you are associated with for this Team. It doesn’t have to be the same Institution as other Teams, and it is not unusual for it to be different. Start to type the name of your institution, and existing options will be displayed for selection. If not, choose the top option to define a new one.
- The Administrators will be notified and you will get an email confirmation.
- Once approved you can add members or others can apply to join. As you created the Team you are the administrator. You can add other administrators.

Adding new BGE collections

Adding a Collection:

- Log into WorkflowHub
- Pick “Collection” from the “+” menu on the topbar.
- Add the Teams associated with the collection. These will only be Teams that you are a member of.
- Add the Maintainers of the Collection. You can add any WorkflowHub member as a maintainer, and they will be able to add and remove workflows from the collection.
- Tag Collections with ”BGE” and if appropriate “ERGA” and/or “BioScanEurope”. Other tags you might consider: “Biodiversity", "Genomics", <what other tags do we recommend?>

Other WorkflowHub features

Standard Operating Procedures

WorkflowHub can also be used to register SOPs. To register SOPs, you must be registered and logged into your account. SOPs can be uploaded in a variety of formats (word files, e-lab notebooks, code, annotated spreadsheets etc), as well as being registered as a remote URL. Typically, projects will define through a template what form their SOPs will take. BGE should define such a template.

SOPs are assets just like Workflows. So anything you can do with a workflow (above) you can do with a SOP.

- Log into WorkFlowHub
- From the top bar, select **'+ Create'**
- From the dropdown menu, select **'SOP'**
- Select whether you wish to upload a 'local file', using the '**choose file**' button, or enter a' Remote URL' to locate your SOP
- Complete the fields, this includes:
    - Teams that the SOP is being registered with
    - Sharing (publicly, with individuals, teams or spaces) - the wider the sharing the better visibility.
    - Attributions links to related SOPs, so workflows can be linked together.
    - Workflows links the SOP to existing workflows in WorkFlowHub
- Click **'Register'** to submit the SOP

Events, Presentations and Publications

WorkflowHub supports metadata and links to Events, and registering Presentations and Publications.

- Select Event from the + Create menu
- Fill out the metadata
-

Events can be linked to Publications and Presentations

- Go to the "Actions" dropdown menu
- Click "Edit Event"
- Scroll to the "Publications" or "Presentations" section and click to expand it
- Select the desired publication or presentation from the dropdown menu
- Click the "Update" button

Presentations, publications and SOPs can be linked to workflows in a similar way.

Further documentation

- [WorkflowHub documentation]()
- [WorkflowHub AMA slides]()