Instituto Superior de Engenharia de Lisboa
BSc in Computer Science and Computer Engineering
Project and Seminary
Progress Update

PHYLOViZ Web Platform

A Modular and Web-Based Tool for Phylogenetic Analysis

Supervisors:

- Cátia Vaz, ISEL
- Alexandre P. Francisco, IST

48089 André Páscoa
48280 André Jesus
48287 Nyckollas Brandão

May 2023

Table of contents

01 Introduction

02 Motivation

03 Architecture

04 Implementation

05 Technologies

06 Progress

07 Demo

08 Next Steps

01

Introduction

What’s **Phylogenetic Analysis**?

Introduction

Phylogenetic Analysis is a field of biomedical research which allow to understand the evolution of bacterial and viral epidemics.

Phylogenetic Analysis

01 Alignment of genetic sequences

02 Application of typing methodology

03 Application of phylogenetic inference methods

04 Visualization

02

Motivation

Related Tools

PHYLOViZ

Desktop application for Phylogenetic Analysis

PHYLOViZ Online

Web application for Phylogenetic Analysis

FLOWViZ

Workflow Manager

PhyloDB

Graph-Oriented Database

Phylolib

Library of Phylogenetic Analysis Algorithms

Problems

- Features in both versions of PHYLOViZ are not the same
- PHYLOViZ Online is not modular, which makes it difficult to extend and maintain
- Current solutions do not scale for large data analysis and visualization
- Results and optimizations are not stored for reuse

Solution: PHYLOViZ Web Platform

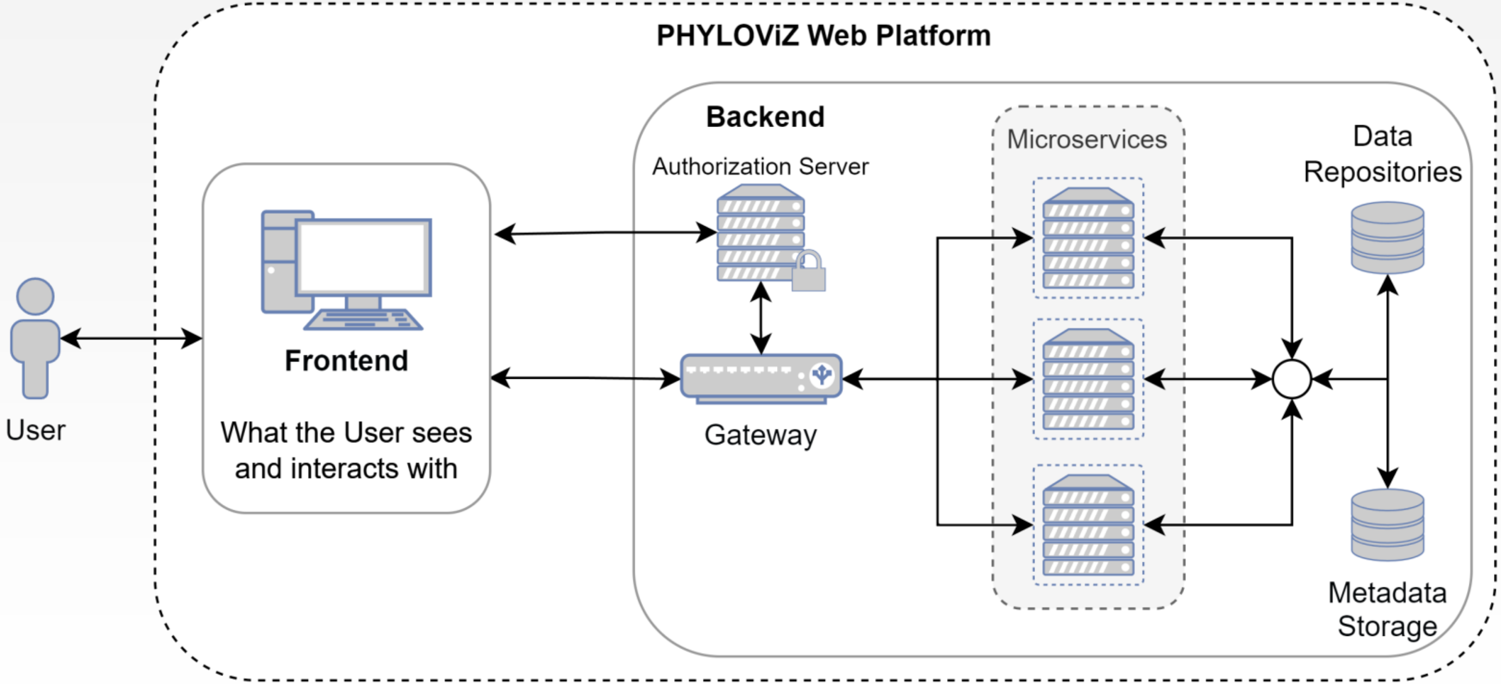
Modular Architecture

Web Based

Advanced Data Management

03

Architecture



PHYLOViZ Web Platform

User

Frontend

What the User sees and interacts with

Backend

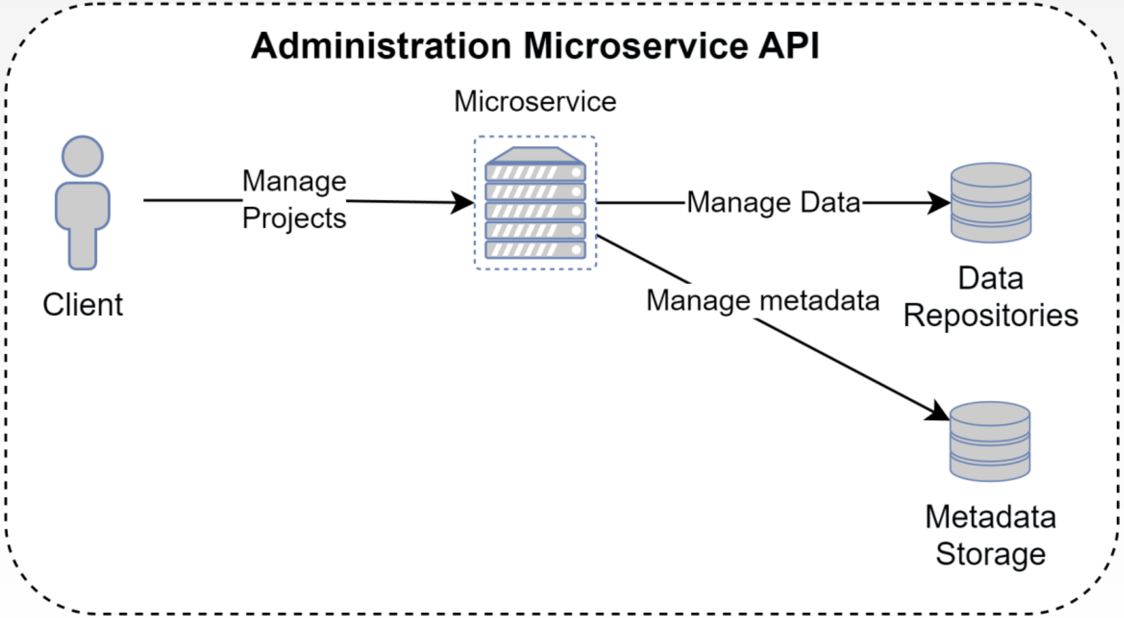
Authorization Server

Gateway

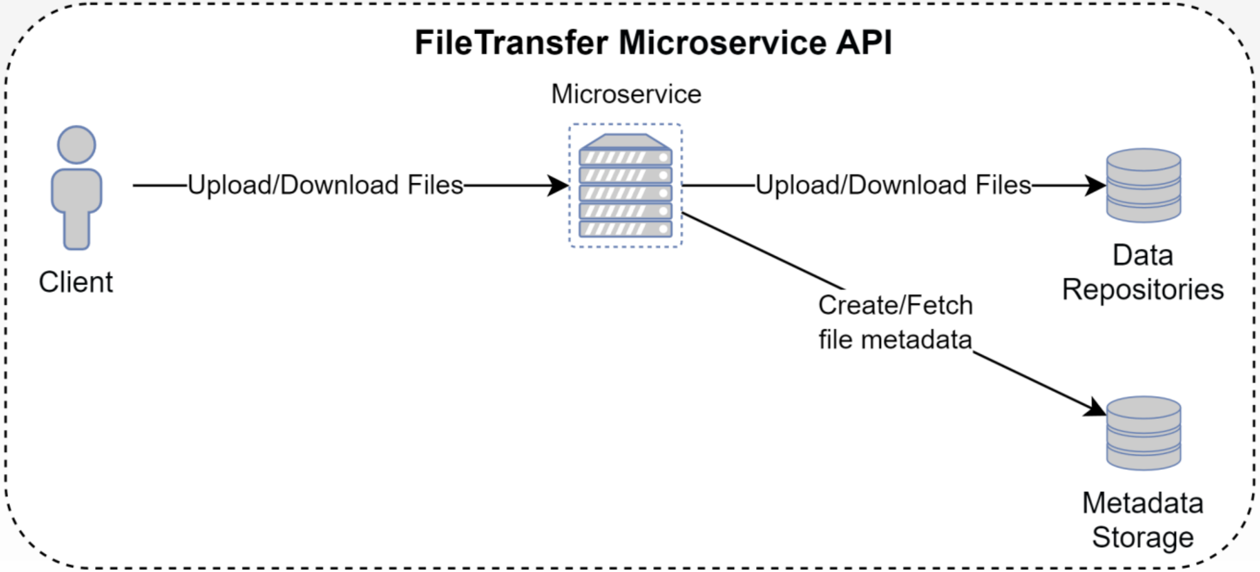
Microservices

Data Repositories

Metadata Storage


Administration Microservice API
Microservice
Client
Manage Projects
Manage Data
Manage metadata
Data Repositories
Metadata Storage

FileTransfer Microservice API



Microservice

Client

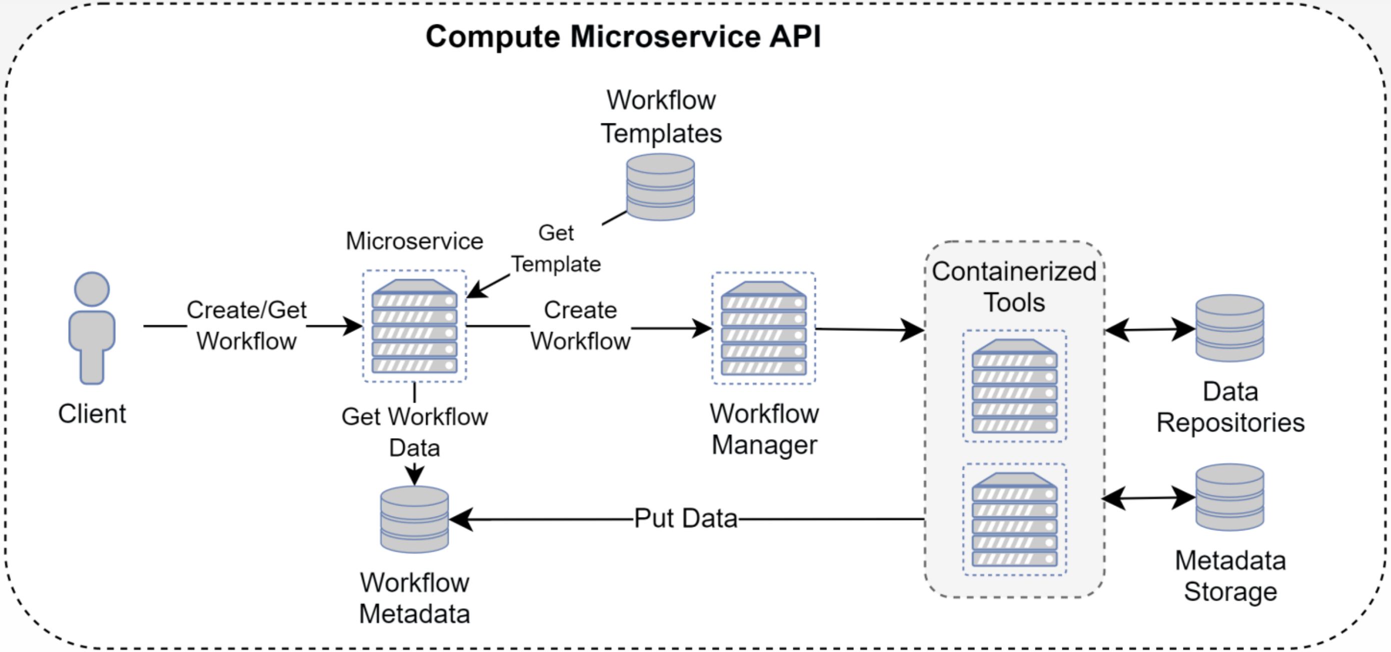
Upload/Download Files

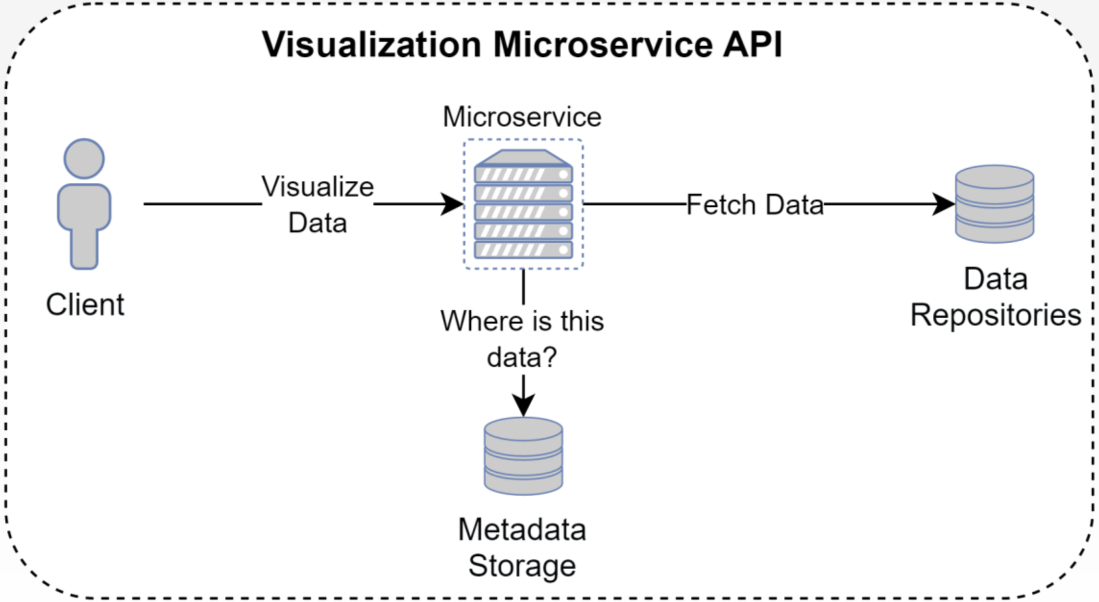
Upload/Download Files

Data Repositories

Create/Fetch file metadata

Metadata Storage


Compute Microservice API
Workflow Templates
Get Template
Microservice
Client
Create/Get Workflow
Create Workflow
Workflow Manager
Containerized Tools
Data Repositories
Metadata Storage
Get Workflow Data
Put Data
Workflow Metadata


Visualization Microservice API
Microservice
Client
Visualize Data
Fetch Data
Data Repositories
Where is this data?
Metadata Storage

04

Implementation

Data Model

Phylogenetic Data

PhyloDB

Model and Database used for first deployment. Uses the graph database **neo4j**.

OpenStack S3

Object storage of OpenStack used for file storage.

Others

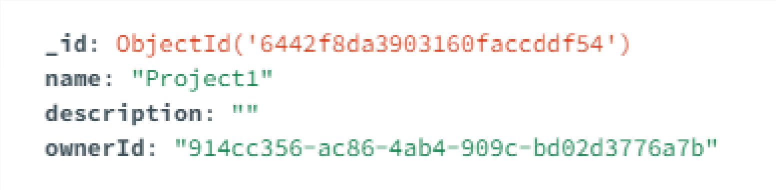
Other repositories can be easily used, even at the same time, because of metadata storage.

Metadata Document based

Abstraction between application and data repositories. Stores general information of each resource and access information for each data representation.

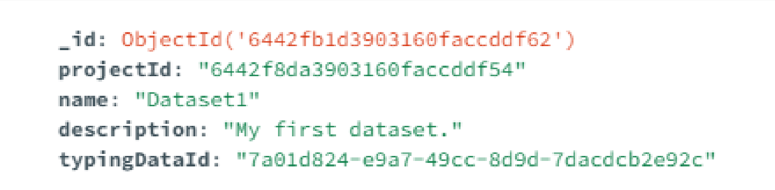
Metadata Documents

Project



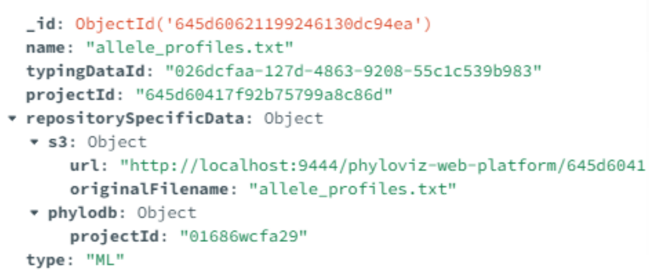
```
_id: ObjectId('6442f8da3903160faccddf54')
name: "Project1"
description: ""
ownerId: "914cc356-ac86-4ab4-909c-bd02d3776a7b"
```

Dataset



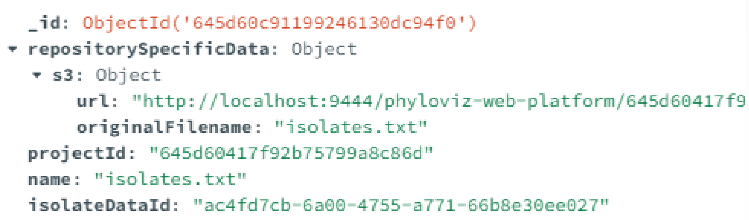
```
_id: ObjectId('6442fb1d3903160faccddf62')
projectId: "6442f8da3903160faccddf54"
name: "Dataset1"
description: "My first dataset."
typingDataId: "7a01d824-e9a7-49cc-8d9d-7dacdcb2e92c"
```

Typing Data



```
  _id: ObjectId('645d60621199246130dc94ea')
  name: "allele_profiles.txt"
  typingDataId: "026dcfaa-127d-4863-9208-55c1c539b983"
  projectId: "645d60417f92b75799a8c86d"
▾ repositorySpecificData: Object
  ▾ s3: Object
      url: "http://localhost:9444/phyloviz-web-platform/645d6041
      originalFilename: "allele_profiles.txt"
  ▾ phylodb: Object
      projectId: "01686wcfa29"
  type: "ML"
```

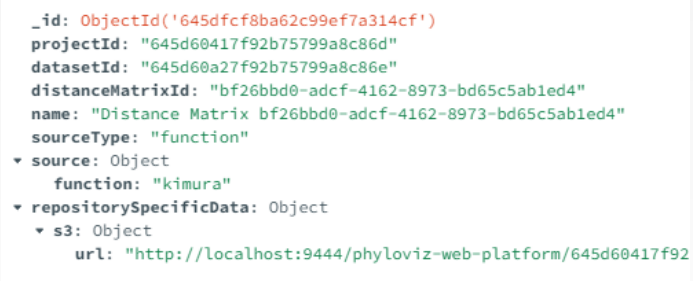
Isolate Data



```
  _id: ObjectId('645d60c91199246130dc94f0')
▾ repositorySpecificData: Object
  ▾ s3: Object
      url: "http://localhost:9444/phyloviz-web-platform/645d60417f9
      originalFilename: "isolates.txt"
  projectId: "645d60417f92b75799a8c86d"
  name: "isolates.txt"
  isolateDataId: "ac4fd7cb-6a00-4755-a771-66b8e30ee027"
```

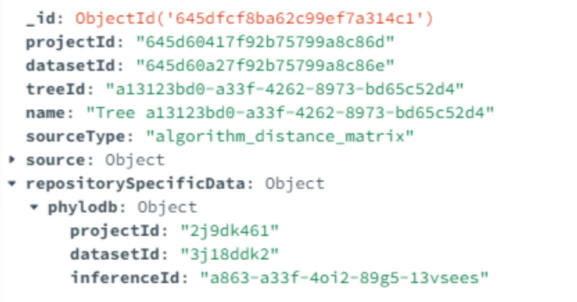
Metadata Documents

Distance Matrix



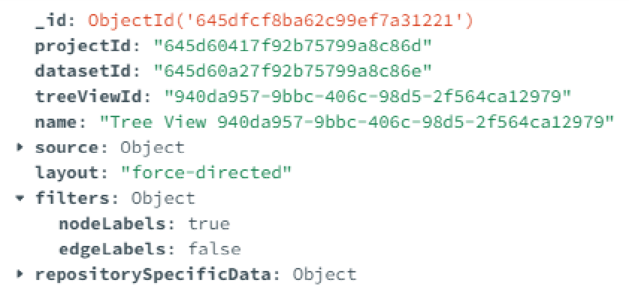
```
  _id: ObjectId('645dfcf8ba62c99ef7a314cf')
  projectId: "645d60417f92b75799a8c86d"
  datasetId: "645d60a27f92b75799a8c86e"
  distanceMatrixId: "bf26bbd0-adcf-4162-8973-bd65c5ab1ed4"
  name: "Distance Matrix bf26bbd0-adcf-4162-8973-bd65c5ab1ed4"
  sourceType: "function"
▾ source: Object
    function: "kimura"
▾ repositorySpecificData: Object
  ▾ s3: Object
      url: "http://localhost:9444/phyloviz-web-platform/645d60417f92
```

Tree



```
  _id: ObjectId('645dfcf8ba62c99ef7a314c1')
  projectId: "645d60417f92b75799a8c86d"
  datasetId: "645d60a27f92b75799a8c86e"
  treeId: "a13123bd0-a33f-4262-8973-bd65c52d4"
  name: "Tree a13123bd0-a33f-4262-8973-bd65c52d4"
  sourceType: "algorithm_distance_matrix"
▸ source: Object
▾ repositorySpecificData: Object
  ▾ phylodb: Object
      projectId: "2j9dk461"
      datasetId: "3j18ddk2"
      inferenceId: "a863-a33f-4oi2-89g5-13vsees"
```

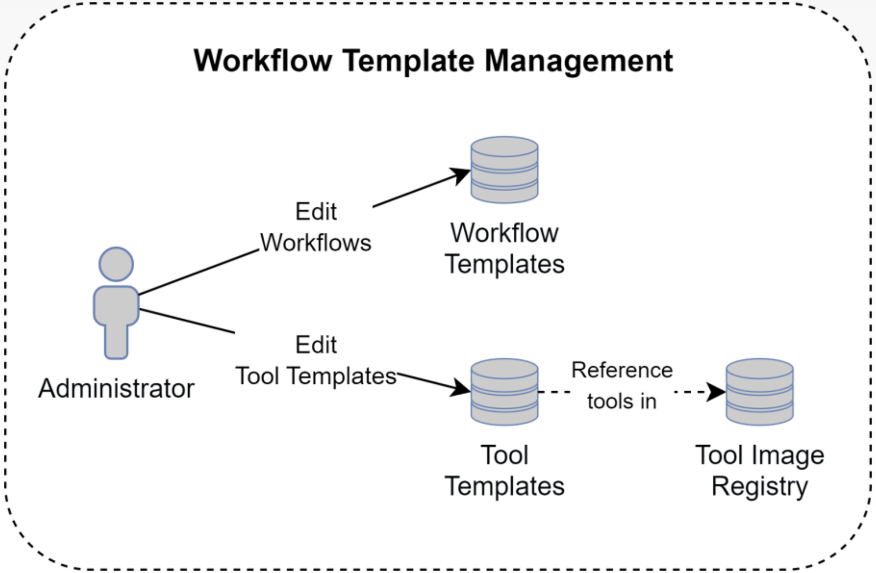
Tree View



```
  _id: ObjectId('645dfcf8ba62c99ef7a31221')
  projectId: "645d60417f92b75799a8c86d"
  datasetId: "645d60a27f92b75799a8c86e"
  treeViewId: "940da957-9bbc-406c-98d5-2f564ca12979"
  name: "Tree View 940da957-9bbc-406c-98d5-2f564ca12979"
▸ source: Object
  layout: "force-directed"
▾ filters: Object
    nodeLabels: true
    edgeLabels: false
▸ repositorySpecificData: Object
```

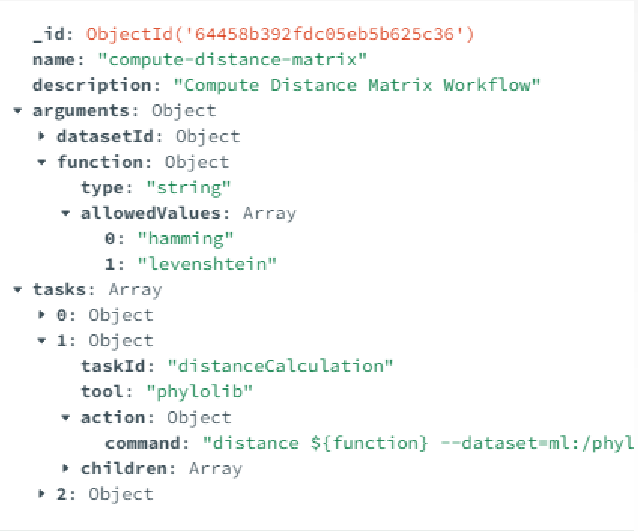
Workflows

- Workflow and tool templates are managed by the system administrator and stored in a document database.
- Workflows and tools can be added, edited and removed during runtime.
- Ideally tool images are previously uploaded to the custom Docker image registry.



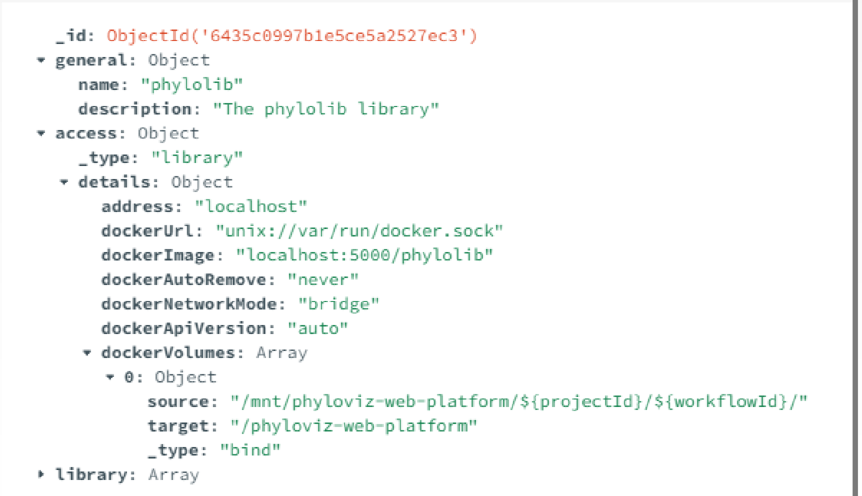
Workflow and Tool Documents

Workflow Template



```
  _id: ObjectId('64458b392fdc05eb5b625c36')
  name: "compute-distance-matrix"
  description: "Compute Distance Matrix Workflow"
▾ arguments: Object
  ▸ datasetId: Object
  ▾ function: Object
      type: "string"
    ▾ allowedValues: Array
        0: "hamming"
        1: "levenshtein"
▾ tasks: Array
  ▸ 0: Object
  ▾ 1: Object
      taskId: "distanceCalculation"
      tool: "phylolib"
    ▾ action: Object
        command: "distance ${function} --dataset=ml:/phyl
    ▸ children: Array
  ▸ 2: Object
```

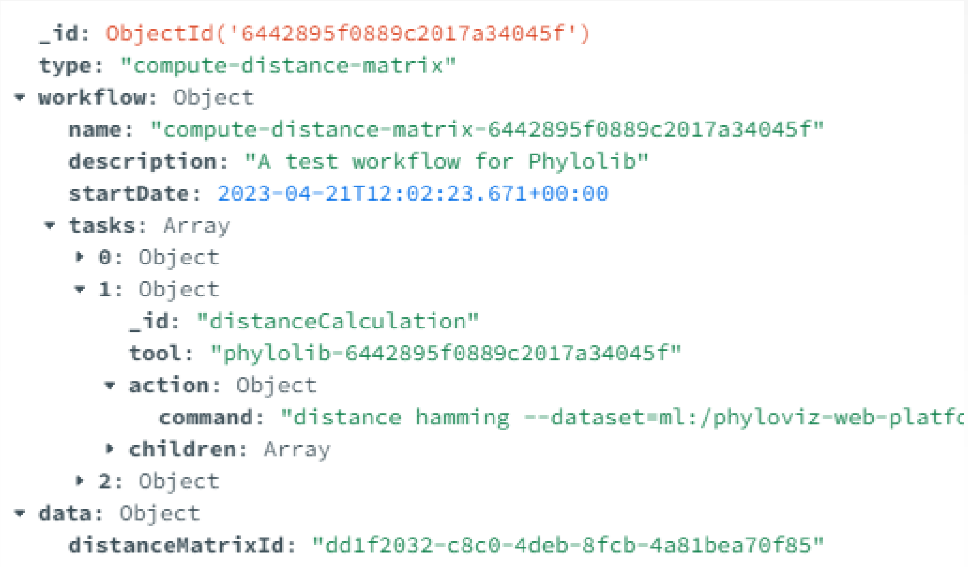
Tool Template



```
  _id: ObjectId('6435c0997b1e5ce5a2527ec3')
▾ general: Object
    name: "phylolib"
    description: "The phylolib library"
▾ access: Object
    _type: "library"
  ▾ details: Object
      address: "localhost"
      dockerUrl: "unix://var/run/docker.sock"
      dockerImage: "localhost:5000/phylolib"
      dockerAutoRemove: "never"
      dockerNetworkMode: "bridge"
      dockerApiVersion: "auto"
    ▾ dockerVolumes: Array
      ▾ 0: Object
          source: "/mnt/phyloviz-web-platform/${projectId}/${workflowId}/"
          target: "/phyloviz-web-platform"
          _type: "bind"
▸ library: Array
```

Workflow and Tool Documents

Workflow Instance



```
  _id: ObjectId('6442895f0889c2017a34045f')
  type: "compute-distance-matrix"
▾ workflow: Object
    name: "compute-distance-matrix-6442895f0889c2017a34045f"
    description: "A test workflow for Phylolib"
    startDate: 2023-04-21T12:02:23.671+00:00
  ▾ tasks: Array
    ▸ 0: Object
    ▾ 1: Object
        _id: "distanceCalculation"
        tool: "phylolib-6442895f0889c2017a34045f"
      ▾ action: Object
          command: "distance hamming --dataset=ml:/phyloviz-web-platfo
      ▸ children: Array
    ▸ 2: Object
▾ data: Object
    distanceMatrixId: "dd1f2032-c8c0-4deb-8fcb-4a81bea70f85"
```

05

Technologies

Backend Technologies

Spring

Gateway and Microservices

Java

Programming Language

Redis

Session Storage

Keycloak

Authorization Server

Backend Technologies

OpenStack

S3 Buckets for Big Data and VMs

MongoDB

Metadata Storage

Docker

Containerization Platform

Python

Language used for some starter compute tools

Frontend Technologies

React

User Interface

Material UI

React Library

TypeScript

Programming Language

Webpack

Module bundling

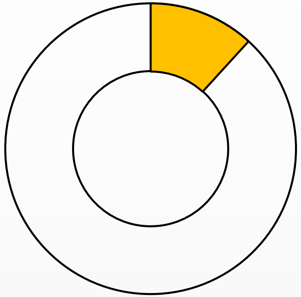
Cosmos

Tree Visualization

06

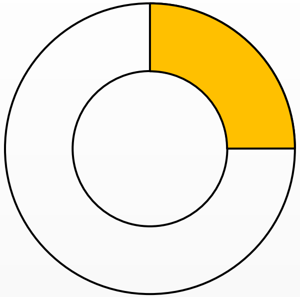
Progress

Progress



Feb 20th to Mar 13rd

- Study and analyze related tools
- Design architecture



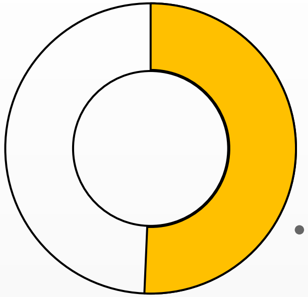
Mar 14th to Apr 10th

- Definition of API
- Definition of metadata schemas
- Implementation of Gateway and Administration Microservice
- Implementation of several pages in the frontend application

Challenges

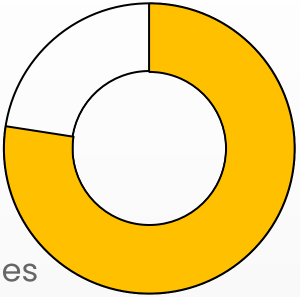
- Learn and work with many different technologies and tools
- Many configurations needed in order to execute the application and tools
- Design a modular architecture

Progress



Apr 11th to Apr 23rd

- Implementation of Compute and Visualization Microservices
- Continued frontend development
  - Beginning of the development of the frontend tree visualization



Apr 24th to May 15th

- Backend improvements
- Continuing frontend development
- Testing

Challenges

- Render the tree view
- Some tools are very expensive on computer resources
- Changes needed to be made in PhyloDB's code to integrate it more easily and allow for more flexibility of use

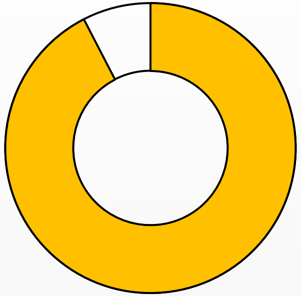
07

Demo

08

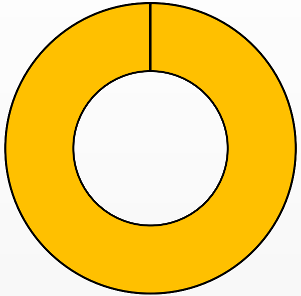
Next Steps

Next Steps



May 25th to Jun 5th

- Partial visualization of phylogenetic tree
- More filters in tree visualization and more layouts
- More tests
- Finish frontend application
- First deployment
- Prepare beta



Jun 5th to Jun 15th

- Final details and delivery

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**