Development of a document management platform

«SciDocCloud»

Ivan Sokolov*, Nikita Balashov

10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)

7 July 2023

Introduction

- Scientific groups collaborate on diverse types of documents:
  - Articles
  - Abstracts
  - Meeting minutes
  - Presentations
  - Reports, etc
- Common problems of collaborative work:
  - Organizing safe and structure documents storage
  - Tracking changes in documents
  - Restricting access to documents
  - Sharing of documents
- Solution is to use a Document Management Software (DMS)

Background and Motivation

- The most common DMS in the HEP community is a **DocDB:**
  - Document storage & organization
  - Version control
  - Sharing access
  - Document exchange
  - Event Management
- Several independent instances of the DocDB are deployed in JINR:
  - SPD
  - BM@N
  - Baikal-GVD
  - Participants of the JINR neutrino program
- But experience with the DocDB has shown:
  - Not fully meet requirements:
    - Functional
    - Security and reliability
  - Problems with access based on groups and roles
  - System support complexity
  - Complex and time-consuming process of refinement system
  - Outdated user interface
  - No client-side validation input mechanisms

Document Database

[DocDB Home] [Administer] [New] [Search] [Last 20 Days] [List Authors] [List Topics] [List Events] [Help]

Instructions
Your Account
Preferences
Apply to Groups
Limit Groups
DocDB Statistics
About DocDB

Administer DocDB

Create or change documents or other information

Search for [   ] (Advanced)

Show Neutrino-doc-# [   ] -v [   ]

Show documents modified in the last [20] days

Calendar of events or **today's** events

List:

◊ Authors   ◊ All documents

◊ Topics

◊ Groups

◊ Keywords

◊ Events

[DocDB Home] [Administer] [New] [Search] [Last 20 Days] [List Authors] [List Topics] [List Events] [Help]

DocDB Version 8.8.12, contact *Document Database Administrators*
Execution time: 0 wallclock secs (0.12 usr + 0.02 sys = 0.14 CPU)

Project Overview

- The DocDB has many drawbacks and feels outdated
- The **goal** of the SciDocCloud project is to develop a document management platform as alternative to DocDB using modern software development tools:
    - Core functionality from DocDB
   - Improved Access management
    - New features

Core Functionality from DocDB

- Authorization and ~~registration~~
- Admin functions:
  - Manage groups, users, roles, authors
  - Manage base meta information for documents (topics, institutions, journals, keywords, document types)
  - ~~But Confirm actions by login and password from DB~~
  - View usage statistics
- View and filter various lists with data (authors, documents, topics and etc)
- Create and storage documents
  - View and update document
  - Download and upload files in document
- Search for documents by various criteria
- Document version control
- Notification by email
- ~~Calendar and Event Management~~
- ~~Administer External DocDBs~~

5
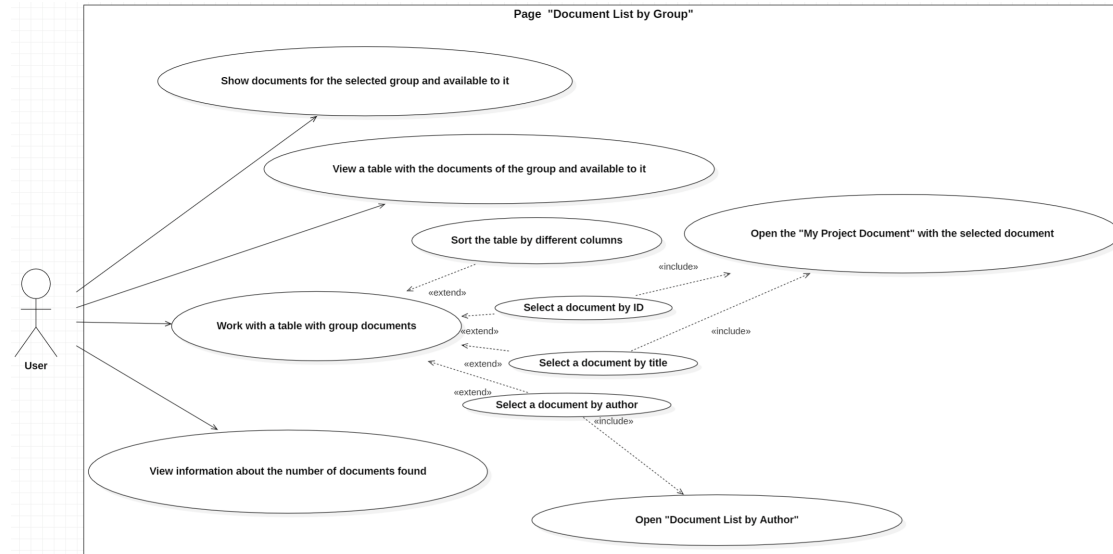
Base Software Stack: Frontend

- Base web technology :
  - HTML
  - CSS
  - JavaScript
  - TypeScript
- *Vue.js* – progressive JavaScript framework for building user interfaces
  - *Axios* – promise-based HTTP client library for frontend JavaScript applications
  - *Pinia* – intuitive, type safe and flexible Store for Vue
  - *Vee-validate* – validation library for Vue.js
  - *Quasar framework* – open-source Vue.js-based cross-platform framework whose motto is: "write code once and simultaneously deploy it as a website, a mobile app and/or an Electron app"
  - *Vite* – build tool that aims to provide a faster and leaner development experience for modern web projects

6
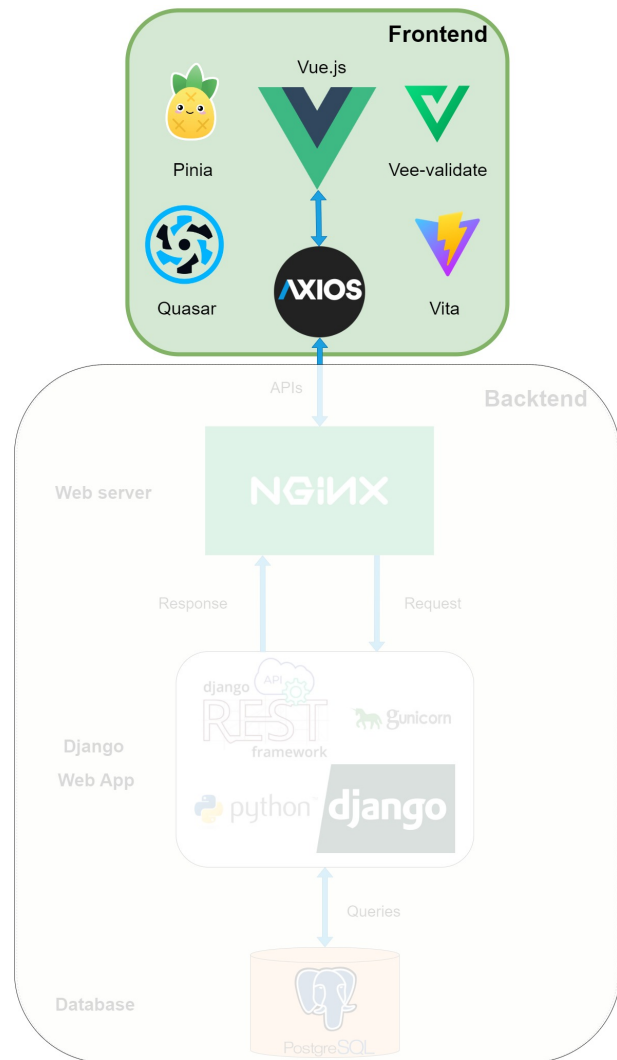
Base Software Stack: Backend

- Web App: **Python + Django**:

  - *Python* – interpreted, object-oriented, high-level programming language with dynamic semantics
  - *Django* – open-source framework for backend web applications based on Python
  - *Django REST framework* (DRF) – powerful and flexible toolkit for building Web APIs
  - *Djoser* – library provides a set of DRF views to handle basic actions such as registration, login, logout, password reset and account activation
  - *Simple JWT* – JSON Web Token authentication plugin for the DRF
  - *Python Social Auth* – allow to easy-to-setup social authentication and authorization mechanism for Python projects supporting protocols like OAuth (1 and 2), OpenID and others
  - *Django-cors-headers* – django App that adds Cross-Origin Resource Sharing (CORS) headers to responses
  - *Django-environ* – python package that allows to configure Django application with environment variables
  - *Django Groups Manager* – django app allows to create hierarchical groups by using django-mptt tree structure
  - *Django-mptt* - reusable Django app which allow to use MPTT with Django models
  - *Django-guardian* – implementation of per object permissions on top of Django's authorization backend
  - *Psycopg2* – PostgreSQL database adapter for the Python programming language
  - *Django-filter* – generic, reusable application allows filter down a queryset based on a model's fields, displaying the form to let them do this
  - *Gunicorn* – python WSGI HTTP Server for UNIX

- Web Server: **NGINX** – open source software for web serving, reverse proxying, caching, load balancing, media streaming, and more
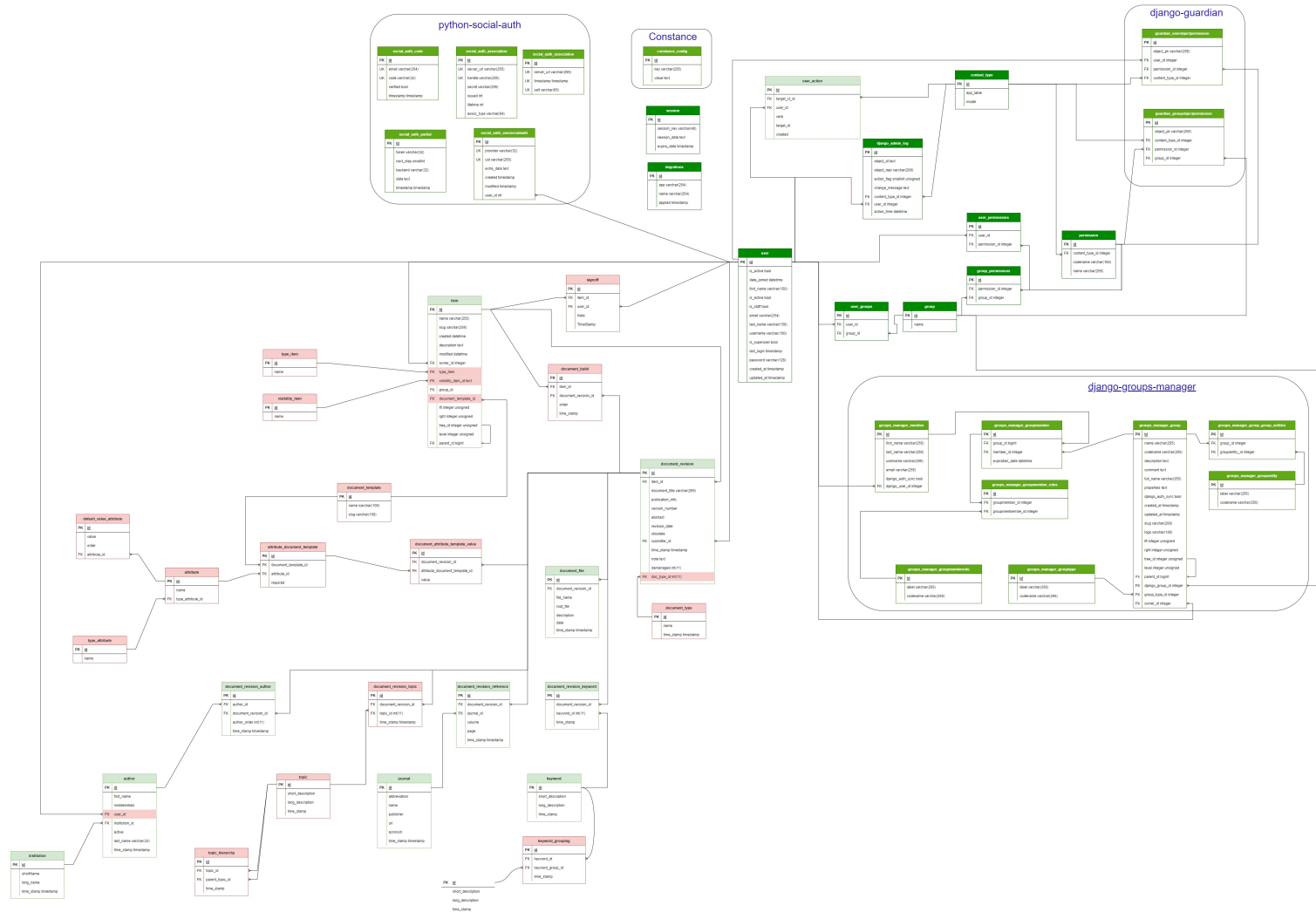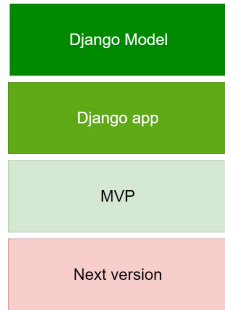
- Database: **PostgreSQL** – powerful, open source object-relational database system

API
+
Authorization
+
Registration

Hierarchical
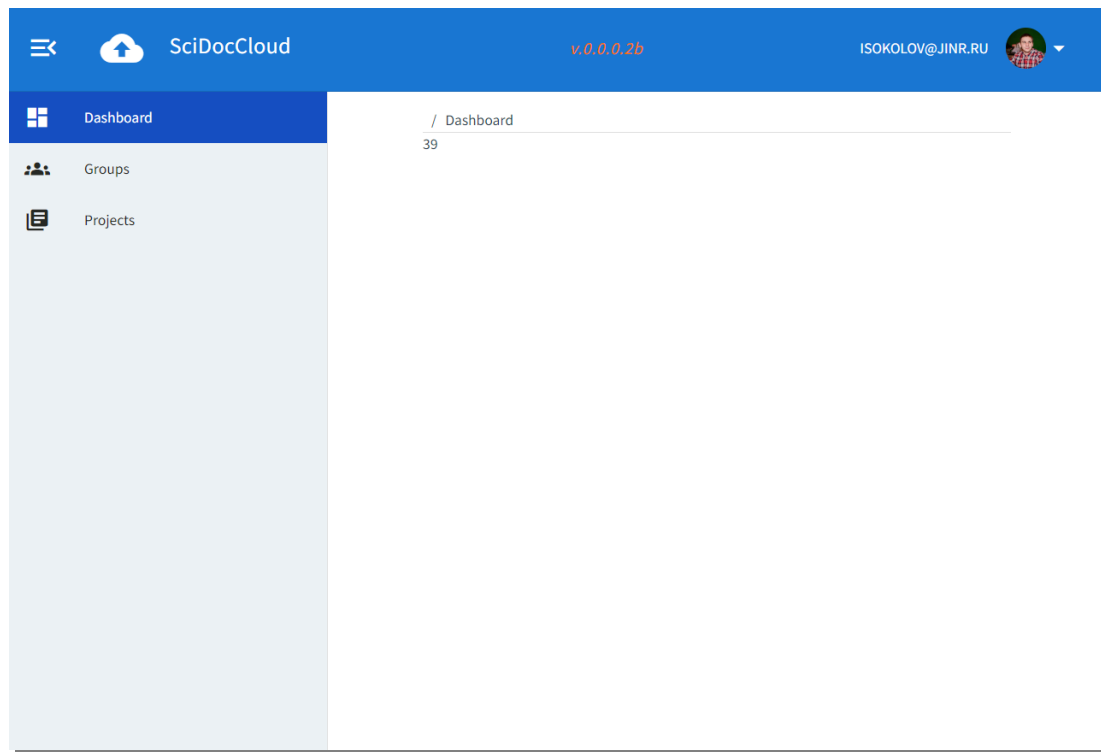organization
+
Role
and
Permissions
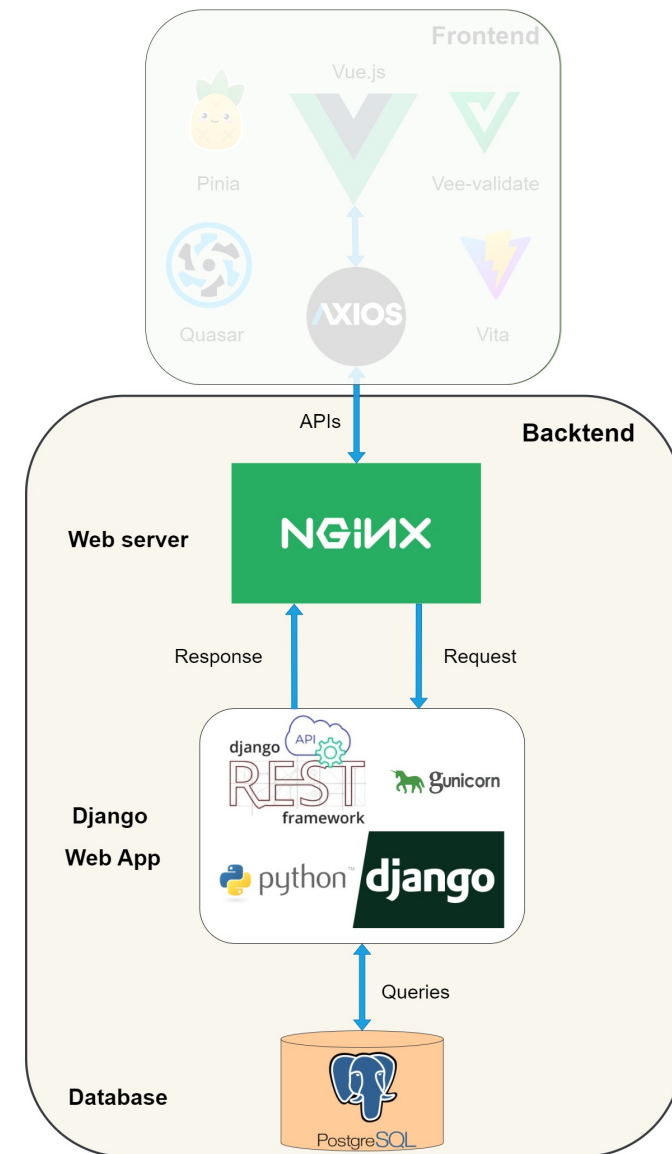
Platform Database Model

Current State of Development

- **The main frame of the platform has been created (Login Layout, Main Layout, routing, navbar, sidebar)**

- Login by JINR SSO and Django account using JWT-token

- Hierarchical organization:

  − Groups and Subgroups

  − Projects

- Manage groups, subgroups and projects

- Prepared the basis for roles and permissions on backend

- Manage members in group (invite, set/change role, remove)

Current State of Development

- The main frame of the platform has been created (Login Layout, Main Layout, routing, navbar, sidebar)

- **Login by JINR SSO and Django account using JWT-token**

- Hierarchical organization:

  - Groups and Subgroups

  - Projects

- Manage groups, subgroups and projects

- Prepared the basis for roles and permissions on backend

- Manage members in group (invite, set/change role, remove)

SciDocCloud

Email
djangouser.example.eu

Password
••••••••••••••••
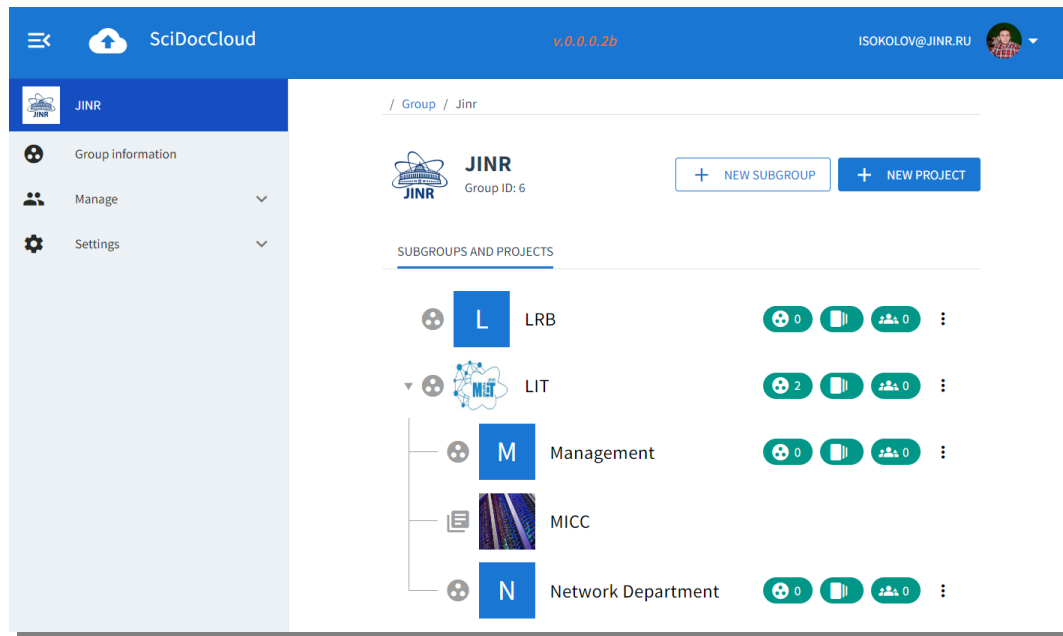
LOGIN   JINR SSO

Current State of Development

- The main frame of the platform has been created (Login Layout, Main Layout, routing, navbar, sidebar)

- Login by JINR SSO and Django account using JWT-token

- **Hierarchical organization:**
  - Groups and Subgroups
  - Projects

- Manage groups, subgroups and projects

- Prepared the basis for roles and permissions on backend

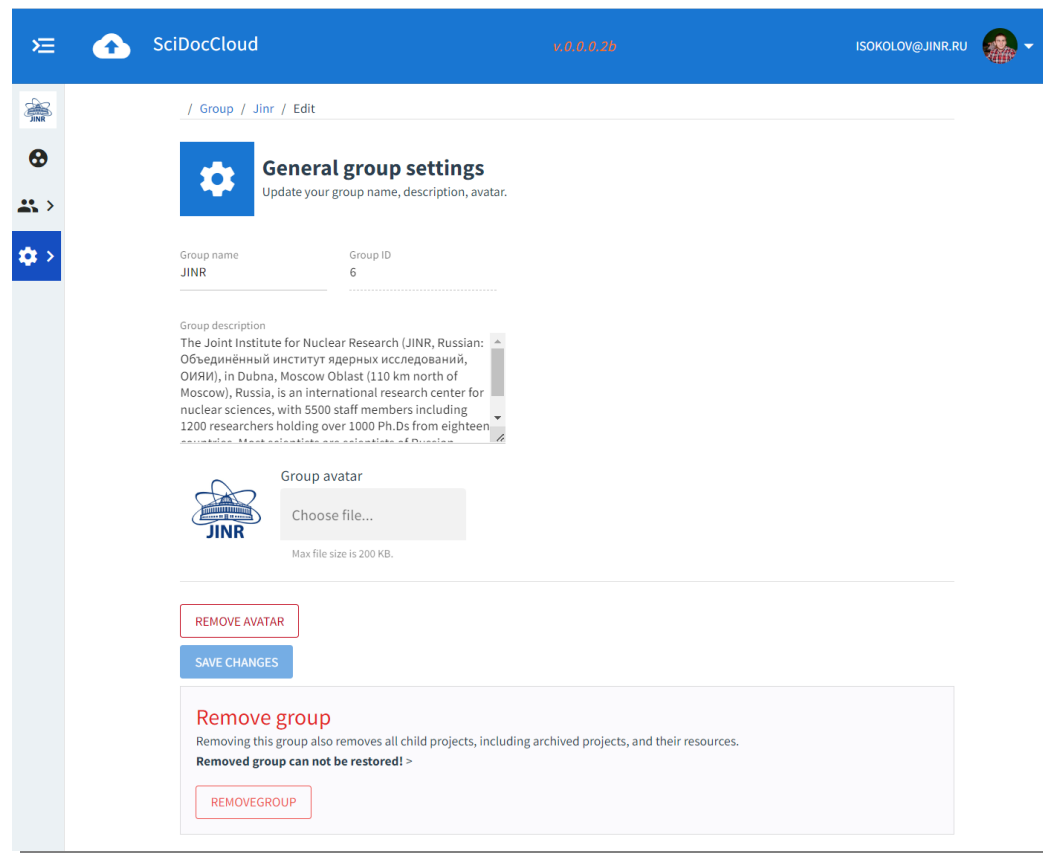- Manage members in group (invite, set/change role, remove)

Current State of Development

- The main frame of the platform has been created (Login Layout, Main Layout, routing, navbar, sidebar)

- Login by JINR SSO and Django account using JWT-token

- Hierarchical organization:

  – Groups and Subgroups

  – Projects

- **Manage groups, subgroups and projects**

- Prepared the basis for roles and permissions on backend

- Manage members in group (invite, set/change role, remove)

Current State of Development

- The main frame of the platform has been created (Login Layout, Main Layout, routing, navbar, sidebar)

- Login by JINR SSO and Django account using JWT-token

- Hierarchical organization:

  - Groups and Subgroups
  - Projects

- Manage groups, subgroups and projects

- **Prepared the basis for roles and permissions on backend**

- Manage members in group (invite, set/change role, remove)

Group members permissions

Actions	Owner	Maintainer	Editor	Viewer
View subgroup	+	+	+	+
Browse group	+	+	+	+
View project	+	+	+	+
Create subgroup	+	+		
Edit group settings	+	+		
Manage group members	+	+		
Create project in group	+	+		
Delete group	+			
Rename project	+			

Project members permissions

Actions	Owner	Maintainer	Editor	Viewer
View a list of documents	+	+	+	+
View document	+	+	+	+
Edit document	+	+	+	
Add document	+	+	+	
Edit project settings	+	+		
Manage project members	+	+		
Delete document	+	+		
Delete project	+			

Current State of Development

- The main frame of the platform has been created (Login Layout, Main Layout, routing, navbar, sidebar)

- Login by JINR SSO and Django account using JWT-token

- Hierarchical organization:

  - Groups and Subgroups
  - Projects

- Manage groups, subgroups and projects

- Prepared the basis for roles and permissions on backend

- **Manage members in group (invite, set/change role, remove)**
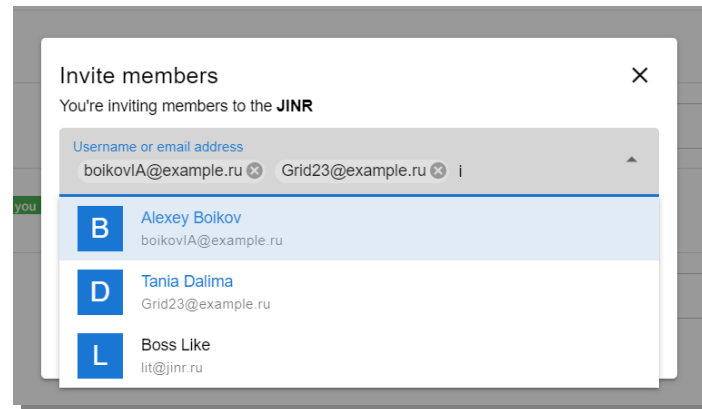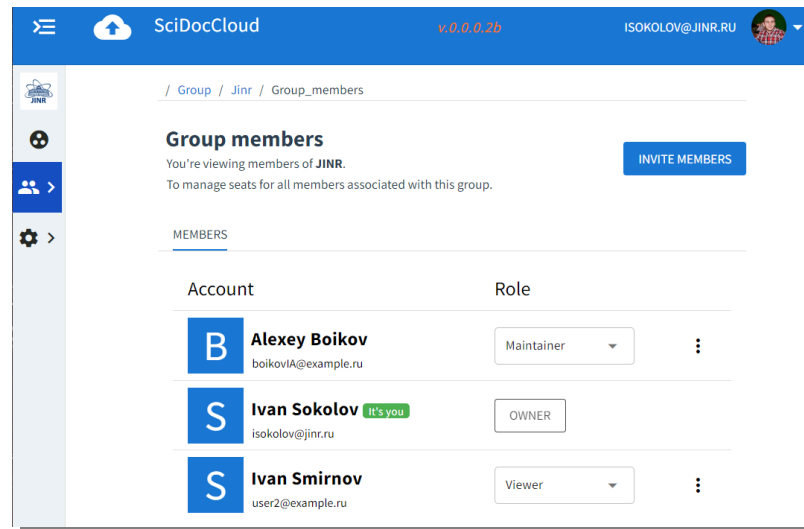
Development Roadmap

- Document creation and storage

- Document version control

- Hierarchical organization:
  - Folder and Document

- Document Template Editor

- History of user actions

- Support for various storage systems (local FS, S3 cloud storage)

- Integration with Indico

- and others...

15

Conclusion

- Development MVP is in process (core functionality)
- Currently development in private JINR GitLab repo (https://git.jinr.ru/scidoccloud/scidoccloud)
- As soon as MVP is ready, we will create a public mirror on GitHub open for contributions
- Waiting for the news on docs.jinr.ru

Thanks!