Storage Service for Scientific Documentation

Ivan Sokolov*, Nikita Balashov

The 12th Collaboration Meeting of the BM@N Experiment at the NICA Facility

16 May 2024

Introduction

- Scientific groups collaborate on various types of documents

    - Articles

    - Abstracts

    - Presentations

    - Reports, etc

- Common challenges of collaborative work

    - Organizing safe and structure documents storage

    - Tracking changes to documents

    - Restricting access to documents

    - Sharing of documents

- Many scientific projects generate a significant amount of documents (files + revisions)

Background and Motivation

- Scientific communities use various software for document storage and collaboration

  - DocDB

  - Xwiki

  - CERN Document Server, etc.

- Several **independent** instances of the DocDB are deployed in the JINR

  - BM@N

  - Baikal-GVD

  - SPD

  - Participants of the JINR neutrino program

- Experience with these systems has shown:

  - Not fully meet requirements:

    - Functional

    - Security and reliability

  - Problems with data integrity

  - System support and improvement are complex processes
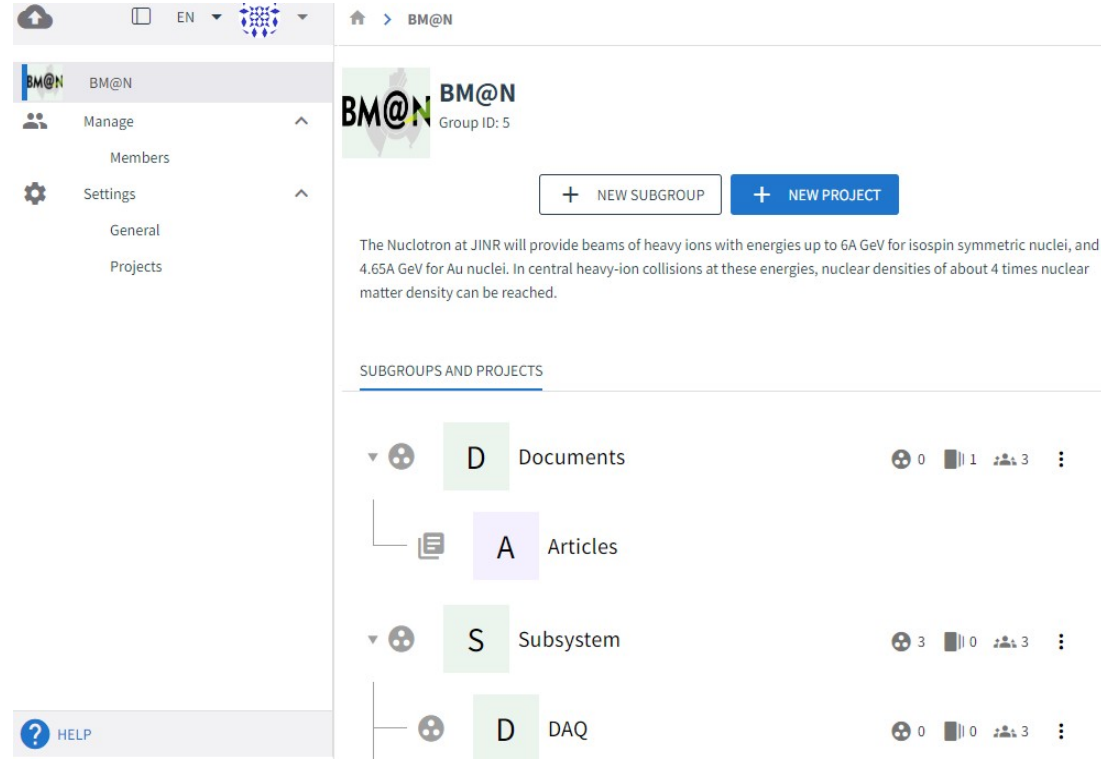
Project Overview

- The goal of the project is to provide a centralized multi-community documentation storage service for collaborative activities

    - Implement the core features of DocDB

    - Expend system to support group and project with flexible access control

- Specific requests from BM@N

    - Organization of collaborate on publication activities and reports

        - Defining per-document editors

        - Document discussion system

        - Document review process

        - Custom meta-data fields in documents

    - Email notification system

- **Hierarchical organization**:

  groups / subgroups → projects → documents

- The idea comes from GitLab

  – Using GitLab terminology

  – Project ≈ GitLab repository

- Project contains

  – Document storage

  – Collaboration tools

- Document as an abstract entity

  – Form with metadata fields

  – Collection of files

  – Version control

Organization Structure in Service

- Hierarchical organization
  groups / subgroups → projects → documents

- **The idea comes from GitLab**

  – Using GitLab terminology

  – Project ≈ GitLab repository

- Project contains

  – Document storage

  – Collaboration tools

- Document as an abstract entity

  – Form with metadata fields

  – Collection of files

  – Version control

Organization Structure in Service

- Hierarchical organization
groups / subgroups → projects → documents

- The idea comes from GitLab

  - Using GitLab terminology

  - Project ≈ GitLab repository

- **Project contains**

  - Document storage

  - Collaboration tools

- Document as an abstract entity

  - Form with metadata fields

  - Collection of files

  - Version control

Organization structure in service

- Hierarchical organization
  groups / subgroups → projects → documents

- The idea comes from GitLab

  – Using GitLab terminology

  – Project ≈ GitLab repository

- Project contains

  – Document storage

  – Collaboration tools

- **Document as an abstract entity**

  – Form with metadata fields

  – Collection of files

  – Version control

Organization structure in service

- Hierarchical organization
  groups / subgroups → projects → documents

- The idea comes from GitLab

  – Using GitLab terminology

  – Project ≈ GitLab repository

- Project contains

  – Document storage

  – Collaboration tools

- **Document as an abstract entity**

  – ***Form with metadata fields***

  – Collection of files

  – Version control

- Hierarchical organization
  groups / subgroups → projects → documents

- The idea comes from GitLab

  – Using GitLab terminology

  – Project ≈ GitLab repository

- Project contains

  – Document storage

  – Collaboration tools

- **Document as an abstract entity**

  – Form with metadata fields

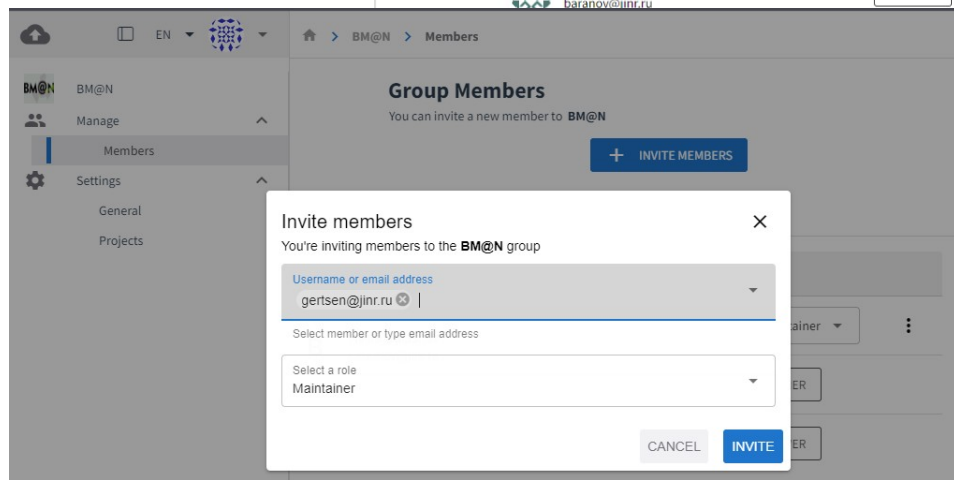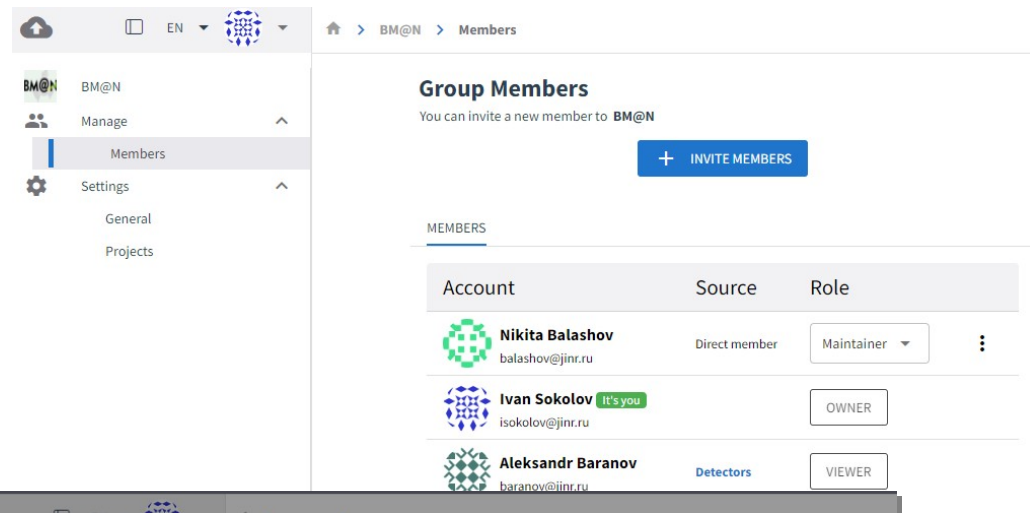 – ***Collection of files***

  – Version control

- Hierarchical organization
  groups / subgroups → projects → documents

- The idea comes from GitLab

  – Using GitLab terminology

  – Project ≈ GitLab repository

- Project contains

  – Document storage

  – Collaboration tools

- **Document as an abstract entity**

  – Form with metadata fields

  – Collection of files

  – ***Version control***

Role-based Access Model

- Access restriction to groups or projects based on roles similar to GitLab

- Currently implemented roles Owner, Maintainer and Viewer
  - Each has their own permission
  - In plan to add new roles

- Group/project membership

- Nested groups membership

Realized Features for BM@N

- *Defining per-document editors →*
  Implemented the "Editors List" feature to manage document editing permissions (only owner)

- Document discussion system →
  Added the comments section to the document (all members)

- Custom meta-data fields →
  Added fields for journal references

Realized Features for BM@N

- Defining per-document editors →
  Implemented the "Editors List" feature to manage document editing permissions (only owner)

- *Document discussion system →*
  Added the comments section to the document (all members)

- Custom meta-data fields →
  Added fields for journal references

Realized Features for BM@N

- Defining per-document editors →
  Implemented the "Editors List" feature to manage document editing permissions (only owner)

- Document discussion system →
  Added the comments section to the document (all members)

- *Custom meta-data fields →*
  Added fields for journal references

Document Review Workflow

Atricle Workflow

Stages

1. **Draft**
2. **Approved**
3. **Sent**
4. **Reviewing and editing**
5. **Accepted**
6. **Published** — Non-Editable

Report Workflow

Stages

1. **Draft**
2. **Approved**
3. **Presented** — Non-Editable

Roles

1. **Editor**
2. **Speaker bureau**
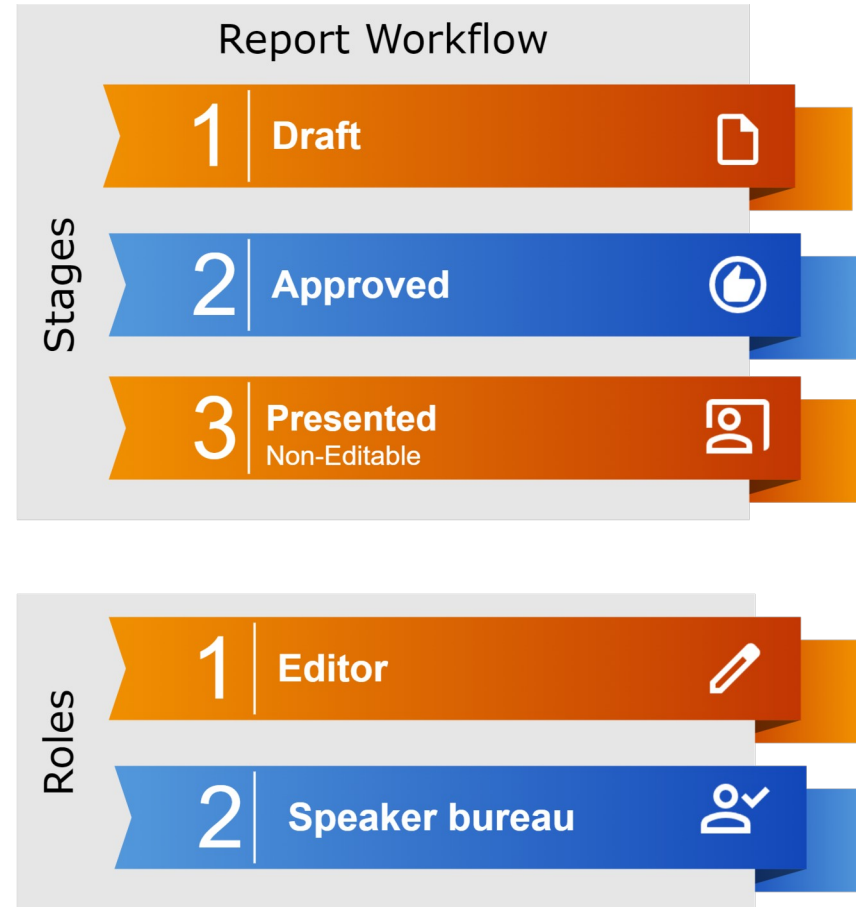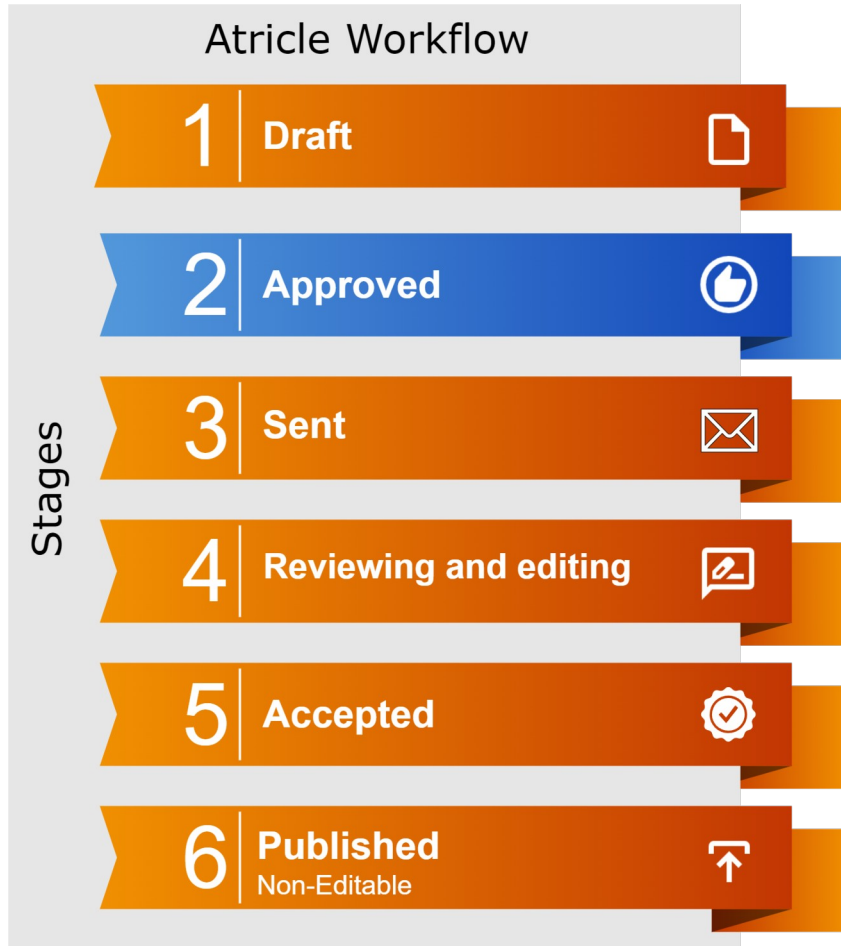
Data Availability and Consistency

- Files for documents are stored in the JINR's Ceph cloud storage

  - Upload and Download files via S3

  - Triple replication provided by Ceph

  - Easily scalable

- Nightly Backups for Service Meta Information Database

- Service is deployed on a VM in the JINR Cloud:

  - Quick recovery after hardware failure

- Automated deployment via GitLab CI/CD → in plans

JINR Ceph cloud storage

The service's DataBase

JINR GitLab Repository
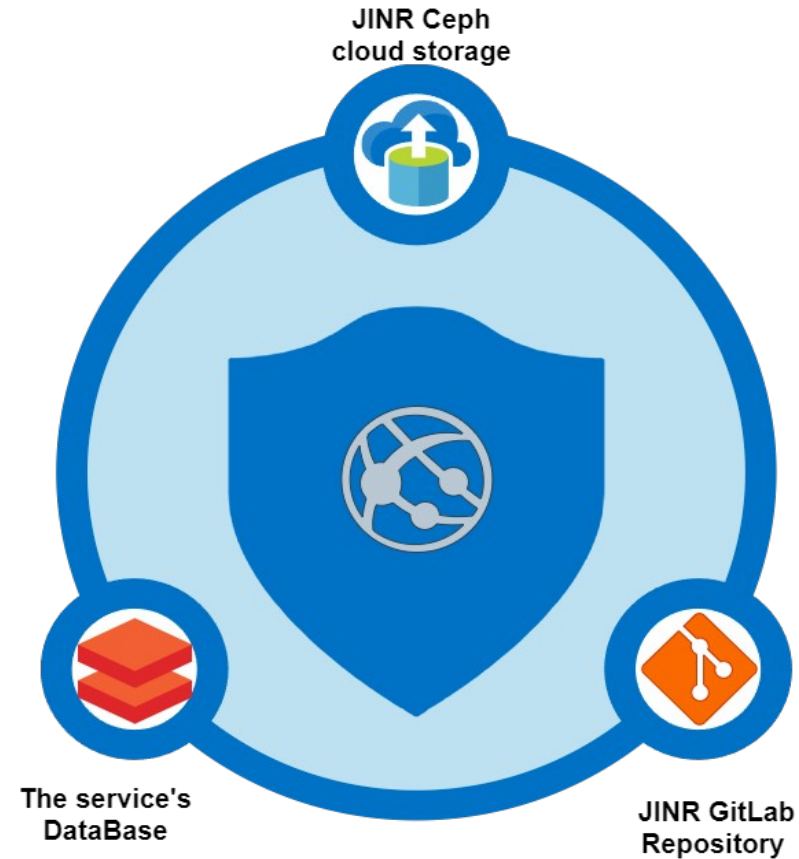
Road Map

- Document meta-data Templates → in progress

- Subscribe/notification system

- Namespace and document visibility levels

- Data migration from JINR DocDBs

- Workflows

- Integration with Indico (Event Calendar)

- The service is available for use →
  docs.jinr.ru

- Please report bugs and suggest improvements via email →
  docs@jinr.ru

Your feedback helps us to improve our service and makes it better!