

Joint Institute for Nuclear Research

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



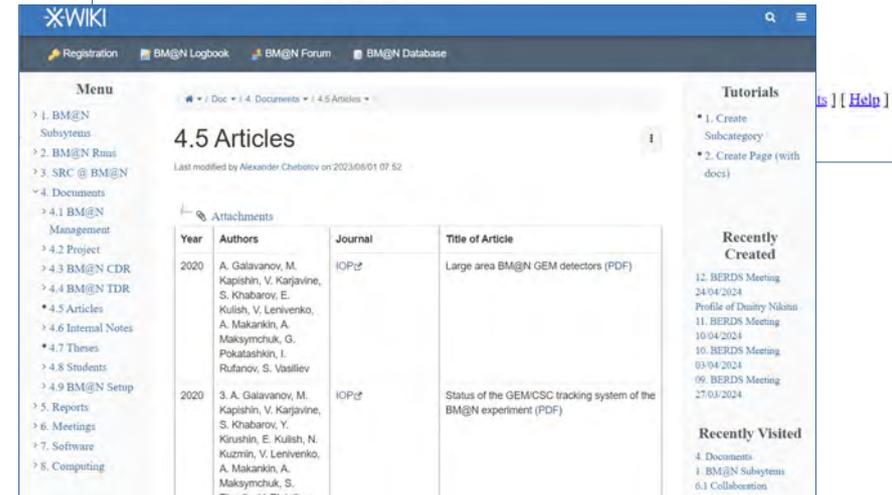
- Scientific groups collaborate on various types of documents
 - Articles
 - Abstracts
 - Presentations
 - Reports, etc
- Common challenges of collaborative work
 - Organizing safe and structure documents storage
 - Tracking changes to documents
 - Restricting access to documents
 - Sharing of documents
- 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

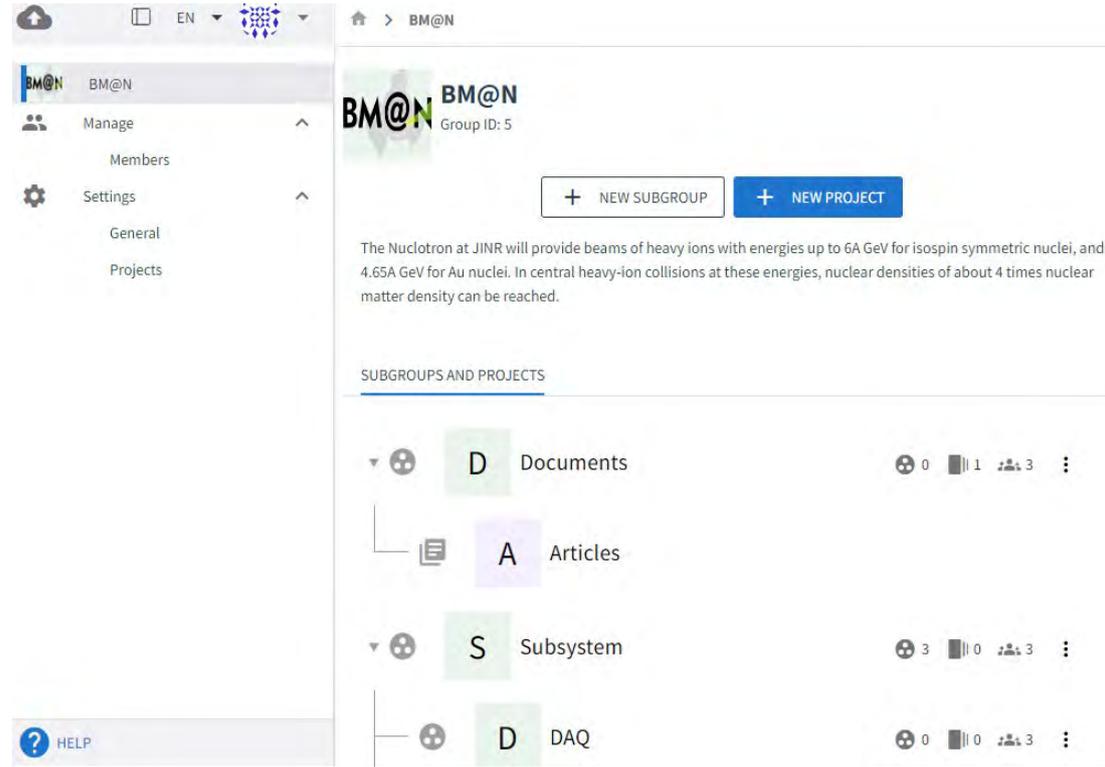


- The goal of the project is to provide a centralized multi-community documentation storage service for collaborative activities
 - Implement the core features of DocDB
 - Extend system to support group and project with flexible access control
- Specific requests from [BM@N](#)
 - Organization of collaborate on publication activities and reports
 - Defining per-document editors
 - Document discussion system
 - Document review process
 - Custom meta-data fields in documents
 - Email notification system

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

The screenshot shows the 'Groups' page in GitLab. At the top, there is a breadcrumb 'Your work / Groups' and a 'New group' button. Below this is a search bar 'Search by name' and a dropdown menu 'Last created'. The main content is a list of groups with the following details:

Group Name	Role	Members	Projects	Repositories
SciDocCloud	Developer	0	2	2
Neutrino Computing Platform				3
jj_nano	Developer	0	4	8
cloud-team	Guest	2	26	11
Configuration Management		0	7	3
htcondor		0	5	2

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

BM@N > Documents > Articles

Articles
Project ID: 15

+ NEW DOCUMENT

Configure this list

All Search

Title Description Abstract Authors Created at Last Update

#	Title	Abstract	Authors	Last Update
3-v1	High-Performance Optimization of Algorithms Used in the BM@N Experiment of the NICA Project	Results of high-performance optimization of BmnRoot software modules are presented. The BmnRoot package used in the BM@N experiment of the NICA projec...	Sergei Merts et al.	5 days ago
4-v0	Large area BM@N GEM detectors		A Galavanov	3 minutes ago
5-v0	Исследование характеристик GEM детекторов эксперимента BM@N	BM@N experiment (Baryonic Matter at Nuclotron) is aimed to study core-core (up to gold-gold) collisions in extreme conditions. High intensity of inter...	C Васильев et al.	1 minute ago

[Development of the BM@N Web](#)

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

The screenshot displays a document management interface for a document titled "Storage Service for Scientific..." (Reports Document 75-v4). The breadcrumb navigation shows the path: JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75 > version 4. The document is submitted by Ivan Sokolov and updated by Konstantin Gercenberger. It was created on May 10, 2024, at 18:09 and revised on May 10, 2024, at 18:11. The latest version is "reports-doc-75-v5" (created at 18:35), and other versions include "reports-doc-75-v3" and "reports-doc-75-v2" (both created at 18:10). The authors listed are Nikita Alexandrovich Balashov and Ivan Alexandrovich Sokolov. The editors are Sokolov Ivan, Gercenberger Konstantin, and Balashov Nikita. The document contains four files: "Notes.odt" (61.48 KB, Document), "sokolov_presentation.odp" (5.43 MB, Presentation), "Workflow_Article.png" (1.64 MB, Image), and "Workflow_report.png" (1.71 MB, Image). A comment by Ivan Sokolov from 29 minutes ago states "Добавил презентацию и заметки" (Added presentation and notes), which has 3 replies, with the last reply by Nikita Balashov 28 minutes ago.

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

The screenshot displays a document management interface for a document titled "Storage Service for Scientific..." (Reports Document 75-v4). The breadcrumb navigation shows the path: JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75 > version 4. The document header includes an "UPDATE DOCUMENT" button. The authors listed are Nikita Alexandrovich Balashov (balashov@jinr.ru) and Ivan Alexandrovich Sokolov (isokolov@jinr.ru). The "Notes and Changes" section indicates: "Added all material for presentation + presentation + notes". The "Files" section lists: Notes.odt (61.48 KB, Document), sokolov_presentation.odp (5.43 MB, Presentation), Workflow_Article.png (1.64 MB, Image), and Workflow_report.png (1.71 MB, Image). The "Editors" list includes Sokolov Ivan (isokolov@jinr.ru), Gercenberger Konstantin (gertsen@jinr.ru), and Balashov Nikita (balashov@jinr.ru). The "Comments (6)" section shows a comment by Ivan Sokolov 29 minutes ago: "Добавил презентацию и заметки", with 3 replies, the last by Nikita Balashov 28 minutes ago. The right sidebar shows document metadata: Document # (reports-doc-75-v4), Submitted by (Ivan Sokolov), Updated by (Konstantin Gercenberger), Document Created (10 May 2024, 18:09), Revised (10 May 2024, 18:11), Latest Version (reports-doc-75-v5), and Other Versions (reports-doc-75-v5, reports-doc-75-v3, reports-doc-75-v2).

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

Storage Service for Scientific...
Reports Document 75-v4

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

Authors:
Nikita Alexandrovich Balashov (balashov@jinr.ru)
Ivan Alexandrovich Sokolov (isokolov@jinr.ru)

Notes and Changes:
Added all material for presentation + presentation + notes

Files:

Notes.odt 61.48 KB File type: Document	Workflow_Article.png 1.64 MB File type: Image
sokolov_presentation.odp 5.43 MB File type: Presentation	Workflow_report.png 1.71 MB File type: Image

Editors:
Sokolov Ivan (isokolov@jinr.ru)
Gercenberger Konstantin (gertsen@jinr.ru)
Balashov Nikita (balashovn@jinr.ru)

Document #:
reports-doc-75-v4

Submitted by:
Ivan Sokolov

Updated by:
Konstantin Gercenberger

Document Created:
10 May 2024, 18:09

Revised:
10 May 2024, 18:11

Latest Version:
reports-doc-75-v5

Other Versions:
reports-doc-75-v5 (Created at: 10 May 2024, 18:35)
reports-doc-75-v3 (Created at: 10 May 2024, 18:10)
reports-doc-75-v2 (Created at: 10 May 2024, 18:10)

Comments (0)

Ivan Sokolov 29 minutes ago
Добавил презентацию и заметки

3 REPLIES Last reply by Nikita Balashov 28 minutes ago

Organization structure in service



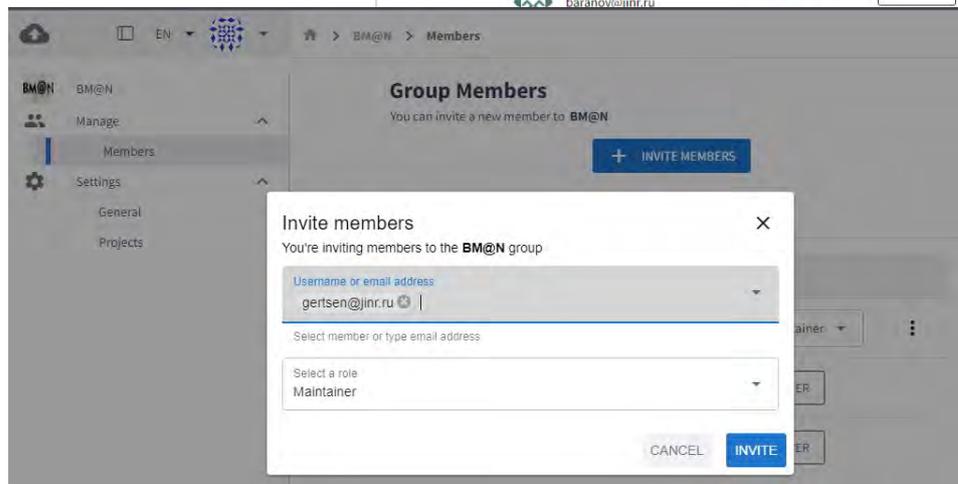
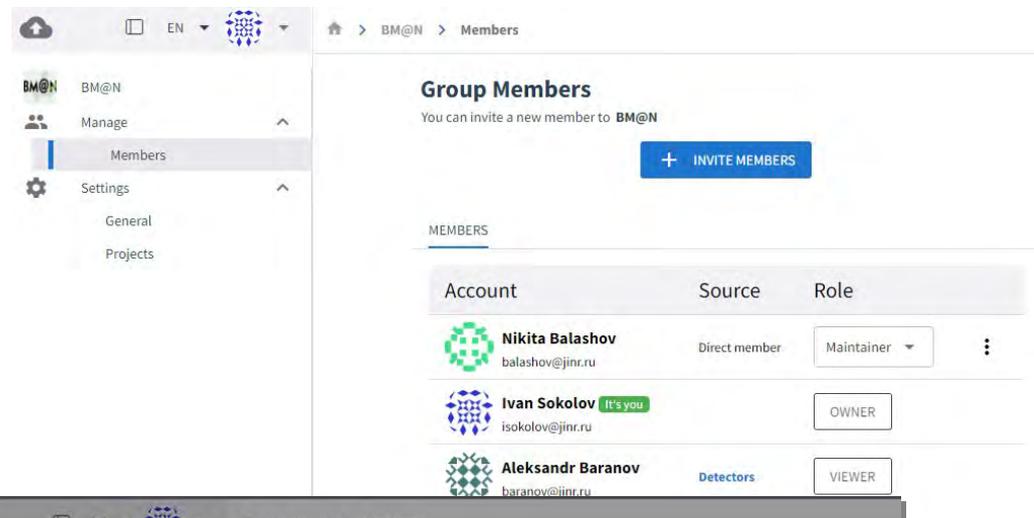
- Hierarchical organization
groups / subgroups → *projects* → *documents*
- The idea comes from GitLab
 - Using GitLab terminology
 - Project ≈ GitLab repository
- Project contains
 - Document storage
 - Collaboration tools
- **Document as an abstract entity**
 - Form with metadata fields
 - Collection of files
 - **Version control**

The screenshot displays the 'Storage Service for Scientific...' interface for a document titled 'Reports Document 75-v4'. The breadcrumb navigation shows the path: JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75 > version 4. The document is a report on the 12th Collaboration Meeting of the BM@N Experiment at the NICA Facility. It lists two authors: Nikita Alexandrovich Balashov and Ivan Alexandrovich Sokolov. The 'Notes and Changes' section indicates that all material for presentation was added. The 'Files' section lists four files: 'Notes.odt' (61.48 KB, Document), 'Workflow_Article.png' (1.64 MB, Image), 'sokolov_presentation.odp' (5.43 MB, Presentation), and 'Workflow_report.png' (1.71 MB, Image). The 'Editors' section lists three users: Sokolov Ivan, Gercenberger Konstantin, and Balashov Nikita. The 'Comments' section shows a comment from Ivan Sokolov and three replies. On the right, the 'Document #:' is 'reports-doc-75-v4', 'Submitted by:' is 'Ivan Sokolov', 'Updated by:' is 'Konstantin Gercenberger', and 'Document Created:' is '10 May 2024, 18:09'. The 'Revised:' section shows a revision on '10 May 2024, 18:11'. The 'Latest Version:' is 'reports-doc-75-v5'. The 'Other Versions:' section lists 'reports-doc-75-v5' (Created at: 10 May 2024, 18:35), 'reports-doc-75-v3' (Created at: 10 May 2024, 18:10), and 'reports-doc-75-v2' (Created at: 10 May 2024, 18:10). A red arrow points from the 'Version control' text in the list to the version history section.

Role-based Access Model

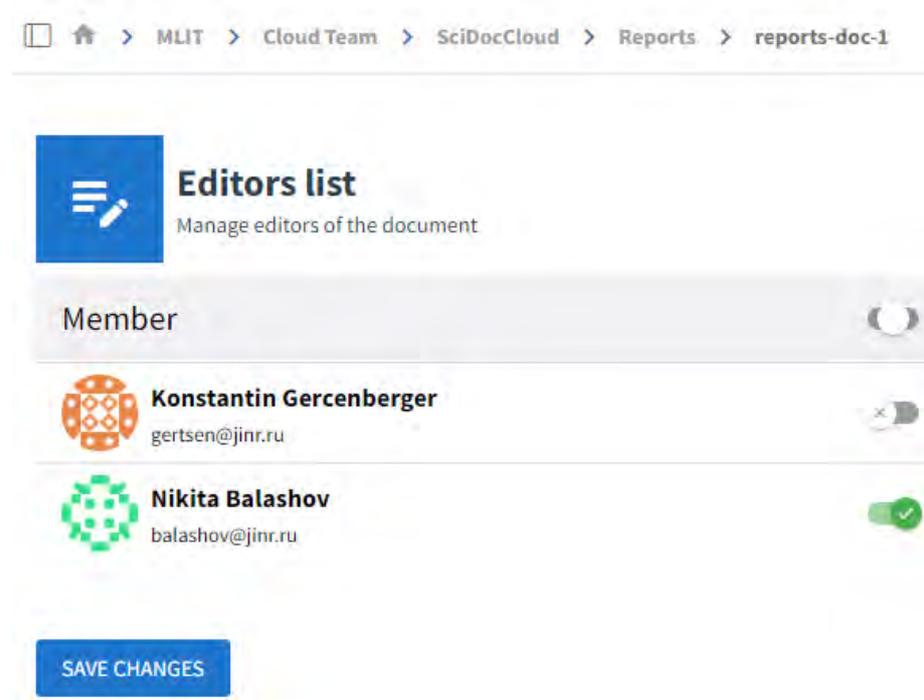


- Access restriction to groups or projects based on roles similar to GitLab
- Currently implemented roles Owner, Maintainer and Viewer
 - Each has their own permission
 - In plan to add new roles
- Group/project membership
- Nested groups membership





- *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

The screenshot shows a 'Comments (6)' section on a document. It features two visible comments:

- Ivan Sokolov** (29 minutes ago): "Добавил презентацию и заметки". Below the comment, it indicates "3 REPLIES" and "Last reply by Nikita Balashov about 2 hours ago".
- Konstantin Gercenberger** (25 minutes ago): "Так как у меня будет суммарный доклад по проведенным работам просьба присылать свои презентации мне, как можно раньше, чтобы у меня было время спокойно сформировать свою презентацию. Крайне желательно не позже этих выходных. Можно присылать черновые еще версии, чтобы я составил свой доклад, а потом уже подменил итоговыми версиями слайдов". Below this comment is a "COLLAPSE REPLIES" button.

At the bottom of the screenshot, there is a text input field with the placeholder "Write a comment here..." and a blue "COMMENT" button. Above the input field is a rich text editor toolbar with icons for bold, italic, link, underline, text color, font color, bulleted list, numbered list, indent, and outdent.

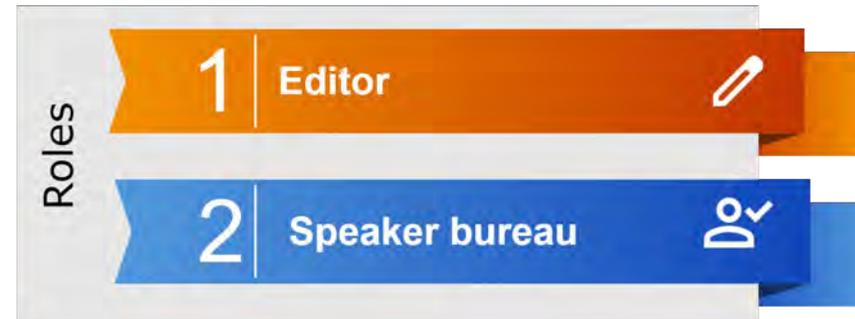
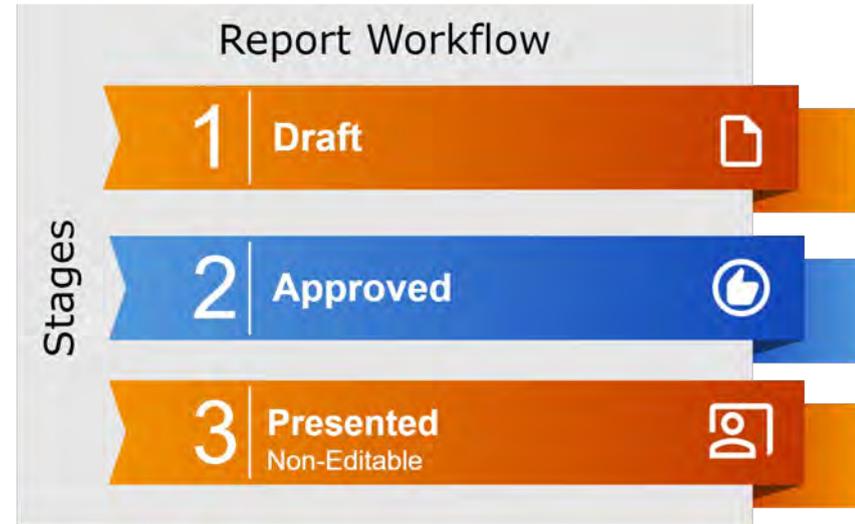
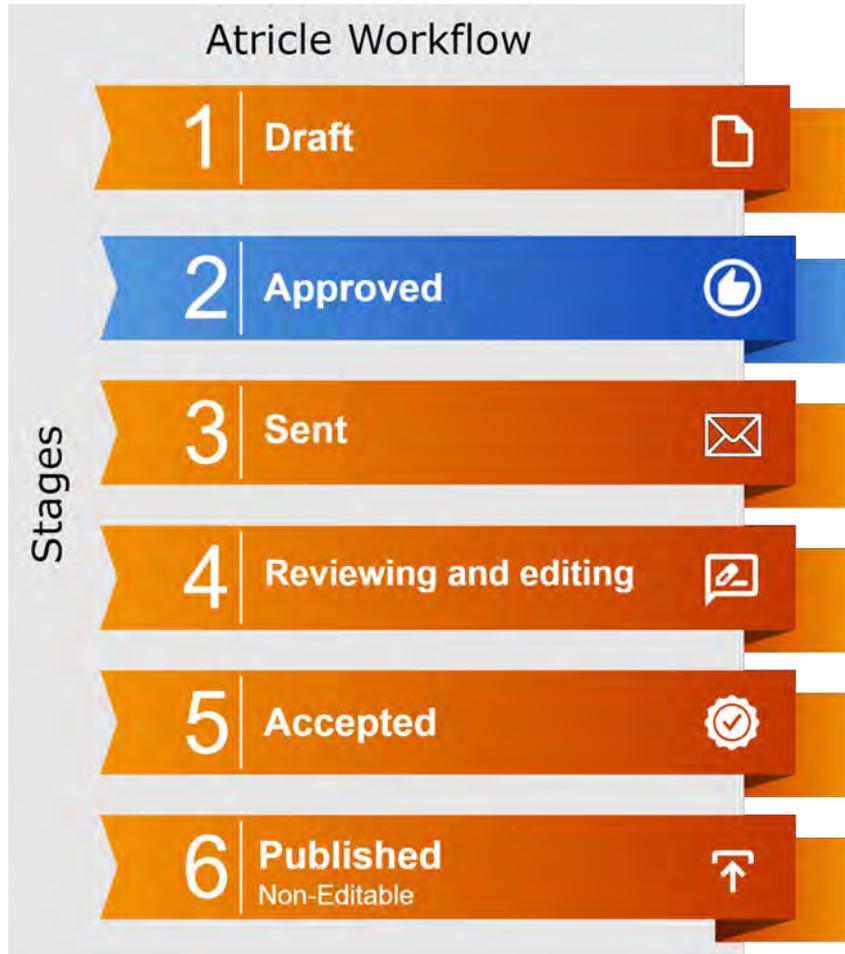
Realized Features for BM@N



- 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**

The screenshot displays a web interface for document editing. At the top, there are sections for "Notes and Changes" and "Other Publication information". The main focus is a "Journal References" section, which is highlighted with a red border. This section contains three input fields: "Journal", "Volume", and "Page". Below the form, there is a "File upload" button with a plus sign and a progress indicator showing "0.0B / 0.00%". At the bottom of the interface, there are "CANCEL" and "+ UPDATE" buttons.

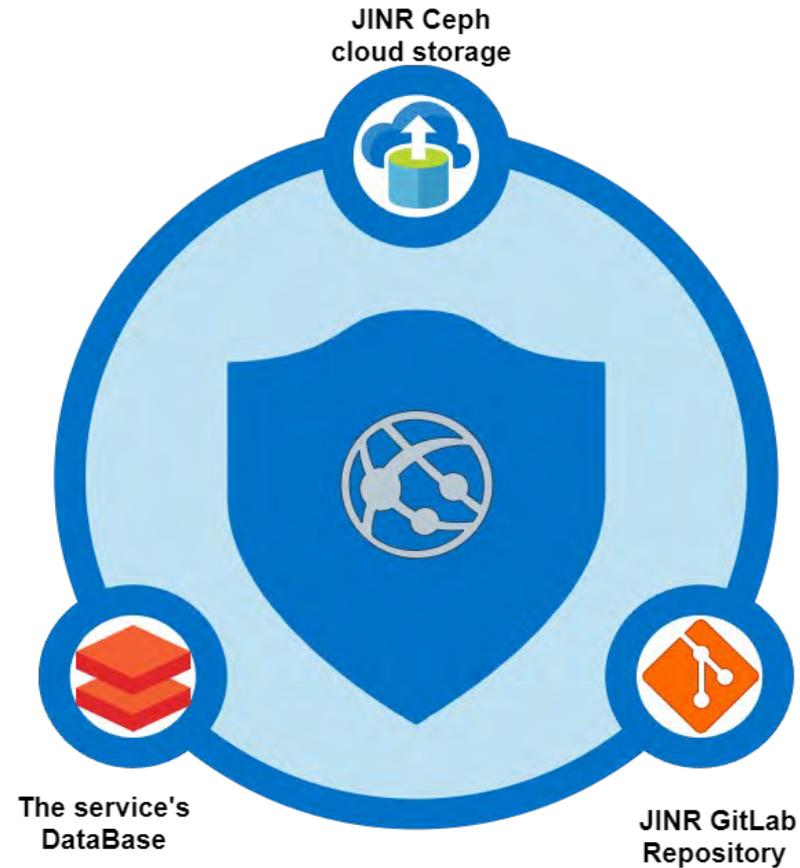
Document Review Workflow



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



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!**

