



Joint Institute for Nuclear Research

Storage Service for Scientific Documentation

Ivan Sokolov*, Nikita Balashov

29 October 2024

Introduction



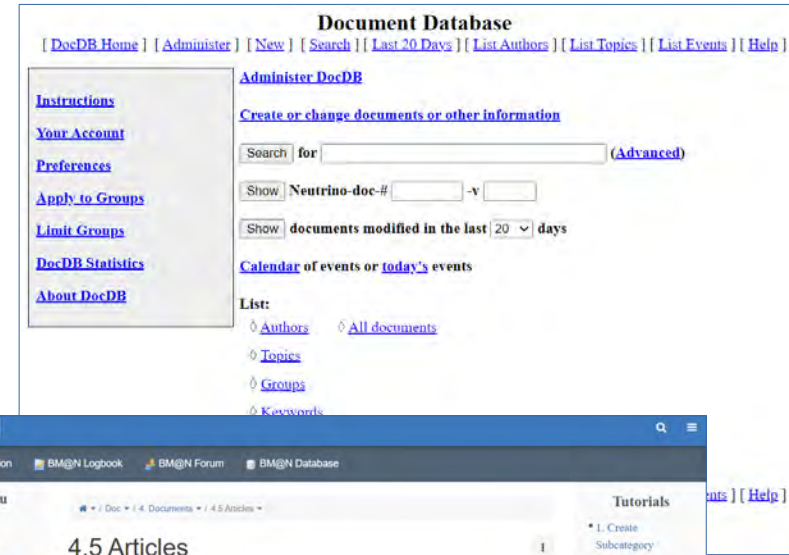
- Scientific groups collaborate on various types of documents
 - Articles
 - Abstracts
 - Presentations
 - Reports, etc
- Common challenges of collaborative work
 - Organizing safe and structured documents storage
 - Tracking changes of documents
 - Restricting access to documents
 - Sharing of documents
- Many scientific projects generate a significant amount of documents (meta-information + files)



Background and Motivation



- Scientific communities use various software tools for document storage and collaboration
 - DocDB
 - Xwiki
 - CERN Document Server, etc.
- Several **independent** instances of the DocDB system are deployed at JINR
 - BM@N
 - Baikal-GVD
 - Shared instance for Neutrino projects at JINR
 - Accelerator Division No.1 of the VBLHEP accelerator complex
- Experience with these systems has shown
 - Not fully met requirements:
 - Functional
 - Security and reliability
 - System support and improvement are complex



Project Overview

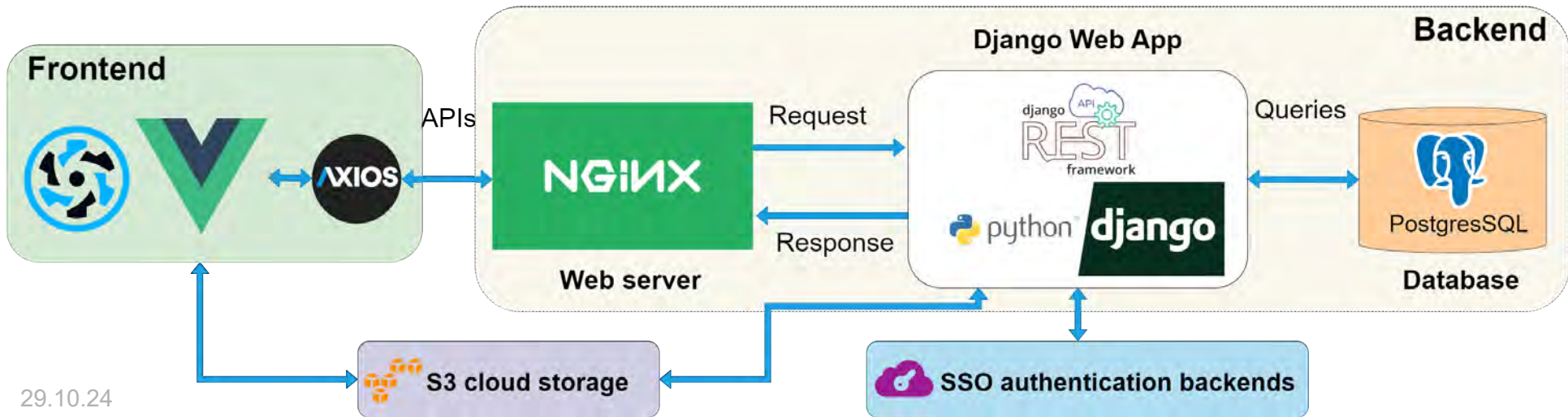


- The goal of the project is to provide a centralized, multi-community documentation storage service for collaborative activities
 - Implement the Core Features of the DocDB system
 - Enhance the access rights system to support groups and projects with flexible access control
- Specific requests from collaborations for organization of cooperation on publication activities and reports
 - Per-document editors
 - Document discussion system
 - Custom meta-data fields in documents
 - Built-in event calendar
 - Notification system
 - Document review process

SciDocsCloud Software Platform



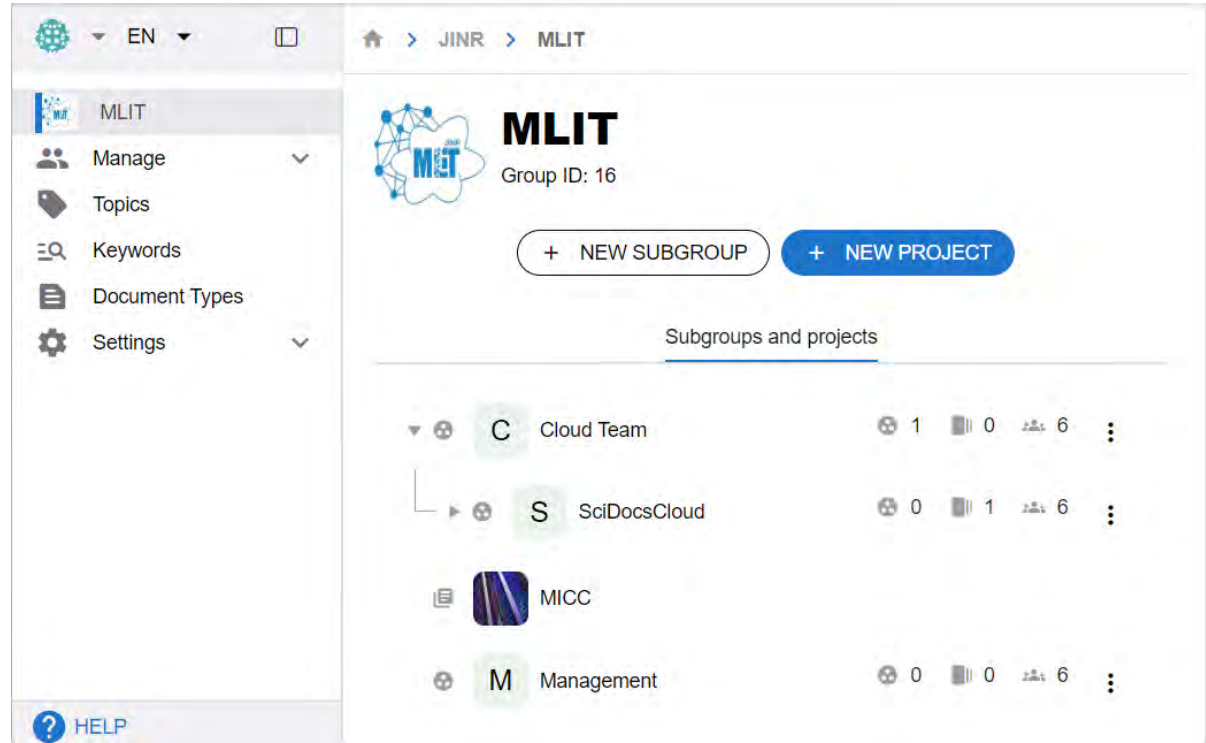
- Development of a more secure, reliable and feature-rich software platform for storing and tracking of documents in scientific collaborations
 - Utilizing modern software development tools
 - Leveraging our experience with DocDB
 - Supporting LocalFS and S3-compatible Cloud storage backends for document files
 - Supporting Single-Sign-On (SSO) authentication
 - Free and open-source software
- Storage Service for JINR research groups is based on this platform



Organizational Structure



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



Organizational Structure



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

The screenshot shows the 'Groups' page in GitLab. At the top, there's a breadcrumb 'Your work > Groups'. Below it, the title 'Groups' is displayed on the left, and 'Explore groups' and a 'New group' button are on the right. A search bar with 'Search by name' and a dropdown menu for 'Last created' are present. The main content area lists several groups with their names, roles, and statistics. The groups listed are: SciDocCloud (Developer, 0 members, 2 projects, 2 subgroups), Neutrino Computing Platform (3 subgroups), jj_nano (Developer, 0 members, 4 projects, 8 subgroups), cloud-team (Guest, 2 members, 26 projects, 11 subgroups), Configuration Management (3 subgroups), and htcondor (2 subgroups).

Group Name	Role	Members	Projects	Subgroups
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

Organizational Structure



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

The screenshot displays the BM@N project interface. The top navigation bar shows 'EN' and 'NICA > BM@N'. The main header includes the BM@N logo and 'Project ID: 183'. A '+ NEW DOCUMENT' button is visible. Below the header, there are filter options: 'Hide Column Filter', 'All', 'Search', and a list of columns: '#', 'Keywords', 'Document Type', 'Topics', 'Description', 'Abstract', 'Authors', 'Created at', 'Last Update', and 'Template'. The main content area shows a table of documents with columns for '#', 'Title', 'Topics', 'Authors', and 'Last Update'. The table contains three rows of document entries.

#	Title	Topics	Authors	Last Update
65-v4	Beam Lifetime and Beam Collimation in NICA Collider Operating in Heavy Ion Mode	Beam cooling, Beam dynamics, Collider, Machine detector interface, Safety, Vacuum	Сергей Иванович Плучков	3 minutes ago
69-v0	ТЕХНИЧЕСКИЙ ПРОЕКТ УСКОРИТЕЛЬНОГО КОМПЛЕКСА NICA, Том IV	Beam instrumentation, Collider, Cryogenics, Design reports, Engineering, Safety	Григорий Владимирович Трубников	23 days ago
68-v0	Present and approved accelerator facilities	Beam optics, Collider, Linac, Synchrotrons	Валерий Анатольевич Лебедев	about 1 month ago

Records per page: 20 1-20 of 65

Organizational Structure



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

Account	Source	Role
Ivan Sokolov It's your isokolov@jlnr.ru		OWNER
Nikita Balashov balashov@jlnr.ru	Direct member	Author
Valerij Lebedev	Direct member	Viewer
Oksana Timoshenko	Direct member	Maintainer

Organizational Structure



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

The screenshot shows a web application interface. On the left is a sidebar menu with the following items: Reports, Manage, Topics (highlighted with an orange box), Keywords, Document Types, and Settings. The main content area is titled 'Topics' and includes a search bar, a '+ NEW TOPIC' button, and a table of topics.

#	Topic	Description
1081	Cloud service Reports	
1083	Infrastructure Reports	
1082	Script Reports	

Records per page: 20 1-3 of 3

Organizational Structure



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

The screenshot displays a document management interface for a document titled "Storage Service for Scientific...". The interface includes a breadcrumb trail: JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75. The document is identified as "Reports Document 75-v11 (latest)" and has an "UPDATE DOCUMENT" button. The document details include: Document Title: Storage Service for Scientific Documentation; Keywords: (empty); Description: Report on the 12th Collaboration Meeting of the BM@N Experiment at the NICA Facility; Authors: Nikita Alexandrovich Balashov (balashov@jinr.ru) and Ivan Alexandrovich Sokolov (isokolov@jinr.ru); Topics: Cloud service; Files (2): Статья (Sokolov_Balashov_SciDocCloud_Article_v3.docx) (260.14 KB, Document) and Архитектура (Development Stac.png) (321.58 KB, Image); Comments (6): A comment by Ivan Sokolov from 5 months ago, "Добавил презентацию и заметку", with 3 replies, the last by Nikita Balashov. The right sidebar shows "Editors(4)": Sokolov Ivan, Gercenberger Konstantin, and Balashov Nikita; "Document #": reports-doc-75-v11 (latest); "Submitted by": Ivan Sokolov; "Updated by": Ivan Sokolov; "Document Created": 10 May 2024, 18:09; "Revised": 11 Oct 2024, 15:35; "Latest Version": reports-doc-75-v11; and "Other Versions" with a list of previous versions and their creation times.

Organizational Structure



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

The screenshot shows a web interface for a document titled "Storage Service for Scientific Documentation". The interface is organized into several sections:

- Editors (4):** A list of users: Sokolov Ivan, Gerckenberger Konstantin, and Balashov Nikita.
- Document #:** reports-doc-75-v11 (latest)
- Submitted by:** Ivan Sokolov
- Updated by:** Ivan Sokolov
- Document Created:** 10 May 2024, 18:09
- Revised:** 11 Oct 2024, 15:35
- Latest Version:** reports-doc-75-v11
- Other Versions:** A list of previous versions: reports-doc-75-v10 (Created at: 11 Oct 2024, 15:33), reports-doc-75-v9 (Created at: 11 Oct 2024, 15:25), reports-doc-75-v8 (Created at: 11 Oct 2024, 15:22), reports-doc-75-v7 (Created at: 11 Oct 2024, 15:21), reports-doc-75-v6 (Created at: 11 Oct 2024, 13:43), and reports-doc-75-v5 (Created at: 10 May 2024, 18:35).
- Document Title:** Storage Service for Scientific Documentation
- Keywords:** (empty)
- Description:** Report on the 12th Collaboration Meeting of the BM@N Experiment at the NIGA Facility
- Authors:** Nikita Alexandrovich Balashov (balashov@jinr.ru) and Ivan Alexandrovich Sokolov (isokolov@jinr.ru).
- Topics:** A list of topics, including "Cloud service".
- Files (2):** A list of files: "Статья (Sokolov_Balashov_SciDocCloud_Article_v3.docx)" (260.14 KB, Document) and "Архитектура (Development Stac.png)" (321.58 KB, Image).
- Comments (6):** A comment by Ivan Sokolov from 5 months ago: "Добавил презентацию и заметки". It has 3 replies, with the last reply by Nikita Balashov 5 months ago.

Organizational Structure



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

The screenshot shows a web interface for a document titled "Storage Service for Scientific Documentation". The interface includes a breadcrumb trail (JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75), a document title, a "UPDATE DOCUMENT" button, and a "Default" tab. The document details section contains fields for "Document Title", "Keywords", "Description" (Report on the 12th Collaboration Meeting of the BM@N Experiment at the NICA Facility), and "Authors" (Nikita Alexandrovich Balashov and Ivan Alexandrovich Sokolov). A "Topics" section lists "Cloud service". The "Files (2)" section, highlighted with an orange box, contains two items: "Статья (Sokolov_Balashov_SciDocCloud_Article_v3.docx)" and "Архитектура (Development Stac.png)". Below this is a "Comments (6)" section with a comment from Ivan Sokolov. On the right, a sidebar shows "Editors(4)" (Sokolov Ivan, Gerckenberger Konstantin, Balashov Nikita), "Document #:", "Submitted by:", "Updated by:", "Document Created:", "Revised:", "Latest Version:", and "Other Versions:".

Organizational Structure



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

The screenshot displays a document management interface for a document titled "Storage Service for Scientific...". The interface includes a breadcrumb trail (JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75), a document title, a description, authors (Nikita Alexandrovich Balashov and Ivan Alexandrovich Sokolov), topics (Cloud service), and a list of files. A right-hand sidebar shows the document's metadata, including its ID, submission and update details, and a list of other versions. An orange arrow points from the "Version control" bullet point in the text to the "Other Versions" section in the screenshot.

Document #:	reports-doc-75-v11 (latest)
Submitted by:	Ivan Sokolov
Updated by:	Ivan Sokolov
Document Created:	10 May 2024, 18:09
Revised:	11 Oct 2024, 15:35
Latest Version:	reports-doc-75-v11
Other Versions:	
reports-doc-75-v10	Created at: 11 Oct 2024, 15:33
reports-doc-75-v9	Created at: 11 Oct 2024, 15:25
reports-doc-75-v8	Created at: 11 Oct 2024, 15:22
reports-doc-75-v7	Created at: 11 Oct 2024, 15:21
reports-doc-75-v6	Created at: 11 Oct 2024, 13:43
reports-doc-75-v5	Created at: 10 May 2024, 18:35

Organizational Structure



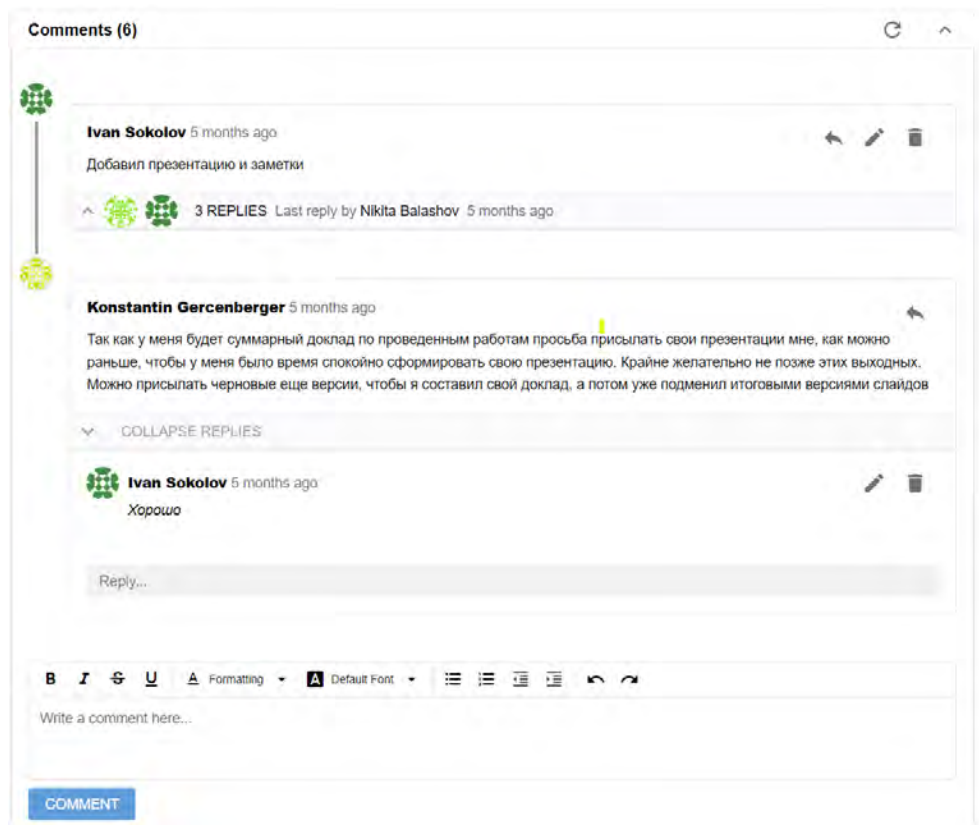
- Hierarchical organization
groups / subgroups → *projects* → *documents*
- The idea comes from GitLab
 - Using GitLab terminology
 - Project ≈ GitLab project
- Project contains
 - Document storage
 - Collaboration tools
 - Meta-data management tools
- **Document as an abstract entity**
 - Form with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

The screenshot displays a document page for 'Storage Service for Scientific...'. The interface includes a breadcrumb trail (JINR > MLIT > Cloud Team > SciDocsCloud > Reports > reports-doc-75), a document title, a description, authors (Nikita Alexandrovich Balashov and Ivan Alexandrovich Sokolov), and a list of files. A sidebar on the right shows the 'Editors(4)' list, which is highlighted with an orange box. An orange arrow points from the 'Editors' list in the sidebar to the 'Editors' list in the main content area. The sidebar also displays document metadata such as 'Document #', 'Submitted by', 'Updated by', 'Document Created', 'Revised', and 'Latest Version'. A 'Other Versions' section lists previous versions of the document with their creation dates.

Organizational Structure



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



Role-based Access Model



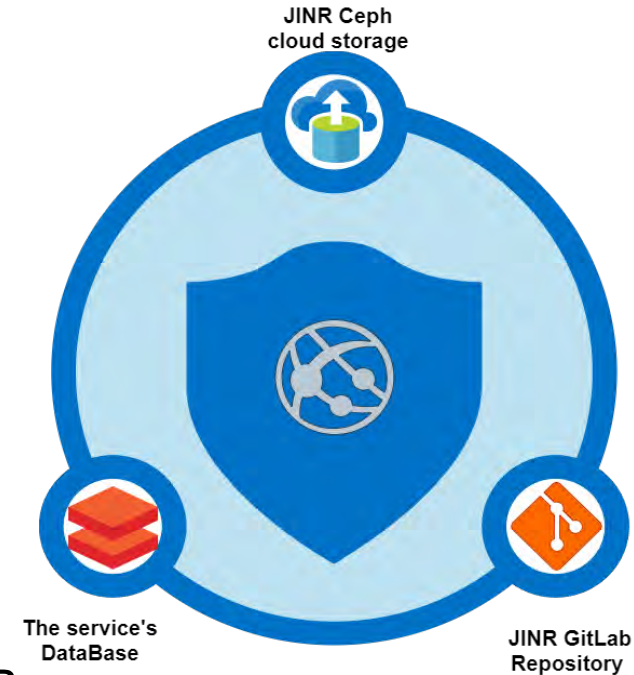
- Access restriction to groups or projects based on roles similar to GitLab
- Currently implemented roles: Owner, Maintainer, Author and Viewer
- Group/project membership
- Nested groups membership

Action	Viewer	Author	Maintainer	Owner
View all	+	+	+	+
Use comment block		+	+	+
Create document		+	+	+
Update document		+ -	+ -	+
Create group			+	+
Create project			+	+
Create subgroup			+	+
Manage members			+	+
Manage topics			+	+
Manage keywords			+	+
Manage document type			+	+
Edit group			+	+
Edit project			+	+
Remove project				+
Remove group				+
Remove document				+
Manage editor list				+

Storage Service for Scientific Documentation



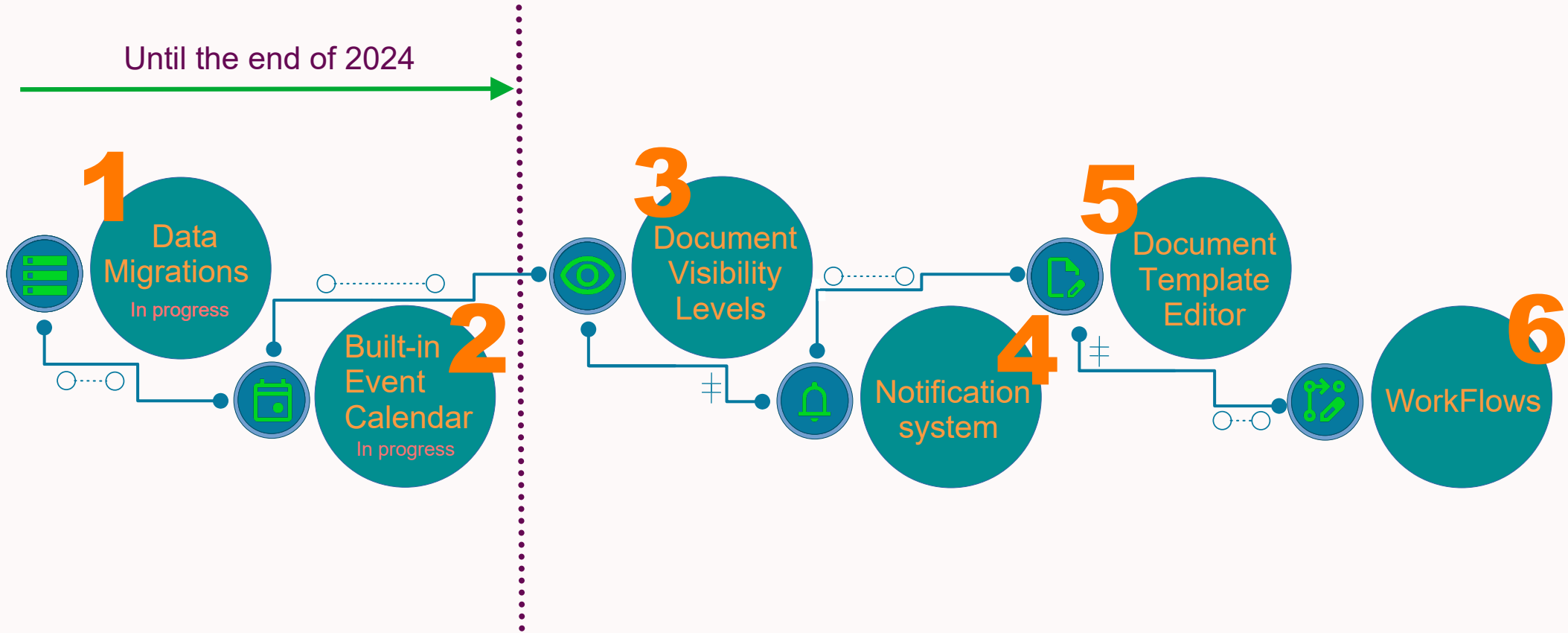
- The platform-based service has been deployed at JINR
 - JINR's Ceph cloud storage
 - Triple replication provided by Ceph
 - Easily scalable
 - Nightly Backups for Service Meta Information Database
 - Deployed on a VM in the JINR Cloud
 - Quick recovery after hardware failure
- Available to JINR employees and collaborators via *JINR Single Sign-On* (SSO) accounts at docs.jinr.ru



Road Map



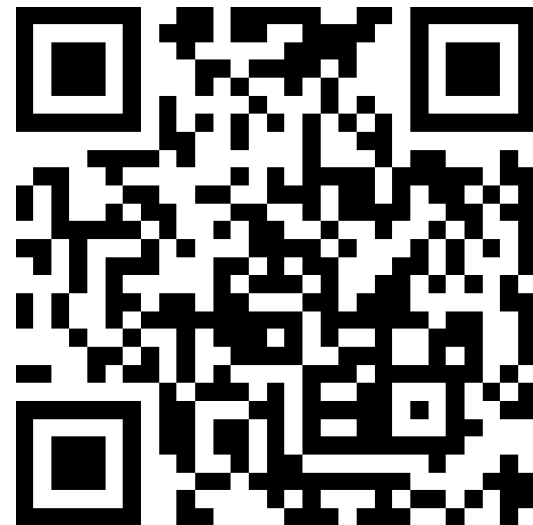
Until the end of 2024



Conclusion and FeedBack



- The SciDocsCloud platform is free and open-source software
 - Documentation in progress
 - Opportunities for contributions coming soon
 - Git-repository will be published soon
- The platform-based Storage Service for JINR research groups has been deployed
- The service is available to JINR employees and collaborators → docs.jinr.ru
- For questions, bug reports, or suggestions contact us at docs@jinr.ru



Thanks for your attention!

Platform Database Model

