PubDB

An analysis database for Belle II

David Koch¹, Thomas Kuhr¹
Belle II Germany Meeting, 2024-10-01

¹LMU







What this is all about

A web-app for Belle II to

- host documents (papers, theses, reports, posters, ...)
- guide users through the whole process of creating public papers
- act as a database of ongoing analyses

1. The software that runs Belle II's **document server** (<u>docs.belle2.org</u>) is outdated and needs replacement.

- 1. The software that runs Belle II's **document server** (<u>docs.belle2.org</u>) is outdated and needs replacement.
- 2. There is no streamlined way to follow the process of creating a paper.

- 1. The software that runs Belle II's **document server** (<u>docs.belle2.org</u>) is outdated and needs replacement.
- 2. There is no streamlined way to follow the process of creating a paper.
 - there are well-defined policies regarding the process of creating a Belle II paper, ie. when and how to form a review committee, how the review process should take place, who is responsible for what, ...

- 1. The software that runs Belle II's **document server** (<u>docs.belle2.org</u>) is outdated and needs replacement.
- 2. There is no streamlined way to follow the process of creating a paper.
 - there are well-defined policies regarding the process of creating a Belle II paper, ie. when and how to form a review committee, how the review process should take place, who is responsible for what, ...
 - no centrally organized mechanism which implements the policy: all e-mail based

- 1. The software that runs Belle II's **document server** (<u>docs.belle2.org</u>) is outdated and needs replacement.
- 2. There is no streamlined way to follow the process of creating a paper.
 - there are well-defined policies regarding the process of creating a Belle II paper, ie. when and how to form a review committee, how the review process should take place, who is responsible for what, ...
 - no centrally organized mechanism which implements the policy: all e-mail based
 - no database that keeps track of the status of ongoing analyses

→ PubDB

PubDB features

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents ✓
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents ✓
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents √
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓
- implement publication checklist with visual guide ✓

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents √
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓
- implement publication checklist with visual guide ✓
- clear overview over ongoing analyses ✓

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents √
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓
- implement publication checklist with visual guide ✓
- clear overview over ongoing analyses ✓
- comment system for review ✓

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents ✓
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓
- implement publication checklist with visual guide ✓
- clear overview over ongoing analyses ✓
- comment system for review ✓
- configurable notifications and alerts on status changes ✓

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents ✓
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓
- implement publication checklist with visual guide ✓
- clear overview over ongoing analyses ✓
- comment system for review ✓
- configurable notifications and alerts on status changes ✓
- automatic versioning of uploaded documents ✓

- replicate docs.belle2.org capabilities
 - upload pdfs and other types of documents √
 - categorize documents into public/private, papers, notes, theses, ... ✓
 - migrate all data from docs.belle2.org to PubDB
- metadata-rich, searchable database of analyses ✓
- implement publication checklist with visual guide ✓
- clear overview over ongoing analyses ✓
- comment system for review ✓
- configurable notifications and alerts on status changes ✓
- automatic versioning of uploaded documents ✓
- history log of every change to the metadata \checkmark

 integration with Gitlab, eg create a repository from a template if a new note / paper is created in PubDB

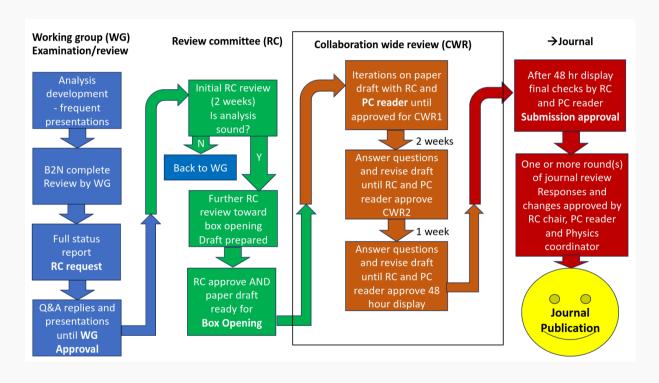
- integration with Gitlab, eg create a repository from a template if a new note / paper is created in PubDB
- project-managing support for analyses
 - search ✓
 - Labels such as status, datasets used, topic, ... ✓
 - · Gitlab-issues-like milestones

- integration with Gitlab, eg create a repository from a template if a new note / paper is created in PubDB
- project-managing support for analyses
 - search ✓
 - Labels such as status, datasets used, topic, ... ✓
 - Gitlab-issues-like milestones
- REST API (partly ✓)

- integration with Gitlab, eg create a repository from a template if a new note / paper is created in PubDB
- project-managing support for analyses
 - search ✓
 - Labels such as status, datasets used, topic, ... ✓
 - Gitlab-issues-like milestones
- REST API (partly ✓)
- HEPData integration

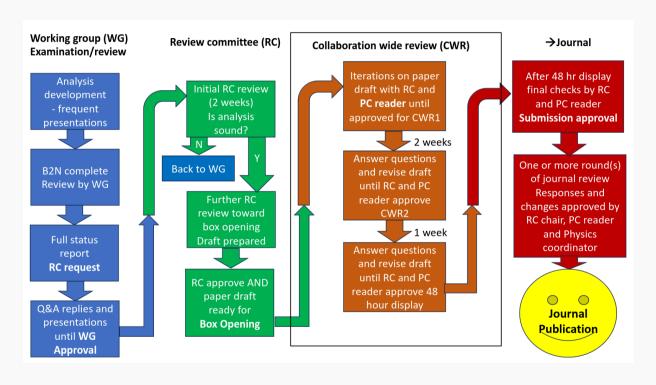
Demo

Demo



https://authors.belle2.org//deliver.pl?F=1

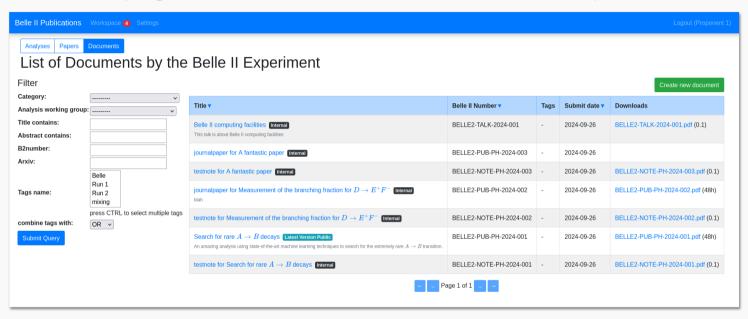
Demo



The following is a series of screenshots that demonstrate how PubDB guides (fictional) users along this flowchart.

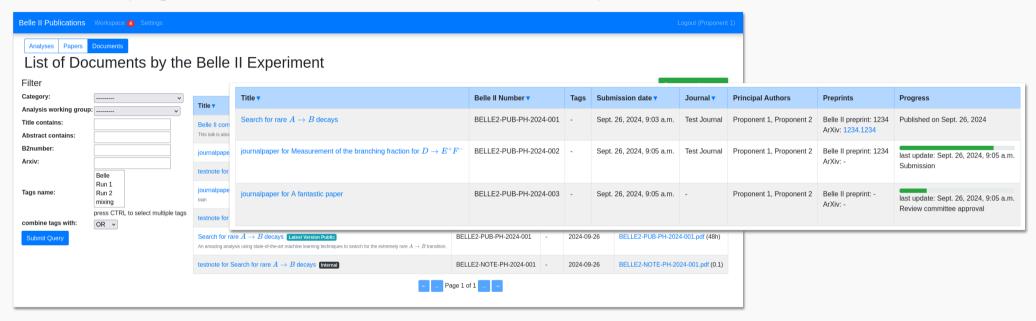
https://authors.belle2.org//deliver.pl?F=1

Welcome page with a list of all documents / analyses



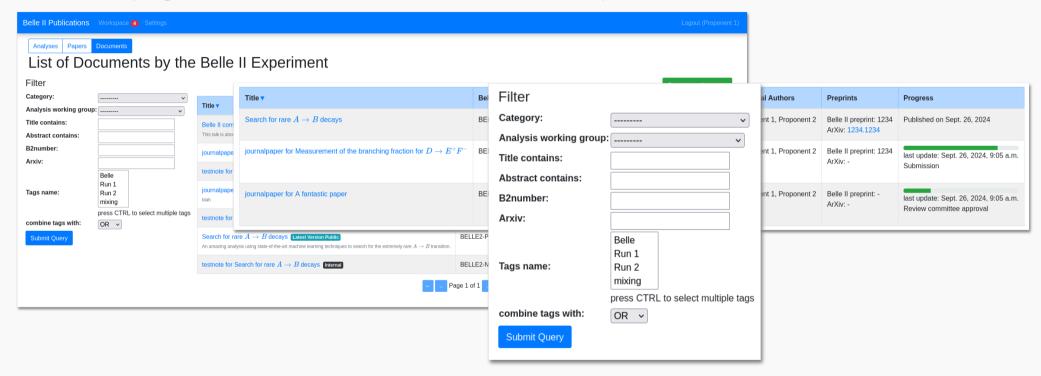
For unauthenticated users, only published papers are listed.

Welcome page with a list of all documents / analyses



For unauthenticated users, only published papers are listed.

Welcome page with a list of all documents / analyses



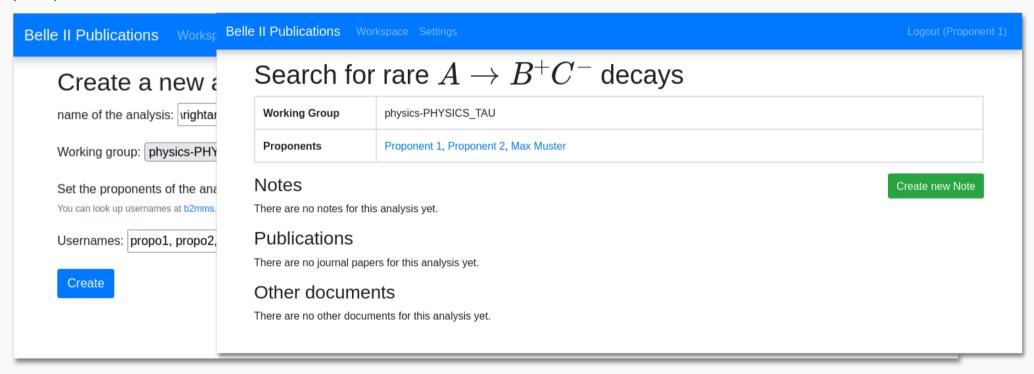
For unauthenticated users, only published papers are listed.

Creating a new analysis just requires a name, working group and a list of proponents:



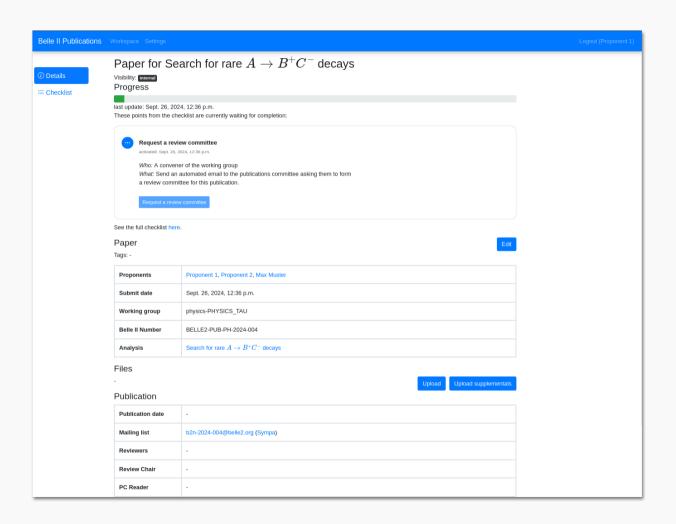
Several documents can be organized under an analysis: notes, papers, talks, ...

Creating a new analysis just requires a name, working group and a list of proponents:



Several documents can be organized under an analysis: notes, papers, talks, ...

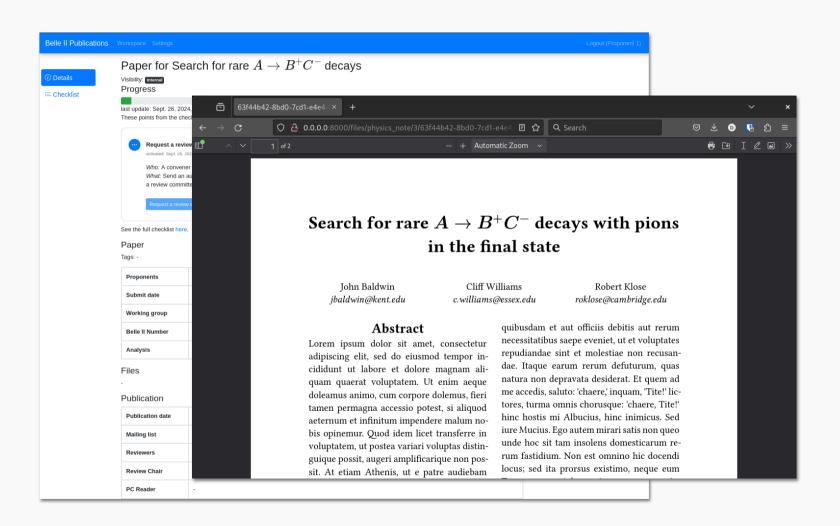
The view of a newly created **journal paper**. The main journey begins here.



The view of a newly created **journal paper**.
The main journey begins

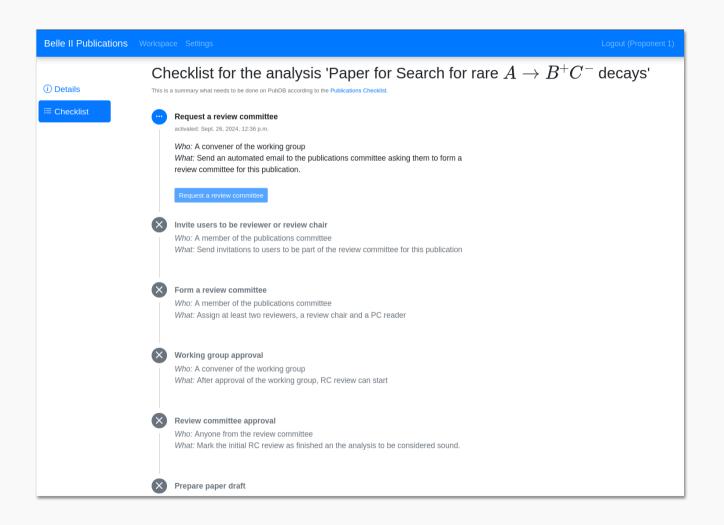
PDFs can be uploaded directly to PubDB.

here.

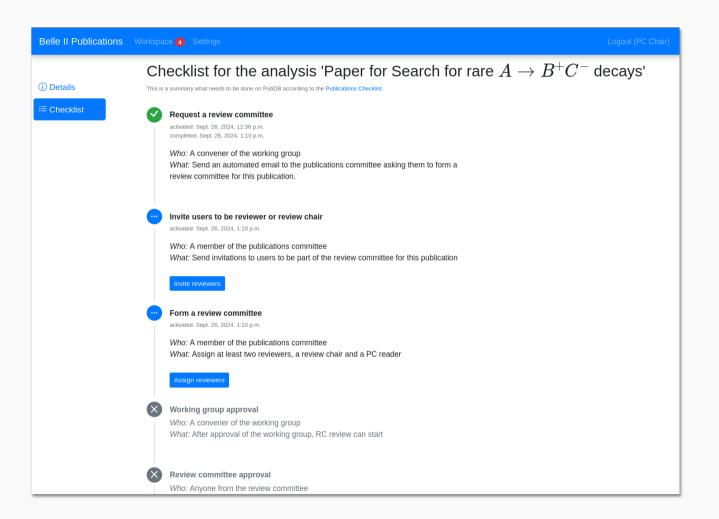


The **checklist**view: a step by
step guide
through the
process of
creating a
journal
publication

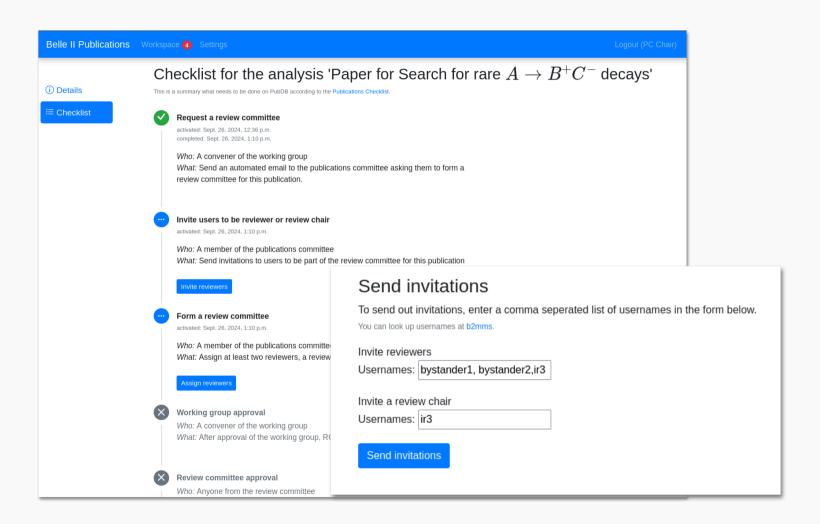
+ timestamps
when each task
was started and
finished



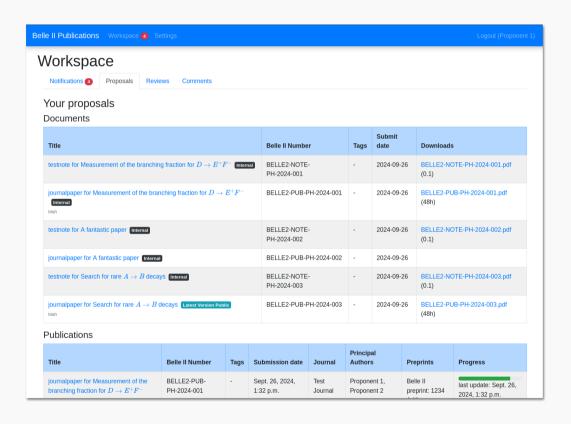
When it's time to form a **review** committee, a member of the publications committee can either invite reviewers or assign them directly.



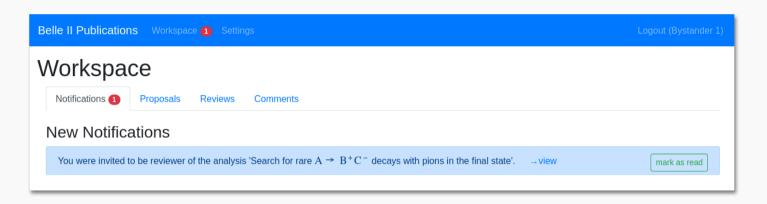
When it's time to form a **review** committee, a member of the publications committee can either invite reviewers or assign them directly.



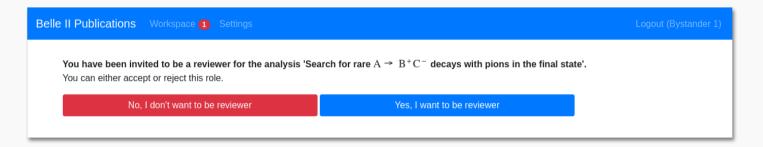
Introducing the **workspace**: an overview of your notifications, analyses you're working on, analyses you're reviewing and comments you've left



speaking of notifications, this is how it looks like if you've been invited to be reviewer:



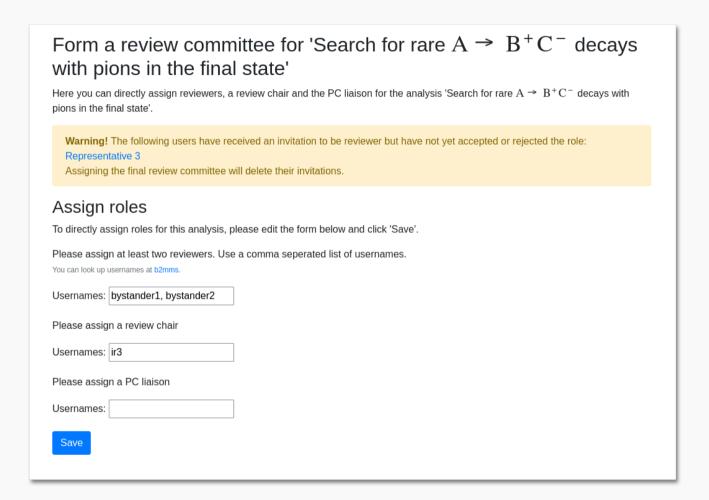
you can either accept or deny:



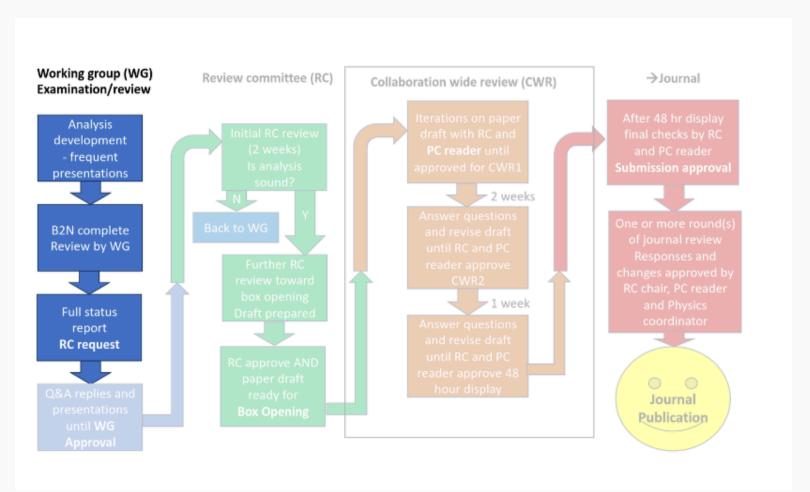
Accepting or denying reviewer invitations will in turn be visible to members of the publication committee:

Invite a review committee for 'Search for rare $A \rightarrow B^+C^-$ decays with pions in the final state' Here you can invite reviewers and a review chair for the analysis 'Search for rare $A \Rightarrow B^+C^-$ decays with pions in the final state'. Users who receive an invitation can accept or reject their role on PubDB and are assigned automatically according to their response. A member of the publications committee may edit the composition of the final review committee later. The following people received an invitation to be a reviewer for this analysis: Bystander 2, Representative 3 The following people received an invitation to be review chair for this analysis: Representative 3 The following people are currently in the review committee of this analysis (either by accepting an invitation or by direct assignment via a member of the publications committee): Reviewers: Bystander 1 Review Chair: -Send invitations To send out invitations, enter a comma seperated list of usernames in the form below. You can look up usernames at b2mms. Invite reviewers Usernames: Invite a review chair Usernames: Send invitations

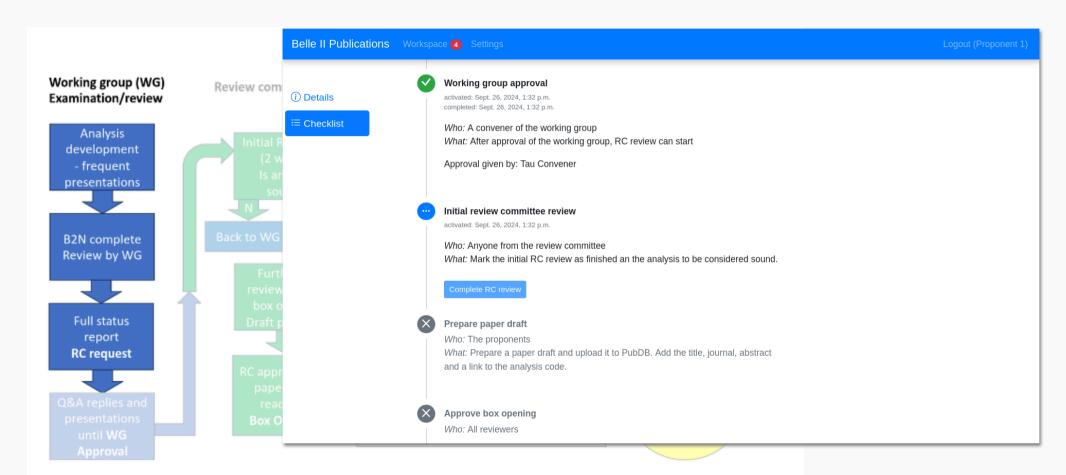
The review committee is settled when a PC member assigns them. The form will be automatically filled with everyone who accepted an invitation.



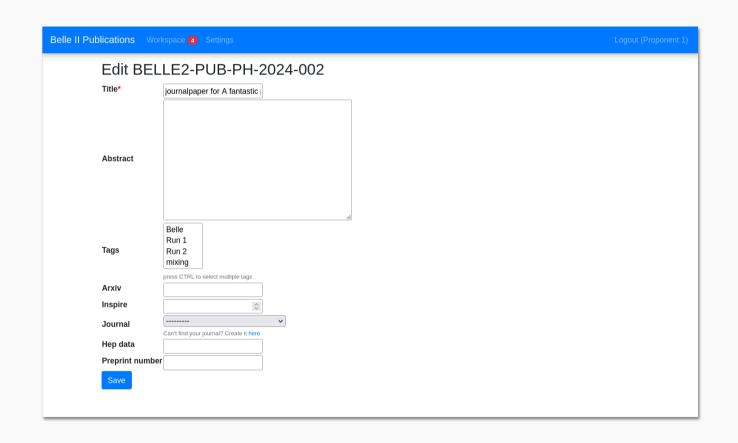
quick reminder of what we did so far...



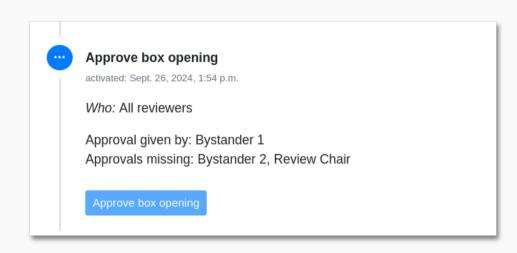
... and how to continue



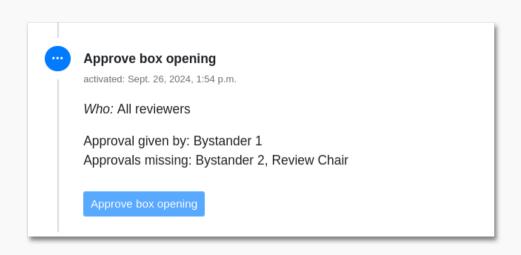
The **edit view** as seen by a proponent.



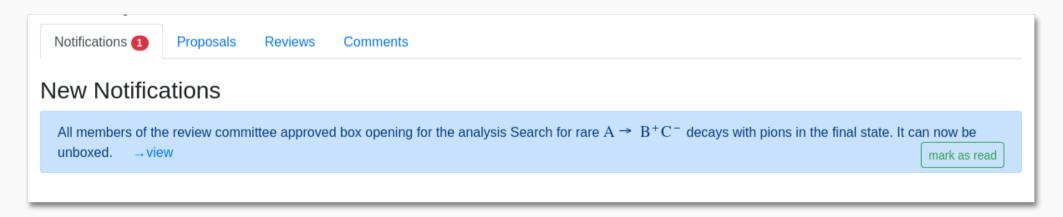
When it comes to approval of **box opening**, all members of the review committee need to tick a checkbox to indicate that they approve box opening.



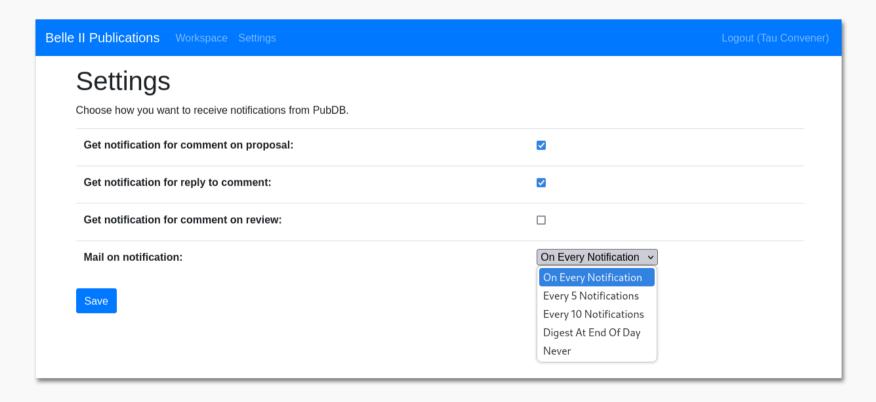
When it comes to approval of **box opening**, all members of the review committee need to tick a checkbox to indicate that they approve box opening.



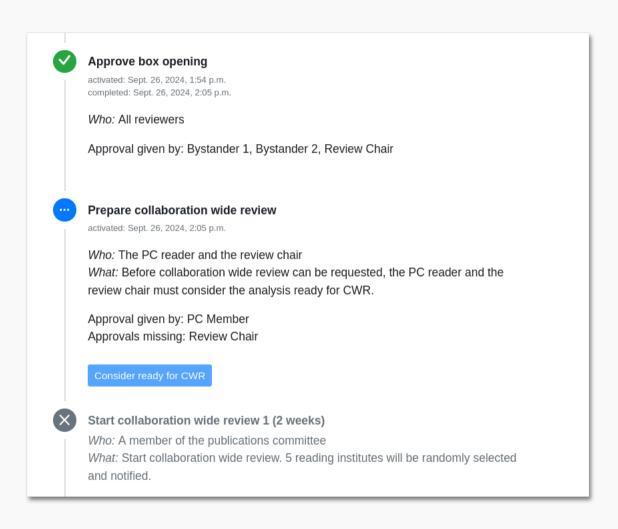
Once every reviewer has done so, all proponents will be notified:



By the way, you can configure if and how you want to get emails for notifications in PubDB:



Before **collaboration wide review** can start,
the draft must be
approved:



After it is approved, an automatically generated email gets sent to the publications committee:

Dear Publications Committee,

The proponents and reviewers confirm that the paper draft is in a shape that it meets the quality criteria for CWR1 and ask to start the collaboration wide review.

Title: journalpaper for email test

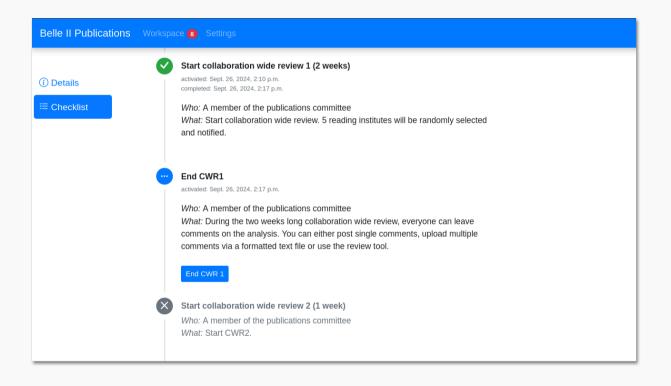
Reviewers: Review Chair <revchair@desy.de> (chair), Bystander 1 <bystander 1@desy.de>, Bystander 2 <bystander 2@desy.de>

PDF: BELLE2-PUB-PH-2024-002.pdf

Supplementary material: -s

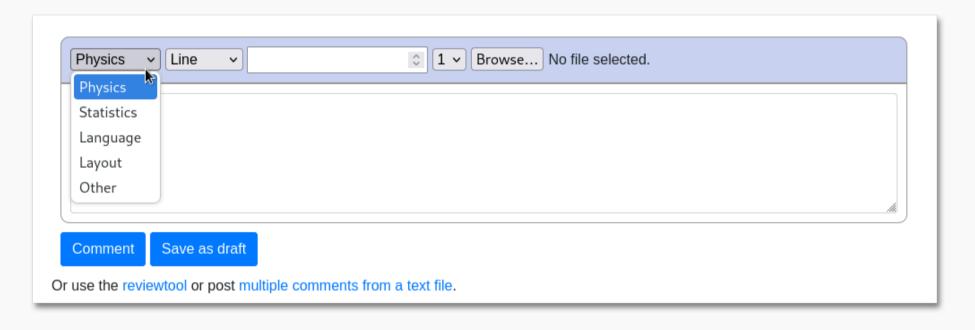
In order to start collaboration wide review, a member of the committee can visit the checklist and press the button "Start CWR1". Thanks!

Best regards, PubDB Bot Now a member of the publications committee can start collaboration wide review 1. A mail will be sent out to **coll-members** and the status of the paper will be indicated as "CWR1"



Everyone is invited to leave comments.

A comment contains (markdown-formatted) text that can optionally refer to a topic (physics, statistics, language, layout, other), a line (or figure or table) number and the version of the draft it is referring to.



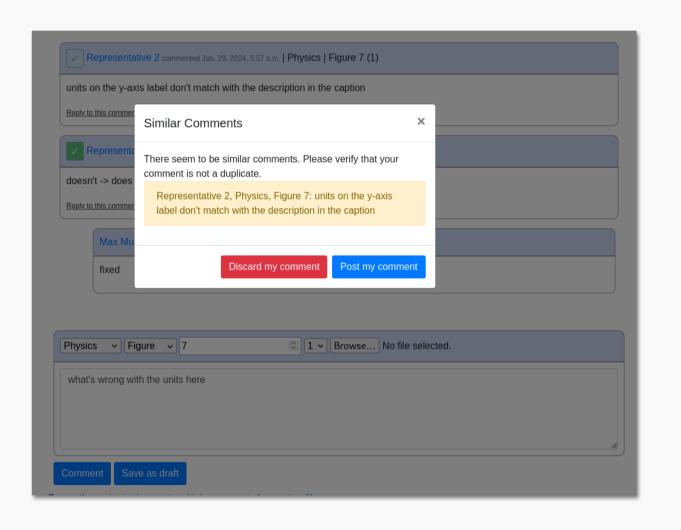
Uploading comments as a file is also supported.

replied to, they can be marked as resolved (by the author of the comment) and they can be filtered to help proponents address them in an organized manner.

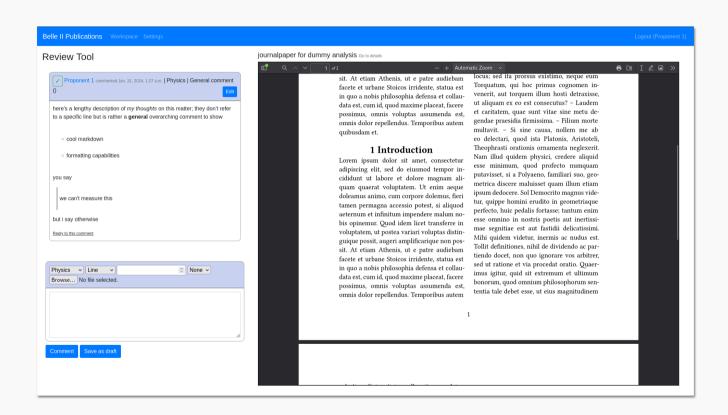
Comr	ments	
Filter con		
Kind:		
Refers to	version: [
Marked a	as resolved:	
Status: -		
User:	v	
Submit (Query	
	Representative 2 commented Jan. 29, 2024, 5:57 a.m. Physics Figure 7 (1)	it
units	s on the y-axis label don't match with the description in the caption	
Reply	to this comment	
	Representative 2 commented Jan. 29, 2024, 6 a.m. Language Line 342 (1)	
	Trepresentative 2 commenced data 25, 2024, 6 a.m. Language Line 042 (1)	
does	sn't -> does not	
Reply	to this comment	
	Max Muster replied Jan. 29, 2024, 6:01 a.m.	
	fixed	

please excuse slightly ugly HTML

PubDB tries to detect similar comments in order to avoid duplicate comments.



The **review tool** view shows the comment form and your own comments side by side with the draft pdf to aid with a smooth workflow for reviewers.



Multiple versions of the paper as well as supplemental material can be uploaded by the proponents.

- version numbers in the format "X.Y" are automatically increased
- proponents can mark an upload as a "major change" which increases the major version
- on CWR1, CWR2 and the 48h display, versions are aliased so that they can be found quickly

Files

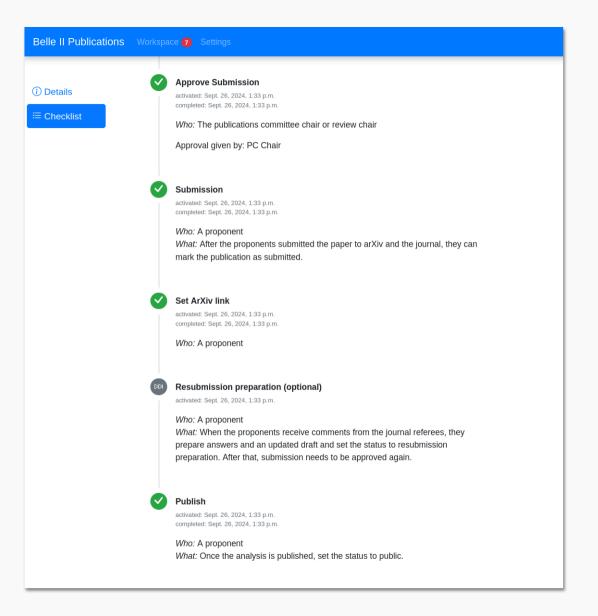
- BELLE2-PUB-PH-2024-005.pdf (versions: 1.3, 48h, CWR2, 1.0, CWR1, 0.1)
- BELLE2-PUB-PH-2024-005-supplemental.pdf (versions: 48h)

Upload

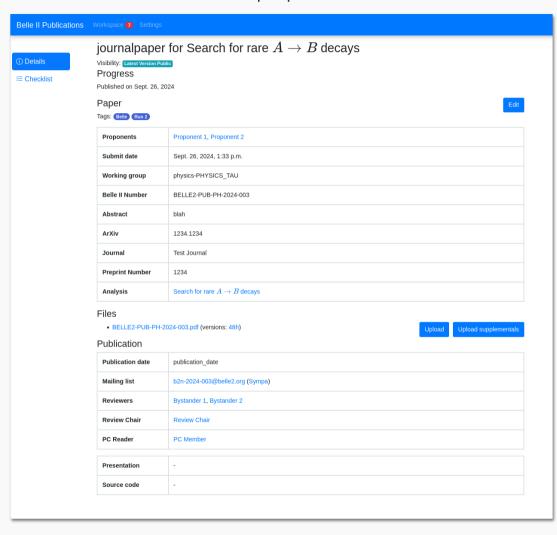
Upload supplementals

After CWR1, CWR2 and the 48h display it becomes straightforward (at least on PubDB): someone from the publications committee can approve submission, a proponent can set the status to **submitted** and finally set it to **published**.

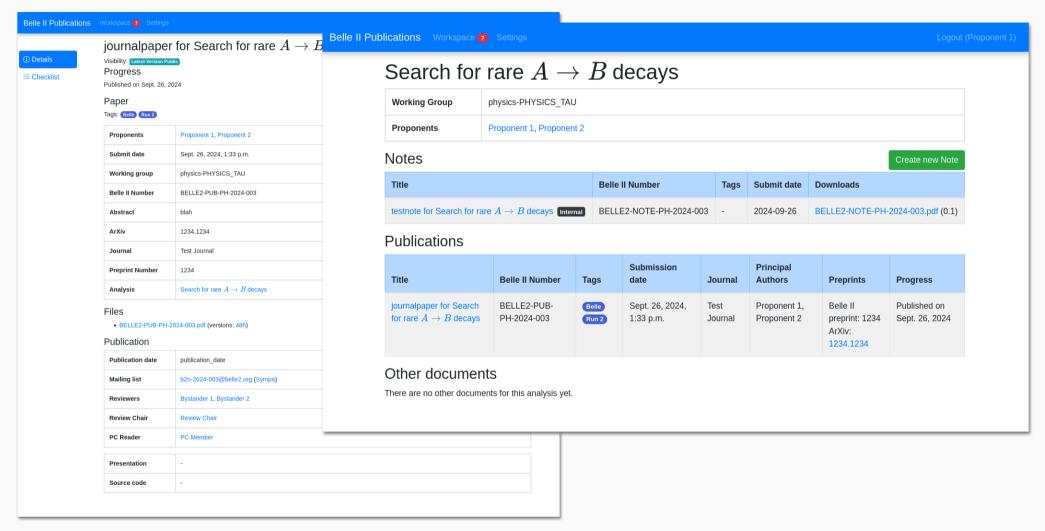
Yay!



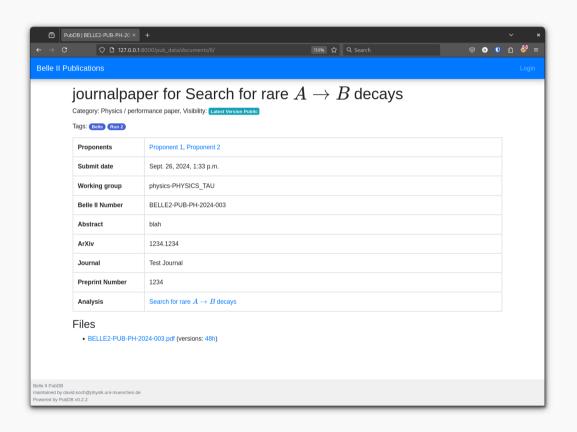
This is how the final paper record looks like



...and the record for its analysis



Public view of the same paper



Development

Timeline

October

- finish beta testing
- implement critical fixes / features
- prepare migration from Invenio to PubDB

November

- migrate
- test migration

December

- collaboration switches to PubDB
- 2.12. put Invenio in read-only mode

Timeline

· 28.2. turn off Invenio



Development takes place on <u>DESY Gitlab</u>.

Try it out yourself! Easiest with docker:



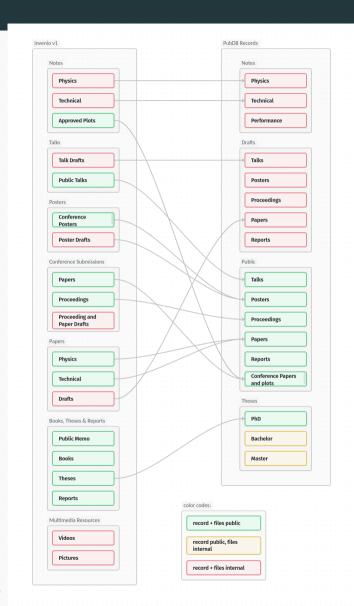
```
docker login gitlab.desy.de:5555 # use DESY credentials
docker run -p 8000:8000 -it gitlab.desy.de:5555/belle2/admin/pubdb/demo:0.2.2
```

then visit http://127.0.0.1:8000 in your browser. You can log in as any test user to visit the site from the point of view of any role (eg. proponents of a testanalysis, conveners, PC member, ...). Find more information in the README.

To report a bug, give feedback or to suggest a feature, please open an <u>issue</u> or drop me a message at <u>david.koch@physik.uni-muenchen.de</u> or/and meet me in the coffee break.

Backup

Migration of categories



Document server

The current docs.belle2.org is used to host and categorize pdfs and plots.

Both server and software (called **invenio** v1) are severely outdated and unmaintained → needs replacement for security reasons, especially the server



Document server

The current docs.belle2.org is used to host and categorize pdfs and plots.

Both server and software (called **invenio** v1) are severely outdated and unmaintained → needs replacement for security reasons, especially the server



Possible strategies:

- 1. update only the server and don't change invenio
- 2. update both server and invenio
- 3. implement document hosting capabilities in PubDB

Updating invenio

Invenio has been completely overhauled, is now at version 3 and is called **InvenioRDM**.

- spent a considerable amount of work trying to get a production-ready InvenioRDM instance and preparing migration from v1 to v3, but did not succeed
- vast complexity for a relatively small aspect of the problem (hosting documents)
- → not feasable for our usecase
- → implement hosting capabilities ourselves in PubDB
- \rightarrow one app to rule them all

It is possible to upload many comments at a time via a specially formatted text file

