

The extended abstract and the corresponding paper are targeted for business session.

SoleCRIS "Archiving"-feature

– How a university utilizes SoleCRIS to promote open access

Niilo Lahti
Product owner
Solenovo

In this paper, I will introduce the "Archiving"-feature and its basic functionalities in SoleCRIS.¹ Afterwards, I will specify how our university customer utilizes and how organizations generally may utilize the feature to promote open access of their publications by self-archiving or by primary archiving them. I will end the paper by discussing plans for further development of the feature.

"Archiving"-feature in SoleCRIS

All publications in SoleCRIS have their general publication information in their respective "General" publication forms. These custom forms, that are classified based on the publication types of the Finnish Ministry of Education,² are used to collect basic publication information. "Archiving"-feature provides an additional form, either self-archive or primary archive -form, for those publications that are meant to be archived in the university's repository. From the repository, the university puts the publications on display in their website thus promoting their open access.

As a necessary background information, SoleCRIS has two logged user roles: Researcher and Librarian.³ Researcher is the basic user who most commonly adds their own research information to the system and searches research information from it. Their options are streamlined to promote easy information collection. Librarian has more alternatives for editing research information, taking reports, and administration.

Before a Librarian role can add an "Archive"-entity to a publication, the publication must meet specific requirements:

- Publication does not already have an "Archive"-entity
- And user (Researcher or Librarian) has set the "General"-publication in a specific process state: "Incomplete", "Approved", or "Not for data collection"⁴

¹ SoleCRIS is a current research information system by the Solenovo company. See further <https://www.solenovo.fi/#1>.

² See e.g., <https://www.aka.fi/en/research-funding/apply-for-funding/how-to-apply-for-funding/az-index-of-application-guidelines2/list-of-publications/> and <https://wiki.eduuni.fi/display/cscsuorat/7.5+Julkaisutyypit> (in Finnish).

³ SoleCRIS has also a Public view (from which a user logs in). See e.g., <https://oulu.cris.fi/en>.

⁴ Librarian uses different process state in their research entity handling workflow. "General"-publication entities/forms and "Archive"-entities/forms have different process states. For "General", the process states are called: "Approved", "Imported", "Incomplete", "Unprocessed", "Not for data collection", "Doesn't belong

- And user has selected one or more of the following choices:
 - User has selected “I wish to self-archive my publication” in the “General”-publication
 - User has added an attachment file to the “General”-publication
 - Librarian has selected “Availability” information as follows:
 - “Open access status at the time of entry” is selected as “1=Open access”
 - And “Open access status of publication channel” is one of following: “1=Fully open access publication channel”, “2=Partly open access publication channel” or “3=Delayed open access publication channel”
 - And “Publisher’s license” is one of following: “1=CC BY”, “2=CC BY NC”, “3=CC BY NC ND”, “4=CC BY SA”
- Additionally, we wanted to prevent publications, that had been archived before the “Archiving”-feature, to be eligible for (re-)archiving
 - Publication is eligible for archiving only if “General”-publication has been updated after 31.10.2023.
 - If “Self-archived”-value of the “General”-publication is “Yes”, then the publication is not eligible for archiving

When these requirements are met, the Librarian can see the corresponding publications in the “Unprocessed” “Archive processing”-queue. The Librarian can add an “Archive”-entity for these publications. The Librarian selects either self-archive- or primary archive -form based on their need.⁵ The Librarian then fills the basic fields in the form which differ to some extent based on the selected form. The common basic fields include among others: “URN”, “Abstract” and “Language of abstract”, “Citation”, “Version”, “Copyright and license information”, “License type”. Self-archive-form specifically has e.g., “Embargo period in months” -field and Primary archive -form e.g., doctoral thesis related information.

Notably, archived publications are meant to displayed with their actual research material. Thus, “Archiving”-feature requires a robust attachment handling. E.g., when a Researcher adds an attachment to a “General”-publication, that attachment shows also in the “Archive”-entity ready to be processed. When Librarian selects an attachment for processing, they can change the attachment’s “Filename”, its “Display order” in the repository, whether it is “Restricted material”, and whether it will be “Published in repository”. The publication (“General” + “Archive” -entities) cannot be transferred to the university’s repository without selecting “Published in repository”: Yes for at least one attachment. Moreover, “Archive”-forms have some data format (schema) and process state validations. Once the Librarian has set the “Archive”-entity’s process state as “Archived” or “Archived but unpublished” and all the validations are met, the publication is transferred to the university’s repository via integration(s).

Summary

- In SoleCRIS, all publications have “General”-publication forms (for the purposes of MinEdu)
- Due to “Archiving”-feature, a publication can be supplemented with either a self-archive- or primary archive -form (in addition to the “General”-form).
 - “General”-publication must meet specific requirements before a user can add an archiving-entity/form to the publication.
- “General”-entities and “Archive”-entities have separate process states and thus separate workflows.
- “Archive”-forms have distinct functionalities for handling attachments.

to university”, and “Only to researcher’s publication list.” For “Archive”, the process states are called: “Unprocessed”, “Archiving incomplete”, “Archived but unpublished”, “Archived”, and “Archiving not possible”.

⁵ Self-archive has already been published somewhere while primary archive is published for the first time.

Utilizing “Archiving”-feature as an end user

This part of the paper focuses on the pros and cons for the end user who utilizes the feature. We have gathered actual user experience of the feature via job shadowing. This part and the previous part of the paper introducing the feature are interwoven to some extent. Our customer organization has different people for handling “General”-publication forms, and, in terms of archiving, self-archive and primary archive -forms. All the handlers use the Librarian role, which always has the same rights.

The organization utilizes a separate application from which an “Archive”-handler checks for potential publications that have recently been added to SoleCRIS (usually by import but sometimes by manual addition). Afterwards, “Archive”-handler checks whether the publications have sufficient information in their “General”-forms in SoleCRIS. If the information is sufficient, then “Archive”-handler likely just changes the process state of the “General”-form (usually from “Imported”) to “Incomplete”, “Approved”, or “Not for data collection” (see previous section). If the information is insufficient, then “General”-handler usually supplies the necessary information. Either handler may ask for a Researcher to provide any missing attachment.

When the “General”-form has all the necessary information for potential archiving, the publication appears in the “Unprocessed”-queue in the “Archive processing”. “Archive”-handler selects publications based on priority and based on his/her job description whether the publication needs self- or primary archiving. “Archive”-handler adds either a self-archive- or primary archive -form to the publication.

When “Archive”-handler processes the “Archive”-form, they often need to look at the “General”-form for specific information. In such cases, they must keep several browser tabs open or navigate between “General”-form and the “Archive”-form.

Another case, when “Archive”-handler must leave the “Archive”-form for a moment is when they must generate a URN-value for the “URN”-field. The handler uses an in-house application or goes to generator.urn.fi and copy-pastes the generated value to the field. They utilize the URN-value to replace the names of attachments in the “Archive”-form with consistent URN-based filenames.

A particular issue pertains to the “Abstract”-field in the primary archive -form. The handler copy-pastes the abstract from another environment. It is important for the handler to know how the abstract lines up to see how it will appear in the repository. Currently, the handler must use other means to gauge this.

Table 1 Usual tasks of different user roles regarding “Archiving”-process

Task / Role	Researcher	Librarian (“General”-handler)	Librarian (“Archive”-handler)
Enables a publication to have an “Archive”-form	X	X	(X)
Adds an “Archive”-form to publication (self-archive or primary archive)			X

Adds attachments	X	X	X
Handles "Archive"-forms			X

Summary

- While an "Archive"-handler has all the necessary functionalities for processing "Archive"-forms, they also have specific issues that slow the workflow:
 - Distinguishing publications that need to be prioritized
 - Necessity to navigate between "General"- and "Archive"-forms
 - Necessity to use other systems
 - Difficulty to handle filenames and "Abstract"-information

Optimizing "Archiving"-feature

We are planning to develop "Archiving" -feature further based on our job shadowing observations and customer feedback. Currently, customer has all mandatory functionalities for their work. Thus, we focus on optimizing the workflow by removing unnecessary steps and otherwise streamlining the handling process. The preliminary solutions that I suggest below refer to the issues that the "Archive"-handler comes across in their workflow, discussed in the previous section.

When the "Archive"-handler selects potential publications from the "Unprocessed"-queue, it is sometimes difficult to determine from the readily available publication data, which publications they should prioritize. This issue is solved by adding the "Closed access"- and "Open access"-status to the search results and the ability to order the publications based on that information. Thus, the handler can easily prioritize the "closed access"-publications.

When the handler processes the "Archive"-form, it is time consuming for them having to navigate between "General"- and "Archive"-forms (or keep several tabs open). We want to solve this by showing relevant "General"-publication information while the handler edits the "Archive"-form. We plan to show "General"-publication information in a sidebar so that the handler does not have to navigate to "General"-publication or keep several browser tabs simultaneously open. Sidebar could have other similar applications in the system. Since the sidebar is often closed, it is generally designed to contain useful additional information which does not immediately affect the user's actions in the current view.

Another distinct time-consuming issue for the "Archive"-handler is when they must use a separate website or application to fill the "URN"-field in the "Archive"-form. We want to streamline this part of the process by creating a button next to the "URN"-field that automatically generates URN-value. Since the "Archive"-handler needs to use the URN-value to systematically rename the attachment filenames, we want to semi-automate this part of the workflow as well. We plan to create a button next to each attachment that is being edited. When the "Archive"-handler clicks button, the system generates a filename (that replaces the current filename) for the attachment in the "Filename"-field based on the URN-value.

Yet another time saver is to automatically transfer specific data from "General"-form to "Archive"-form and vice-versa as some data is collected in both forms. E.g., "Publication year" could be gotten from "General"-form to "Archive"-form when the latter is being created. On the other way around, "Link to online publication" and "Link to self-archive version" -field values could be overwritten in the "General"-form once they have been updated in the "Archive"-form.

“Archive”-handler has more versatile needs for viewing and editing “Abstract”-field when they fill the information. We plan to provide a separate modal where the handler has basic text editing tools for retaining text formatting. Moreover, we plan to enable stretching of the modal for the handler to check how the text lines up.

Sometimes the “General”- and “Archive”-handlers need to communicate with each other and researchers regarding certain publications. Currently, SoleCRIS contains commenting-feature but not a more comprehensive means of communication. We want to enable this by creating a system-wide pinning and notification features. With these features, user can pin users to research entities⁶ and message them, what they want them to do with those entities. Users would receive information about pinnings and messages to a notification view and potentially to email as well.

Summary

- We want to optimize “Archive”-entity workflow by
 - Making it easier to distinguish publications that need to be prioritized
 - Decreasing the necessity to use other systems
 - Decreasing the navigation within SoleCRIS
 - Optimizing the information handling
 - E.g., URN, filenames, and abstracts
 - Enabling pinning and communication within SoleCRIS

⁶ Publications, research activities, conferences, and in the future, research datasets.