

Description

This specification clarifies RUcore policies regarding PDFs submitted to the repository.

Specification

In consideration for our need to provide long-term preservation to academic works deposited in RUcore, and in keeping with Open Access policies, the repository must only accept PDFs that have no file-based security restrictions embedded in the document. Such restrictions include those which prevent modification, printing, cutting and pasting of text or other elements, or which encrypt the PDF file using a password in order to control who may access or print the contents of a PDF document.

Deposit modules and the Workflow Management System should be configured such that the platform will detect PDF files with these restrictions in place at the time they are uploaded, and decline to accept such files, informing the depositor or user about the presence of these security restrictions on their PDF file and our reasons. The wording for this message to the user should be as follows:

The PDF you have submitted contains a software-based restriction that prevents certain actions, such as printing, editing or copying. Unfortunately, we cannot accept PDFs which contain such restrictions, as they prevent RUcore from effectively preserving them in the long-term. To continue, you will need to deposit a PDF which does not contain any such restrictions.

If you did not put this restriction in place, it is possible that it was added earlier in the life cycle of the document by a publisher or other source. You might need to contact that source to request an unrestricted copy.

If you need additional information or assistance, please contact us at rucore_support@rmail.rutgers.edu.

Background

We have encountered an issue where a certain number of faculty-submitted PDFs could not be modified such that a cover sheet could be added, per our SOAR/Open Access specifications. We determined that this problematic set of PDFs had a software-based restriction placed on them, preventing a user from changing the contents of the document. Certain PDF authoring software packages have the ability to place this and other restrictions on PDF documents, which can include the following conditions:

- Prevent the extraction of pages from a PDF document
- Restrict addition or removal of comments, or modifying certain electronic form fields
- Restrict the insertion, deletion or rotation of pages
- Restrict copying and pasting, or searching of text or other elements
- Require a password to open the document, or to print pages

By and large, such software-defined restrictions are unreliable, as some PDF tools can bypass these limitations relatively easily. In fact, a number of commercial programs exist solely to disable these restrictions, and still other software packages simply ignore any such restrictions, occasionally (but unreliably) permitting full functionality regardless of what is specified. Although this could be exploited to circumvent such restrictions, there is a concern that if

an author of a faculty submission has knowingly placed such a restriction on a PDF, it may be interpreted as implying additional rights and access restrictions on their work beyond any deposit agreements the depositor has endorsed.

Another potential concern is that the restriction may have been placed without the original author's knowledge or permission. Embedded metadata from some questionable PDFs suggests that the file may have been generated by an academic publisher (e.g. Elsevier or EBSCO), and the publisher may have added the restriction to the file as part of its file processing pipeline. A related concern is that such restrictions, if added intentionally, may imply licensing conditions which run counter to the Faculty Deposit Agreement that each author agrees to prior to depositing their work. Regardless of origin, the specific reasons and intentions behind such restrictions are often unknown to us, potentially limiting our ability to act.

Beyond simple cover sheet generation, such files present additional concerns. From a preservation standpoint, these files are at risk should the RUcore team be required to undertake a format migration, or other required preservation-level event. An example of such a possible instance is if a new document format supersedes PDF/a, rendering our existing files obsolete from a format perspective. It is for these reasons that restricted PDFs cannot be accepted as archival preservation datastreams for inclusion into RUcore.