# Digitally signed PDFs with attachments

So far, OL Connect had the ability to digitally sign PDFs, and to add attachments to PDFs. However, creating a signed PDF with attachments was not possible. With version 2020.2, we now have a way to do this that works for regular PDF, as well as for PDF/A-3. This enables use cases such as the Thai e-Tax invoice.

This article explains how to combine the capabilities of OL Connect to create digitally signed PDFs with attachments.

What was improved and why

Creating a digitally signed document, which can be done with Connect's Output Creation, and then adding attachments to it, will cause the digital signature to no longer be valid. This is because a digital signature is meant to guarantee that a document was not altered since it was signed.

So, to be able to produce a signed PDF with attachments, it is essential to add the attachments before digitally signing it. Unfortunately, Connect ignores attachments in PDF input files, so attachments added with Workflow's Add PDF/A-3 Attachments task, would not be included when signing that PDF with Connect's Create Output.
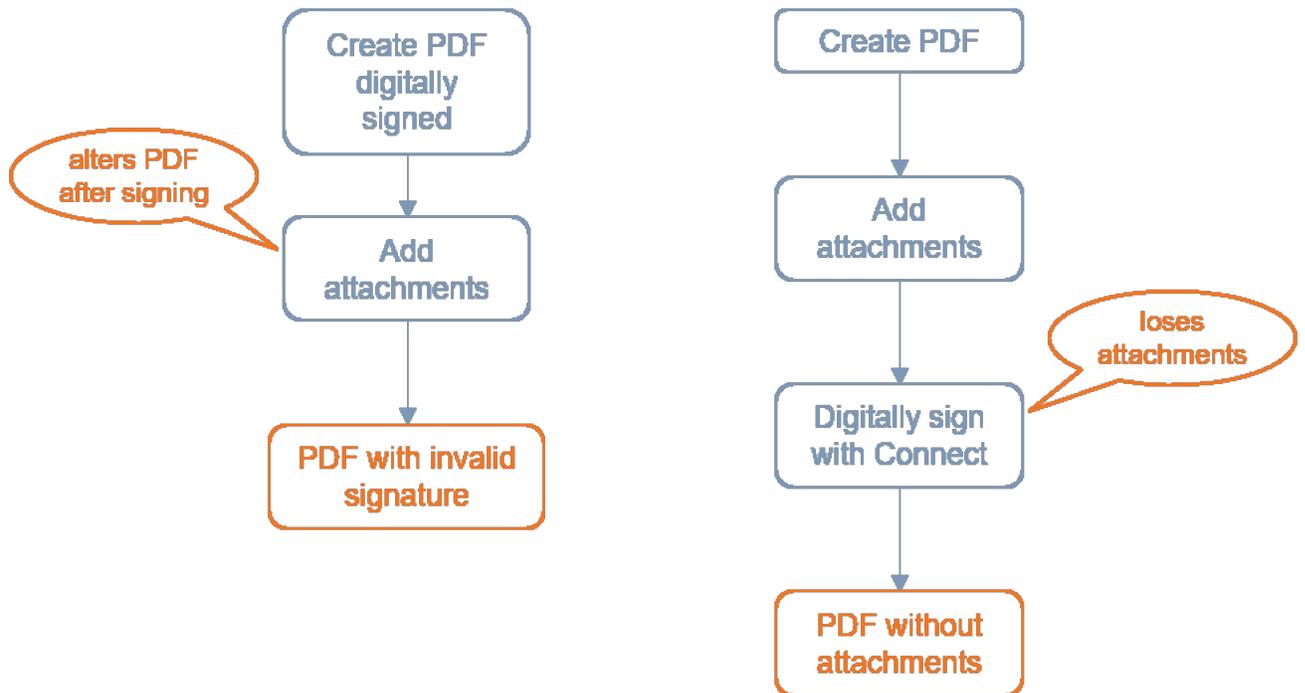
Changing signed PDFs is something you should simply not do. But Connect losing attachments is something that can be fixed.

Both Content Creation and Output Creation would not retain attachments when processing PDF. For Content Creation this kind of makes sense because it uses the input PDF as a background; it's not obvious that such a background would bring along some attachments. However, Content Creation can now be skipped. So, if we support attachments in Output Creation, we can produce the desired result. So that's what we did for 2020.2.
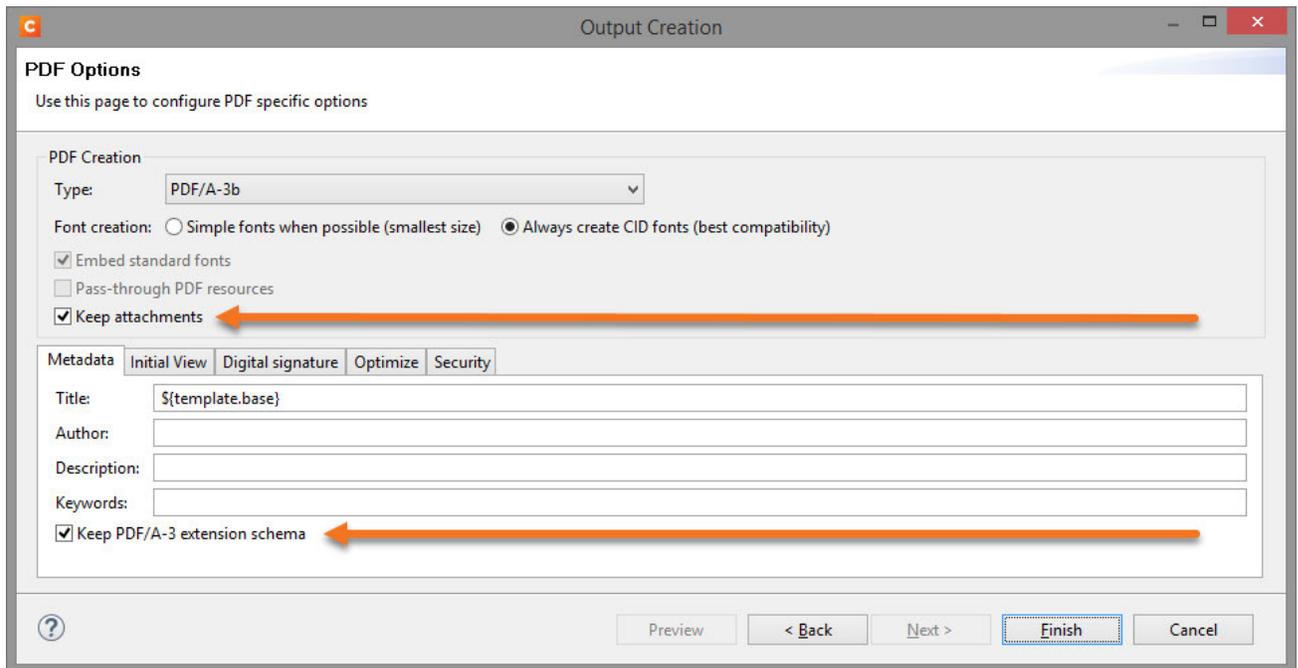
OL Connect 2020.2 has settings for keeping attachments, and, in case of PDF/A-3 output, for keeping metadata.

Overview of the process

With the improvements we made, creating a digitally signed PDF with attachments is a multi-step process:

- use OL Connect to create PDFs;
- add attachments to these PDFs, using the PDF/A-3 Attachments task in Workflow;
- create content items from these PDFs, by data mapping and skipping content creation;
- create digitally signed PDF's from these content items with OL Connect Output Creation.
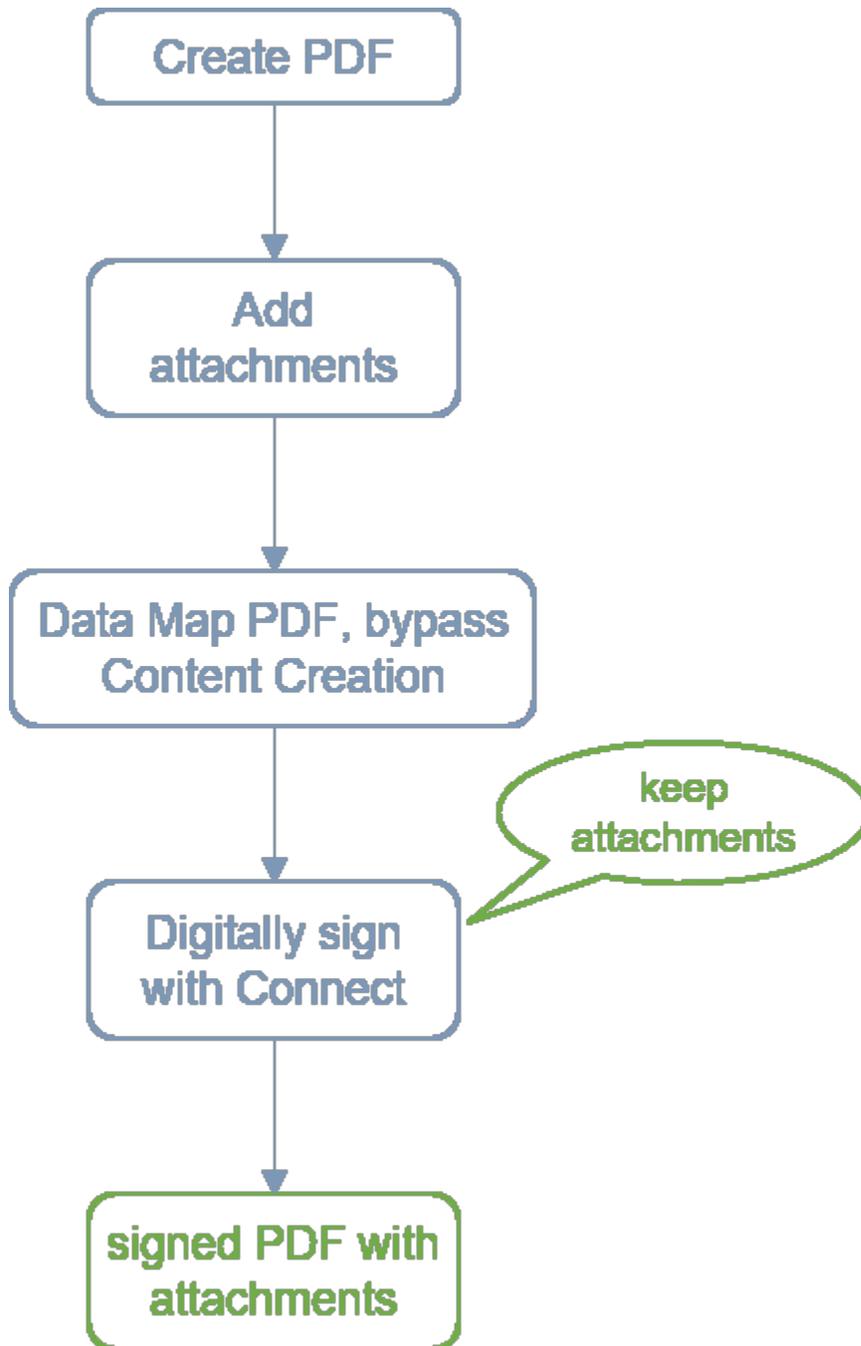
To do this efficiently, it's best to handle records and documents in groups where possible. Of the steps above, only the PDF/A-3 Attachments task has to work with one document at a time.

The third step requires scripting a REST API call to OL Connect.

Even though four main steps sounds simple enough to do in Workflow, making it all come together in the right way still requires thought, and takes time; in the end, it took me most of a day to create a good working example. To prevent others from having to spend the same time, we publish it here, and explain what it does.

PDFA-3 Attachments SampleDownload

In addition to showing how to create digitally signed PDFs with attachments, this sample also uses runtime parameters in Job Creation to control file locations more easily.

Running the sample

Download the zip file and unpack it. If you do this in `C:\OL Connect\Configurations\`, it should run without modifications. To run it from a different location, modifications are needed in the Workflow configuration and the output presets.

Open the Workflow configuration `pdfa3 attachments sample.OL-workflow` located in `C:\OL Connect\Configurations\PDFA-3 Attachments Sample\Workflow\`

Import all Connect resources located in `C:\OL Connect\Configurations\PDFA-3 Attachments Sample\Resources` in Workflow. Skip the `.pfx` file; it's the signing certificate and doesn't need to be imported.

Change the username and password to the credentials you have set in the Connect Server Configuration tool. Since 2020.2 no longer works with default passwords, this is something we will have to get used to. Change it in the Workflow Preferences and in the scripting step of the process (step 11).

You can now run the process in Debug mode. There's a single record data file in `C:\OL Connect\Configurations\PDFA-3 Attachments Sample\Data\sample-1rec.xml`.

To run automated, use Send Local to deploy the Workflow configuration, and make sure the service is started. You can now kick off a run by dropping either `sample-1rec.xml` or `sample-50rec.xml` from in `C:\OL Connect\Configurations\PDFA-3 Attachments Sample\Data` in `C:\OL Connect\Workspace\In\PDFA-3_attachments_signed`.

The output will appear in `C:\OL Connect\Workspace\Out\PDFA-3_attachments_signed`.

Changing the project location

To deploy and run from a different drive or folder, change the following:

- In `pdfa3 attachments sample.OL-workflow`, change the global variables `workspace_root` and `configurations_root`.
- In both output presets (`intermediate PDF.OL-outputpreset`, and `signed PDFA3 keep attachments and metadata.OL-outputpreset`), change the Job Output Folder.

- Change the path to the keystore file of the digital signature in the last output preset (`signed PDFA3 keep attachments and metadata.OL-outputpreset`.
- Don't forget to Import / Send to Workflow both presets and Send the Workflow configuration.

That's it.

Stepping through the process

To explain how it works, we will go over the most relevant steps of the process.

1. Folder Capture - this uses the process name for the folder to pick up files.
2. Set Variables - `run_id` is set to a unique string; this is used throughout the process to make sure that only the files and records related to this run are processed.

Steps 3 to 6 could also be a single All-in-One.

3. Data Mapping
4. Content Creation - this determines the looks of the output PDFs
5. Job Creation - this sets some metadata properties that are used by Output Creation
6. Output Creation - generate PDF's to an intermediate location, unique to this run; every document becomes a separate PDF
   - it doesn't really matter if this step produces regular PDF, or PDF/A-3. Step 16 is responsible for the final output format, and the attachments task (step 10), will work with either.
7. Branch - if performance is too low, it may be an option to put step 9-12 in a separate self-replicated process. The way that step 8 helps uploading the Data Mapper config only once would have to be changed in that case; this could be moved to a separate script in a startup process.
8. Set Variables - these are used in the script of step 11,
   - `data_mapper_config` is set here to not have to edit the script if that configuration needs to be changed
   - `data_mapper_id` is cleared here, and then only set once when iterating over the separate PDF's in step 9-12.
9. Folder capture - pick up an intermediate PDF from this run (note the usage of `run_id` for the intermediate folder.
10. Add PDF/A-3 attachments - adds the attachment. This is also set to add e-Invoicing metadata.
11. Run Script - run Data Mapper on the PDF and have Data Mapper turn the input PDF into a content set.
    - This will turn the input PDF (which now has an attachment) into a content set without modifying it, so both the attachment and the e-Invoicing metadata *stay intact*.
    - The script also sets `run_id` as a property on the content set, so it's possible to retrieve it later.

After step 9-12 have iterated over all intermediate PDF's, we have a content set (of a single content item) in Connect for every PDF, and each of these content sets has a property `run_id` to mark them as part of this run.

13. External Program - this removes the intermediate folder since these are no longer needed.
14. Retrieve Items - retrieves a list of all content sets with the property `run_id` set to the unique run id of this run.
15. Create Job - creates a single job from all content sets that were retrieved. As in step 5, this also sets some metadata properties for use in Output Creation.
16. Output Creation - create separate PDF/A-3 output files, while retaining the e-Invoicing metadata and attachments, and digitally sign them.

What is needed for production

A signing certificate

The sample uses a self-signed certificate to sign PDFs. When viewing the output in Adobe Reader, it will report problems with the signature.

Signature validity is UNKNOWN.
- The Document has not been modified since this signature was applied.
- The signer's identity is unknown because it has not been included in your list of trusted identities and none of its parent certificates are trusted identities.

Warnings like this are normal for self-signed certificates. A real application would need an official signing certificate from a trusted authority. If you need to learn more digital signatures, you can start here.

Create real attachments

The sample only shows how to add an attachment to a PDF, and get it signed. It uses a fixed XML file that is attached to all PDFs. In real OL Connect applications, it's likely that a different file has to be attached to each PDF.

Creation of these attachments is not covered by the example, because there are simply too many ways to do this.

Clean-up

The process is designed to leave no intermediate files. However, if there is an error between step 6 and step 13, the intermediate folder will stay behind.

Workflow and Connect Server on different systems

This sample was created while running Workflow and Connect Sever and its engines on the same system. If these are not on the same system, some more modifications will be needed. The easiest solution would be to deploy the project on a shared drive that is available on all servers (using the same drive letter everywhere).