



TAKE BACK CONTROL OF YOUR ALFRESCO CONTENT

Executive Summary

Organizations often choose to implement a collaboration and content management platform like Alfresco after losing the control of their enterprise content. This content and its multiple versions often end up being dispersed through the entire network of the organization, and its management and evolution depends on the goodwill of every single user. The implementation of Alfresco is not sufficient to solve this problem. If you don't enforce rules around the use of Alfresco, it will end up being a new type of "Content Trash Can", just as your network folders were.



This whitepaper will discuss the following concepts:

- Unified Capture : Identical content creation whatever the type and format of your input content is
- Intelligent Capture: Automatic metadata extraction, automatic creation of the document folders hierarchy, format conversion and business rules
- Traceability, security and conformity

Unified Capture

There are multiple ways to produce content. And for each of those distinct ways, there are multiple techniques to add the content to Alfresco. For example, you could scan paper documents from a network scanner, and then add them manually in Alfresco with their metadata. Or create a web service that will allow transferring electronic documents to Alfresco from an external system. In such a context, it's important to have a clear strategy in regards to how your content will be added to Alfresco, strategy that will enforce your business rules. Whatever the source of your content is, you should always make sure that your document types, your metadata, your formats and your folders hierarchy are defined globally and in a constant way.

Hence, it's important to choose a single point of entry to Alfresco. This entry point could be a simple Alfresco folder from where your metadata would be validated, your business rules would be applied and your content would be channeled. Another option would be to create a complete capture API (enforcing your business rules) that would be called from all different sources of content. Or, you could use a specialized capture tool, which will allow you to reach those different goals.

Your objective should be to try to capture your content right from the initial steps of your global process, as soon as it gets to your organization. If possible, we even recommend letting your

customers enter the metadata by themselves, using for example online forms (that will enforce your business and metadata rules). Keep in mind that your goal is always to produce content that will be well structured and easy to use within your internal processes.

If you adopt an integrated and unified approach for your content capture, it will be a lot easier to improve your processes and to push the content creation upstream in your process. As the content will be uniform whatever the source is, going from one source to another will be much easier and won't require heavy reengineering of your processes and of your Alfresco infrastructure.

Intelligent Capture

An electronic document with bad metadata is like a paper document in the wrong box or folder. In fact, it's probably a lost document. Assigning the right metadata becomes a key operation, which will allow you not to lose any document in addition to finding them as quickly as possible. If your data entry is done manually by each user, the door is open for human errors. In this context, it becomes necessary to automate as much as possible the data entry and validation operations, especially in a scenario where you have heavy loads of content.

To automate data entry, many techniques are available, like optical character recognition



(OCR), barcode recognition or extracting keywords from an electronic document. Additionally to those tools, document management strategies can be put in place to favor quick and uniform data extraction. A good example would be to add barcodes on your documents, and to add in this barcode the document type and the customer number.

Once the metadata is extracted, it's also recommended to do a counter-validation of its integrity. For example, if a customer number and an invoice number are extracted from a scanned paper invoice, it could be interesting to call a web service of your ERP or invoicing system and to confirm that those two reference numbers are a good match. With this type of technique, you raise the quality of your metadata and you catch potential errors at the beginning of your process.

With this high quality of metadata, it's then possible to automate some of your content processes. For example, if you always extract the customer number, its last and first names and the document type, you can create automatically a constant and uniform Alfresco folders hierarchy that will respect your naming policies. If not, you're at risk of having duplicates and folders and documents with wrong names. The following figures show examples of hierarchies with manual or automatic folders naming.

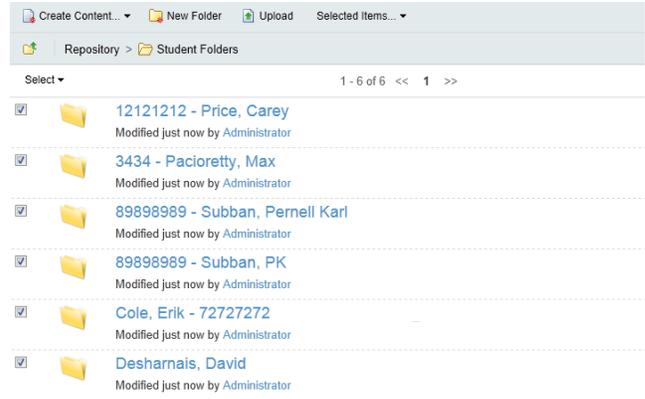


Figure 1- Manual/Irregular Folders Hierarchy

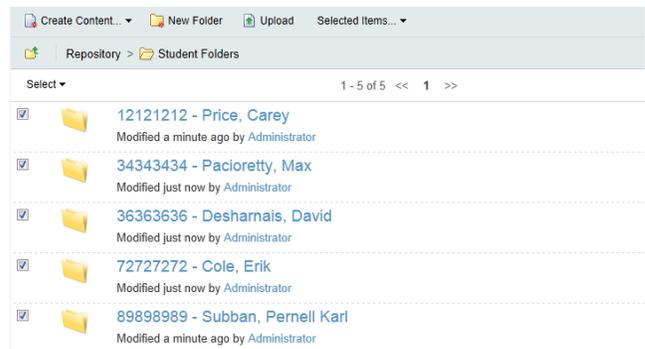


Figure 2 – Automatic/Uniform Folders Hierarchy

Also, such a strategy will clarify the debate “Folders against Metadata” that can exist regarding the best classification technique to use in Alfresco. By having well-structured folders AND good metadata, all different types of users will be happy and work with their favorite navigation or content search option.

Traceability, Security and Conformity

In addition to enforcing the content creation rules, you should also enforce the permission schema for the operations available to your users. For example, if a user is only allowed to create documents of types A, B and C, he shouldn't have types D and E available when uploading a document. This way, the system will be simplified for your users, and there won't be as many documents with a potentially wrong content type.

It's also important to have good traceability on the operations that were executed in the system. To achieve this, you should make sure to have clear reports and a good audit trail showing who did what and when in the system, and giving to complete lifecycle of your content.

Finally, you shouldn't neglect to respect your legal obligations. For example, you should make sure that your documents are kept in the PDF/A format when reaching their long-term archival state.

Conclusion

Alfresco allows you to get back the control of your enterprise content. But the system is not sufficient, and without the will to enforce business rules around your content creation processes, you at risk of living the same frustrations that you were experiencing before implementing Alfresco.

Avoiding this risk is possible by unifying your capture operations and by simplifying the metadata entry. By following those principles, you will keep a clean, traceable and secured Alfresco system.

About Corium

Corium is the publisher of Librex, a capture tool that allows you to capture your content in a simple, structured and unified way. It offers different capture methods like paper scanning, fax transfer, email capture, virtual printing and electronic documents automatic importation. Also, its metadata extraction tools like optical character recognition, barcode extraction and external systems integration allow you to automate your capture processes. Librex offers an intelligent connector to Alfresco.



For more information, visit us at www.coriumsoft.com



WWW.CORIUMSOFT.COM