How to design your enterprise file conversion platform.

A checklist, full of things to consider when selecting an enterprise wide transformation platform.

Introduction

For 40 years now, we have been creating content digitally. Documents, spreadsheets, databases, presentations, videos, audio files… We created 2.5 quintillion data bytes every day in 2018. That number is only increasing exponentially and expected to reach 175 zettabytes of data in the global datasphere by 2025.

Much of this content will need some form of format conversion at some stage. This can be for various purposes: compliance, standardization, archiving, communication and many more. It can also simply be because the existing file format is no longer supported.
As an example, PDF has become a de facto standard for many outgoing documents, while PDF/A is widely used for archiving.

But what should you look for when selecting an enterprise conversion platform? What are some key requirements you should consider for all your enterprise transformation needs?

We put together a **checklist of requirements** that can serve as a guideline when selecting the right conversion platform for your needs. The list is by no means exhaustive. Different industries and different organizations will have their own unique perspective.

1. User interface: How do you configure the conversion software?

Mass file format conversion ideally happens in the background, as an 'invisible process'.

Therefore, the idea is to set it up once - and let it perform the conversion in the background.

Below are a few critical characteristics a true enterprise conversion software should have:

- The software has a simple to use web-based interface, so it can be accessed from anywhere. No thick client or plug-ins required.

- The incoming file formats are automatically recognized by the system, no need to tell the system what files to expect.

- Workflows can be created to tell the system what to do (options for the conversion) via a simple drag and drop workflow interface.

- Maximum reusability. Workflows and their configurations can be reused to facilitate the inclusion of new use cases. Conversion workflows can be exported & imported for future use, exchanged between systems, and facilitate troubleshooting with the vendor.

- Depending on their importance or time sensitivity, conversions can be prioritized accordingly.

- Logic can be defined in the workflow to take specific routes- with different conversion settings- depending on different criteria (e.g., the type of file, the value of specific metadata, the template the document is using, etc.)

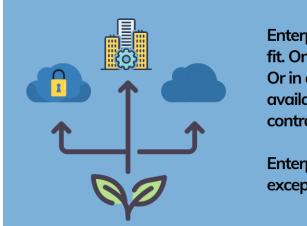## 2. Connectivity: How can you connect your enterprise systems?

File format conversion is used by many user groups throughout the enterprise: sales and marketing, human resources, operations, delivery, support etc.

Company-wide file format conversion is a centralized function that reduces the need for desktop conversion tools:

- The software can connect to most commonly used enterprise content management systems via a native integration.

- Metadata can be pulled from these systems and combined with the content in the final conversion (e.g., to include the author's name in the header of the document).

- The software supports RESTful API to read and write files to/from other content systems that do not have a native integration with the conversion platform.

- The software supports connections to email systems via POP3, IMAP, Exchange, or Office 365 to allow the conversion of emails and attachments as well as to allow automated notifications of system activities (job failures, etc.).

- It is possible to connect to standard network shares.

3. (Cloud) Deployment: What deployment options are available?

Enterprises are deploying software as they see fit. On premise or in the cloud.
Or in a hybrid cloud. Depending on network availability, geographic location or legal contraints.

Enterprise-wide document conversion is no exception.

- The conversion platform can be deployed either locally in your data center, or in the cloud: Microsoft Azure, AWS, Google Cloud etc.

- Both physical and virtualized servers (VMWare, Hyper-V) are supported.

- Can be installed on either Windows or Linux.

- Container technology support for massive scalability: Docker & Kubernetes deployment.

4. Security: What security options are available?

Security is a main concern for companies all around the globe.

When you are converting documents that contain sensitive data, you want your conversion software to always remain secure.

Hereunder are a few critical characteristics:

- Access to the system must be restricted via controlled usernames & passwords.
- Support for encrypted SSL or TSL?
- Are there any files stored in the system? If so, are these files encrypted?
- Can all passwords that might be used in the system be encrypted?

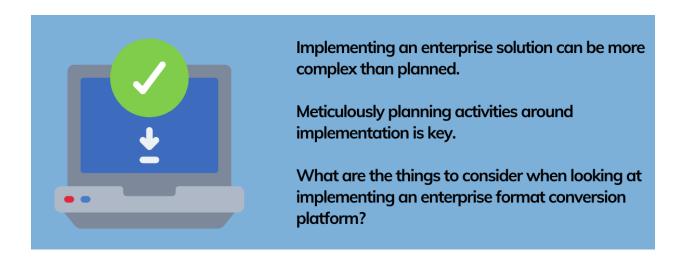## 5. Scalability: How scalable is the platform?

As different divisions use the system, potentially in different countries, the conversion process needs to be able to cope with peak loads. Periodic peaks (end of month, launch of new products, ...) can place the conversion software under stress.

What scalability criteria should you keep in mind when designing your enterprise conversion architecture?

- The software can automatically scale up and down based on your needs.
- The software can be deployed on Docker & Kubernetes to allow maximum scalability.
- Horizontal scaling is possible during peak loads instead of vertical scaling (which demands more resources).
- The software can deal with peak moment if your document conversion stream is not steady (e.g., more documents during the week than the weekend).
- The conversion performance should be high enough to minimize the need for scaling across multiple servers, and reduce the costs associated when running multiple instances (whether in the cloud or on premise).
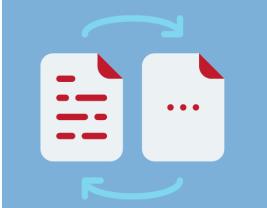
6. Implementation: What about the implementation?

Implementing an enterprise solution can be more complex than planned.

Meticulously planning activities around implementation is key.

What are the things to consider when looking at implementing an enterprise format conversion platform?

- Are there sufficient skills in the market to do the implementation?
- The software can be implemented easily without needing any additional external resources.
- The pre-requisites for the software do not require any additional specialists to be involved.
- Companies specialized in implementation and training are available for the software.
- The software implementation can easily be done remotely.
- The software can be implemented within 1 to 6 months; depending on the complexity of the project.
- Full validation services are available either from the vendor or one of their partners.

7. Conversion options: What can an enterprise file conversion platform do?
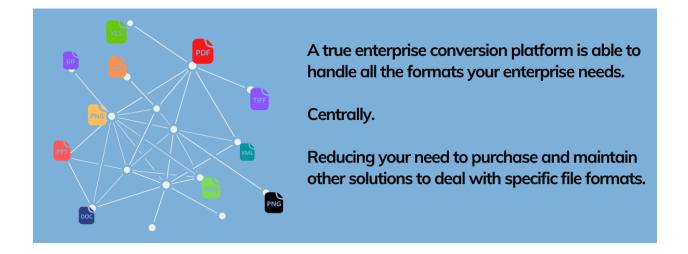

The focus of the enterprise conversion software should be on automating the file format conversion process while reducing the costs associated with manual file format conversion.
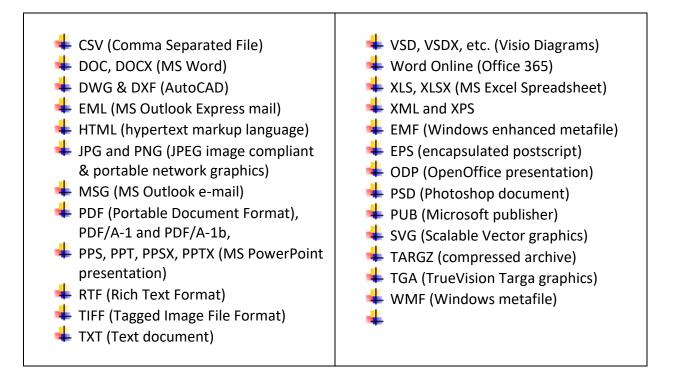
- The software doesn't need Microsoft Office, Adobe technology or other authoring applications to run and perform conversions.

- The software returns a clear message (and detailed logs) for any conversions that do not complete successfully.

- The software has a queueing mechanism that jobs are added to and documents are then processed sequentially from, based on date and priority.

- If the input document contains any macro & executable code, the software is able to ignore or choose not to execute it.

- If the input document is password protected, the conversion is rejected, and an error message is shown.

- The software is able to process different types of files in specific ways via automated routing in the workflow.

- It is possible to generate multiple output files, of different formats or with different configurations, that can be saved in one or more supported storage location.

8. Supported formats: What file formats are supported?

A true enterprise conversion platform is able to handle all the formats your enterprise needs.

Centrally.

Reducing your need to purchase and maintain other solutions to deal with specific file formats.

The conversion software is capable of handling at least the following input formats:

- CSV (Comma Separated File)
- DOC, DOCX (MS Word)
- DWG & DXF (AutoCAD)
- EML (MS Outlook Express mail)
- HTML (hypertext markup language)
- JPG and PNG (JPEG image compliant & portable network graphics)
- MSG (MS Outlook e-mail)
- PDF (Portable Document Format), PDF/A-1 and PDF/A-1b,
- PPS, PPT, PPSX, PPTX (MS PowerPoint presentation)
- RTF (Rich Text Format)
- TIFF (Tagged Image File Format)
- TXT (Text document)

- VSD, VSDX, etc. (Visio Diagrams)
- Word Online (Office 365)
- XLS, XLSX (MS Excel Spreadsheet)
- XML and XPS
- EMF (Windows enhanced metafile)
- EPS (encapsulated postscript)
- ODP (OpenOffice presentation)
- PSD (Photoshop document)
- PUB (Microsoft publisher)
- SVG (Scalable Vector graphics)
- TARGZ (compressed archive)
- TGA (TrueVision Targa graphics)
- WMF (Windows metafile)

9. Document & output validation: Does the output meet specific requirements?

Checking if the output meets the detailed requirements is typically performed manually today. Does the output meet the necessary technical and formatting needs?

An enterprise conversion platform should be able to automate much of this validation process, further reducing risk-inducing manual steps and the need for additional tools.

- The software has built-in validation functionality and doesn't require any external applications to achieve this.
- The software is able to fix the output by re-processing the file if any of the initial validation checks do not pass.
- The software is able to validate the output against a wide range of industry guidelines & other internal rules (e.g. correct PDF version, optimization set, fonts embedded correctly, etc.)
- The software is able to identify issues in the original source document and automatically fix them, or generate a report that can be used to correct the issues manually later.

10. Regulatory compliance: Can the software generate output that complies with the standards set by your industry regulators?

Depending on the industry you are in, the regulator mandates how information should be provided to them.

Make sure you check that the software you are selecting can output in the specified format & specifications set by your regulator.

A few examples below.

- Can your documents be converted to EMA, PMDA, and FDA required PDF formats (each slightly different but all based on ICH guidelines)?
- Can your documents be converted to archive-ready PDFA or TIFF formats that meet the compliance requirements of your archive system?
- Can you fully automatically check that the documents meet these requirements before they are shared?
- If not, can the tool show a clear report highlighting the issues, and/or fix them automatically?

11.Reporting: Clear management dashboards?

You want to know what is going on with your systems. Therefore, dashboard and reporting capabilities are key for an enterprise conversion platform. It should be able to clearly report how many transformations are made, how many have failed and much more.

- The software is able to return a detailed report on successful and failed conversions.

- The software is able to return a detailed report on average file size for conversions.

- The software is able to show how many documents were processed with each workflow.

- The software is able to show how many documents were processed for each department.

- The software is able to generate high-level executive dashboard views of statistics.