



THE XPRIORI REPORT

Xpiori Offers the Keys to Successful Document Migration and Classification

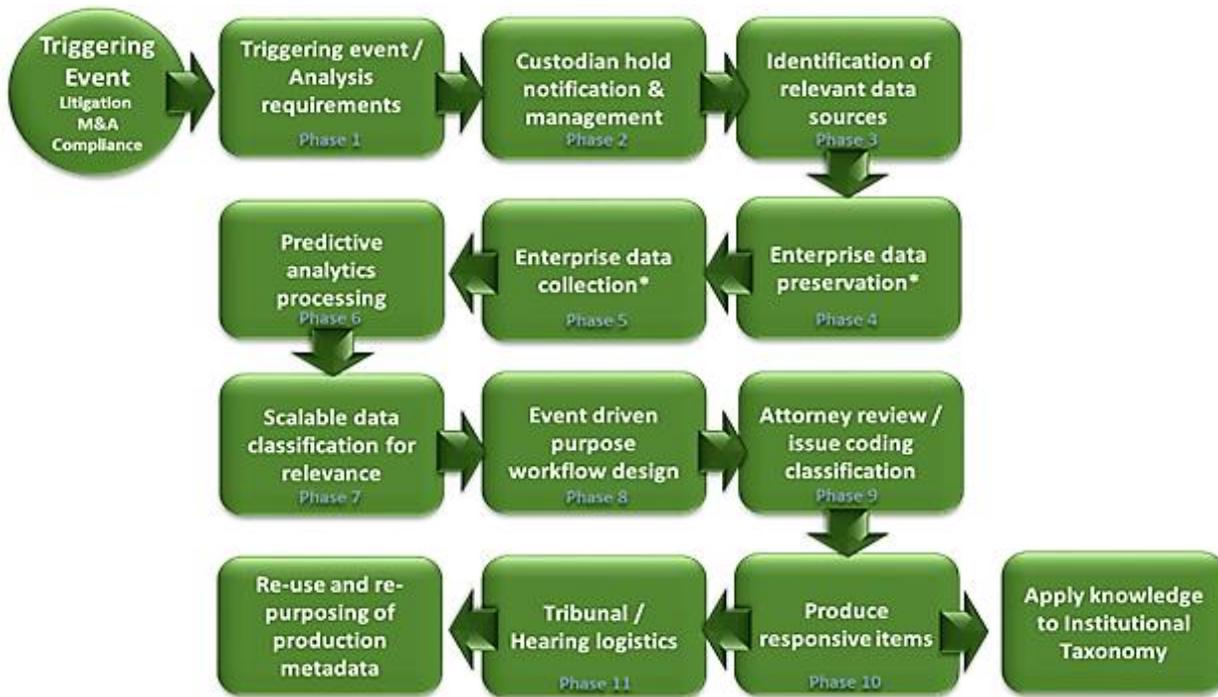
I. What is “Document Migration and Classification”?

With significant amounts of information being created on an ad hoc basis, it is not unusual for organizations to find that they have it spread to disparate places. Gathering it and organizing it into a rational storage program or even for current project use is a challenge. If you gather it, you are migrating it. If you are organizing it, you are classifying it.

II. Where do you start?

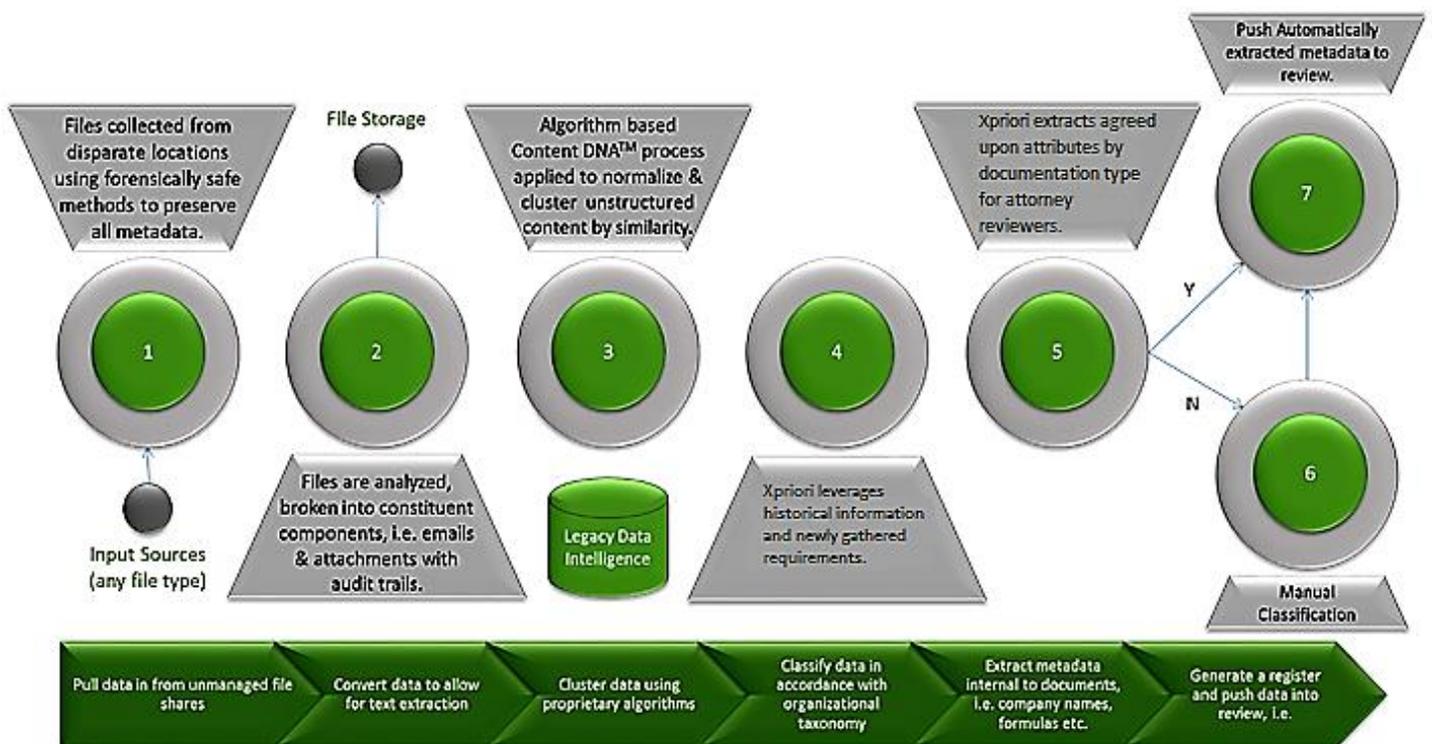
You start by gaining an understanding of what you have and what you need. This requires a GAP analysis, in the graphic below, you will see the methodology that we deploy and ... IT IS EFFECTIVE!

People always need information in response to a “triggering event”, and at the end of the day you need to organize to such events. The taxonomy that you deploy – the organizational structure that you should deploy – should reflect these needs.



III. What should your “Document Migration and Classification” Work Flow look like?

We deploy to the workflow described in the graphic below and have found it easily understood by IT, Legal and various Stakeholders as a data classification method. It is a baseline representation of the methods that should be used.



IV. What processes should be reflected in the Document Migration and Classification workflow and the Technologies that you deploy?

1. To pull data from the various input sources, you will need to identify those sources, together with all stages of document creation and management that they reflect. Your goal is to develop a holistic solution that includes repositories and classification systems that reflect your business processes and needs along with special treatment for unusual or industry specific data types;
2. With the tremendous magnitude of data that your business is creating, you must use technology to speed up the ability of relevant stakeholders to discern the relevance of information and present that information in the context of legal, information security, regulatory or other requirements. To do this, you must be able to limit the amount of information that has to be actually read by a human being and deploy human judgment to larger document sets rather than individual documents.
3. To this end and at each stage of a workflow, the process should be supported by technologies that:
 - Identify and collect from all network addressable IP's of storage devices and locations and do so in a non-intrusive manner so as to not disrupt operations;
 - Find and help cull duplicates, including documents expressed in a variety of different files types, e.g. a document that appears as html, pdf and word etc.
 - Automate the parsing and classifying of all unstructured information and present clusters of similar information for human review (there are several technologies that we apply depending upon their respective fits to the circumstances);
 - Enable the clustering documents and information on an iterative basis so that it can be adjusted to reflect various percentages of similarity and degrees of association to tune the process for optimal results;
 - Enable a clustering process that starts with the content and/or structure of the information and not its file type or metadata -- actual content should inform about what is important form the basis for placement of the information within the taxonomy
 - Enable algorithms that provide a high degree of self- organization, given the massive amounts of information to be reviewed;
 - Collect and generate all possible metadata associated with each document or file;
 - Enable attribution that will reflect the classification that has been made as well as any other requirements;
 - Deal with all of the information presented – textual and non-textual – regardless of format or amount and not just the results of keyword searches and other guesses about the content that might miss something; without this approach, you are unlikely to effectively contain risk;
 - Scale to the magnitude of the task;

- Apply machine learning so that the code associated with developing various clusters can be used to deal with newly presented information efficiently – it will adhere to a cluster or present itself as an outlier;
- Enable this replication process as a force multiplier to speed consideration of legacy information in the Document Migration and Classification Workflow, and later as a part of a solution that includes managed file shares to largely automate classification of information from operations on an ongoing basis;
- Automate the delivery of information once classified to the appropriate place within the storage taxonomy or for use in a project.

V. Xpiori deploys the technology and tools with all of the above characteristics and with a team that accomplishes the project in a forensically sound and fully auditable manner – on time and on budget.

Our processes always include best-of-breed:

1. **Content profiling** – we can deal with hundreds of millions of objects (files and folders containing, unstructured, semi-structured and structured information) to generate statistics and extract metadata. This high level activity provides stakeholders with the ability to make informed decisions related to assessing the environment and to prioritize activities with respect to discrete information silos;
2. **Content collection** – this process may involve the actual moving of content from one storage location or environment to another using highly secure transfer protocols and encryption. It may also, depending on the requirements of the initiative, involve merely changing the access rights to content to prevent alteration which satisfies “legally defensible collection in place” requirements. We deploy tools that automate the copying and loading of information from any source on the network without disrupting current operations. Point, click and load. The logistical problems associated with personal devices and paper documents remain, but those are a question of process management. Scanning can be done directly to the process;
3. **Processing** – this term refers to de-duplication, near de-duplication, clustering (grouping content based on some threshold of visual and textual similarity), generation of indexable text (OCR) and “rendering” of content to normalized formats, i.e. conversion of engineering drawings to PDF A renditions, for archival along with the original files.
4. **Classification** – this process involves the application of a RIM or other taxonomy, as augmented by us, to clustered content;
5. **Attribution** – different document types have different attribution requirements. Using regular expression, literal and fuzzy searching techniques, we can populate unlimited metadata fields for specific document types should those documents contain the sought after or desired attribute;
6. **Metadata and content loading** – once content has been sufficiently classified and attributed, we can automatically generate the XML or other load file data for ECM Archive ingestion in accordance with whatever specification the client system requires;
7. **Value extraction** from the face of documents to facilitate analytics; and

8. **Carry forward** of code and methodology to an automated process for classification of information as it is created by or received as part of day to day operations.

We have very substantial references and use cases that we can share with you. Give me a call to discuss your needs. We can deploy in North America, Asia and Europe with experienced teams.

For more information about Xpiori and our approach to IG, Compliance and eDiscovery: call me at 719-210-5318 (cell) 719-425-9840 Ext 2 (office); email

tdix@xpiori.com

Our website has significant information for you as well:

www.xpiori.com

Best regards,

Tim Dix, CEO

Xpiori, LLC

