

FAQ on PowerHouse AutoClassification

How does AutoClassification work?

AutoClassification uses pattern-recognition technology to understand the type of documents and files in question, and then pulls out salient data elements from each file to create client-custom metadata. The three step process goes like this: Data Mine (Index), Disposition, Display. Data mining extracts out all the DNA of each file, Disposition uses a custom-configured Rules engine to decide what to do with (how to treat) each file, and Display determines where and how the file will live henceforth.

Does PowerHouse crawl a corporate network or file share to find files to AutoClassify? How does it locate files to process?

PowerHouse either ingests data directly or processes files in place. Generally, instead of crawling, it is told where to look – given access to file shares, drives, servers, etc. and polls those areas on a regular or scheduled basis. Once a file is detected, PowerHouse uses the native/extracted/OCR text to make determinations about the content, such as whether PII is present, whether there is duplicative content (or whole files via hash matching). A metadata record is then created for each file that “tags” the information with a rich set of metadata and dispositions. Next either PowerHouse can move/migrate data by itself, or leave that step for authorized admins.

How are inactive or duplicate files detected?

Identical dupes are detected via hash value comparison. Near duplicates are detected via a comparison of text, as well as created/extracted metadata. File inactivity is detected by calculating last modified date against the current (or last crawl/index) date.

How is file metadata or other information presented?

Results are presented as fielded data in a metadata database format, and/or within Valora’s BlackCat product, which is a data visualization front-end that expressly presents such data.

What actions can be taken by authorized admins? (for example, delete, move, apply policies)

All actions can be undertaken automatically by PowerHouse in bulk, at the time of each file’s analysis or scheduled. Or these actions can be taken directly by admins. One clever use case is for PowerHouse to identify which actions to take for each file, and label each file accordingly, but not actually execute the action, so that the administrator can authorize the action in bulk.

How does PowerHouse extract or enrich metadata from files?

PowerHouse is exceptionally strong at creating/enhancing metadata. We support upwards of 600 potential fields, with specific fields for specific document types. Any existing metadata is extracting using a metadata extraction process similar to ESI processing for eDiscovery. This process covers off the file-level metadata, such as: last modified date, file name, path, custodian, etc. Next all textual content is evaluated for clues and indicators of additional metadata, for fields such as: Brand Names, geographic locations, parties present, government entities, PII, etc. Finally some of the Rules dispositioning results in additional metadata tags, such as retention periods, end-of-life handling, and Legal Hold status.

How do authorized users add additional metadata to files?

Authorized users may add additional metadata or other information on a single file basis, across file subset groups or as Rules to apply on a broad scale. In general, the more metadata available, the richer the search and data presentation available. Furthermore, the easier it is to automate disposition steps such as determining content security level or retention period.

Does PowerHouse support AutoClassification as part of file ingestion or migration into an ECM, DMS or other repository?

Yes, this is what Valora and PowerHouse are all about. AutoClassification is supported either as part of ingestion/migration or while processing in place (no ingestion). PowerHouse supports myriad fields of automated metadata creation. Common examples include: Document Type, Subtype (hierarchical nesting is supported), Authors, Recipients, Copyees, Date created/executed/revised/stored (or any combo), Employee Names, PII (many forms), Recommended File Name, Recommended File Location (folder), Recommended File Security, etc. Those fields with “recommended” in them are derived by the Rules Engine, that works in tandem with the AutoClassification engine.

How are security and/or retention policies applied to files?

PowerHouse supports written policies for information handling by first automatically determining any/all metadata that will go into a retention or security decision. Next, a weighted confidence rules engine determines the best statistical match for retention/security/placement or location/duplication/etc. If the client already has a retention policy, that will become the basis for “data-driven rules” inside PowerHouse. These are machine-language self-derived rules, rather than manually configured.

What packaged integrations/connections exist for other enterprise applications (For example, eDiscovery tools, archiving platforms, search engines, other)? Describe APIs available for partners or customers to develop own integrations

Although PowerHouse has a lot of built-in eDiscovery functionality, we also support integrations with other applications such as Relativity, LAW, and Nux for eDiscovery; OpenText, Alfresco, and SharePoint/O365 for corporate file storage; and, iManage and NetDocs for law firm use cases. In addition, PowerHouse supports back-end integrations to pull assessment data (files to be AutoClassified) and back-end “guidance data” (information that informs or generates rules in the Rules Engine). The list of specific connector integrations changes almost daily, please contact us for up-to-date information.

Does PowerHouse have its own API available for partners or customers to develop their own integrations?

Valora expects to publish our own RESTful API for PowerHouse in 2018. For now we support direct querying of the underlying SQL database.

How is PowerHouse used by customers during M&A or divestiture and other corporate reorganization efforts to classify and move content to appropriate owners/stewards?

This is emerging as a powerful, new use case for PowerHouse. Valora’s expectation is that we would “point” PowerHouse at the files to be AutoClassified and then let it do its thing (of course some prior configuration would already have been performed). PowerHouse will automatically join or separate content based on the current configuration.

Does PowerHouse move files to content repository services (on-premises or cloud)? Which repositories are supported?

Yes, PowerHouse supports this type of bulk movement/migration of files. When desired, it performs the movement/migrations directly. In other scenarios, files are just tagged for movement/migration and the client’s administrator performs this task. Many repositories are supported. See the answers above, plus several cloud options within Azure, AWS and third-party cloud solutions.

How is PowerHouse deployed? What options exist for on premises vs. SaaS or hosting models?

PowerHouse can be deployed on premise, within a private cloud, or as a SaaS option with Valora, or one of our strategic partners, managing the deployment and maintenance directly.

What are the top three differentiators of PowerHouse?

1. True AutoClassification – machine-driven metadata creation for hundred of fields of attribute information
2. Rich Rules Engine for many types of file/content disposition ranging from retention period tagging to security tagging to Legal Hold to recommended file names and storage locations
3. Valora is also a services provider of similar services and has been for many years. We understand the crazy and difficult situations our clients face with their real-life data stores. Chances are we've seen it, too, and so our software is both heavily battle-tested, and yet highly configurable, specifically to accommodate the vagaries that every scenario brings.

Which languages is PowerHouse able to analyze?

PowerHouse supports native analysis in English and English-like languages. This includes the romance languages (French, Spanish, German, etc.), the Scandinavian languages (Swedish, Dutch, etc.), as well as other English-esque languages such as Cyrillic (Russian) and Greek. For other languages we offer native AutoTranslation to English. AutoTranslation is built into PowerHouse and supports 100 languages as of June, 2017.

What type(s) of PowerHouse support are offered for deployment and troubleshooting? Is this support included with purchase or is it available for an additional cost? Are professional services required for deployment?

Currently, Valora provides all of its own support during extended business hours of 8am – 8pm, weekdays. We are available, and typically very active in any client installation/deployment. All support, trouble-shooting and training is included with the license fee. Professional services fees (beyond software license fees) are required for deployment.

What are some typical challenges or concerns that you encounter from your customers? How do you help to address these challenges/concerns?

Although Valora has been using our software for over 15 years internally, we have only recently started licensing it to others for use in their environments. Because of this, a challenge we face is having few reference accounts to point to, other than ourselves. We are addressing this chicken-and-egg problem by encouraging clients to begin with smaller pilot projects, and easing into full roll-outs. This keeps costs down and expectations in check. We are also extending our Beta Program through the end of 2017. This program provides early access and discounted pricing to the technology, in exchange for feedback and some leniency in the performance or the relationship.

How is your solution priced? Describe the most common models (i.e., per user, by volume, annual subscription, etc.)

Our solution is typically priced in three stages: pilot/configuration (which is a one-time professional services charge), backfile (which is a flat, capped, all-in fee to process whatever backfile exists), and ongoing license (which is an annual fee for ongoing, perpetual processing/re-processing). We do not charge per user or per volume fees (by file count, GB/TB, etc.). Valora also supports outsourced services provision on the same PowerHouse and BlackCat software.

How many customers does Valora have? What are the top three vertical markets? What are the top three geographic marks?

Valora has several hundred services customers over the course of many years. We are just beginning to have product customers and currently count 8 in that category. Our primary vertical is legal, whether RIM/IG or eDiscovery. We support large corporate legal and records departments, as well as law firms, government agencies and consultancies. Our top geographic markets are US, EU/UK and Australia, in that order.

How would you describe your key enterprise buyer? What role/position are they typically in?

The typical role is in some way responsible for Information Governance at the organization. This could be the CIO, CIGO, Director IG/Security/Compliance, sometimes Records & Info Mgmt if they have high enough profile.

Which vendors do you encounter as your main competitors?

We are still discovering our competitors on the product side. Early competitive analysis suggests StoredIQ, Nuix, Active Navigation and M-Files.