

SUBSCRIBE
to Data Informed

A New Model for File Cleanup

by [Kon Leong \(http://data-informed.com/?s=Kon+Leong\)](http://data-informed.com/?s=Kon+Leong) | June 23, 2017 5:30 am | 0 Comments

In *The Life-Changing Magic of Tidying Up*, Marie Kondo argues that rather than repeatedly tidying your house, a better strategy is to clean it only once, properly. In fact, by learning how to clean it right the first time and shifting your state of mind, you can adopt new behaviors that will enable you to keep it clean long-term.

Though cleaning enterprise file shares and repositories is somewhat more complicated, similar principles apply. It doesn't make sense to return to file cleanup every so often, indefinitely. It's simply not practical or cost-effective — however, this is the way far too many organizations look at file management. What these companies need is not another annual spring cleaning; they need a new strategy.

With that being said, here are some suggestions that may prove helpful for maintaining file systems in the long run.

Declutter

Though the idea of cleaning the entirety of enterprise file shares can be daunting, the currently available analytics capabilities make it much more attainable than many would expect. Experience has shown that 80% of insight from file analysis comes from metadata alone. When analyzed together, information such as creation date, author and file type can direct organizations to delete large volumes of redundant, obsolete and trivial data (ROT). This is a great first step for long-term management of enterprise files. Getting rid of unused files can also reduce unneeded liability presented by security breaches, insider threats, and ransomware attacks such as WannaCry. In this case, the old IT mentality may be appropriate: The less you have, the less you have to protect.

Consider Legal

In a recent [survey by Osterman Research \(http://www.insidecounsel.com/2017/04/12/challenged-by-the-old-and-new-corporate-e-discovery?slreturn=1495561849\)](http://www.insidecounsel.com/2017/04/12/challenged-by-the-old-and-new-corporate-e-discovery?slreturn=1495561849), only 39 percent of organizations were prepared to store, retain or produce word processing, spreadsheet or presentation files. With such documents so often becoming unofficial business "records," these employee-created files can be a major blind spot for legal departments that don't have access to them. Organizations that are able to classify, govern and ultimately discover these documents give counsel a major leg up in early case assessment and eDiscovery.

My advice is to therefore give special consideration to your in-house counsel when devising a management strategy for enterprise files. Organizations should have a process in place for identifying, classifying and governing business-related documents, which often lie scattered across enterprise systems, and ensure they are made searchable for legal. As a crucial component of the initiative, organizations may wish to implement automated policies for remediation to ensure important documents are accessible and to prevent clutter from accumulating over time. This goes hand in hand with having an ongoing governance strategy, which should be clearly defined before organizations begin a cleanup initiative.

Think About the Future

With the General Data Protection Regulation (GDPR) approaching, organizations that process EU resident data will be accountable for a new set of standards. Among these is an obligation to obtain explicit consent from subjects for the processing of personal data, to respond to subject requests for deletion, and an obligation to delete personal data once it is no longer being used for its original purposes. With fines of up to 20 million EUR, or 4% of annual worldwide turnover, GDPR is a top priority for 92 percent of US organizations, according to [PwC research \(http://www.pwc.com/us/en/press-releases/2017/pwc-gdpr-compliance-press-release.html\)](http://www.pwc.com/us/en/press-releases/2017/pwc-gdpr-compliance-press-release.html). However, organizations focusing on structured database systems that typically hold consumer information can forget to look in less obvious spots, such as file shares and SharePoint sites.



Kon Leong, President, CEO and Co-founder, ZL Technologies

Search

Featured Downloads

- **Upcoming Webinar on 7/25: [Advanced Analytics – Transform Your Data into Business Value \(http://event.on24.com/r.htm?e=1444877&s=1&k=0E2E0A6FCF738B3E0B1893E5856CC6DF&part](#)**
- **[Why Business Analytics in the Cloud? \(http://info.data-informed.com/tableau-whitepaper-why-business-analytics-in-the-cloud\)](http://info.data-informed.com/tableau-whitepaper-why-business-analytics-in-the-cloud)**
- **[Interactive Analytics, Visualizations and Data Modeling on Large Hadoop Data Sets \(http://info.data-informed.com/looker-for-hadoop-whitepaper\)](http://info.data-informed.com/looker-for-hadoop-whitepaper)**

Categories

- [Advanced Analytics \(http://data-informed.com/advanced-analytics/\)](http://data-informed.com/advanced-analytics/)
- [Big Data Glossary \(http://data-informed.com/glossary-of-big-data-terms/\)](http://data-informed.com/glossary-of-big-data-terms/)
- [Cloud Computing \(http://data-informed.com/big-data-cloud-computing/\)](http://data-informed.com/big-data-cloud-computing/)
- [Customer Analytics \(http://data-informed.com/customer-analytics/\)](http://data-informed.com/customer-analytics/)
- [Data Visualization \(http://data-informed.com/big-data-visualization/\)](http://data-informed.com/big-data-visualization/)
- [Internet of Things \(IoT\) \(http://data-](http://data-informed.com/iot/)

For these more elusive unstructured systems, in-place advanced sampling can illuminate areas that hold personal data and demand more attention. From there, indexing and content analysis allows organizations to perform enterprise-wide searches for personally identifiable information, health information, and credit card information. The ability to find such sensitive data on the fly will become increasingly valuable once GDPR hits in May of 2018, but organizations that want to be ready by then should get an early start.

Look at the Bigger Picture

Even now, many organizations still forgo file cleanup, an instrumental reason simply being the fear of deleting important documents. While understandable, this outlook has risks of its own. For instance, having to find important relevant documents in a sea of ROT in a timely manner during eDiscovery can be like trying to find a needle in a haystack. Neglecting management of enterprise files adds undue pressure to your legal team, and can significantly increase already-astronomical review costs.

Furthermore, when GDPR goes into effect, organizations will be obligated to delete personal data when it's no longer needed, as well as to find and take action on personal data upon request. There will be an expectation of best efforts, and an organizational policy to not delete files is unlikely to cut it.

Meeting these requirements doesn't happen overnight. Organizations may be best served by taking a holistic governance approach that accounts for both email and files, while implementing "privacy by design" (<http://www.zlti.com/blog/can-privacy-survive-the-era-of-analytics/>) at the architectural level. In this approach, organizations can begin to consolidate data silos and regain control of their data, while ensuring they have the framework to properly manage personal data.

In Marie Kondo's philosophy of cleaning, she says that learning how to tidy up properly can spark lifestyle changes that lead to a long-lasting, positive impact. It's these changes, not the initial cleaning, that keeps the house in order over time. Her reasoning has resonated with a lot of people, and for good reason. It may be due time for organizations to adopt a similar philosophy.

Kon Leong, President, CEO and Co-founder, ZL Technologies, is responsible for managing all aspects of the business, including strategy, finance, sales and marketing. Earlier, Kon was co-founder and president of GigaLabs, a vendor of high speed networking switches. Prior to that, Kon was First Vice President of Mergers and Acquisitions at Deutsche Bank. He was at the General Motors Treasurer's Office in New York City, where he managed GM's venture capital investments in high tech. He also spent eight years in various IT engineering and management positions at Burroughs, Philips and Union Bank. Kon earned an MBA with Distinction from the Wharton School and received an undergraduate degree in Computer Science from Concordia (Loyola) University, after completing a year at the Indian Institute of Technology.

Subscribe to Data Informed (<http://info.data-informed.com/subscribe>) for the latest information and news on big data and analytics for the enterprise.

(https://cta-service-cms2.hubspot.com/ctas/v2/public/cs/c/?cta_guid=35fecf6c-3f93-468b-aa0f-3defc82ce544&placement_guid=4892be79-794a-48ac-b88c-60079a665837&portal_id=215445&redirect_url=APefipFuArO3zQNe0LQS43aWWI6-GRoE_OizFIVL2UVAvwKU3SZA5Y_L-rwEwSoDXkuOHDdEpdapQ5Dwyp5zxpM0ZY56qeS7XX88G9ofVMAMjTB4ya4oMuhdmUSEzHkwb8UibVtG6hRnR3eXz9zKzZc6aQc6mhmhmx_taNGuXFPLN09nBGqVana8Lnr&hsutk=&canon=http%3A%2F%2Fdata-informed.com%2Fa-new-model-for-file-cleanup%2F&_hstc=248701080.e98c688ee9f321cacef9ee0596271054.1498510580153.1498510580153.1498510580153.1498510580153)



Tags: [data governance](http://data-informed.com/tag/data-governance/), [data management](http://data-informed.com/tag/data-management/), [data strategy](http://data-informed.com/tag/data-strategy/), [file management](http://data-informed.com/tag/data-strategy/), [General Data Protection Regulation \(GDPR\)](http://data-informed.com/tag/file-management/), [legal departments](http://data-informed.com/tag/general-data-protection-regulation-gdpr/), [metadata](http://data-informed.com/tag/legal-departments/)

- [informed.com/internet-of-things/](http://data-informed.com/internet-of-things/)
- [Management](http://data-informed.com/big-data-management/)
- [Operational Analytics](http://data-informed.com/operational-analytics/)
- [Skills](http://data-informed.com/big-data-skills/)
- [Technology](http://data-informed.com/analytics-technology/)
- [Use Cases](http://data-informed.com/use-cases/)

Join the Conversation



(https://cta-service-cms2.hubspot.com/ctas/v2/public/cs/c/?cta_guid=74807e22-84ef-4e77-baa9-1551b00e72d6&placement_guid=659de830-867b-439a-809e-872d132cc6ec&portal_id=215445&redirect_url=APefipFNiORJdfuHtkHz1nGt8Bh5M0qWl1k90A_YoHeBX2PG4wUl1pQPWskZcf_informed.com%2Fa-new-model-for-file-cleanup%2F&_hstc=248701080.e98c688ee9f321cacef9ee059)



(https://cta-service-cms2.hubspot.com/ctas/v2/public/cs/c/?cta_guid=e912b2ca-34ff-4226-829c-40a28d7bea07&placement_guid=ae05c418-e4a9-484a-898f-7ec0af3fdc05&portal_id=215445&redirect_url=APefipFY1ydmJl8s_8V0crNXMKlpePwXKJ2Hxudb0WAAdT_E_tDb4dQT7viEd468telvCC15-8ztlIn0y4x5J8iRNBm8bvF-SeIWWPK8-SKuLTn-OxdRrcMvjXz65Tx4G2c05WJMb8oP3QgGg&hsutk=&canon=http%2F%2Fdata-informed.com%2Fa-new-model-for-file-cleanup%2F&_hstc=248701080.e98c688ee9f321cacef9ee059)



(https://cta-service-cms2.hubspot.com/ctas/v2/public/cs/c/?cta_guid=cbc6249e-9d39-4ab8-951b-35cfaa5a04e6&placement_guid=8f45ef68-49df-4455-8a88-5222115317da&portal_id=215445&redirect_url=APefipF9GyAwfHb3kUJH9suEnCcat6tHjzYASBqgB76nqisYD0BjHwLcq8lCJGhC_lzna-xmCrESdB4uESX8RT35THtGnBY9xe_030SY9bVFXL1Tgaegik_informed.com%2Fa-new-model-for-file-cleanup%2F&_hstc=248701080.e98c688ee9f321cacef9ee059)



(https://cta-service-cms2.hubspot.com/ctas/v2/public/cs/c/?cta_guid=bbdf54b8-57e2-473a-8104-6a5a3656fb3d&placement_guid=728a0fcd-e7aa-498b-9d96-98f31426118c&portal_id=215445&redirect_url=APefipGDAENybGxuab222hisGAG8HEOR2sEJKY5BcJtiX1IBSx-9BT6mOT8sn4hiYRoBYKYVnoNEwYTyr4O_SagmlUF5A-OpC4I4Bml-gS3LscIKa4sx4Xs-aiBvXqJNxnSzN61M90JcbQKyDTstsp-G2ShaFNIDvRdRlKp3L9ctcsNgU&hsutk=&canon=http%3A%2F%2Fdata-informed.com%2Fa-new-model-for-file-cleanup%2F&_hstc=248701080.e98c688ee9f321cacef9ee059)



(https://cta-service-cms2.hubspot.com/ctas/v2/public/cs/c/?cta_guid=bbdf54b8-57e2-473a-8104-6a5a3656fb3d&placement_guid=728a0fcd-e7aa-498b-9d96-98f31426118c&portal_id=215445&redirect_url=APefipGDAENybGxuab222hisGAG8HEOR2sEJKY5BcJtiX1IBSx-9BT6mOT8sn4hiYRoBYKYVnoNEwYTyr4O_SagmlUF5A-OpC4I4Bml-gS3LscIKa4sx4Xs-aiBvXqJNxnSzN61M90JcbQKyDTstsp-G2ShaFNIDvRdRlKp3L9ctcsNgU&hsutk=&canon=http%3A%2F%2Fdata-informed.com%2Fa-new-model-for-file-cleanup%2F&_hstc=248701080.e98c688ee9f321cacef9ee059)