Avoiding Vendor Lock-in When Transitioning Email to the Cloud

BY: NICK CAVALANCIA

IT Pro™ WINDOWS®

SPONSORED BY

TransVault™

When migrating to cloud-based systems such as Microsoft Office 365, your first thought is, of course, end-users' 'live' mailboxes. But when making a complete move to the cloud, there also needs to be a plan about what to do with a huge repository of leaver's mailboxes, email journals and on-premises legacy archives, including third-party archives and the ubiquitous personal archives (PST files). After all, with the decision to move email services to the cloud, and with the right Office 365 subscription plan enabling

essentially unlimited storage, why would you have all these legacy email records remain on-premises, clocking up on-premises storage and software support costs, not to mention an eDiscovery nightmare?

The role of a true data custodian should encompass a plan that covers the entire lifecycle of the data, from creation all the way through to obsolescence or deletion, not just how to migrate it effectively to the environment that meets the current needs of the organization. In industries where regulatory compliance dictates email retention, policies can stretch as long as a decade (or indefinitely, as is often decided by risk-averse corporate governance), this means the custodian has to be thinking a very long way in advance about how data is going to be stored effectively and efficiently, regardless of the platforms used over time.

Some vendors of SaaS-based platforms charge a subscription for legacy data storage, often as a line item on an overall invoice that covers mailbox provisioning as well. In some instances, hidden in small print, these same vendors insert contractual cost obligations around data extractions (should the customer ever wish to stop their subscription and have their data returned to them). When you consider that terabytes of legacy data can amass very quickly in enterprises, and that vendors can charge thousands per terabyte to extract this data, the cost to businesses of migrating can be eye-watering. This is where we start to encounter the term "vendor lock-in"; where the total cost to migrate to new systems becomes so prohibitive that customers have little choice but to stay with the incumbent vendor.

This concept of vendor lock-in applies to both the vendor hosting your primary messaging, as well as the one hosting your email archive. In both cases, you need to be aware of just how difficult it could be to part ways.

Whether your organization is concerned about either compliance or legal requirements (or both), the issues listed below should be covered when planning your migration. The focus on both compliance and legal requirements is important during and after the move as well. Depending on corporate policy, applicable regulations, and legal concerns, the lifecycle of current and archived email may range from years to decades – a duration of time so material, that you should be considering how to avoid vendor lock-in now.

In this whitepaper, we'll address various facets of vendor lock-in when it comes to moving legacy email data, keeping in mind the data's entire lifecycle. These are:

- Problems caused by the format of data extractions
- Quality/reliability of extraction
- Speed of extractions, or the lack thereof
- Cost of data extraction

You'll notice that all of the concerns above revolve around issues related to leaving a vendor. While the cloud services you're selecting today may meet your needs and budget, remember that it's not unreasonable

Why Maintain an Archive?

Let's begin by talking about why your company maintains an archive in the first place. The archive isn't just a long-term backup copy of your organization's email; it's a permanent legal record of communication. And that communication can come under scrutiny – whether due to a question of adherence to various compliance standards, or should the company need to address a legal issue.

While some organizations may not initially consider themselves subject to compliance, regulations like the General Data Protection Regulation (GDPR) in the European Union, and various regulations in the US protecting personally identifiable information apply to any business that maintains customer data. Additionally, every business has personnel data (which can include bits of healthcare information), many take credit cards, etc. So, there are plenty of examples where organizations can be subject to compliance.

Many organizations also seek to reduce the risk of successful legal disputes by maintaining a copy of communications. Sexual harassment lawsuits, other HR issues, cases of data or IP theft, all may require the use of historical email as part of the legal process.

to envision a future where costs are ramped up, service levels are changed, etc., resulting in your organization needing to move on to another vendor. When you're signing the deal, though, it's all daisies and rainbows. So, it's important to be thinking about each of the considerations above up-front, before committing to an agreement with any vendor, as it is critical to the success of the organization's ongoing management of legacy data, from various perspectives including legal, governance, budgetary and IT.

Lock-In Issue #1: Problems Caused by the Format of Data Extractions

There are a number of potential issues that arise when you extract email from a given vendor.

Vendors generally only provide the formats they believe will be beneficial while you work with them – and not necessarily those that would be particularly helpful when you move onto the next email or archive platform.

- **It's likely per-mailbox** – most extraction is accomplished with the idea that IT is trying to export out a single mailbox. This means data is the mailbox, and not email system-centric. For example, if an email went out to 20 users in a Bcc, there is no detail in the exported message about the other users that were sent the email.
- **It's flat** – some vendors export email out to what essentially is a flat list of messages; folders or a message hierarchy of any kind isn't maintained, making the organization of messages completely unviable and giving a compromised end-user experience.
- **It's (possibly) in an unsupported format** – Google supports MBOX and Office 365 supports exporting to PST - but does the new vendor support those formats? If the "old" and "new" vendor's formats don't match, you're looking at further conversion of data that can corrupt, strip out metadata, etc.
- **Corruption is ignored** – email corrupted at the source will simply be extracted as such and posted to the export format – and you won't find out about this until after you've tried importing it into the new system.

Lock-In Issue #2: Quality/reliability of extraction

This is a tough one to determine, as most vendors simply provide a way to export the data – either by using a web interface, via some kind of eDiscovery, or using scripts to

scale well past a one-by-one extraction. But these methodologies provide no visibility into the specifics around data integrity, whether messages were skipped due to corruption, whether the extraction is comprehensive over just being completed, etc.

And then there's the issue of chain-of-custody.

One of the main reasons you should keep a long-term archive is to retain a copy for legal reasons; which means the data needs to be able to stand up in a court of law. Chain-of-custody dictates that it must be demonstrated that at no time during a migration have any messages been altered. The simple act of moving data from one format to another as part of an extraction potentially breaks the chain. And if the data is, in fact, preserved, the extraction formats do nothing to audit access, modification, or deletion, thereby breaking the chain-of-custody. If you maintain an archive for legal purposes, you need to map messages 1:1 from the old to the new vendor, with some way of ensuring messages aren't modified in any way in transit.

Lock-In Issue #3: Speed of extractions… or the lack thereof

Let's say a part of your organization gets sold and half your mailboxes and archives need to be extracted to be placed into another platform. You already know the "let's do this one at a time" method simply isn't going to cut it. Sure, some PowerShell scripts will help export data to PSTs in bulk (sort of in a one-by-one fashion), but the reality is when it comes time to part ways with a vendor, you want out as quickly as possible (without sacrificing integrity). Ultimately, if you're reliant on the archive vendor's own timeline to extract the data, then the timeline for your migration project may end up being compromised. This also goes for ingestion of data into the new archive vendor's platform. Shipping drives containing legacy data to vendors for them to ingest is a prime example here – that drive might sit on a shelf, in a queue for ingestion for a considerable period of time.

Lock-In Issue #4: Cost of data extraction

While vendors like Microsoft and Google make exporting data relatively easy (putting aside the previously mentioned issues around data quality, chain-of-custody, etc), there are 3rd party archiving solutions you may be considering in addition to either an on-premises or cloud-based email platform. Also, because these platforms are far more proprietary, getting out can be a rather expensive proposition – professional services, one-off custom coding, etc. can all create such a monetary headache that it is simply more cost-effective to stay.

Tips for Ensuring an "Exit Strategy"

With your eyes wide open around the potential pains that can come with trying to leave a cloud vendor, there are a number of tips to ensure the smoothest of exits, should the need arise to move onto another vendor.

Before you commit to a cloud email or archiving vendor:

- ✓ **Understand the supported extraction methods** – ask about the supported email formats, whether email is extracted on a per-mailbox basis, how are errors logged and addressed, and how are compliance and legal mandates addressed.
- ✓ **Understand the cost of extraction** – ask about whether extraction can be done internally by IT or whether professional services are required, and what the cost structure is to extract data (e.g. per-mailbox, per-GB, etc.).
- ✓ **Get a "pre-nup"** – it *is* possible to make being able to leave with your data at *no cost* part of the deal. *Ask for it* as part of the deal and get it in writing. The time to negotiate is *before you do the deal* - when everyone wants it to work out and no one thinks there will be a problem, so it's far easier to include a zero-cost exit.

Transitioning Without Lock-In

With so many vendors focused on getting customers into their messaging or archiving solutions, it's not uncommon for organizations to find their data locked into proprietary archives and storage systems, or trapped in proprietary and unusable formats, at a time when they want or need to move on.

The key to unlocking the future for your data is to recognize the possibility that there someday may be a need to leave the current vendor. Marry that with the assumption that it's going to be one messy divorce, and you'll quickly come to realize you need to plan today, ensuring there's a clear path out, should you need it.

If you're currently stuck in a "messaging marriage" you need to get out of, but can't due to the issues mentioned in this paper, consider looking at a third-party solution that specializes in migrating email and archives between platforms. In many cases, the concerns around cost, legal, and compliance can all be minimized, if not eliminated. ●