


Frequently Asked Questions

Why do we need to archive in the first place?



US Federal legal recordkeeping requirements dictate keeping HR + Payroll data and documents for anywhere from 1 year for basic documents and data to up to 30 years (for OSHA specific data). With multiple systems, countries and retention policies, long term data management becomes even more complex. [See recommended legal retention periods.](#)

Outside of the U.S. the EU and other countries are imposing strict data security & retention policies which also mean applying more proactive controls over documents and data globally.

Should we build something in-house?

Building in-house most often results in a loss of functionality in data access, security, and reporting options. Building in-house simply adds another system to your internally managed landscape where it could be offloaded to the cloud. Worse, doing nothing at all typically costs enterprises \$100,000-500,000+ if not in licenses alone, then in personnel costs and infrastructure.

Lastly, legacy system security protocols are outdated. In addition, organizations too often consider Excel, Access, and “flat files” in shared drives as potential options for archiving data which is highly insecure and not suitable for sensitive PII in HR and Payroll data. Google “hack [xls, pdf, mdb] password” to discover how many tools are readily available to access these files even when secured with password protection.

IT departments are often great with solutions, but what is often occurring in many organizations is that every system decommission results in a different and unique archiving strategy across HR and Payroll and every country. So organizations are still delivering a patchwork of solutions and methods for saving historical data often in scattered locations and unsecured formats that don't meet the global compliance standards of today.

Fuse is ready made with global data models and updated security standards and offers a place to consolidate and “fence off” sensitive HR and personal data. See the detailed solution brief for the full functionality Fuse provides to enable companies to deploy proactive data retirement strategies, benchmarking, reporting and analytics.

New legislation is being passed each year changing what records companies must keep and for how long and now even how they must secure it in very precise terms:

[Cybersecurity regulation impact on hr and payroll](#) ➡



Should we use the built-in native transactional system functionality for archiving?

The problem with this approach is setting yourself up for indefinite vendor lock-in and potentially never being able to sunset a legacy application. Instead, consider a system agnostic approach that frees you from both vendor lock-in on an expensive legacy infrastructure AND buffers against the loss of system specific knowledge and role requirements in-house.

Why wouldn't we use a generic archiving solution for all enterprise data (not a stand-alone HR and Payroll platform)?

This is a “data lake” strategy. Data lakes are highly insecure for personal data due to the complex security configuration and maintenance required to keep data segregated.

[Gartner Says Beware of the Data Lake Fallacy](#) ➡

We have deployed Fuse within a custom data lake for a customer and experienced these issues firsthand. HR data is not suited for a data lake strategy due to the global compliance implications of retaining personal data. Just one example of many; EU-based companies soon have to start providing employee access to their own data for disclosure due to recent Global Data Protection Regulations. Obviously giving employees access to an enterprise data lake is not an option. The result is HR then must handle all requests for data manually.

Isn't this too expensive for storing just static data?

The solution is cloud-architected on a modern, highly scalable technology which permits it to be used for companies starting around 1k employees up to hundreds of thousands while remaining cost effective at all levels between.

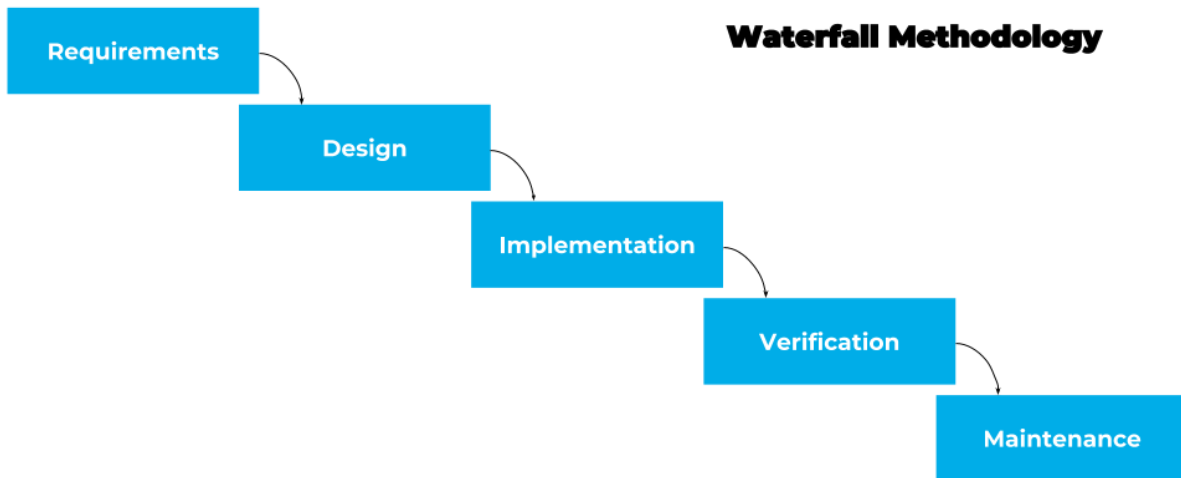
Shouldn't we look at archiving after the conversion is completed to our new system?

Speaking from experience, don't wait until HR and Payroll resource bandwidth is tied up in implementations, year-end activities, and other competing initiatives to handle data archiving and validation efforts. Most ASP, SaaS, and hosting contracts only allow 60-90 days maximum for data reclamation before systems are taken down and data access lost. Customers often get caught between the outgoing contract expiration and not having a solution.

Archiving tools and resources overlap heavily with new ground and cloud system conversion efforts and are best done while there is budget in place for conversions. The two efforts go hand-in-hand and archiving later means redoing much of the conversion logic already used in new system implementations.

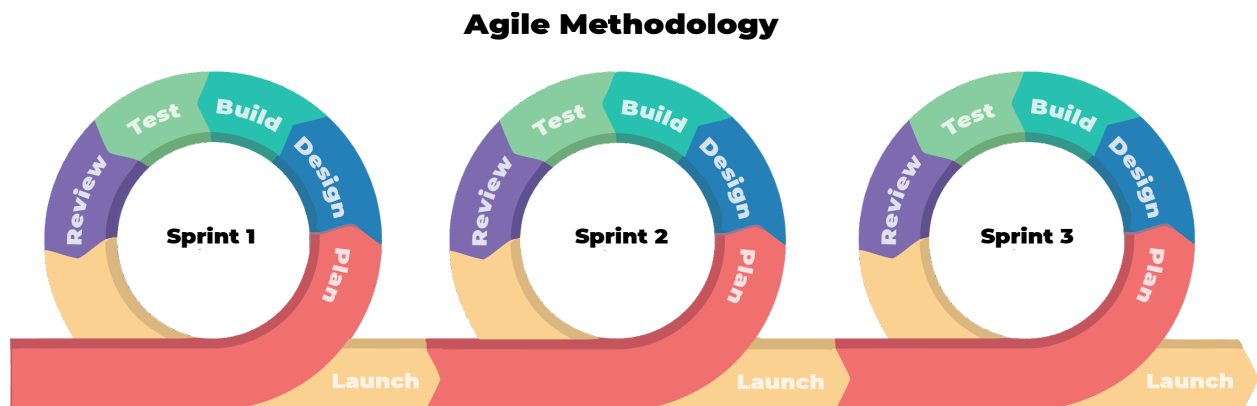
Most organizations are on a transformation journey from "ground" (on-premise) to cloud applications. There is generally not a cloud ROI realization to be seen while the historical landscapes continue to burden HR and IT departments. When legacy landscapes are kept up long term you are not "replacing" your systems, you are only adding to the landscape (and in turn, the costs).

Typical System Implementations are performed with vendors using a **Waterfall** methodology where design and sign off are done in full up front.



There are some drawbacks which make this unsuitable for Archiving efforts:

- Each phase must be fully completed before continuing to the next.
- Emphasizes “Big Design up front”, not iterative re-work each phase.
- Requirements may not be met completely.
- Resources must be fully engaged on average for a fixed amount of time.



Archiving and Enterprise Landscape Decommissioning needs to be run with an **Agile** Approach. This ensures data is complete and accurate in its end state and the goal of decommissioning systems can be fully achieved to realize ROI. Specifically, it allows the business longer timelines to validate data in parallel before the legacy systems are taken down.

- Iterative or Cyclic, allows revision of requirements during configuration and/or development *cycles*.
- ensures conformity with actual user requirements since data is re-worked over and over to clean and format
- Usually a downside is *Scope Creep* with this approach. However, with Fuse Paas especially, this does not cause major cost impacts for archiving additional data or documents.

Waterfall projects take fixed timeframes (calendar days) and 100% dedicated resources (effort days)

Agile projects for archiving may take longer to accomplish (over more calendar days), but fewer effort days to accomplish.

Planning ahead with plenty of buffer for decommissioning and archiving projects is required.

How much data do we actually need to retain?

See the article [Compliance retention periods](#) ➡ for an outline of the typical US only scope involved. In addition to the datasets described, employers are often required to archive the supporting documents as well. Many of these are required for more than 6-7 years.

Additionally, data generated by Payroll and Time systems can be millions of rows even for smaller companies which is not suitable for databases like MS Access.

For example, (in the U.S.) if an organization has just 1,000 active employees with 5 years of history on a biweekly pay schedule, this amounts to roughly 4 million rows of paycheck detail data:

Number of annual pay periods: 26

Number of payroll runs over 5 years: $26 \times 5 = 130$

Number of unique employee paychecks for 1,000 employees = $130 * 1000 = 130,000$ paychecks or deposits.

Average 30 line items per employee pay check: $130,000 * 30 = \mathbf{3,900,000}$ rows

Time data being daily for hourly workers can be similar in terms of number of rows of history:

Hourly/weekly employees = 1000

Annual working days = ~260

Average Clocking times per day = 3

$1000 * 260 * 3 = 780,000 * 5$ (years) = **~3,900,000**

Scaling that for employee populations and pay frequencies the rows of data required scale exponentially:

Rows of Payroll Data for 5 years history				
Active Employees:	10,000	25,000	50,000	100,000
Semi-Monthly Paychecks	36,000,000	90,000,000	180,000,000	360,000,000
Biweekly Paychecks	39,000,000	97,500,000	195,000,000	390,000,000
Weekly Paychecks	78,000,000	195,000,000	390,000,000	780,000,000

This does not even include additional bonus and correction runs throughout the year (+10-20%) and does not take into account any Year to date values usually stored in triplicate per period above for Month-to-date, Quarter-to-date, and Year-to-date.

So the actual number can be 3 or 4 times the count above- it's not suitable for typical legacy databases!