Strategies for Archiving HR & Payroll Data

2022 edition

CHRO's guide

FUSE HR

FUSE HR

Table Of Contents

Why do we need to archive? 3

We will just build something in-house. 3

Cybersecurity regulation impact on hr and payroll ➠ 4

We will use the built-in native transactional system functionality for archiving 4

Why wouldn't we use a generic archiving solution for all enterprise data (not a stand-alone HR and Payroll platform)? 4

Gartner Says Beware of the Data Lake Fallacy ➠ 4

CA Policies and Procedures for Wage Claim Processing ➠ 5

Cybersecurity Regulation Impacts ➠ 6

It's probably too expensive for storing just static data. 6

We will look at archiving after the conversion is completed to our new system. 6

We probably don't have that much data we need to keep around (at least not for more than a year or two). 9

Cloud is not an option for us 11

How do we build a business case for archiving? 11

Strategies for dealing with legacy systems ➠ 11

Licenses 12

Hr and payroll data retention archiving compliance ➠ 12

People 13

Hardware 14
https://awstcocalculator.com ➠ 14
Compliance & Audit Risk 15
https://www.ibm.com/security/data-breach/ ➠ 15

Why do we need to archive?

Legal Compliance US Federal legal minimum 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 & retirement policies which also mean applying more proactive controls over documents and data globally for maximum retention periods..

Data analysis Having a historical record of your data in a reportable manner can provide significant insights into the future of your business and it's hiring practices. Without this data, long term trending analysis is not possible.

Common Objections

We will just 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 also 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 or more annually if not in licenses alone, then in personnel costs and infrastructure.

Lastly, legacy system security protocols are outdated. 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 global HR 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, and 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 ➠

We will use the built-in native transactional system functionality for archiving

The problem with this approach is 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 infrastructructure AND buffers against the loss of system specific personnel 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 Oracle Exadata 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.

Secondly, generic archiving solutions and in-house solutions are not built to serve HR and Payroll needs. Using one common example; they are not designed to support the types of retrospective Labor audits companies are likely to face. The State of California is increasingly active in these kinds of hour and wage audits:

CA Policies and Procedures for Wage Claim Processing➠

The scope of subpoenas submitted by the state counsel typically include details requesting 5 years of history including:

- Full employment history
- Time card details
- Paycheck details (Pay stubs)

These can be difficult to manually reconstruct from raw CSV files or generic solutions.

Lastly, traditional data lakes & warehouses were not built with data privacy in mind. Unfortunately organizations are commonly unable or unwilling to invest in added or retro-fitted security features such as encryption, role based access and PII governance solutions demanded by data privacy legislation globally today.

The Marriott story of a breached data warehouse of personal data is not uncommon➠

In the case of Mergers and Acquisitions the buyer will often be on the hook for all of the above risk.

How do GDPR and other global national Data Privacy trends impact our legacy strategy?

GDPR may affect you even if you don't have major business operations in EU countries as these laws typically protect any individuals who are 'citizens' of member countries. If you have one single EU citizen employed, you may still need some of the same processes & features in place to avoid the risks.

EU GDPR covers the following in detail:

1) what is personal data?
2) what rights a person has with regard to their personal data
3) what responsibilities the data "processor" or holder of personal data has

While EU GDPR outlines data privacy "principles", emerging US and other national data privacy legislation detail "facts" about what can and should be done to ensure responsible handling of personal data.

There are some differences in the various global and national laws being passed regarding handling of personal data, but there are also some global themes to help you build a future-proof strategy.

See the blog post here for the typical measures which help employers establish "reasonable efforts" to secure and protect personal data:

Cybersecurity Regulation Impacts ➠

It's probably 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.

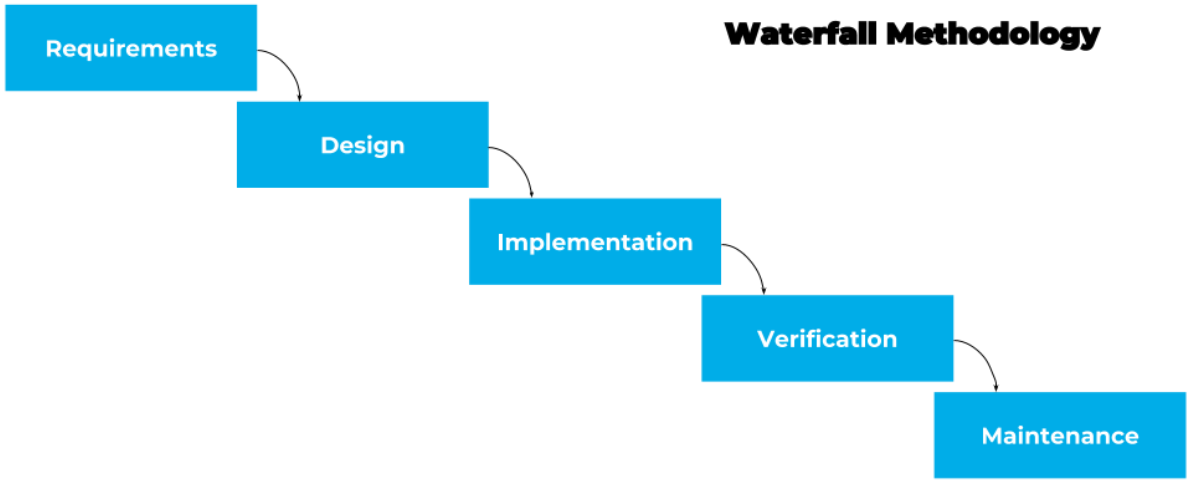
We will look at archiving after the conversion is completed to our new system.

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.

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).

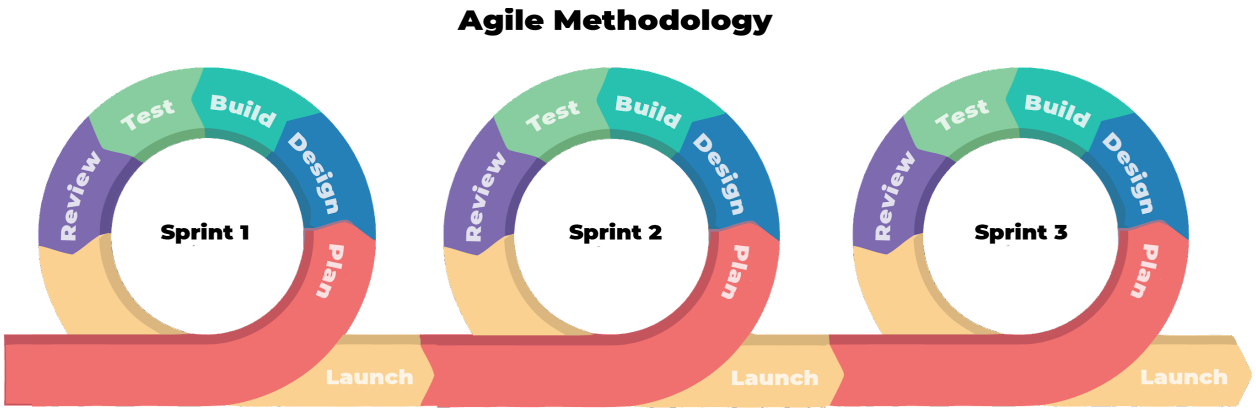
Typical System Implementations are performed with vendors using a ***Waterfall*** methodology where design and sign off are done in full up front.



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.

We probably don't have that much data we need to keep around (at least not for more than a year or two).

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. Data supporting areas such as Pensions, Workers Compensation and Hazardous Materials handling can be 20-30 years or more.

Additionally, data generated by Payroll and Time systems can be millions of rows even for smaller companies which is not suitable for many low cost database options..

For example, (in the U.S.) if a company has 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 * 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 =* ***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 grow exponentially:

Rows of Data needed for 5 years history				
Active Employees:	1,000	5,000	10,000	20,000
Semi-Monthly Paychecks	3,600,000	18,000,000	36,000,000	72,000,000
Biweekly Paychecks	3,900,000	19,500,000	39,000,000	78,000,000
Weekly Paychecks	7,800,000	39,000,000	78,000,000	156,000,000

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! Many companies are already looking for archiving solutions for R3 having seen significant slowdown in performance during large processing jobs like Payroll or Financial postings as well as Organizational Management reporting. ***The data sizes have outgrown the transactional R3 system in many customer installations.***

Cloud is not an option for us

Deployment can be flexible with multiple options:

1) Provider hosted cloud
2) Customer hosted cloud
3) Customer on-premise

However it should be noted that "cloud" has some very subjective definitions from organization to organization.

It is possible to deploy in the cloud while maintaining the instance as a secure extension of your corporate VPN. No traffic can access instances without first logging into your VPN and network.

Alternatively you may also whitelist IP ranges into a VPC (Virtual Private Cloud) to prevent all other traffic from reaching resources at all.

By centralizing your sensitive data you can focus on securing and monitoring access to one repository instead of many across the landscape.

How do we build a business case for archiving?

A business case for archiving is often easy to make for customers running large global SAP, PeopleSoft and other legacy systems. The below posting is great place to start thinking in principle about accumulation of 'technical debt'.:

Strategies for dealing with legacy systems ➠

Often organizations think only of licenses and ignore the costs of people to support and maintain old systems and infrastructure as well as the cost of hardware used to keep them alive. However when adding up the costs you should be sure to incorporate the annual costs of at least the following expenses:

- Licenses / Maintenance
- People

- Hardware
- Compliance & Audit Risk

Licenses

License costs are the easy data points to collect- these should be readily available to IT (or HRIT) if not HR. They can vary widely depending on the vendor landscape. Be sure not to limit your focus only on one single core HR or core Payroll system, chances are you have others in your landscape that contain any of the compliance related data & documents typically covered in retention strategies here:

[Hr and payroll data retention archiving compliance ➠]

Years of Required Data Retention per Policy

Policy		FMLA	FLSA	ACA	EEOC	Worker Comp	IRS	COBRA	H-1/H-2	OSHA	USERRA	HIPAA	FDA	BSA	Min Years	Max Years	Reccomended
Employee Data	Name	10	3	6	3	10	4	6	6	5	3	6	5	5	1	10	10
	Address	3	3	6	3	10	4	6	6	5	3	6	5	5	1	10	10
	Job Title	3	3	6	3	10	4	6	6	5	3	6	5	5	1	10	10
	Job Attributes	3	3		3	10			6		3				1	10	10
Compensation	Pay Rate	5	3		3	10			6		3				1	10	10
	Other base compensation		3		3	10			6		3				1	10	10
	Supplemental compensation		3		3	10			6		3				1	10	10
Payroll	Payroll wage details		3		3	10	4				3				1	10	10
	Working Hours		3	6	3	10	4				3				1	10	10
Time	Leave dates/hours	3	3	6	3	10				5	3				1	10	10
	Leave reasons	3	3	6	3	10				5	3				1	10	10
	Absences/ Attendances	3	3	6	3	10				5	3				1	6	6
	Payable time	3	3	6	3		4			5	3				1	10	10
	Clock times	3	3	6	3					5	3				1	6	6
Benefits	Benefits Coverage data+elections	3		6				6				6			3	6	6
	Benefits premiums	3		6				6				6			3	6	6
Training	Training data									5			5	5	5	5	5
Safety	Accident reports					30				5			5	5	5	30	30
	Materials Handling					30				5			5	5	5	30	30
Recruiting	Applicant Interview notes, feedback				1										1	1	1
	Applicant data				1										1	1	1
	Job Openings, postings				1										1	1	1
Other	Records of dispute				1										1	1	1
	Diciplinary actions, notes				1										1	1	1
	I-9 (Citizenship)								6						6	6	6
	Military Status Information										3				3	3	3
	Resumes												5		5	5	5

Most cloud providers often have little incentive for keeping customer history and on contract termination so there will often be a termination window of 60-90 days in which to reclaim your data via reporting extracts.

When dealing with on-premise legacy systems the costs may be more nuanced. Some providers may be prepared to charge no license at all while others may not negotiate down a full license price to retain historical access.

We typically see a range of anywhere from no charge at all (for limited access and no software or security patches, etc.) up to as much as $500,000 per year for more difficult legacy vendors and large clients.

People

The cost of people associated with maintaining databases, security, applications, networking, SAP R3 Basis functions and compliance/audit should be factored in even if they only spend a fraction of their time dealing with legacy systems it all adds up to effort not spent advancing more strategic initiatives and sometimes even remaining completely redundant for years to come.

Some example IT roles are shown below with estimated average allocation per system:

Prod Support/Maint headcount
DBA (.5)
Security Admin (.1 FTE)
Application Developer (1 FTE)
Functional Support/ Report Analyst (.25 FTE)
Networking Support (.1 FTE)
Overhead Fringe for FTE headcount reduction (Tax, benefits, other employer costs)

Average annual salaries for these roles range from from $US 100,000 to $US 150,000 per year. If you want to use a low average IT salary, $US 100,000 should suffice and be a fairly accurate average of the mix of roles above.

Also you may also need to consider any volume of ticketing internally that is associated with historical data and lookups and IT support requests if that is applicable in your environment via applications such as Service Now. This applies additionally to the time HR resources spend on tickets associated with historical data lookups and how time is wasted on traversing multiple legacy systems and document management solutions to find information requested by both employees (and former employees) AND the business stakeholders.

Be sure to add not only the salaries (by percent allocation applied), but also the overhead associated with all these roles as well which should be distributed as part of costs. That percentage is company and industry specific and may range from 35% up to 60% of salaries. This figure can usually be provided by Finance.

Hardware

Be sure to include the cost allocation for use of database space, servers, etc. While these are often negligible in looking at one system, note that a) this adds up over time and across systems and b) some servers and databases may only be in existence to satisfy the requirements of a single legacy application which means potentially shutting down and/or repurposing hardware. At a minimum it can mean trimming the infrastructure by several systems (application servers, web servers, database servers, etc.).

Amazon AWS has a handy tool for providing data points to compare on-premise server to cloud server TCO:

https://awstcocalculator.com ➠

A basic application server configuration to use for the parameters included therein would include the following type of parameters as an example. However, much larger global firms may have significantly increased capacities and on-premise expenses for systems like SAP R3 which make for an even more compelling case.

Sample TCO for legacy enterprise application would consist of at least 2 servers if not 3, with the following parameters:

1) Application server (4 CPU cores, 16 GB memory)

2) Web server where employee self service is allowed (4 cores, 16 GB memory)

3) Database Server (500GB)

A basic on-premise R3 installation can easily cost $ 100,000s annually even after it is retired. There is a significant savings to be realized in decommissioning permanently.

Compliance & Audit Risk

Risk is a bit more elusive of the costs however, but also should be considered. If you are maintaining 5-10 (or more) different legacy systems containing sensitive data versus one single repository which can be centrally secured there is a corresponding increase in data privacy risk.

Recent studies put the cost of a data breach at between $150 and $350 per employee record depending on the type of data and industry.

According to IBM, across all segments of data and industries,

"This year's study reports the global average cost of a data breach is down 10 percent over previous years to $3.62 million, due in large part to a strong US dollar. The average cost for each lost or stolen record containing sensitive and confidential information also decreased from $158 in 2016 to $141 in this year's study."

https://www.ibm.com/security/data-breach/ ➠

It follows that having this personal data replicated in multiple systems rather than 1 system, the risk factor is exponential. Compliance and Audit usually assesses and determine the value of consolidation in terms of risk reduction.

If you have 1,000 employee personal data records replicated in 10 systems you have 10,000 records which is a significant increase in potential breach costs. Post-mortem remediation and damage control also means analyzing a hopeless tangle of systems traffic where consolidation has not taken place yet.

Many legacy systems represent a significant portion of unnecessarily replicated personal data and often these ageing applications do not keep up with current security protocols and are quickly outdated and exposed to considerable data privacy

breach risks:

Oracle Squashes 53 Critical Security Bugs in one month ➠

Major Vulnerabilities in PeopleSoft Application Server ➠

SAP Patches Critical Vulnerability in Business Client ➠

These are illustrative examples - they occur on a monthly basis for all major applications at every organization globally. Therefore ***if you are collecting legacy applications you are in turn collecting all the aggregate risk as well.***

New legislation imposing much more significant fines has not taken effect and therefore does not count toward that average. Starting in 2018 we expect those per record penalties to jump significantly due to EU GDPR fines being levied. The GDPR regulation has hefty potential fines for non-compliance;

> ***"[organizations] can be fined up to 4% of annual global turnover or €20 Million (whichever is greater)"***

The types of gaps which impact the cost of a data breach and the potential penalty amounts are:

- number of records and systems exposed
- type of data (health, personal, etc)
- whether or not data is encrypted at rest and in-transit
- whether there is role based access controlling who can get to what data
- whether proper data governance and classification exists in all systems
- whether the organization employs data destruction policies and tools to purge personal data no longer required for business purposes

These are also drivers for evaluating global best practices in the upcoming EU GDPR legislation as well as recent cybersecurity legislation passed by NY State Department of Financial Services (DFS).

Applying these controls to several systems versus one consolidated repository can be cost prohibitive and in many cases impossible to implement and monitor effectively.

FuseHR is an HR & Payroll SaaS for Data Warehousing
brought to you by Fuse Analytics.

Fuse Analytics
Perimeter Pointe
1155 Mount Vernon Highway NE
Atlanta, GA 30338
Telephone: +1 212.377.6018
sales@fuseanalytics.com