

BASE LIFE SCIENCE
An Infosys company

DATA MIGRATION FACTORY

What makes the Data Migration Factory better over time?

How tools, processes and people compound with every migration - and what that means for cost and timelines.

56%
Drop in cost

1–2 mo.
Shorter project timelines

200+
Veeva migrations behind the factory model

56%

Up to this drop in cost per document

1–2 mo.

Shorter project timelines

200+

Veeva migrations behind the model

2

Working factories at two Top10 pharma companies

16

Migrations at our two factories across clinical, quality, and regulatory

OVERVIEW

From project to capability

With a Data Migration Factory from BASE life science, you can start aiming for up to a 56% drop in costs per document and reducing project timelines by 1–2 months. This goes for Veeva migrations in both the clinical, regulatory, and quality domain.

It doesn't happen instantly, however. Project by project, the factory becomes more efficient. But how does that happen? What makes the Factory so much better over time?

Building on more than 200 Veeva migrations and two working factories at two Top10 pharma companies, we have seen improvements in several areas. In this white paper, we give you an overview of all the things that play a role in making Migration Factories efficient over time. As with everything else, the solution is a combination of tools, processes, and people.

What is a Migration Factory?

> "A Migration Factory means taking the standard migration approach used for system implementations, making it specific to the client, and standardising as many things as possible to make them repeatable – just like in a factory.
>
> **In the end, the migration factory becomes more of a capability than a series of projects."**

Essentially, each migration becomes a learning iteration, allowing you to improve between each project. A new set of TMF document classifications can be re-used in another migration, meaning SMEs don't have to repeat the mapping exercise. Spending time deploying an AI Metadata Mapping Engine benefits all projects in the future. BASE getting to know your internal GCP-, biostatistics-, Vault Admin-, and IT Security-colleagues on one project means we can operate smoothly and efficiently on future projects – maybe even faster than your own project managers, because we solved the specific issues before, in your organization. Over time, all of this, and more, stacks up.

01 Tools

The tools used in a Data Migration Factory are the same as those used during a migration for a system implementation. The gain lies in the re-usability and the potential for improvement specific to your organization.

A home for your operations

It all starts with a migration server where we can store and analyze data before writing the ETL pipelines that eventually bring data into your target system. This sounds like a trivial thing, but having a home for your operations without risk of losing access or running out of space or compute, and the ability to onboard/offboard people as needed, while everything is aligned with IT Security and your cloud team – that's foundational and ensures that gravel doesn't enter the well-oiled machinery and slow us down.

Everyone who has ever onboarded a new IT vendor knows that internal alignment and setup take time. When we do migrations on our migration factories, we sit on the same server for years, and we can get right into the real work of moving data. Yes, the first project has to pay for setup. But subsequent projects don't.

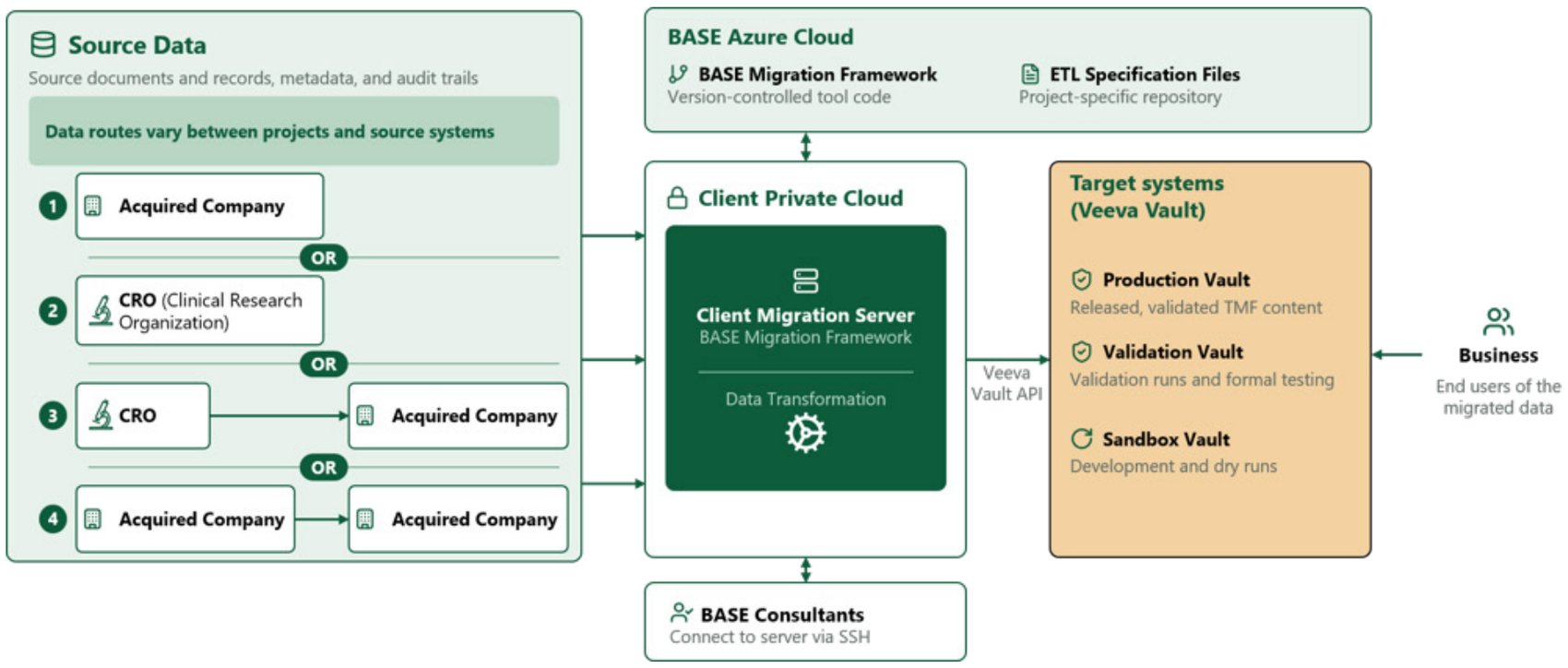


Figure 1: Migration infrastructure and server

Data profiling

15 standard data quality checks before accepting a data extract

The second tool we deploy is a data profiling tool we built to perform some of the 15 standard data quality checks we do before accepting a data extract. While nothing fancy, it quickly gives us an early indication of data issues, and as such helps ensure overall quality.

The tool is also used on regular migrations, but we customize it a little for each Migration Factory. Perhaps one of your CROs has systems that at times deliver data in a way not suited to your target system? Or maybe you have higher-than-average quality requirements that have to be factored in early? Either way, as we learn about – and with – you, we build these checks into the tool and the process: from MD5 checksums to ensure file integrity, to data/metadata/audit trail matching to ensure we have everything we need, and something as simple as whether files are password protected.



Figure 2: Data profiling tool with sample data. The example shows file types (right), which is relevant because most Vaults don't render all file types and thus should not go in. Another example is biostatistics files, which might belong in another system than Veeva.

AI in the factory

Three use cases

Finally, on the AI front, we have found use cases in three areas.

1 · Personal productivity. Our people use Claude for personal productivity like coding, project reporting, creating scripts to sample data for spot checks, and reviewing sanitized validation documentation.

2 · Automated field mapping implementation. Once we have agreed on field mappings with Line of Business, our Mapping Agent can implement them for us. This is a typical example of moving effort from tedious work to the part of the process where human input cannot be substituted. Agreeing on what data goes where, under which circumstances, the edge cases, and the rules that govern these transformations is not work for AI, it's work for humans. But once all these decisions are in place, implementing them as .yml code can easily be outsourced to AI, and depending on the size and complexity of a migration, this can save anywhere from a day to a week or two of work.

3 · Fully automated mapping and enrichment. In some cases, we have automated the entire field mapping and enrichment process using AI. This essentially means a set of agents read, extracted, and mapped thousands of documents, hitting an accuracy level well above 90%. We also did it 33% faster than it would have taken using regular methods (from 6 to 4 months) and with 3 fewer client team members to help us compared to the baseline.

>90%

Accuracy across thousands of documents read, understood, and mapped

33%

Faster: from 6 to 4 months compared with regular methods

3

Fewer client team members needed compared to the baseline

It’s possible – so why not on every factory?

1

If metadata exists, having AI re-extract it is sometimes slower than simply working with what you have.

2

Token costs can outweigh the cost in human hours.

3

AI cannot do a migration end-to-end. AI has amazing potential for migrations, and we use it everywhere we can – but we also think hard about the limits of AI.

02 Process

Having worked on migrations since 2019, we have developed an approach that is both standardized enough to bring best practice, and flexible enough to be tailored to each new customer’s QMS and IT. When working with M&A and CRO intake, you get a chance to standardize even more, because you work with the same setup, people, QMS, and requirements, potentially over years. In turn, this brings down costs and reduces timelines, while still keeping quality high.

Field Mappings

Two very obvious examples are Field Mappings and Value Mappings. While the source data might be unknown, the target field scope, their configuration, and the transformations expected are known and understood after the first migration. That means that already on the second migration, you can skip the mapping workshop and proceed directly to implementation, aligning with the domain SMEs only on edge cases.

By the time you perform the fifth migration on a factory, you have collected enough edge cases that the migration team can start to operate almost independently, giving Line of Business/SMEs a review role, and much less of an active decision-making role. That’s time saved for everyone.

WHAT THIS LOOKS LIKE IN PRACTICE

1 — Baseline
1st migration
Mapping workshop; target fields, transformations, and values become known

2 — Skip the workshop
2nd migration
Implement directly; align with SMEs only on edge cases

3 — Add to the library
4th–5th migration
Most source values are already in your library, but you keep expanding

4 — Work independently
5th migration
Migration team operates almost on its own; SMEs review and approve

Value Mappings

The Value Mappings follow the same logic in a sense. A good example is TMF classifications that often follow slightly different taxonomies depending on the source. You then need to map the source TMF classification (value) to the target system's taxonomy. Each migration is an opportunity to expand your library of custom value mappings, and by the time you get around to the fourth or fifth migration, there's a good to fair chance you can look up the source values in your library. Again, saving time and effort.

Validation starts on day one

As with anything in the pharmaceutical / life science industry, so with data migrations: validation and compliance start on day one. BASE has developed a standard migration validation framework that we use as the basis for all migrations. On a factory, you can take it a step further.

While the first migration requires alignment and customization, the second, third, fourth, etc. can rely on the migration that came before: the specific text describing the target systems, infrastructure, and process; the list of reviewers and approvers; Change Request (if needed) and in what systems. You can't re-use the documentation 1-to-1 because there is always something unique about each migration that has to be documented. But you can significantly reduce the effort.

80% less validation effort in some cases – from 100 to 20 hours.

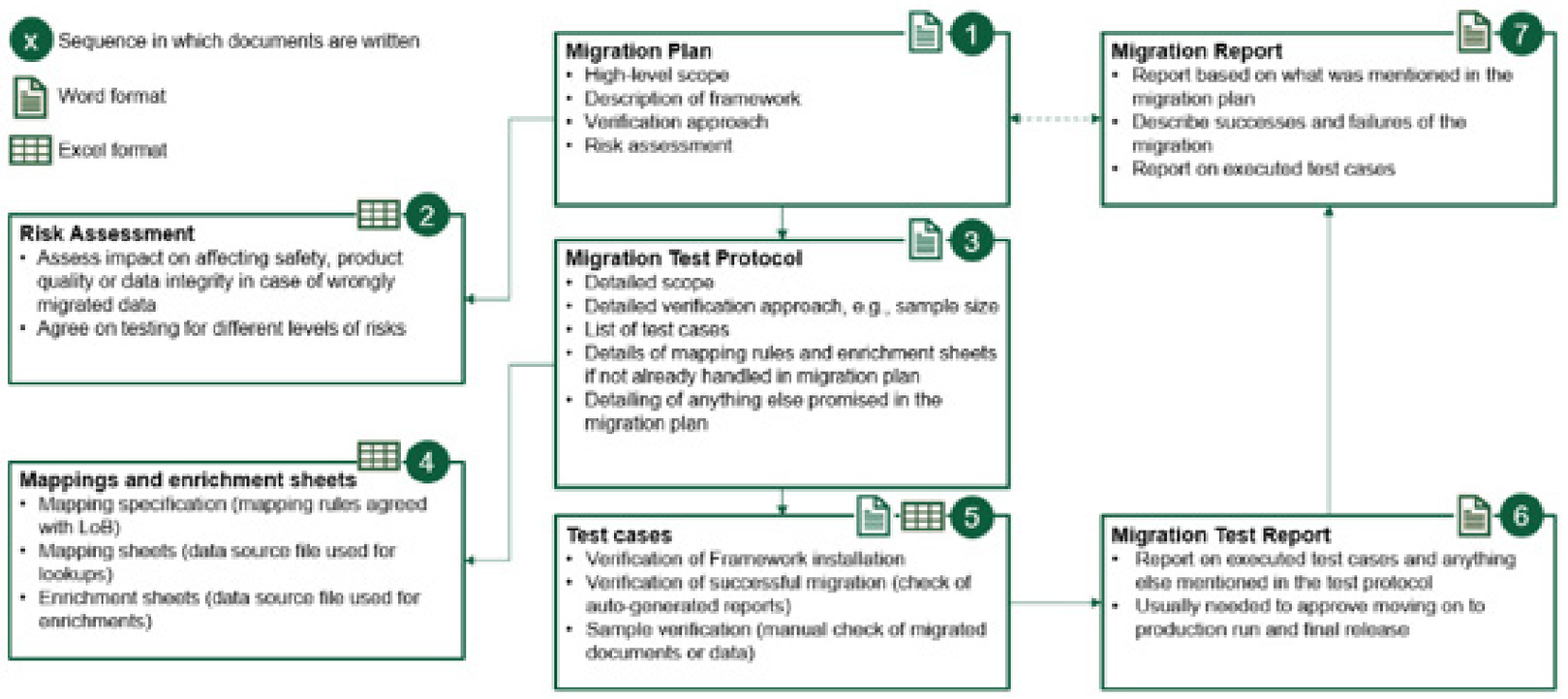


Figure 3: BASE standard validation documentation package. This is customized and then standardized further on a Migration Factory.

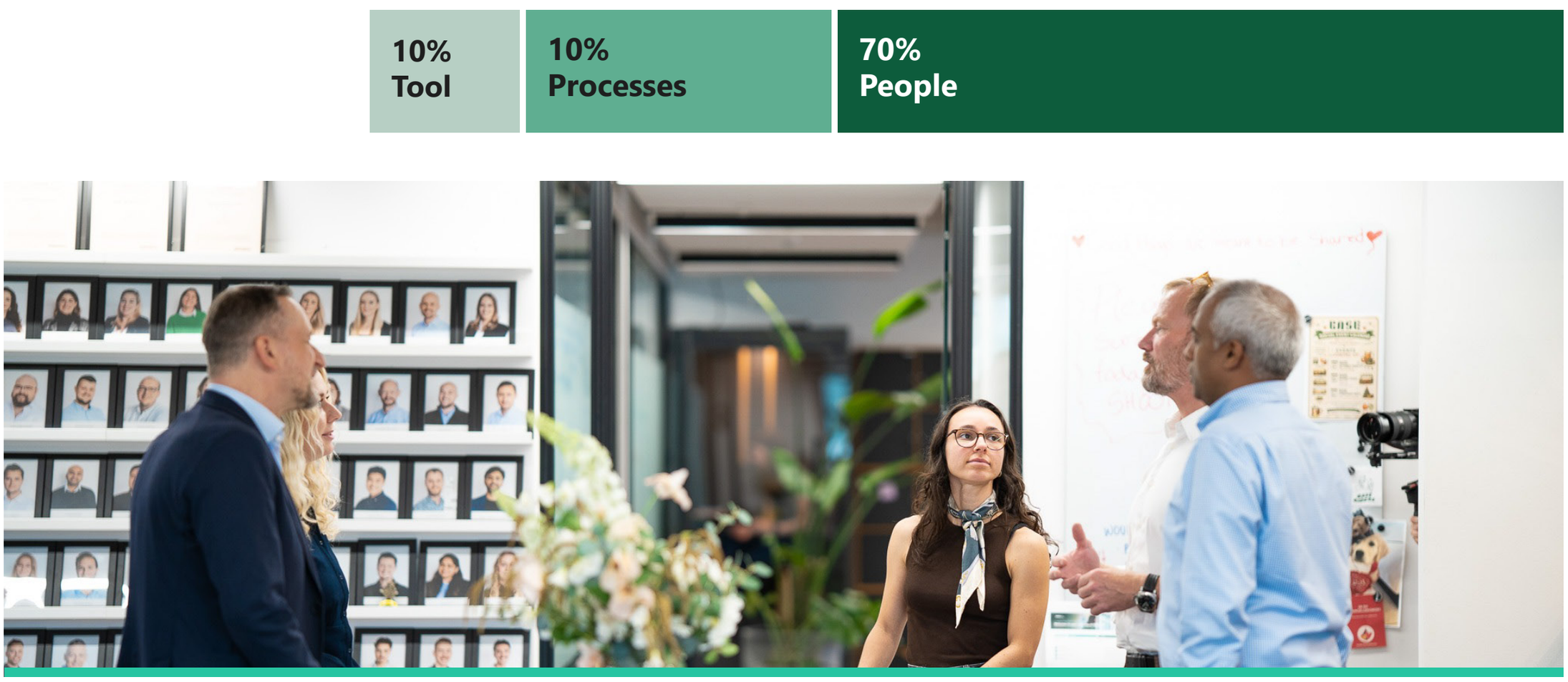
Standardizing environments

The above examples are a lot about templates, how they are used, and how we collect knowledge. Another part of the process to standardize is environments; which, who, and how. Usually, a migration team will have a sandbox for themselves, a validation environment that is used by others, incl. configuration teams, and the production environment. Sharing requires coordination with other teams, but also strict processes to ensure configs are aligned. With some customers, what we have done is simplify the setup and bring responsibility over to the Migration Factory team.

At one customer, we have a dedicated sandbox, and this sandbox is also used for validation of the migration. To mark the transition from sandbox to validation, the environment is simply refreshed by a system admin (obviously planned and pre-communicated as part of the project planning that takes place at the beginning of the project). While a small change, it has yielded a great deal of certainty and control, resulting in a slimmer process with fewer people involved.

03 People

When I started in BASE life science, I was told that any project was 10% tooling, 20% process, and 70% people. Maybe that has shifted in the age of AI, but our experience is still that AI cannot do a migration end-to-end, so for now at least, having the right people is key to good migration projects.



Embedded in your organization

Because a Migration Factory runs for an extended period of time, sometimes years, while a company goes through a wave of acquisitions, there is a chance to embed the consultants with the client organization and almost make them part of it. This comes with a wave of benefits: the Migration Factory team gets to know a lot of people in the client's organization, from GCP experts, Subject Matter Experts, QA, biostat team, and archival personnel, to sponsors, process owners, and system admins.

This makes the Factory team incredibly capable at navigating the organization and contacting the right people whenever certain issues arise, without having to be dependent on internal project managers who may themselves be new to migrations in a certain domain. The Migration Factory team becomes a lean operation from a client perspective.

Learnings that stick

Another great benefit is that the same team members pick up the learnings from project to project. Yes, you can document learnings, but it will never stick the same way it does when it's a person facing the same issue as he or she did on a previous migration with that customer. Whether it's data, process, people, system, or something different, knowing the issue and the customer does make the projects better.

Planning for the long run

Finally, there is a different sense of motivation for setting things up for the long run when you know you have to do a migration for the same customer in the same domain now, in 6 months, and in 12 months. We regularly hold sessions for each customer to see what can be improved. For some it's documentation, others infrastructure. Some customers have very aggressive AI agendas, and testing AI in migrations becomes easier. Either way, the Migration Factory teams for each customer have a reason to plan ahead, because they will live with the consequences in the future.

04 Concluding remarks

The numbers presented at the beginning of this white paper come from one of our factories, where we have been at work for close to four years and completed 12 eTMF migrations, three quality and one regulatory – all spawned by acquisitions and CRO intake. We didn't see a drop in cost of 56% per document between the first and the second project, but over time, it's possible.

In future white papers, we will dive into how AI is used in a factory, but also how under-hyped and unsung heroes like re-used decision logs and field mapping specifications play a role in making everything faster and more cost effective.

The lesson so far is simple:

A Migration Factory gets better with every migration – because tools, processes and people all learn together.

And that turns a series of projects into a capability.

Want to know more?

If you're interested, please reach out - we'd be happy to explain more!



Ulrik Oppenlaender

Manager · Data Migrations and Integration
uopp@baselifescience.com

www.baselifescience.com