#### PROJECT INITIATION DOCUMENT

(Version created to support DAP605)

|  |  |
| --- | --- |
| **Project name** | SSATS to LabVIEW ETL Data Transposition |

|  |  |
| --- | --- |
| **Release** | Draft/Final FIX AMERICANISMS  APPENDICES  REFERENCES  PHOTOS  FIGURES  Date: |

|  |  |
| --- | --- |
| **PRINCE2** | Based on a reduced version of the PRINCE2 PID documentary requirements |

|  |  |
| --- | --- |
| Author: | Jack Sargeant |
| Owner: | Jack Sargeant |
| Client: | Spellman UK |
| **Document Number:** | 1 |

#### Document History

|  |  |
| --- | --- |
| **Document Location** | This document is only valid on the day it was printed.  The source of the document will be found in the Control section of the Project File. |

|  |  |
| --- | --- |
| **Revision History** | Date of next revision: |

|  |  |  |  |
| --- | --- | --- | --- |
| Revision date | Previous revision date | Summary of Changes | Changes marked |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |
| --- | --- |
| **Approvals** | This document requires the following approvals.  Signed approval forms are filed in the project files. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Signature | Title | Date of Issue | Version |
| David Mason |  |  |  |  |
| Matthew Ford |  |  |  |  |
|  |  |  |  |  |

|  |  |
| --- | --- |
| **Distribution** | This document has been distributed to: |

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Title | Date of Issue | Version |
|  |  |  |  |

#### Purpose

#### To define the project, to form the basis for its management and the assessment of overall success.

Contents

[Background 4](#_Toc158041388)

[Project Definition 4](#_Toc158041389)

[Project Objectives 4](#_Toc158041390)

[Business Case 4](#_Toc158041391)

[Options Analysis 4](#_Toc158041392)

[Defined Method of Approach 4](#_Toc158041393)

[Project Scope 4](#_Toc158041394)

[Project Deliverables 4](#_Toc158041395)

[Exclusions 4](#_Toc158041396)

[Constraints 4](#_Toc158041397)

[Assumptions 4](#_Toc158041398)

[Project Organisation Structure 4](#_Toc158041399)

[Communication Plan 4](#_Toc158041400)

[Project Quality Plan 4](#_Toc158041401)

[Project Controls 4](#_Toc158041402)

# Background

Spellman High Voltage manufacture high voltage power systems used across varying markets. These include but are not limited to applications in medical, research, and security sectors. The organization is based in Hauppauge, New York, and employs 2300 staff across facilitates worldwide. Locations include Mexico, China, Germany and the United Kingdom.

Historically each site has operated somewhat independently, being responsible for their own customers and practices. Recently however, corporate want to begin to centrally control practices, procedures, and data. This initiative aims to achieve a holistic view of global operations and enable the identification and sharing of best practices across the company. Ultimately, this could lead to the streamlining of manufacturing and procurement processes, empower data-driven decisions, and make the movement of product lines to other sites far less ad-hoc.

A critical manufacturing step for Spellman is the testing of products. This step ensures customers receive their goods as expected, and such operate as agreed. Testing procedures and software have been historically developed independently on a per-site per-unit basis, with each site utilizing their own approach. Corporate have instructed all sites to streamline testing practices by grandfathering current solutions and replacing such with solutions that articulate Test Stand[[1]](#footnote-1) and LabVIEW[[2]](#footnote-2).

The UK site have developed their own test platform called “SSATS”. SSATS is a C#.NET derived form-based application, that articulates excel based test specifications to initiate tests and collect resulting test data. These excel workbooks are equipped with embedded automated testing procedures, achieved by implementing a native software framework using VBA. This enables the UK site to make the testing of products almost automatic, and only require an operator to enter product details and press start. This approach has proven suitable since its inception 10 years ago; however, it is rife with problems regarding adaptability, centralized control, and platform centric limitations.

Data from SSATS is collected once a test specification is closed. A worksheet named “Database” states all the fields to be lifted from the workbook once closed. The SSATS GUI maps this information to a DTO (Data Transfer Object) in which writes its contained contents to the SSATS database. SSATS was designed to test one serial number at a time, however, through business need, the UK team needed a solution to test more than one unit at a time using a single workbook. The team developed “Squirrel”. Squirrel is a API centric service responsible for lifting results for more than one serial number from a single workbook. This means that there are two data inception points for information stored in the Mustang Database.

Whilst the UK site have started to develop test solutions using the outlined software packages, it will take some time before solutions can adequately replace tests currently carried out using SSATS. However, corporate wish to start collating test data as soon possible. Because of this, a requirement exists to copy, sanitize and reformat data from the UK’s own SSATS result database into a format that matches the globally agreed, new database schema. In doing so, UK test data can be used in global business intelligence initiatives and visualizations and enabling such data to be present in their ERP system and material movement software Scanstar[[3]](#footnote-3).

# Project Definition

## Project Aims and Objectives

### **Aim**

* **Prototype a method to migrate Spellman UK's product testing data to a standardized format and prepare such for integration with the global system by June 2024.**

##### **Objectives**

* **Implement a data cleaning process and procedure that addresses identified data quality issues in UK test data extracted from the SSATS database to achieve 90% data completeness and consistency by May 15, 2024**
* **Implement a dynamic, remote and highly available distributed data integration service that can access and process test data from different test data inception points by June 2024**
* **Identify an appropriate solution for transposing test data from one database schema to another, in which 95% of all test data captured is transferrable and useable by April 2024**

## Business Case

Spellman’s new global initiative focuses on enabling organization wide collaboration, but also formalizing data collection, and ensuring a common format for all similarly shaped data collected. The collection of test derived data will be crucial for Spellman to build business continuity and security, as such can be used to understand product performance, highlight and forecast any problems, and can also be used to calculate throughput and yield, factors that affect the larger business paradigm.

Corporate have also highlighted how data like the test data will be ingested into the global data lake[[4]](#footnote-4). This centralized repository can empower deeper understanding of the organizations current and future stance through data cubes[[5]](#footnote-5), enabling multidimensional analysis across various parameters, kinds of data, locations, and timeframes.

Building on this strategic data initiative, Spellman require the UK’s test data, and other site-specific data within the global data lake. As the UK contributes a share to global operations and revenue, comprehensive representation of the UK’s data becomes critical. In collating such, Spellman can ensure data-driven initiatives include insights into all locations under their umbrella. By readily integrating UK data into the data lake, Spellman will be able to gain a clearer picture of the UK’s contribution and performance and reinforce organizational security by painting a far more accurate picture of this site’s performance.

It is evident that a business need exists for this data, and UK site should immediately start transposing test data into a format suitable for this global initiative, to begin fostering an age of improved global business intelligence.

While the outlined global initiative addresses the standardization for broader analysis, there's also a crucial need for it at a more local level in the UK. Having historically experienced challenges with the quality and consistency of data written their SSATS database, the UK team sees the data transposition requirement as a golden opportunity to cleanse, standardize, and strengthen all locally collected data. Improvements to data and associated visualizations can be used to look at product performance, identify bottlenecks and issues, and look at equipment performance. Most importantly however, such could assist the UK site in bolstering the business case for automation initiatives and stand as concrete proof to improvements made regarding performance. visualizations of this kind are Ad-Hoc currently, and the data collected is not in a fit state to be prudently used, due to the volatile nature of the UK’s test platform and data collection method(s).

## Options Analysis

Spellman UK have a few options available to them relative to their requirements. It would be important to add that **the organization shows no interest in additional investments to technologies at this time.** The requirements outlined infer the use of a remote service, responsible for moving results from one database to another, whilst remaining dynamic and automatic as specified. Some of recommendations would incur a cost, and have been stated for the interest of both the project board, and stakeholders:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Options | Benefits | Drawbacks | Feasibility | Impact | Compliance |
| Do Nothing | No cost or time spent required.  No additional risk.  Frees up development staff. | No standardization to data.  Data in no state to be moved or utilized by Spellman UK or corporate.  No means of corporate retrieving test data in an expected format. | Not very feasible. This approach very much misaligns the requirements by corporate to centralize and collect data. | Corporate will not have full scope of all data required from the UK site to create meaningful conclusions. | Would not affect or violate any compliance ruling regarding regulation, or company policy. |
| Developed in-house, self-hosted | Full control.  Tacit knowledge applicable from both developers and IT staff.  Flexibility.  Security.  Customizable SLAs with IT regarding service uptime | Time and cost on development and hosting.  Complexity and scalability regarding infrastructure management. | Completely feasible following review with IT. Developers have knowledge of practices and requirements to deliver solution. Infrastructure also exists and can support. | Would broaden the portfolio of UK IT, as would need to upkeep service.  Development time spent takes away from another project.  May require additional training. | Does not affect or violate any policies. Data collected is not personal so does not breach GDPR or similar. IT staff and developer would be responsible for adhering to relevant policies, however. |
| Developed in-house, cloud hosted | No reliance or requirement on in-house hardware/infrastructure.  Scalability | Security concerns regarding handling of sensitive data  Vendor-Lock In  Less control than self-hosted.  Potentially limited by SLAs agreed with IT.  Additional Costing | Not as feasible as all in-house. Organization have voiced there is no interest in investing with cloud or “as a service” providers. Developers and IT staff also lack knowledge to control such.  This suggestion is also the only one that would incur additional cost, outside of investments made by the organization in terms of time and allocating owned infrastructure. Appendix 1 demonstrates a quote from AWS (2024), which denotes a monthly cost of $55.40 based of current organizational activity. | Would require ongoing costs and negotiations with cloud provider. | Would likely breach any policies regarding data safeguarding under the organization’s ISO:27001 centric policies. Organization very much wants to keep data “In-House” |
| Database centric solution | Efficiency: Can leverage DBMS engine.  Ease of management.  Can be integrated into current DBMS.  Limits movement of data. | Limited functionality.  Costing for additional cores.  Limited scheduling flexibility.  Known volatility with currently deployed DBMS.  Added complexity compared to what could be possible in code. | Whilst somewhat feasible, and seeming the organization already use databases, this could work. However, the UK have experienced with varying database configurations and problems and may not wish to tie critical activities to such. | Would require load balancing and throttling to ensure DBMS can process all information.  System would also require reporting mechanisms, meaning additional development time. | Would not affect or violate any compliance ruling regarding regulation, or company policy as data is being kept in-house. |
| Attach Writing events to Excel Test Specifications | No need for a service to be hosted or controlled.  Data writing could be tied to specific tests and rolled out incrementally. | Would need to be attached to every single test specification individually.  Limited by VBA code.  Would need to be event driven.  Very much becomes tightly coupled with Office, and any updates that may hinder operation | Not too feasible. Whilst it makes to write data from the source itself, by nature all test specs remain somewhat independently controlled, meaning each would need to be manually updated and configured to move such. This would be a huge task, and one not likely to be chosen by the organization. | Would require Test Equipment staff to begin manually changing every test specification, taking them away from other responsibilities.  Also tied to Office and automatic updates that have previously hindered the organization. | Whilst it could be possible to adhere to policies set by IT, XLSM file or any VBA code written in office documents have notoriously been associated as a cyber-attack vector. An automatic write to a database on a network via VBA could raise problems. |

## Defined Method of Approach

Any projects success hinges on the identification and selection of a suitable PMLC (Project management lifecycle), and as this is a software centric project, also a SDLC (Software Development Lifecycle). Choosing the right methodologies doesn't just ensure clear direction and efficient resource allocation, it directly impacts the quality, functionality, and timeliness of the project deliverables. Both PMLC’s and SDLC’s should also be concerned by dictating project control measures, risk mitigation and defining project phases.

This project will be carried out in phases, that will be populated later in this document. For this project, the PMLC will very much serve as a top level, being concerned with overarching project controls and documentation, whereas the SDLC will be concerned with controlling development activities once that stage is reached. The following figure attempts to demonstrate such, where orange dictates PMLC stages, and purple SDLC stages; it attempts to show their relationship:

A diagram of a process

Description automatically generated

Figure 1- demonstrating the relationship between a SDLC and PMLC within the context of this project.

To understand both a suitable PMLC and SDLC for this project, we must understand the organizations current stance ability to meet the projects aim, and subsequent objectives:

* **The development team that will be assigned to this project have a solid grasp of the theory involved regarding data cleaning and mapping.**
* **Whilst the development team are technically capable of delivering a solution, it is more than likely this approach could involve technologies not found in their current stack, indicating some level of learning and prototyping will be required; this kind of project is different from what they have worked on previously.**
* **The organization have a plethora of self-owned and controlled technologies suitable for deploying a solution fit for their requirements.**

Previous sections of the document and this synopsis of the organizations current stance ultimately indicates that whilst the goal and business case for this project are clear, there may be some friction regarding delivery. This is due to potential unfamiliarity with required tooling, and the subsequent need to learn, research and prototype solutions.

In his book “Effective Project Management” Wysocki (2019, p.23) denotes 5 different PMLC/SDLC approaches, with each suitable for different kinds of project. These approaches are iterative, linear, adaptive, incremental and extreme. Wysocki proceeds to explain (2019, p38) four different kinds of process groups, and how the PMLC/SDLC approaches are best mapped to such. These process groups very much define the characteristics of a project, and what a project of that type may entail. Wysocki mapped these based on the certainty or uncertainty of a project’s goal and/or solution.

A diagram of a solution

Description automatically generated

Figure 2- Wyscoki (2019, p39). This model demonstrates how different project approaches can be mapped to different kinds of project.

The process groups and corresponding PMLC/SDLC’s have been mapped as such:

* **TPM (Traditional Project Management):** Projects within this quadrant are ones in which both the end goal and solution are both certain.
* **Agile Projects:** Agile projects generally have a clear goal, but uncertainty about regarding the solution.
* **Extreme Projects:** Extreme projects do not have either a clear goal or solution.
* **Emertxe (ee-MURT-see) Project:** Emertxe projects do not have a clear goal, but do have means of delivering a possible solution, whatever it may be.

It is clear the goal for this project is clear. Whilst the development team will more than likely need to develop some new skills and understand new approaches, this should be something factored in to the phases of the project, rather than classifying the solution as “uncertain”. Projects of this caliber, and phases within any SDLC will generally include a phase for identifying requirements, prototyping and theory crafting, furthermore, it has also been identified that the development team are more than capable of meeting the goal stated. Because of this, this project falls under TPM.

As demonstrated with Wysocki’s classification, the selection of a TPM classification to this project highlights three different PMLC/SDLC approaches:

* **Linear**

A grey rectangular sign with black text

Description automatically generated

Figure 3 - The Linear PMLC (Wysocki, 2019, p34)

* **Incremental**

A diagram of a system

Description automatically generated

Figure 4- The Incremental PMLC (Wysocki, 2019, p44)

* **Iterative**

A diagram of a system

Description automatically generated

Figure 5- The Iterative PMLC (Wysocki, 2019, p49)

*State the projects team current stance based on what is required.*

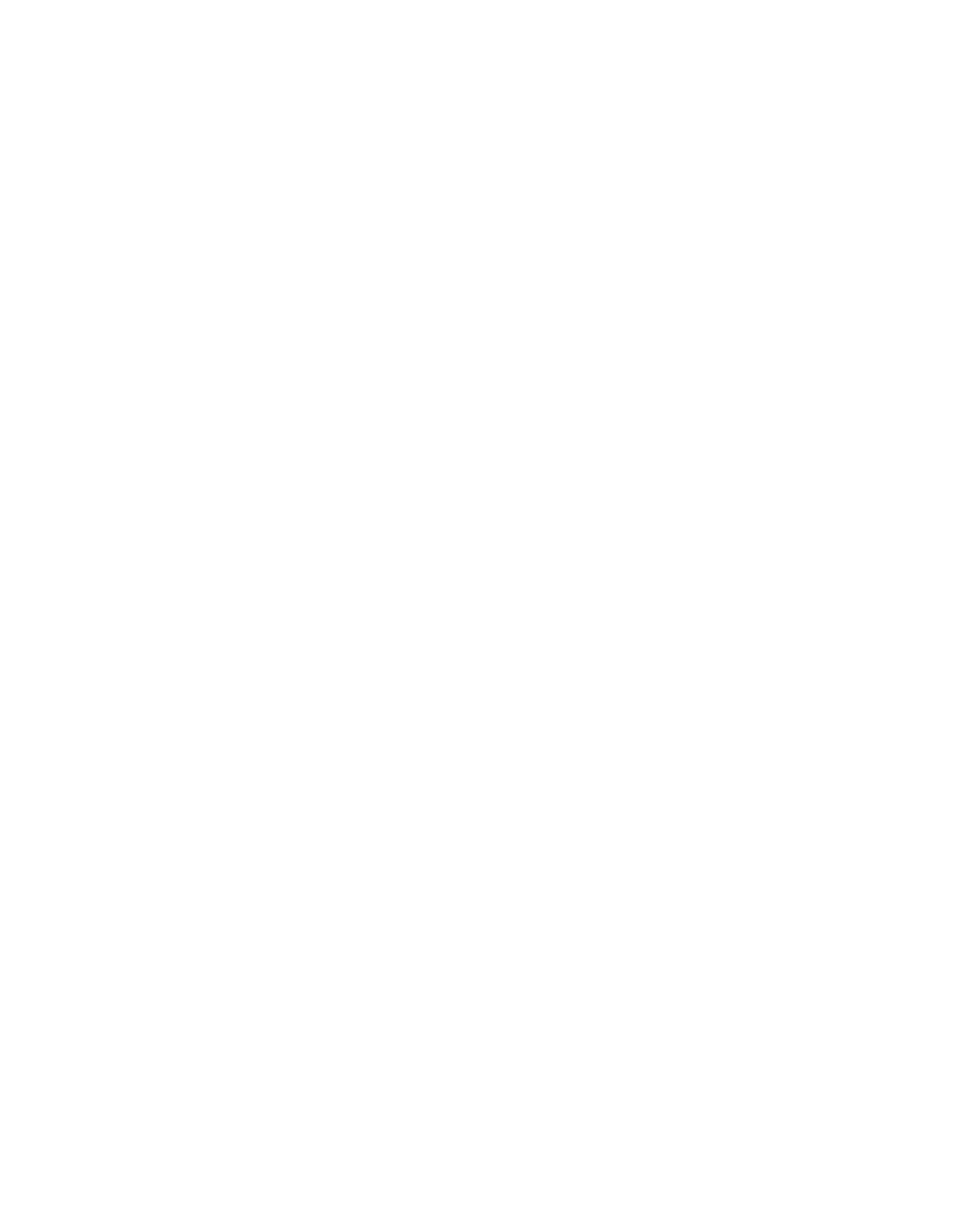
*Compare all options*

*SDLC – being a software project, the SDLC is coupled very tightly with a PMLC. As the project progresses into the development stage, deliverables, content and whatever should change to reflect such.*

Pros and Cons of Iterative SDLC’S

SDLC Phases – Where does it fit into PRINCE 2 (Probably phase 3)

Wysocki, RK 2019, Effective Project Management : Traditional, Agile, Extreme, Hybrid, John Wiley & Sons, Incorporated, Newark. Available from: ProQuest Ebook Central. [13 February 2024].



## Project Scope

Define the scope of the project in terms of any boundaries to it. These may include, for example; geographic boundaries, organizational boundaries, process boundaries, system boundaries, whether you are going live or just developing a prototype etc.

## Project Deliverables

A precise description of the deliverables from the project as expected by your customer. Again this is critical because you are in effect setting the expectation of what will be delivered and gaining your customer’s agreement to this. This may include the system, documentation, training, identifying an ISP etc.

*Service Orientated Architecture.Enables an approach to start such*

## Exclusions

Include here a list of anything that you will specifically not deliver or undertake in order to clarify the expectation. If there is nothing you wish to raise, then ignore it using ‘Not Applicable’.

## Constraints

Include here any constraints to the programme, for example, constraints on your availability, availability of the customer, when the work has to be completed by, architectural standards to be followed, software packages to be used, budget constraints, travel constraints etc. – Not just a bullet point list – may need to be justified. You could limit the number of iterative development cycles here.

## Assumptions

Detail anything that you need to have or must be available in order for you to undertake a successful project, for example; access to specific members of staff, regular reviews of progress, quick decisions, access to the organizations intranet etc. Be specific e.g. if you need decisions or sign-offs to be turned round in 3 working days, say so.

# Project Feasibility Assessment

SWOT

# Project Organisation Structure

It is often useful to include a picture of the organization, showing the team, there roles and who your principle contacts are. This section should have a clear statement on the roles and responsibilities of each member of the team. Customer roles should also be specified.

# Communication Plan

*Identify who the key contacts are for the project together with their contact details. Also identify how often you intend to review and communicate your project progress and who to. It is useful to include some more formal reviews that follow PRINCE2 practice.*

*Mendelow stakeholder mapping matrix*

# Project Quality Plan

*The project quality plan details how you intend to meet the customer’s quality expectations. You need to begin by defining these. You should also state the acceptance criteria, who is responsible for quality, any quality standards you will adopt e.g. covering documentation and verification and validation.*

# Project Tolerances

*This section identifies any agreed tolerances on the project. Typically these would relate to time, cost, quality and functionality. So for example, the project could be delayed by up to one week without reference to the Project Board but beyond this would have to seek approval. You may note that all essential requirements have to be delivered but there may be some tolerance re desirable requirements.*

# Project Controls

*This should document the project controls you intend to put in place. It is likely to include statements on the following: that you will be following the PRINCE2 methodology, how and where you will maintain and keep your project files, how you will manage and requests for change, configuration management controls etc.*

*Controlling Phases using PRINCE2 – following the methodology*

*File Storage and Management*

*Software Development Controls*

*Documentation*

#### Attachments

|  |
| --- |
| Initial Benefits Case *It is worth thinking through the potential benefits of your piece of work from the perspective of the customer. You have detailed the objectives but if you deliver on these, how might the business benefit? Understanding this may help you to identify your key project themes and ideas to pursue.* |
| Initial Project Plan *Ideally, this should be a Gannt chart showing the tasks to be undertaken and when you expect to do them. You should also include critical review points and milestones. It should build on the SDLC phases defined. This will evolve during the project.* |
| Initial Risk Log *You should think through any of the potential risks that might prevent you completing the project to plan and expectation. Issues may be difficulties in contacting people, arranging interviews, unplanned absence etc. Document these risks and identify any actions that you can take to try and prevent them occurring in the first place. Monitor risks on a regular basis*  *Product Description*  *Project Brief.*  *Captured Requirements* |

# References

<https://www.ni.com/en-gb/shop/product/teststand.html>

<https://www.ni.com/en/shop/labview.html#:~:text=LabVIEW%20is%20a%20graphical%20programming,and%20fully%20integrated%20user%20interfaces>.

<https://aws.amazon.com/what-is/data-lake/>

<https://www.thedataschool.co.uk/sherina-mahtani/data-cube-an-introduction-to-the-concept/>

<https://www.ibm.com/topics/rpa>

<https://www.fbi.gov/news/stories/melissa-virus-20th-anniversary-032519>

A screenshot of a computer

Description automatically generated

1. Test Stand is test management software that enables the development, debugging, and deployment of test systems, it also provides full visibility into testing process and results. (NI, 2024) [↑](#footnote-ref-1)
2. LabVIEW is a graphical programming environment that provides test system development options, such as an intuitive approach to programming, connectivity to any instrument, and fully integrated user interfaces. (NI, 2024) [↑](#footnote-ref-2)
3. [↑](#footnote-ref-3)
4. A data lake is a centralized repository that allows for the storage of unstructured and structured data at any scale. (AWS, 2024) [↑](#footnote-ref-4)
5. A data cube refers to a multi-dimensional data structure. Data within a data cube is shown by specific dimensional values. (Mahtani, 2024) [↑](#footnote-ref-5)