MYRROR

Briefing Note

August, 2025

Re- Documenting Complex Processes in Wholesale Financial Services

Introduction

Re-documenting complex processes across a legacy environment in the wholesale financial services vertical is non-trivial. Typically system documentation is out of date, if available at all and often much of the system expertise is "implicit" meaning it is retained in the minds of the system experts.

These experts are usually widely distributed across the organisation, and only have silo'ed "pockets of expertise" of the system's operation, relating to their specific areas of interest and responsibility.

Re-documenting the end-to-end processes executing across the legacy implementations used in this vertical has three objectives. The first being to bring whatever system documentation there is up to date, the second being to document the "implicit" expertise retained in non-written format and the third being to generate and maintain documentation detailing the operation of the system to a standard that meets the regulatory demands.

Automating this process contains much of the frustrations associated with the resource hungry, error-prone manual procedures being employed to achieve this task.

Keywords: Complex systems, Wholesale Financial Services pockets of expertise

Definitions

Process - an assembly of interlinked, atomic activities each of which is represented by a system interaction or "binding". These "bindings" generate observable system behaviour that may be observed, captured and compiled into a representation of the system semantics.

Process Digital Twin - a Process Digital Twin is a continuously updating graphical representation of domain specific observed system bindings. This ontology is best expressed using a meta-model as a standard notation, providing a machine-readable description of the syntactic parameters of the system behaviour.

Digital Twin of the Organisation - the Digital Twin of the Organisation is an end-to-end meta-model representation of the organisations complete and live executing environment, designed to provide timely, accurate and complete information on transaction activity from transaction "capture" to transaction "completion".

Background

Re-documenting system processes requires the system semantics to be identified and extracted. Capturing the system semantics as the system executes reveals significant details of the system architecture and its implementation.

Automating the capture of system semantics, in turn, enables the documentation describing the system architecture and its implementation details to remain current, with the resulting documentation updating automatically as system changes are introduced.

This can be described as creating The Digital Twin of the Organisation a representation that precisely mirrors the live executing environment. This may also be described as the Process Digital Twin[1].

Capturing the System Semantics

Automation Limitations

Whilst automating the capture the system semantics is the preferred approach, it will be the case that some activities will be difficult or impossible to capture using a purely machine-based technology.

These activities will most usually be human-to-human interactions, which, by definition, can be subject to unexpected, unpredictable and non-logical variation.

Automating the capture of the system semantics is thus limited to human-to-machine and machine-to-machine interactions. That is, observing and capturing system semantics is limited to "machine-observable" end-point behaviour executing on the live environment.

A second limitation is the use of structured and unstructured data[2]. For a machine to capture the semantics of a live environment, it must be able to observe and track the system's message bindings as these execute. This is achieved by tracking specific data elements contained in the body of the message(s). Tracking these elements is only possible if these element(s) are located in precisely the same location in message body every time.

Machine-based capture of system semantics is thus limited to those semantics executing using structured data.

This is not a significant limitation. Most large organisations, particularly in the financial services vertical are already implementing AI and ML based automation practices, that, by definition, demand the use of structured data.

"Always-On" System Activity

The accelerating adoption of AI and LLM to fully automate repetitive processes, is facilitating a migration from batch-activities to continuous steaming "always-on" data telemetry[3].

Continuous streaming results in enhanced processing efficiency, increased system hardware utilisation, reduced storage demands and optimised process control.

The elimination of batch-processing will require continuous real-time system observability coupled to continuously updating ontology of the executing semantics to facilitate near-continuous process rectification and optimisation.

Organisational Structure

The wholesale financial services vertical is characterised by collections of isolated, silo'ed domains of system expertise with processes typically being domain-specific batch-activities.

---

[1] For the purposes of this paper the expression "Digital Twin of the Organisation" shall be used to describe the ontology describing end-to-end processes executing across multiple organisational domains whilst the expression "Process Digital Twin" shall be used to describe the ontology representing domain specific activity

[2] It is important to note that "unstructured data" does not mean data without structure, In this context "unstructured data" means data that is structured in a manner that renders it unsuitable for process discovery. This data may be in the form of audio files, emails, images, video, documents, social media and other formats that are structured for the medium intended by not suitably structured for machine based process mapping technologies. It may be possible to transform such data into a format suitable for process discovery but generally these approaches are time-consuming and error-prone, requiring extensive data transformation, integration, enhancement, enrichment, and cleaning before a representative data sample can be extracted

[3] See Appendix 1

The end result is “pockets of system visibility” without organisational wide end-to-end system process visibility from transaction “capture” to transaction “completion”

Documenting End-to-End Processes by Increment

Documenting the end-to-end process activity across legacy systems is achieved by leveraging the “pockets of domain visibility”.

Each individual domain is re-documented by capturing the messaging bindings within that domain, and creating the Process Digital Twin specific of that domain.

Once the domain specific Process Digital Twin has been generated, the process “hand-shakes” to each neighbouring domain may be identified

The domain specific Process Digital Twins are then aggregated. In this manner to build an end-to-end Digital Twin of the Organisation is built by increment.

The Re-Documentation Procedure

The re-documentation procedure is a two-phase process.

The First Phase is comprised of examining the existing documentation, static system representations, reviewing the source code and interviewing the domain experts.

From this may be compiled an understanding of the system’s characteristics, a dictionary of application names, and details of data dependencies, time stamp formats and other system characteristics, as well as providing an initial understanding of the system’s executing semantics. This process will only provide an overview of the existing implementation as almost invariably the existing system documentation will be out of date.

The Second Phase is the deployment of distributed AI Agents across the production environment. These agents observe the activity between the interlinked co-operating system components and report the observed interactions to a central utility. This utility then generates a graphical formalism of the system semantics.

From this domain specific system representations may be extracted. These in turn may be aggregated to generate the Digital Twin of the Organisation.

“Simplicity” is the Key

Capturing a domain specific Process Digital Twin results in the capture of a significant quantity of system information.

The documentation objective should be one of simplicity by producing a documentation portfolio meaningful to technology users and consumable by business users. This can be achieved by using tiers of aggregating meta-models, with each “upper tier” displaying fewer details than the preceding layer whilst providing greater end-to-end system visibility.

Process Mining is not Task Mining

Process actions divide into two types: tasks and activities. Tasks relate to actions executed outside of the information and communications technology system.

These will be human-centric activities such as reviewing emails, making entries onto spreadsheets, answering enquiries and so forth. Although these steps form part of the overall “process” they are not activities captured in the Digital Twin ontology.

Task Mining is a separate discipline supported by screen and keyboard activity monitoring software. Capturing and documenting these human-centric tasks is a separate function to Process Mining.

Process Mining relates to the extraction of the system activities that define the system semantics. These semantics are documented into an ontology detailing system and data state in a single representation.

Capturing the system semantics is the capture of machine-generated process activities, reliant on software code, and machine-based rules engines.

Treating System Semantics as Data

Treating the system semantics as data enables a clear delineation between process “tasks” and system “activity” to be drawn.

Once documented into a Process Digital Twin, system activity may be subjected to the same mechanisms used to manage data.

This enables specific sequences to be categorised and defined in the same way as data is compiled into data dictionaries; ownership can be assigned, glossaries and taxonomies constructed and validation rules developed.

Once this is achieved, in the same way as data management has matured, the management of system semantics can become subject to re-usability, modularisation, standardisation, quality control, compliance with standards, and governance.

Summary

Capturing domain specific Process Digital Twins and progressively aggregating these into a Digital Twin of the Organisation has the objective of :

- breaking down implementation silo’s
- formalising the language used to describe the system semantics
- visualising the migration of data over sets of co-operating system components
- surmounting the barriers that separate “the business” from “the technology”

Once built the Digital Twin of the Organisation will facilitate better insights into the business inefficiencies, create synergies between organisational domains, accelerate business delivery, and expedite the acceleration of revenue accrual.

Businesses are defined by their processes. The capture of the Digital Twin of the Organisation ensures more timely, accurate and complete information of the actual functioning of the business.

Capturing the Digital Twin of the Organisation needs to be both iterative and incremental with the progressively maturing output being modified and augmented by a continuous feedback loop from the target users.

Simplicity governs the presentation of the resulting output. The Digital Twin of the Organisation needs to comprise sets of sub-ontologies, each presenting only that information required by the target users of each sub-ontology, with each sub-ontology able to be tailored to meet the requirements of a specific set of circumstances and a specific user-group.

The HELIXsystem Process Assembler™ is protected technology by US patent grant No 8,554,594 “Automated Process Assembler”

Appendix 1

Batch-Processing to Continuous Steaming Data Telemetry

The migration from batch-processing to continuous steaming data telemetry within the wholesale financial services vertical correlates with Citibank's vision that a similar migration will be witnessed within the corporate treasury sector.

In two opinion pieces, *Insights from Client Advisory — The Real-time Treasury Evolution: A Shift to Modernisation* September 2024 and *Treasury 2030 Modernise or Risk Irrelevance* December 2024 Citibank expresses the view that the adoption of AI, LLM and other new technologies has reached an inflection point, that is transforming the corporate treasury function and its supporting banking services.

In Citibank's opinion, new technology adoption is driving a migration to real-time, always-on, 24/7/365, multi-bank, multi-jurisdictional banking services, with

"banking days, batch processing, cut-off times, end of days, and weekends (to be) left behind as relics of the 20th century."

Citi is of the view that "aways-on" corporate treasury streaming services will soon become "a standard requirement".

This in turn will require "dynamic real-time solutions" as real-time liquidity becomes the norm.

As the migration to continuous streaming treasury services accelerates so will the demand for continuous real-time visibility over digital identities, cash flow management, on-demand liquidity, and risk management.

This demand can only be met by the wholesale financial services vertical migrating it's current fragmented, isolated, individually silo'ed, batch-activity based systems to organisational wide end-to-end system capabilities delivering process visibility from transaction "capture" to transaction "completion".