Whitepaper

The Anatomy of a Single Source of Truth Platform

Strategic Portfolio Management

By

Allan Schwarb
allan.schwarb@gmail.com
@allanschwarb
Linkedin.com/in/schwarb1

There were 5 Exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days.
—Eric Schmidt, CEO Google.

Contents

Abstract

The Internet is the first thing that humanity has built that humanity does not understand, the largest experiment in anarchy that we have ever had.
— Eric Schmidt, CEO Google.

In this paper I present an architecture that allows a modern Development Team to digitally transform from a distributed dataset to a single source of truth (SSOT) platform[i]. This Development Team can then access a collocated ordered dataset if dispersed data is migrated within a system boundary to an SSOT. The KPI/OKR drivers include requirements management, real-time visibility and ability to ultimately execute DevSecOps sprints. This business strategy was executed within budget and time constraints.

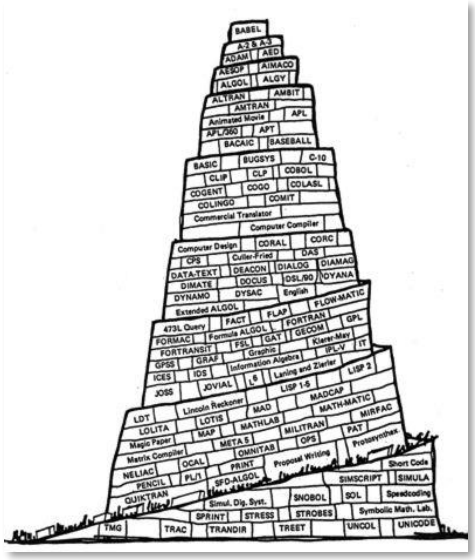

Figure 1. Data & Language Tower of Babel. Communications of the ACM (1/1961) [Association of Computing Machinery – Cover illustration]

To engineer an SSOT strategic portfolio management (SPM) solution is a demanding task. Accomplishing this within personnel, time and fiscal limitations only adds to the complexity.

Business is facing an existential data crisis. This paper is about a way to end today's data anarchy predicted in the 1960s[ii] [Figure 1]. From desk drawer as shown in Figure 2 to immense data centers, data is often everywhere except where you need it: At your fingertips to measure and execute your strategic initiatives. A successful SPM approach is crucial to a winning execution.

In an average enterprise, the CTO is often unplugged from a staggering amount of critical data generated every day[iii] in an enterprise. Meanwhile, middle management attempts to use tools such as Microsoft Outlook while hosting Excel and PowerPoint on various corporate instances of Microsoft SharePoint. Much of this enterprise data is not crawlable because it is elsewhere. Microsoft Outlook is firewalled from your teammates. Of the 1,000s of emails, can a Product Owner see the System Administrator's email account? Further, which email is actionable? Finally, every week one's staff scrambles to update their organization's PowerPoint slides with fabricated data reflecting C-suite expectations instead of presenting real-time live data.

Figure 2. USB flash drives in desk

Keywords: Single Source of Truth, Executive Self-Service, Strategic Portfolio Management, Enterprise Agile Planning, Real-Time Visibility, Roadmap Alignment, Blockchain, Kill Excel

1 Problem

The world is one big data problem.
— Andrew McAfee, Co-director MIT Digital Economy Initiative.

Your data is critical to your business success. To that end, how long is your average data access time (DAT) [Table 1], especially on a firewalled multi-data center platform? In this case, how long (in seconds) would it take leadership to access program-critical data in the depths of multiple data oceans? Perhaps, a week or even a month's time might not be a shocking turnaround time. In this Time Point context, the delay would be Bαα2 to Bα1, in other words: Forever.

How long can your business model sustain discreet lengthy data transformations like this?

Data Access Time (DAT)	Second(s)	Time Point (TP)
1 ms	10^{-3}	Aαα
1 cs	10^{-2}	Aα1
1 ds	10^{-1}	Aα2
1 s	1	Aα3
1 m	60	A1
1 h	3600	A2
1 d	28800	Bαα1
1 w	144000	Bαα2
1 m	624000	Bα1
1 y	7488000	Bα2
@allanschwarb | linkedin.com/in/schwarb1		

Table 1. Average data access time (DAT)

You may not be in healthcare where HIPAA-data[iv] noncompliance can result in profound legal and financial consequences, however, your data spawl can significantly impact the viability of your enterprise as illustrated in microcosm [Figure 3].

Figure 3. Laptop repair center: Data sprawl?

Figure 4 represents just a few data tail-related questions?

Where did you find the data?
How old is this data element?
Is the dataset complete and accurate?
What version are we looking at?

Figure 4. Data questions

First, there is no way to execute data analytics without collocated programmatic data.

Second, only 34% of organizations are 'very confident' in their Data & Analytics (D&A) insights[v]. Further, perhaps one of the most pernicious issues is corporate culture.

A common laissez-faire data governance policy may not manage elements as described in Figure 5.

It is your data.
You can store it where and in whatever format you wish.
Only you are responsible to maintain your dataset.
Email? Good luck:
Cognizant of and
Converting to an actionable process.

Figure 5. Modern corporate data policy

Third, the above does not even include issues such as multiple SharePoints and numerous firewalls.

Finally, the human factor is clearly the most challenging as described in Figure 6[vi].

Implementing a new process and technology is perhaps the easier of the three. However, your People need to become believers in a SSOT solution. It is the responsibility of an SSOT-evangelist to inspire the team to develop and embrace technology and processes to win at all three.

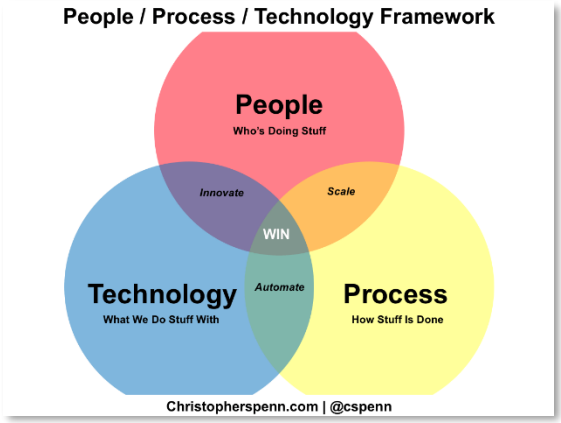

Figure 6. People are key to success

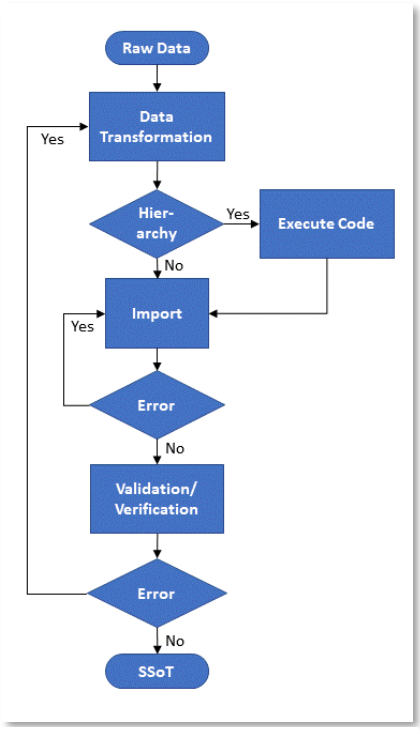

Figure 7. SSOT Process Flow

2 System Features

[The single source of truth] is a … holy grail[vii] that we've been trying to put together for a long time.
– Mark Benioff, CEO Salesforce.

This SSOT solution consists of several essential features including requirements management, real-time visibility and ability to ultimately execute DevSecOps via sprints. The requirements management phase consisted of developing data, establishing a hierarchical structure and validating the completed dataset as shown in Figure 7. Where data was not historically available or accessible from an archive, a new dataset was decomposed and developed from high level system requirements.

This facilitates capturing the opportunity value of managing the team's product's disseminated data at scale. That is, in a modern enterprise, a product's critical data is available at one's fingertips (i.e., executive self-service). The hierarchical single source of truth provides immediate access to time-critical data by the Development Team and leadership, at scale.

3 Related Work

The most valuable commodity I know of is information.
– Gordon Gekko, Wall Street.

My research and development (sandbox) involved developing and maturing tooling especially in the areas of hierarchy, custom fields and import syntax.

To achieve a functional hierarchy, a relative relationship was established within the dataset as shown in Figure 8. In this example, Level 2 represents a Parent and each Level 3, a child. In this case numerous children are identified. Developing the VBA code to format these levels so hierarchal links were achievable within the platform is a significant tooling accomplishment.

Level	Issue ID	Vertical	Issue Typ	Summa	ID	Input Da
2	1	Network	Epic	Global con	24.24.001	24-Oct-20
3	2	Network	Story	Remote te	24.24.002	24-Oct-20
3	3	Network	Story	Global con	24.24.003	24-Oct-20
3	4	Network	Story	Low-cost,	24.24.004	24-Oct-20
3	5	Network	Story	Flexible co	24.24.005	24-Oct-20
3	6	Network	Story	High speec	24.24.006	24-Oct-20
3	7	Network	Story	Low-cost,	24.24.007	24-Oct-20
3	8	Network	Story	Make sate	24.24.008	24-Oct-20
3	9	Network	Story	End users I	24.24.009	24-Oct-20

Figure 8. Hierarchical CSV data

Further R&D included developing custom fields, field configurations and field configuration schemes. This established the platform relationship to the columnar raw data. Some of the custom fields included checkboxes, radio button and select list (cascading) lists. Others consisted of date, URL and various text fields.

Finally, the syntax for each data type was matured to ensure successful migration.

4 System Anatomy

Consumer data will be the biggest differentiator in the next two to three years. Whoever unlocks the reams of data and uses it strategically will win.
– Angela Ahrendts, SVP Retail Apple.

First, I will provide a high level discussion of the architecture as seen in Figure 9. As discussed above, a lack of data governance and conventional technology has led to an unprecedented data sprawl antipattern. In my associated use case specification[viii], I identified the primary actors on a modern agile Development Team consisting of Product Owner, System Administrator and Scrum Master. Assuming, each of these individuals maintains an email “blackhole” and other data scattered across a global enterprise over decades, data is scattered.

In the flow of events, the basic flow first step is identifying that the team’s data is distributed.

The second step establishes the definition of system-critical and locates and/or develops this dataset. Some of the sources of data include: dependency emails, Microsoft Word, Excel, PowerPoint, PDF, custom BI reports, stand-alone enterprise systems (e.g., individual PC(s), servers, data centers, uncatalogued storage device(s), custom solutions, Microsoft Azure SharePoint(s), ticketing systems, test automation, CI/CD, cloud ops, custom fields, ALM(s)+bolt on(s), wiki(s), PPM Solution, time tracking,

roadmap tool(s), transformation assessments, etc.) Once defined, locate and capture the data within the system boundary sourced from across the enterprise.

Step three requires the technologist(s) to develop and mature the dataset per the system engineering parameters.

The fourth step consists of transferring all system-critical data from across the enterprise to (a.) within the system boundary as shown in Figure 9.

Step five establishes a hierarchy and includes acquiring/developing the enterprise dataset storage solution/code to order issues per system-driven hierarchy.

The next step ensures the data is clean, formatted and maps correctly. This enables dataset migration launch to the single source of truth platform (b.), also shown in Figure 9.

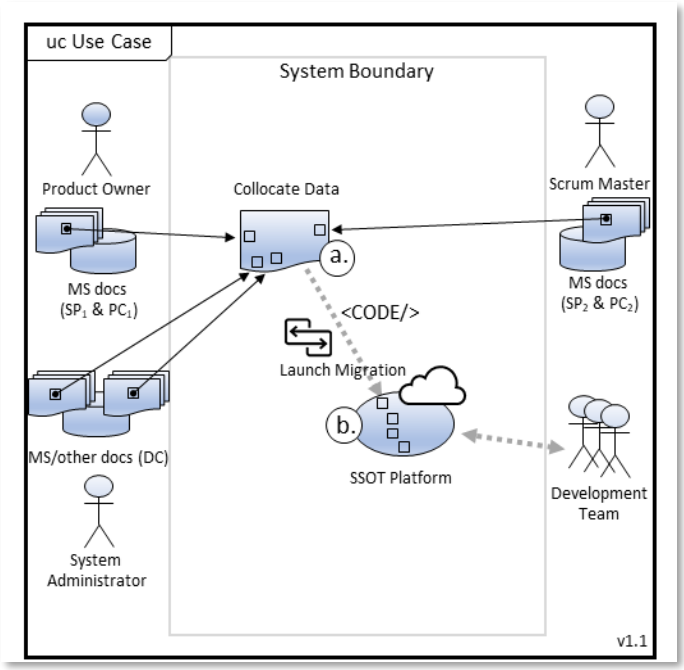

Figure 9. High Level SSOT Architecture

The last of the basic flow steps consists of the technologist(s) validating and verifying the hierarchical single source of truth dataset. At this point the use case ends.

Although there can be alternative flows, I have included these in Appendix A below.

5 Results and Performance

Data scientists are involved with gathering data, massaging it into a tractable form, making it tell its story, and presenting that story to others.
– Mike Loukides, Editor O'Reilly Media.

The most important measure of an SSOT platform is real-time visibility at scale of program-critical data. Figure 10 includes representative linked data after migration to a platform within the system boundary.

In this case, there are three levels represented including parent, child and grandchild of various issue types. This displays only a representation of thousands of system requirements integrated across multiple projects.

All stakeholders including end-client, Systems Engineering, DevOps, and third tier suppliers have access to system specifications and can execute accordingly. Which custom screens, workflows, boards and

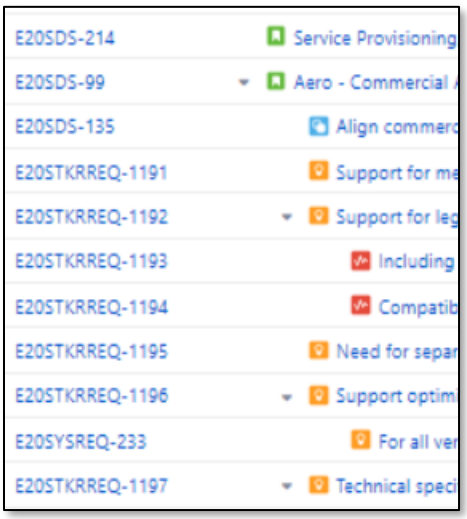

Figure 10. Hierarchical data platform

training in place, these executions may now simultaneously include quarterly program reviews (waterfall), hardware engineering (Kanban) and software DevOps (scrum).[ix]

The end-client is very positive and working to influence all stakeholders to "kill" their own Excel worksheet data which exists outside the established system boundary.

6 Conclusions

Big data is at the foundation of all the megatrends that are happening.
– Chris Lynch, American Writer of Books.

After significant R&D, developing this architecture and executing a digital transformation, a modern Development Team has now successfully transitioned from a distributed dataset to a single source of truth (SSOT) platform. This Development Team has 24/7 access to a collocated ordered dataset within the specified system boundary. This single source of truth platform was delivered within budgetary and time constraints.

Now decision-makers have the data they need at their fingertips including the following benefits:

- Entire Development Team knows where data is located.
- No longer incorrect data element version.
- Dataset is complete and accurate.
- Dataset version is current.
- Substantially reduced time spent identifying which recorded data is correct.
- Iteratively improved data intelligence capabilities of end-client.

Organizations seeking to achieve a single source of truth must consider execution time, complexity and have architectures that are scalable per their data specifications. By leveraging business intelligence and designing for requirements management, real-time visibility, ability to ultimately execute DevSecOps sprints and organizations' existing datasets they are positioned to actively develop and influence Development Teams to embrace these new platforms.

Finally, is your strategic initiative's average data access time consistently exceeding Aα3 (>1s)? What is the real opportunity cost to you and your enterprise?

7 Future Work

Establishing a single version of truth is about separating the strategic signal from the operating noise of your business.
– Brent Dykes, Entrepreneurs Contributor Forbes.

The following are areas of R&D where I plan to further develop my skills and develop the SSOT platform:

- Roadmap alignment – Continuing to execute from C-suite to DevSecOps team.
- Issue collectors – Gather feedback on any website in the form of issues, even for users without accounts.
- WebHooks – Develop listeners.
- Mail handlers – Added/edited via new incoming mail.
- LexoRank management – Determine if rebalance can be run; and actual rebalance of Rank field in database.
- Automation – Develop rules, audit log, library and usage
- Plugin/App – Build new app.
- Blockchain – Use for developing single source of truth platform.
- Document repository plan – Support end-client develop enterprise plan.
- Communication policy – Support end-client develop enterprise policy.
- Jira Align – Next EAP execution.

Vitae

Allan Schwarb[x] was born in Ft. Lauderdale, Florida and received a B.S. in Engineering at the University of Central Florida and an MBA at the Rollins College Crummer Graduate School of Business.

His digital strategy and transformations include:

- Orchestrated & launched 1st Agile tools on Boeing Global Systems (BGS) platform including Jira
- Launched custom DOD Jira instance including Zephyr, ServiceNow & Salesforce plugins
- Hands dirty launching Jira Align, Portfolio and Structure for Jira solutions
- Extensive DevOps Agile coaching & training international enterprise teams

His interests in technology includes:

- PUBLISHED Use Case Specification — Hierarchical Single Source of Truth Platform (Mo 10/12/2020)
- Fully integrated tactical & strategic understanding of Agile at Scale integration/visibility issues
- Diverse technology certifications earned including cloud & ML/AI technology transformation tools
- Accomplished using Visual Basic for Applications (VBA) for data transformation/migrations

8 Appendix A: Scalability

I have designed a single source of truth platform to be scalable in the near term to a goal of 60,000+ issues. As bandwidth, application, storage and intentional deprecation solutions develop over time, there is no reasonable upper limit to the scalability of a similar platform.

9 Appendix B: Use Case Specification Flow of Events | Alternative Flows

Some of the alternative flows related to Figure 9 include:

1.1 CANNOT LOCATE.
At basic flow step LOCATE DATA, the technologist(s) may be unable to locate historic/archived data related to system parameters. Therefore, the use case resumes at basic flow step DEVELOP DATA.

1.2 DISCOVER HEALTH.
At basic flow step COLLOCATE DATA, the data health may not be discoverable until basic flow step LAUNCH MIGRATION. Therefore, the use case may require an additional sub-step to test data health/syntax between the LAUNCH MIGRATION and the ultimate single source of truth platform. Thereafter, the use case resumes at ESTABLISH HIERARCHY.

1.3 CONFIRM MATURITY.
At basic flow step ESTABLISH HIERARCHY, first data must be ordered, verified and validated between each parent and child(ren) relationship(s). Next, code is developed per system parameters to ensure dataset migration is successful. Thereafter, the use case resumes at LAUNCH MIGRATION.

1.4 REVEAL SYNTAX.
At basic flow step LAUNCH MIGRATION, errors may be encountered if the single source of truth platform recognizes syntax issues. These errors must be corrected in the COLLOCATE DATA step prior to migration. Thereafter, the use case resumes at VALIDATE DATASET.

1.5 PLATFORM CREATED.
At basic flow step VALIDATE DATASET, the single source of truth platform has been created within the system boundary, is accessible to the Development Team and the use case ends.

10 Appendix C: Enterprise Agile Planning Applications

According to Gartner's Magic Quadrant in Figure 11, some of the Enterprise Agile Planning (EAP) tool leaders include:

- Atlassian Jira Align
- Broadcom Rally
- CollabNet VersionOne
- Planview
- Targetprocess

Figure 11. Gartner, Magic Quadrant Enterprise Agile Planning (1Q2020)

11 Appendix D: Strategic Portfolio Management Applications

Forrester presents Wave™ Strategic Portfolio Management applications in Figure 12. Those with stronger strategy and larger market presence include:

- Atlassian
- BMC Software
- Cherwell Software
- IBM
- Microfocus
- ServiceNow
- TOPdesk
- USU Software

Figure 12. Forrester Wave™: Strategic Portfolio Management Agile Organizations (Q4 2019)

References

- [Atlassian 20] "Unlock the agility of your enterprise," Atlassian.com, San Francisco, 2020.
- [Bronstein 19], Bronstein, Molly, "Atlassian leads the pack again in strategy on the Forrester Enterprise Service Management report," Atlassian.com, San Francisco, 2019.
- [Jones 20] Jones, Capers, "Allan Schwarb | Use Case Specification | Hierarchical Single Source of Truth Platform_v1.1 | Agile Use Case," email, Siesta Key, 2020.
- [Kaufman 20], Kaufman, Barak and Merson, Phil, "3 ways to contain SaaS sprawl and spend in a post COVID-19 world," blog.shi.com, New York City, 2020.
- [Kudva 16] Kudva, Santosh, "The Difference Between System of Record and Source of Truth," LinkedIn.com, Atlanta, March 31, 2016.
- [Mendes 16] Mendes, Natalie, "Establishing a single source of truth across the product organization," Atlassian.com, San Francisco, May 5, 2016.

---

[i] [Schwarb 20] Schwarb, Allan, "Use Case Specification — Hierarchical Single Source of Truth Platform," St. Louis, October 12, 2020.

[ii] [Sammet 00] Sammet, Jean E., "Programming languages: History and Fundamentals," Virginia Polytechnic Institute and State University, Blacksburg, March 22, 2000.

[iii] [Desjardins 19] Desjardins, Jeff, "How much data is generated each day?," World Economic Forum, Geneva, April 17, 2019.

[iv] [Bond 20] Bond, Robert, "The 5 Elements to Effectively Managing HIPAA Compliance," secureops.com, Québec, June 16, 2020.

[v] [Olavsrud 16] Olavsrud, Thor, "Executives still mistrust insights from data and analytics," CIO.com, New York City, November 2, 2016.

[vi] [Penn 18] Penn, Christopher S., "Transforming People, Process, and Technology, Part 1," ChristopherSPenn.com, Framingham, January 10, 2018.

[vii] [Salesforce 20] "Build a single source of truth for your customers," Salesforce.com, San Francisco, 2020.

[viii] [Schwarb 20] Schwarb, Allan, "Use Case Specification — Hierarchical Single Source of Truth Platform," St. Louis, October 12, 2020.

[ix] [Rehkopf 20], Rehkopf, Max, "Kanban vs. Scrum," Atlassian.com, San Francisco, 2020.

[x] [Schwarb 20], Schwarb, Allan, "Profile," LinkedIn.com, St. Louis, 2020.