Fixing Data

Executive owner(s): Andi Karaboutis, Group CIDO
Name of forum: Group Executive Meeting May 2020 | Report Date: April 28th, 2020

EXECUTIVE SUMMARY

CONTEXT

The current state of Data at National Grid poses significant challenges for the company. It exposes us to security risks, has already impacted our major programs, and necessitates costly manual efforts to deliver our daily operational services. "Fixing Data" holistically (and no longer as one-off efforts) is imperative and urgent. This paper seeks to build on our agreed actions to hire a Chief Data Officer, continue to improve our Data Governance & Stewardship building off the progress of the Data BMS, and proposes additional steps to "Fix Data." We would appreciate your feedback and commitment to arrive at a clear, unified, and decisive path forward.

QUESTION

1. What is our Data problem and what is it costing us?
2. What are the components to "Fixing Data"?
3. What steps should we agree on to rapidly and effectively "Fix Data"?

CONCLUSION

Deep dives into the domains of Employee and Customer Data demonstrate a substantial problem, and are illustrative of similar problems across all major Data domains. Unless we "Fix Data," our security, major programs, and financial and operational performance are at risk. In addition, clean, organized Data is like a common language and underpins business analytics and Digital Transformation. To overcome our Data problem, the key components are (i) building upon the progress of Data BMS to improve capabilities of Data Governance & Stewardship; (ii) implementing standardized Data Tools & Platforms; and (iii) driving Master Data Solutions in each major domain (Employee, Customer, Asset, Product, Vendor, and Finance). To ensure rapid action and coordination, it is imperative to agree on a "Hub-and-Spoke" operating model, an owner to drive Master Data for each major domain, and a holistic funding model.

WHAT IS OUR DATA PROBLEM AND WHAT IS IT COSTING US?

Based on recent work by Boston Consulting Group, National Grid rates "low" or "very low" across the three major components of Data (Master Data Solutions, Data Governance, and Data Tools & Platforms). BCG's study, that looked into a subset of the Data domains at National Grid, estimated **the costs of bad Data are $73-98M**. Given the limited scope of their review, they concluded that the full cost is likely much higher."[1]

[1] *Analysis of the Current State of Data at National Grid by BCG dated April 7th, 2020.*

Diagram 1: Current High-Level Assessment by BCG of Data Components at National Grid

Components	Current Maturity Level (1 = Low 5 = High)	Details
Master Data Solutions	Very Low (1 of 5)	• ***Master Data generally does not exist and has not been defined.*** *Data* is stored in multiple disparate places causing issues • ***Lack of Operating Model and Funding Model*** *to drive sustainable enterprise approach to meet business needs for Data* • ***Data and the Business Processes that use it are generally not clearly designed together*** *- instead Programs focus on technology implementation without clarity of end-to-end data/process needs*

Data Governance	Low (2 of 5)	• ***Data BMS is an important step in the right direction****. Some areas demonstrating adoption.*
		• ***Major gaps remain, including****:*
		  • *Lack of strong Data processes, methodologies and documentation on "how" to adopt Data BMS*
		  • *Lack of capabilities to map and design such processes.*
		  • *Lack of organizational understanding of Data as an Asset*

Data Tools & Capabilities	Very Low (1 of 5)	• ***Lack of enterprise platforms and tools to manage Data*** *results in different approaches within each Program, maintaining siloes* • ***Lack of Data Professionals*** *to sufficiently manage our Data, e.g., Data Architecture, Data Engineering, DataOps*

1. Deep Dive: Employee Data Domain

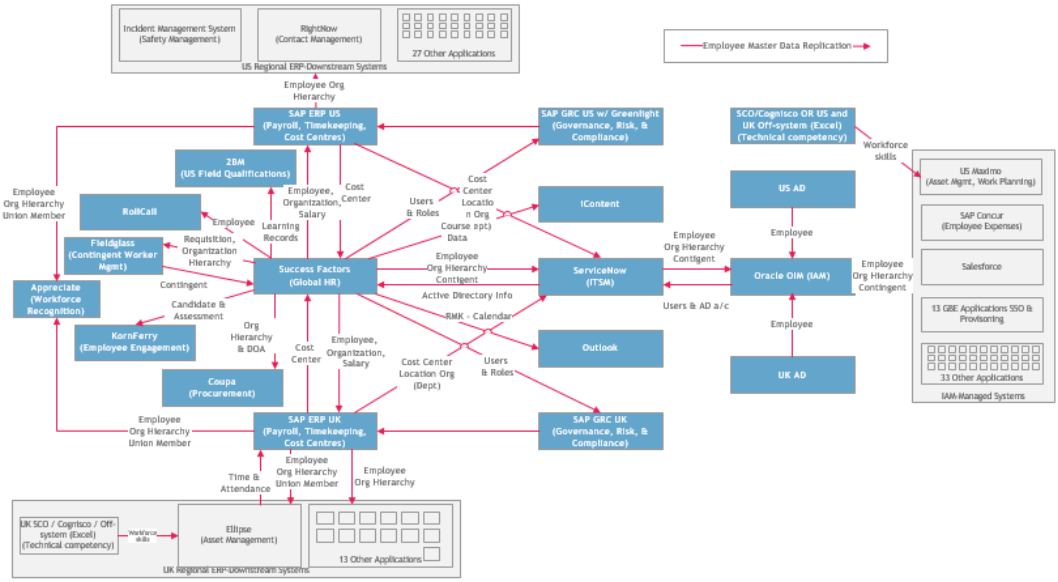
Employee Data (e.g. employee name, personal data, etc.) is at the core of several key end-to-end business processes, and is used by over 100 applications at National Grid. Per the discussion at the December Group Exec, a high-level assessment of the Employee Data domain has been completed. Results show critical deficiencies that require high-cost, high-risk manual interventions to deliver core business capabilities. Business capabilities affected include but are not limited to (i) securely managing user access; (ii) determining location of resources (to optimize scheduling for GBE and other systems, as well as overall facility utilization / cost, etc.); (iii) performing total workforce management and lowering costs (i.e. driving optimal skill and efficiency outcomes), including decreasing dependencies on contractors and third-parties, measuring the effectiveness of different roles, and driving best cost for commoditized skillsets; and (iv) optimizing spans and layers.

The data deficiencies include:

- No Employee Data Master / "single source of truth." Employee Data is being replicated across more than 40 systems (see diagram below) without good organization, hygiene, and controls.
- These myriad systems handle Employee Data in inconsistent ways, even at times overwriting critical data in another system. The result is a tremendous amount of manual work by HR, Finance, and Business Services to attempt to keep this data accurate, often with limited success. Those responsible to clean Data complain of "too many cooks in the kitchen."
- Programs (like MyHub) are developed in a siloed manner, without a full understanding of how employee Data is used by other systems. The result is major cross-application deficiencies in data availability and quality issues. Examples include:
  - Multiple inconsistent employee organizational hierarchies. There is a disconnect between org hierarchies (in MyHub) and cost centre hierarchies (in SAP) that causes massive Data differences and requires extensive manual clean-ups.
  - Employee types (e.g. full time, contractor, contingent) with different definitions across systems. There is no overarching definition to drive consistent reporting and analysis, leading to missing Data in reports, different outputs for management, and confusion.

- Employee attributes (e.g., skills) with different definitions across systems (e.g., SuccessFactors does not contain information about 'skills.'). Missing data is often vital for some downstream systems and business capabilities, such as 'skills' data being required for field force job scheduling. As another example, employee location can be set to 'MASSHAMP' – requiring downstream manual re-work for payroll to process.

Diagram 2: Replication of Employee Data in over 40 Systems



2. Deep Dive: Customer Data Domain

At National Grid, simply put, we cannot and will not be able to provide an outstanding level of customer service until our Data issues within the Customer Data Domain are fixed. Why?

Because Customer Master Data is divided across 17 systems in the US (versus the target end state of only 3 systems[2]). Recent work by IBM revealed that:

- There are 227 business capabilities at National Grid that make up the delivery of customer service and operations.
- 77.5% (176 of the 227) are accessing Data in an inconsistent or inaccurate manner.
- 65.6% (149 of the 227) pull data from multiple sources, meaning a high likelihood of inaccurate and incomplete information to serve the customers.

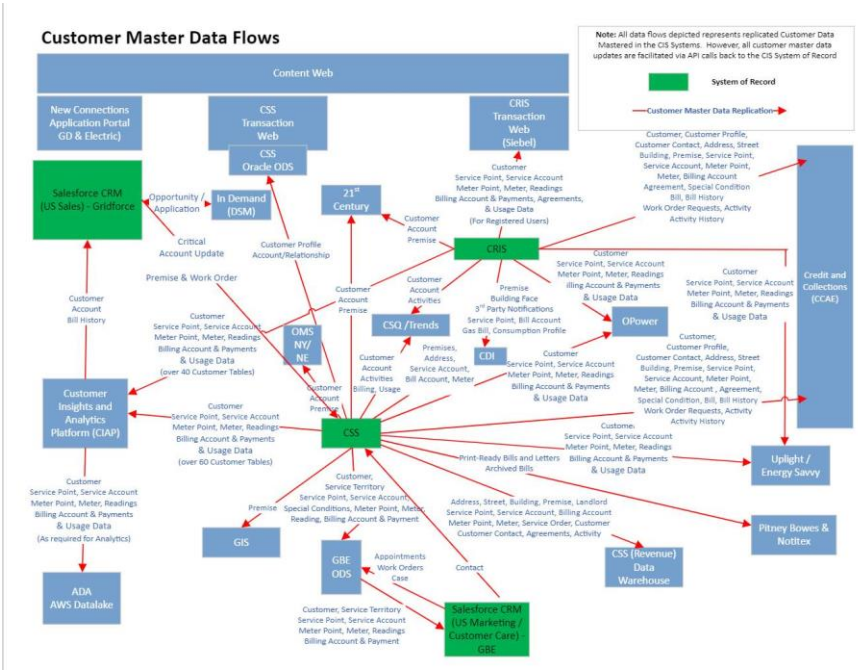
What does that mean in practical terms:

- We cannot speak to customers with one voice across the customer lifecycle, because each team (sales, customer service, marketing) has different information.
- We cannot effectively market our products to customers, because:
  - there are multiple, conflicting sets of product information, so that if the same customer contacts us, products may be explained in differing ways;
  - there are multiple sets of contact information, meaning outbound marketing is costly / inefficient, and contact information is often not accurate.
- We cannot offer customers the ability to contact us in the manner of their choice, because unified customer data is not available across all communication platforms.

- We cannot customize the customer experience, because we lack the data and advanced analytics to tailor offerings and communications to customer preferences.

Finally, the issue of poor Customer Data extends beyond the Customer domain. Programs like GBE also use Customer Data. Both programs are attacking parts of the Customer Data problem in different, sometimes conflicting, and duplicative ways, driving up costs and risks for both programs.

Diagram 3: Customer Data in the US spread across over 17 systems



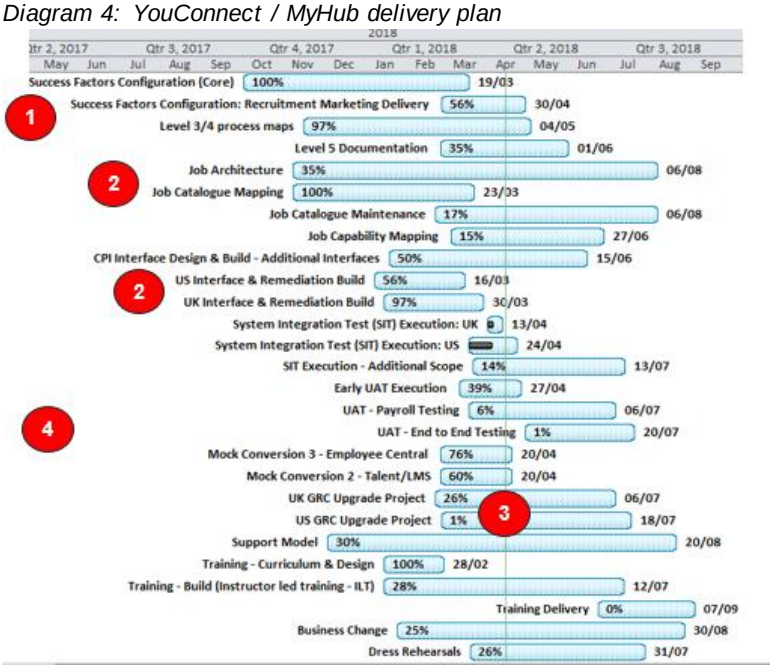
WHAT ARE THE COMPONENTS TO "FIXING DATA"?

"Fixing Data" is not currently being done at the quality and pace required to meet our objectives (and large programs are being reset or failing), because we do not have a common framework, operating model, and funding model.

Programs are driven by implementing "software applications" and meeting release dates, and not being led by driving business outcomes, redesigning processes and fixing data at a domain level.

A good example of this "software-application-centric" approach is evidenced by an early version of the YouConnect / MyHub delivery plan.

Diagram 4: YouConnect / MyHub delivery plan



Concerns:

1. The project appears to start with configuring SuccessFactors, without sufficient planning work. There is no "future business" visioning to define desired business outcomes. Rather, it's a "software is the solution" vision.
2. There are no tasks related to process redesign nor fixing data. Without clearly desired business outcomes and processes, even the job architecture was not done properly. For example, the program did not create a Job Role level, which has caused major issues for access and identity management as well as an inability to understand performance and efficiency by role.
3. It appears too many activities are being run in parallel with a "Big Bang" approach, versus layering into segregated deliverables of Minimum Viable Product ("MVP").
4. There appears to be no work to determine interactions with other business processes and systems.

…all of which has resulted in substantial data and downstream issues caused by the Employee Domain, as previously discussed. Below is one more example from December 2019 (from the Gas Ops team):

All,

It has come to our attention that there was an issue when changes were made to the HR hierarchy for Gas Operations on Thursday December 12th.

Issue: All the Rhode Island MA Electric employees were inactivated in WFS – this means their time completed in Salesforce did not get sent to Workforce Software and therefore could never be sent to ATD. Because of this SAP Hierarchy change, vehicles and overtime were not captured, unplanned absences couldn't be reported, and future absences couldn't be requested.

To addresses these issues, we need three key components to "Fix Data":

1) **Data Governance & Stewardship:** Using the Employee Data Domain example, HR should own and steward this data and determine:
   - What comprises an employee or person record and what are the attributes (e.g., a "person" record = employees, contractors, and third-party workers (who access our locations or applications. Attributes include name, address, role, location,

department, Inland Revenue Number or SSN, etc. Not everyone should be able to see or access all attributes.)?
- Who can "create" employee data (HR only, supervisors, managers, business services?)?
- More specifically, who exactly should have permission to create, replicate, update, delete ("CRUD")? When this ownership is not overtly defined, we end up in situations like the one where an automated process in GBE regularly overwrites part of the HR org hierarchy, requiring constant manual reworks by the HR team.
- More specifically and technically, can we define the desired business outcomes, Data processes, Data uses?

NOTE: Data BMS has rapidly progressed this area, raising awareness and driving ownership. However, there are still critical gaps, as Data owners are often without experience and the required professional capabilities and/or are operating in silo's.

2) **Data Tools & Capabilities**
   - What are our standardized tools and platforms based on the business processes and uses, so that we can have one, common approach to enterprise Data solutions? IT needs to accelerate the roll-out of these tools.
3) **Master Data Solutions**
   - For each domain, do we have a **"single source of truth" (a.k.a. Data Master)**, so that multiple applications can go to one location for the definitive and correct Data?

WHAT STEPS SHOULD WE AGREE ON TO RAPIDLY AND EFFECTIVELY "FIX DATA"?

To ensure rapid action and cross-functional coordination to "Fix Data", **it is imperative that we agree to:**

1) Build upon the good work of the Data BMS, by **agreeing to a "Hub-and-Spoke" operating model** (similar to Digital) wherein:
   - Hub: IT defines the enterprise Tools & Platform (no shadow purchases by programs or BUs) for secure Data storage, cataloguing, cleansing, and robotic process automation ("RPA") – currently NG teams have acquired 3 different RPA tools that we need to standardize to one
   - Hub: IT defines and assesses required capability levels for Data Stewards, who reside in the Business
   - Spoke: each Business not only owns Data Governance & Stewardship, but ensures highly capable resources who meet IT standards
   - Spoke: For analytics, each BU owns their own analytical / reporting teams (but use the defined enterprise Tools & Platforms)
2) **Define an owner to drive implementation of Master Data for each major Data domain**:
   - Employee: Andy Doyle
   - Customer: Greg Knight (US) & Chris Bennett (UK)
   - Finance: Andy Agg
   - Vendor: Vivienne Bracken
   - Asset: Electric = David Wright (UK) & Chris Kelly (US); Gas = Phil Sheppard (UK) & Cordi O'Hara (US) Transmission = Rudolph Wynter (US)
   - Product: Owners to be defined

3) **Agree to a holistic Funding Model, so Master Data solutions can be fully funded** (vs. partial one-offs embedded in programs). Appoint Charles Zentay [VP of Digital and IT Strategy] (working with Finance and the respective Data Domain owners above) to lead effort to report back to Group Exec next month on current vs. required funding for the Data Domains for Employee (Global) and Customer(US), with other domains to follow shortly thereafter.