**Context:** We are standing up elements to “Fix Data” at National Grid and are seeking to achieve consensus on the nature of the “Data Problem” and endorsement of a comprehensive approach to solve the problem. A number of aspects of this solution have begun, and to keep up the momentum, it is important that we agree on a unified approach, including the cross-functional collaboration and a strategy to fund.

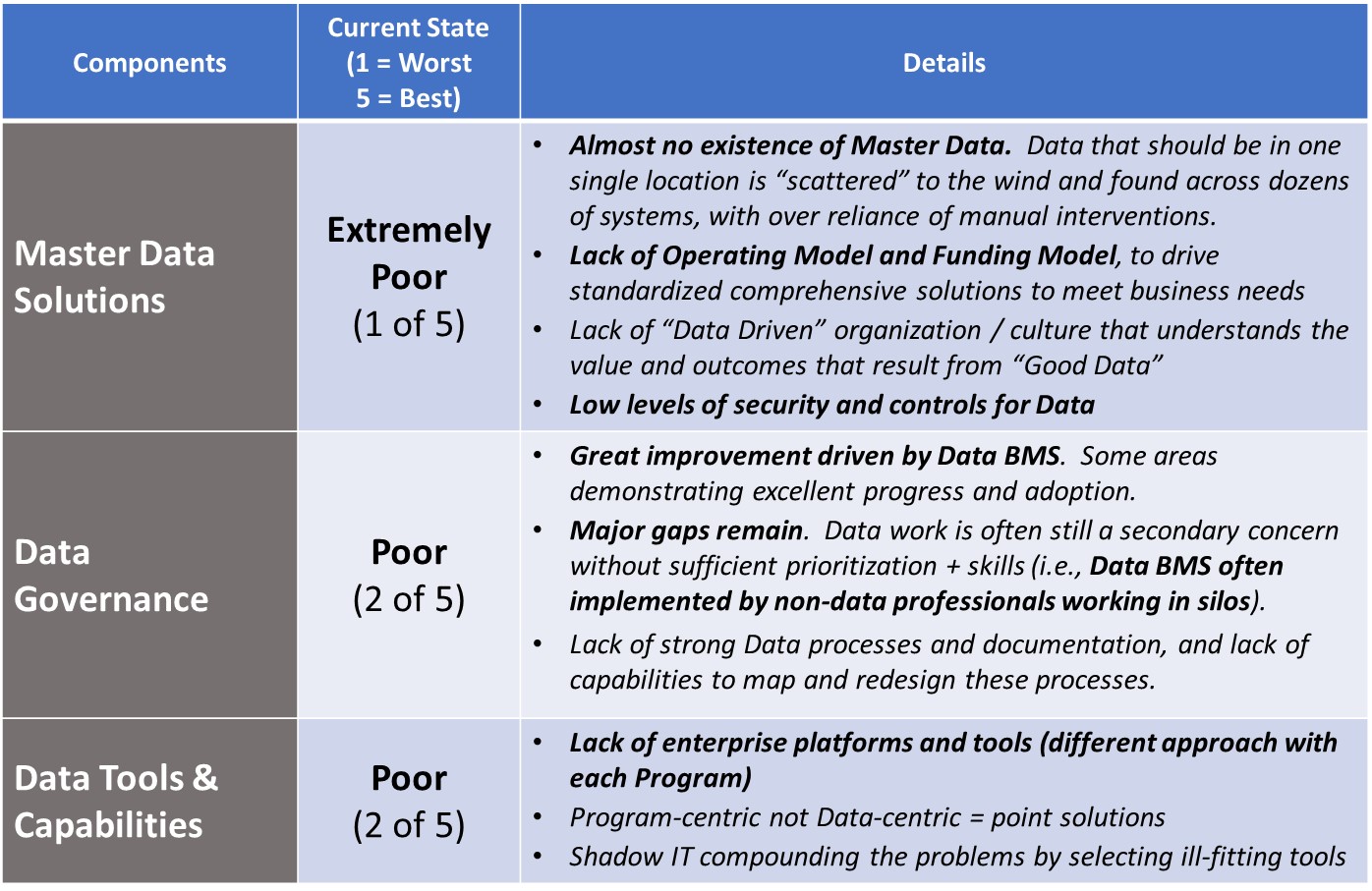
**Questions:**

1. **What is the nature of the “Data Problem” at National Grid? How does that impact a given domain (e.g. Employee Data Domain)?**
2. **What are the shortcomings with how National Grid addresses data in major programs today?**
3. **What is the value that is lost, and risk that is increased, by not addressing our data management issues across our data domains?**
4. **How can we rapidly remediate in-flight programs to rapidly achieve the best data results? And how do we build an enduring organizational capability to continually and sustainably “Fix Data” and leverage “Data as a Strategic Asset”?**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

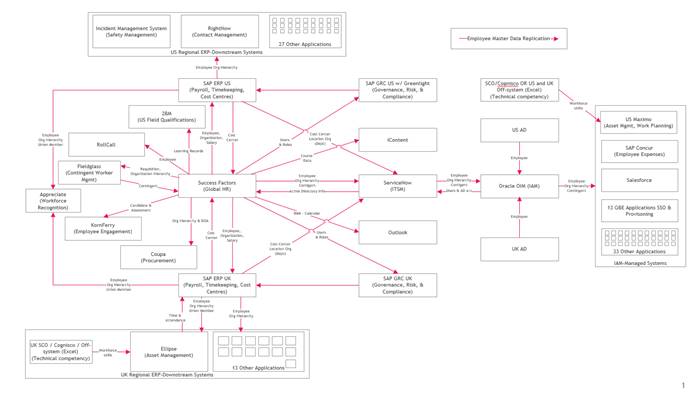
1. **What is the nature of the “Data Problem” at National Grid? How does that impact a given domain (e.g. Employee Data Domain)?**

*High-level assessment of Data Components at National Grid*



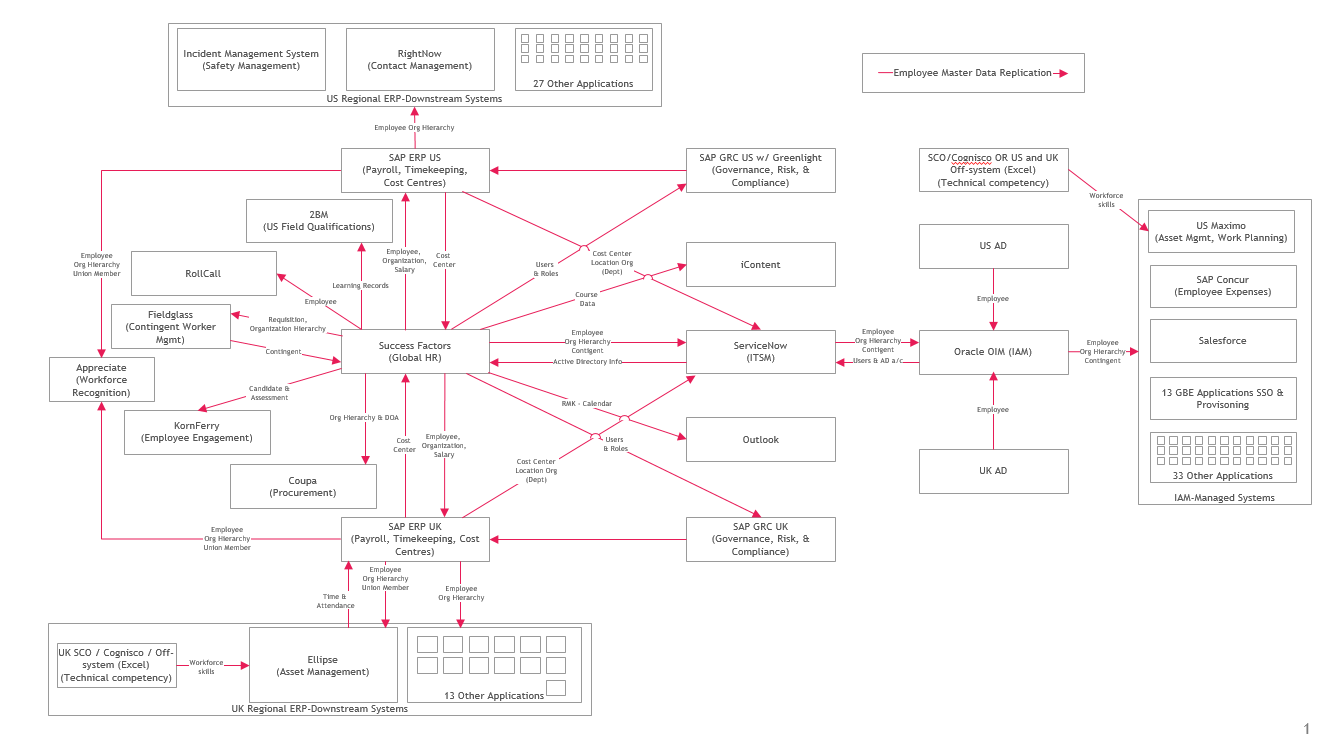
* There is consensus across National Grid that there is a significant problem with data quality and accessibility, which has a substantial negative impact on efficiency/productivity, quality of decision making, and level of risk. This problem exists across all six data domains (employee, asset, customer, vendor, finance, and legal) and is expressed in a number of ways (inaccurate or missing data, manual steps to aggregate data, etc.).
* However, there is less consensus and understanding around the root causes of our data issues and the resulting difficulties this creates, both for our “day to day” activities and for our ability to successfully implement large programs (and realize their benefits).
* Example of how this applies for a given domain (Employee Data Domain)
* Employee data (e.g. employee name, personal data, etc.) is a vital asset of the company, is at the core to several of our key end-to-end business processes, and is used by 100+ applications at National Grid
* Five root causes are driving the data problem within the Employee Data Domain (and are similar across other data domains as well):
  1. Solutions are developed without a full understanding of how employee data is used across the organization: Applications (like MyHub) are often developed with a siloed view regarding employee data (e.g. focused on employee data attributes relevant for a specific project or application without fully understanding how employee data is used for other applications)
  2. There are many sources of partial truths (vs. one single source of truth): We’ve deployed [40+] applications that each handle employee data in inconsistent or competing ways. Data is not stored in a single location with a standardized definition.
  3. There is a lack of people capabilities and skillsets to sufficiently fix and manage our data: This problem manifests both in limited number of Data Professionals (data architects, data engineers) within National Grid and lack of sufficient data capability, skills, and understanding across the company. In addition, we keep standing up resources within programs and then collapsing that capability without building an enduring capability. Finally, the company has not yet stood up an enterprise platform and toolset to support its data capabilities.
  4. We lack a unified, agreed-upon operating model of “who does what” regarding data at a business unit, program, and enterprise levels: Rather, we have a patchwork of immature capabilities scattered throughout the organization, with business units and programs at different levels of capability and taking different approaches to fix data.
  5. We have gaps to implement and maintain our data standards: While we’ve articulated data standards and governance principles, there are gaps to implement and maintain these standards (lack of process clarity, ways of working, and tools)
* The above root causes result in Grid being challenged to perform basic employee related tasks, such as:
  + Inability to perform total workforce management (to drive optimal skill and efficiency outcomes)
  + Inability to appropriately manage user access and control risk
  + Payroll and tax issues (e.g. accuracy, timing) across the US and UK
  + Limited ability to determine mix between full time employees, contractors, and vendors at National Grid locations (to “in-skill” / drive down contractor and 3rd party dependencies)
  + Inability to accurately determine location of resources (required to optimize scheduling (e.g. GBE, etc.), facility load factors, etc.)
  + Inability to quickly understand spans and layers for any given part of the organization
* Technical explanation of how this impacts a given domain (Employee Data Domain)
* Our systems/applications do not share a single coherent definition of employee master data, in some cases operating with “competing” definitions:
  + Multiple organizational hierarchies exist that are not consistent, which result in inconsistencies around employee reporting relationships flowing into multiple systems
  + Employee types (e.g. full time, contractor, contingent) may exist in one system, but not others; and furthermore have different definitions and naming conventions across systems
  + Employee attributes (e.g. skills) are not defined or treated consistently *across* systems
    - (e.g. SuccessFactors has learning and development activity, but does not have ‘skills’ as an attribute which is important for some applications, such as field force job scheduling)
  + Employee attributes that are not fit for purpose or clearly defined *within* any one system
    - (e.g. employee location can be set to ‘MASSHAMP’ – requiring downstream manual re-work for payroll to process)

See Exhibit 1 for an illustrative diagram of how Employee “Master” Data is replicated across various systems.



* Our typical fix in absence of clear employee domain master data and data model definitions has been to build translation mechanisms between applications – some are automated (but often error prone); others are highly manual (and thus also error prone):
  + Translating different org hierarchies for US and UK ECC instances into a single SuccessFactors instance to drive supporting system workflows has been done ‘automatically’ with code, but many errors of business logic occur in practice.
    - (e.g. a manager will get a request to approve IT equipment for a new hire that he/she does *not* have in his/her group...which results in wasted time by multiple parties to track down and resolve how to approve the request and fix the underlying org hierarchy)
  + Translation of contractor information from Fieldglass into SuccessFactors is manual and one-way. If a contractor’s status changes in Fieldglass, then as a result errors may exist in SuccessFactors
  + Manual data file export and modification is required to create the reporting dashboard for monthly report to Function and Department managers on different employee statistics
  + Manual processes are required to translate cost centers to HR organization hierarchy, creating risk of reporting errors
* Lack of capabilities/skill sets of data organization, and lack of clear operating model, has resulted in Grid employees exacerbating data challenges:
  + Having formal roles like Data Domain Owner with responsibility on paper for ensuring they hold up data standards for their domains, without having the skills and knowledge and guidance to do this properly.
  + Attempting to solve for localized data model and master data gaps using resources who lack the right skillsets and decision rights (e.g. Finance leveraging Business Services to exercise Data Design authority for finance domain)
  + Developing their own ‘manual’ workarounds as users of our systems of record to address challenges, further complicating the data (problem e.g. Creating duplicate records of the same position or employee to expedite hiring processes or fix short-term hiring pains, which results in a proliferation of non-existent “positions” in the organization)

Exhibit 1: Multiple replications and uses across systems of Employee “Master” Data



1. **What are the shortcomings with how National Grid addresses data in major programs today?**

* Our major programs address data problems in silos (because of program scope that is limited to a single application or organization). They often do not solve root causes and inadvertently exacerbate our data problems and operational complexity for other applications / organizations.
* Programs also develop data solutions in different ways, at different levels of maturity, and using different capabilities, architectures and tools. The result increased risk of program failure, budget overruns, and loss of benefit realization, as well as increases in post-implementation process complexity and maintenance costs. There are four drivers of this problem:

1. Programs tend to be technology centric, without sufficient consideration for process changes or data changes required.
   * For example, MyFinance is largely about migration to SAP’s newer S4/HANA architecture, while important process and data management topics like master data management, data mgt operating model and governance, or data quality management solution, are not fully addressed in the project scope/delivery
   * For example, MyHub was focused entirely around the implementation of SAP’s SuccessFactors as a software solution, without consideration of how to properly design master data or how to adopt new business process, data, and new technology together
   * For example, EA analysis of the GBE program indicates that while high-level technology integration approach was well considered and documented, a lack of clear approach to data governance and management approach has created considerable risk
2. Program design does not fully consider the level of data maturity for domains to which the program connects or on which it relies, nor how/when master data will evolve as the domain matures
   * For example, GBE is heavily reliant on the quality of data from Employee domain (along with Customer and Asset domain) for building solutions like field force scheduling/optimization – but the data quality is not available. So, while an optimization model for field force deployment has been developed, the model remains unused across the business because the data required is of low maturity/quality.
   * For example, MyFinance is reliant on Employee, Organization, and Vendor data that is mastered in systems not within the scope of the program. While MyFinance can deliver on its scope successfully, errors in these other data sources can still negatively impact the data quality within MyFinance’s scope
3. Where they do address master data model for a given domain, they do not consider all the ways that data is used at National Grid and the frequency by which that data changes. Examples include:
   * For example, Employee skillsets are crucial for many field management applications “downstream” from SuccessFactors. However, data about skillsets, and their changes over time, were not considered in MyHub despite their importance to Employee domain.
4. Programs are not scoped nor required to improve root cause data issues
   * For example: MyHub scope did not include creation of true Employee master data
   * For example: EA analysis of MyFinance scope indicates that MyFinance does not cover all areas of Finance Data domain change required
5. **What is the value that is lost, and risk that is increased, by not addressing our data management issues across our data domains?**

* Implications to National Grid value associated with our data management broadly fit into three buckets. Several examples of value loss are included below, and these examples alone sum to an estimated $73 - 98M cost of poor data management. As these are only examples, the full cost to Grid across the whole enterprise is likely much higher**:**
* **Direct cash costs:** Incremental cash costs (opex and capex) associated related to program ‘cost to achieve (CTA)’ and operational ‘run the business (RTB) costs’.  Program CTA occurs through additional employee, contractor, and system/tool required to address data complexity for each domain.  Because National Grid does not take a consistent approach to solve data complexity for each program – there is not scale benefit achieved across programs.  RTB costs occur through incremental employee costs to manage manual data workarounds and/or fixes, incremental IT support (both employees, contractors, and licenses) required to support point solutions developed by each program.
  + Recent or current examples
    - $12M planned for data clean-up costs to resolve issues for Grid Mod program
    - $3-5M to resolve data issues left over from the original MyHub program
    - $1.5M on data clean-up in support of MyFinance program
    - $1.5-2M per year on manual data clean-up for GBE program
    - $1-1.5M per year for manual Finance data processing
    - $5-8M per year on manual syncing of data for Customer, HR, Finance systems
* **Lack of benefits realization:** Due to our data complexities, we are not able to realize the full benefits from our various programs, which materializes through: 1) Inability to achieve planned efficiency because Programs either to not fix problems process/domain wide, or “shift” data problems outside of their defined scope into other parts of business processes and data domains; 2) inability to achieve planned efficiency because solutions (e.g. predictive algorithms, visualization, etc.) are not correct, driving inefficiency amongst workforce; 3) inability to apply solutions to the entirety of National Grid because of data incompatibility (e.g. customer solution applicable to CSS, but not CRSS; or solution applicable to UK but not US); 4) Rapidly evolving external energy landscape changes and our solution design is not nimble enough to change with the market; and 5) many of our best minds (e.g. in data science) are preoccupied with doing rudimentary data clean up vs. spending time creating innovative solutions that drive value for National Grid
  + Recent or current examples
    - $40-50M of annual benefit delayed by at least one year due to data problems affecting GBE
    - $5-9M of lost value from Workforce Dispatching and MyAccount programs in US Electric business
    - $3-7M lost opportunity of higher satisfaction in UK Customer Satisfaction incentive
    - $1-2M of Data Scientist time being spent on remedial data cleaning rather than Data Science
    - Increased risk that National Grid will fail to make the critical leap into a “modernized grid” in which data and data/analytics/digital capability are critical strategic assets
* **Risk to the business:** Modernization of utility sector (e.g. grid modernization, smart devices in front and behind the meter, proliferation of distributed energy resources, etc.) is driving exponential data growth, ~10x terabytes increase over last 5 years for typical utilities.  It will only increase in the decade ahead.
  + This data provides an opportunity to transform operations – but also poses risks to utility if data is not managed correctly.   There is going to be increased accountability around incidents – where the utility will be held to account if either: 1) data existed to prevent incident, but utility did not use data for pre-emptive action or 2) data was used incorrectly by the utility, resulting in a non-optimal operation decision.  This is currently the case with PG&E’s wildfire response in California.
  + Data will also enable greater competition to our utility business by companies who are nimbler to extract insights deliver better outcomes more efficiently for customers
  + Regulators are increasing accountability in utility performance – beginning to request ‘productivity’ metrics around approved investments, which will require a more comprehensive understanding of fiscal spend and output achieved.

1. How can we rapidly remediate in-flight programs to rapidly achieve the best data results? And how do we build an enduring organizational capability to continually and sustainably “Fix Data” and leverage “Data as a Strategic Asset”?

* Going forward, we have three recommendations for the Group Exec:
  1. Complete the filling of the data leadership and governance roles agreed to at the December 2019 Group Exec meeting:
     + Chief Data Officer (CDO) & Data Domain Owners: Data leadership accountable to develop and implement our data strategy
     + Data Governance Council: UK and US councils to support CDO enterprise data approach across organizational and program entity boundaries
  2. Build upon the December 2019 actions with the rapid development of a full set of data capabilities, tools, and standards that ensure rapid remediation of inflight programs, including:

Capabilities

Data architecture:

* + - Business data architecture
    - Data & information architecture
    - Integration architecture

Data services:

* + - Master Data Management
    - Data management system administration
    - Database administrators
    - Product service ownership (broader capability than just data)
    - API development
    - Platform support (post go-live maintenance function)

Data analytics:

* + - Data visualization
    - Report design
    - Data science

Data governance (at a program level):

* Data structure, talent, and governance assurance (pro-active)

Tools

Purchase of 5 tools to enable:

* + Enterprise Data Architecture and Modeling
  + Master Data Management
  + Data Cataloging
  + Data Integration (ETL & Data Orchestration)
  + Data Quality Management
  1. Launch a program to triage and remedy our enterprise master data model and root cause issues over FY2021, taking a domain by domain approach with a business value and business process orientation, coordinated across all of our major programs – with two specific outcomes
* High quality and well-managed data that is truly an asset for National Grid
  + Data across all our Domains well-managed as a “single source of truth” for operational, analytical, and business purposes.
  + Data Platform(s) and technology that allow us to make full use of our Data to drive business value and to mitigate risks to the enterprise
* Establish rigorous Data Management as part of our daily operations
  + Build the 3 capabilities listed above (business architecture, future state process design and change, program data management)
  + Organizational ways of working across these 3 and our other capabilities to ensure rigorous Data Management across the Enterprise

**Appendix:**

**Proposed approach for “Data Triage and Remedy Program”**

Goal is to develop a domain centric approach. For each domain, run three phases of agile sprints, shown below. Underlying the full approach are a set of key principles?

1. **Phase 1: Rapid design & visibility value POC (Hack a Future approach)**

Outcome is a practical design for master data implementation and build of a POC for a solution that can be delivered in an iterative, agile fashion. The POC must demonstrate clear business value available and drive business and IT engagement.

1. Preliminary work:
   1. Align on the **priority domain** for focus – Employee domain for illustration here.
   2. Align on the **technology implementation approach** (iterative throughout Phase 1)
2. Refine the Employee Domain Data map to **support 2-3 end to end representative temporal journeys for the employee domain**. E.g., onboarding journey, learning and development, etc. while driving engagement and buy-in from business units
   1. The journeys will be selected based on:
      1. Maximization of business value especially in context of ongoing Programs
      2. Feasibly of implementation of business process change, technology and data
      3. Spread and impact to data model (variety of fields and attributes)
   2. Identify the systems and interfaces, primary logic principles and temporal flow
   3. Engage the business to document process flows, the information architecture and existing primary business processes.
   4. Map the primary business processes (activities, roles) associated with the domain
   5. Define the 7-15 attributes that will drive most short term value
   6. Evaluate the spread of distribution for key attributes across systems today
3. Design a v1.0 of a comprehensive data model for the employee domain
   1. Complete list of fields and key parameters including systems of record source, relationships, “ice/water/steam” view of change frequency, field usage
4. Create POC data platform to allow rapid prototype modeling and validity assessments
   1. Ingest core domain data into the POC data platform (batch, basic automation)
   2. Create prototype custom logic to merge data and create registry PoC version of data
   3. Build data quality checks controls and references for the data being processed
   4. Implement basic IDQ dashboards to extend to this domain
   5. Supplement with alerts and scripts based on needs
   6. Identify a number of key metrics for the domain, e.g. Employee counts, etc.
   7. Implement a number of dashboards to present and visualize this information
5. For major ongoing transformation programs, pressure-test the master data model and the program data needs. Identify edits to master data model *or* changes to Program.
6. In tandem with the work, catalog a backlog of rules and configurations in preparation for the MDM implementation of these rules (see phase 2, point 2).
   1. **Operating and engagement model for enterprise center of excellence group**

Support the operating model for an IT data management center of excellence designed to centrally support enterprise wide data management functions. This stream will run in parallel with Phase 1.1

* + 1. Align on the vision for the enterprise data services team
    2. Design the engagement model between the CoE, program teams, data governance structures and the business including funding model, process support flows and technology decision rights
    3. Define and outline the skill capabilities, roles and responsibilities of the center of excellence team
    4. Define the right organizational constructs to support the enablement of the business to become an effective data driven and enabled organization

1. **Phase 2: MVP for 2-3 business process use cases in the domain.** This phase will be refined during Phase 1 but at a high level will convert the 2-3 business processes or into robust MVPs:
2. Configure the MDG (or alternative tool) to support rules identified in Phase 1, activity 5.
3. Develop the UIs to enable manual intervention and overriding of data when necessary
4. Implement the most critical 3-5 integrations with downstream systems for MVP purposes
5. Align the organization’s governance model with the new tactical processes defined in this stage. Implement and “practice” end-to-end governance including roles, processes, tools
6. Evaluate / redesign front-end input for source data ingestion into MDM solution
7. Throughout phase 2 continue with the refinement and the utilization of the data platform
8. Carry out change impact assessment and training needs analysis required to support new data assets and processes
9. Define the additional employee journeys across broader domain ahead of Phase 3
10. **Phase 3: Full domain model build.** To be further refined during Phases 1, 2 but at a high level:
11. Complete the integrations necessary to finalize the domain model
12. Provide access to data within the domain through well-defined API services for consumption
13. Integrate the golden master record in the enterprise data platform and keep updated to support further analytical processes
14. Industrialize the analytics and reporting from the base data for the organization
15. Incorporate the processes within the enterprise data governance model
16. Stand up new / change existing data management processes and new enduring organizational capabilities (IT and business)