Adaption of Data Governance to a data driven and Agile organization

pwc

Index

Summary

As organizations strive to become data driven, easy access to trustworthy data handled according to policy and regulatory requirements is becoming an imperative. In addition, the standard approach for handling digitalization efforts has shifted from sequential to Agile development.
It is now apparent that present Data Governance capabilities in most organizations are not meeting requirements.

Our observation is that the present standard approach to establish and evolve Data Governance is considered to be ineffective as it focuses on governance of data for regulatory purposes (defensive use). The data sets in scope are often well defined, relatively limited and with legal accountability for accuracy and timeliness held at the C-level. Efforts to develop defensive Data Governance capability have been developed sequentially with full sponsorship and direction provided and overseen by leadership. Data requirements are more or less static after completion.

This approach is by nature unfit to accommodate requirements from big data initiatives and Agile development. Big Data development is characterized by the need to handle massive amounts of data in various formats and an evolving insight of data requirements. Agile is focusing on short development cycles that are handled by self-sufficient teams.

This initiates a need to redesign the Data Governance capability. The new capability must support both defensive (assuring compliance to regulations, ethics and risk minimization) and offensive (supporting insight generation and innovation) use of data and integrate with Agile development. We see that this can be achieved by an approach to Data Governance that adopts and supports Agile principles.

The initial Data Governance capability will be a scalable Minimum Viable Capability (MVC). The MVC comprises the basic structures and capabilities required to provide Data Governance for the first use cases and prioritizations in the data strategy. By applying this approach organizations will be able to incrementally build, manage and exploit reusable data assets directed through business strategy, Agile planning and regulatory requirements.

This Point of View outlines this approach and describes how it was applied at a large Swedish client.

Key concepts

Data Governance consists of a number of interrelated data handling capabilities. They need to be synchronized and governed to enable the accomplishment of the objectives in Data and Analytics strategy. PwC has created a Data Governance framework that depicts the required components and interrelationships.

Data & Analytics Strategy:
"Business plans to use information for competitive advantage and to support enterprise goals. Data strategy must come from an understanding of the data needs inherent in the business strategy: what data the organization needs, how it will get the data, how it will manage it and ensure its reliability over time, and how it will utilize it." - DAMA International

Data Governance:
"The exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets." - DAMA International

Data Governance provides the mechanisms that assure Data Management capabilities to support the strategic priorities. Without Data Governance, there is no defined structure for exercising accountability and control of data and to assure that data management activities are performed in a way that supports strategic objectives.

Information Lifecycle Management
Information Lifecycle Management includes every phase of a "record" from its creation through its distribution, storage, sharing, archiving and potential destruction. During its existence, information can become a record by being identified as documenting a business transaction or as satisfying a business need.
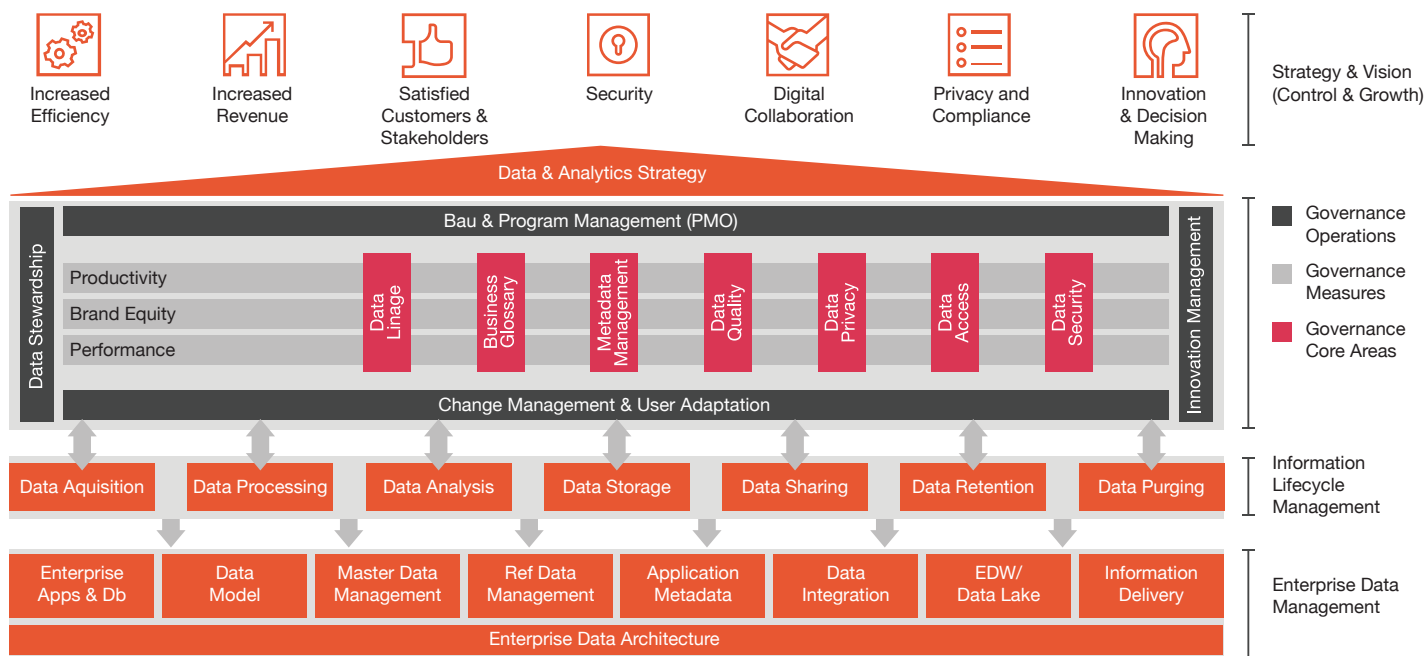
Enterprise Data Management:
Enterprise Data Management is the development and execution of programs and practices that deliver control, protect and enhance the value of data and information assets through their life cycles.

Agile:
Agile is a development approach under which requirements and solutions evolve through the collaborative effort of self-organizing and cross-functional teams and their customer/end user.

Data driven:
Data driven organizations are organizations that value decisions that can be backed up with verifiable data. The success of the data driven approach relies upon the quality of the data gathered and the effectiveness of its analysis and interpretation.

Strategy & Vision (Control & Growth)

- Increased Efficiency
- Increased Revenue
- Satisfied Customers & Stakeholders
- Security
- Digital Collaboration
- Privacy and Compliance
- Innovation & Decision Making

Data & Analytics Strategy

Data Stewardship

Bau & Program Management (PMO)

Productivity
Brand Equity
Performance

Data Linage · Business Glossary · Metadata Management · Data Quality · Data Privacy · Data Access · Data Security

Change Management & User Adaptation

Innovation Management

- Governance Operations
- Governance Measures
- Governance Core Areas

Information Lifecycle Management

Data Aquisition · Data Processing · Data Analysis · Data Storage · Data Sharing · Data Retention · Data Purging

Enterprise Data Management

Enterprise Apps & Db · Data Model · Master Data Management · Ref Data Management · Application Metadata · Data Integration · EDW/Data Lake · Information Delivery

Enterprise Data Architecture

PwC Data Governance Framework

Data Governance needs to be redesigned in order to provide value in a data driven and Agile environment

Digital transformation brings new demands on the way businesses are run. Powered by new technology, new business models and ways of working are emerging. Disruption is reaching almost all organizations who now aspire to become data driven. Regulators are trying to keep up with the change and thus the legal demands are changing as well.

Agile has become the development approach of choice among most organizations. However, Agile development culture is believed to be hampered by traditional centrally controlled Data Governance putting adoption of Data Governance at risk.

Traditionally Data Governance is often implemented and using a top-down approach. Common sequence is maturity assessment, definition of target state, policy, operating model and detailed procedures before implementation proceeds on a data domain. A comprehensive data dictionary, describing data for regulatory reporting or data for a selected domain is produced. In parallel a dedicated Data Governance function, detailed processes and systems support are designed and implemented.

After a while, it becomes apparent that this approach carries an imbalance between effort and reward. Focus is on enforcing regulatory policies, collaboration between Data Governance function and Agile development teams is weak. The underlying cause is that forms for collaboration between Agile and Data Governance teams have not been established. Traditional Data Governance is not flexible enough to take on requirements from Agile development teams. Agile development teams consider data aspects too late, resulting in mutual frustration, unsettled data standards and risks of poor data quality.

The initiative is questioned and the willingness to contribute evaporates. Eventually, the initiative only proceeds for a limited range of data where there is a stipulated regulatory requirement for Data Governance or where the benefits of improved Data Governance brings tangible value to a group of direct stakeholders.

Data driven organizations add considerations that increase the challenge of succeeding with Data Governance:

• Extreme volumes of data require automated policy enforcement
Automated ingestion and distribution of data containing sensitive information requires automated protection of data by applying techniques like data masking based on data protection policies.
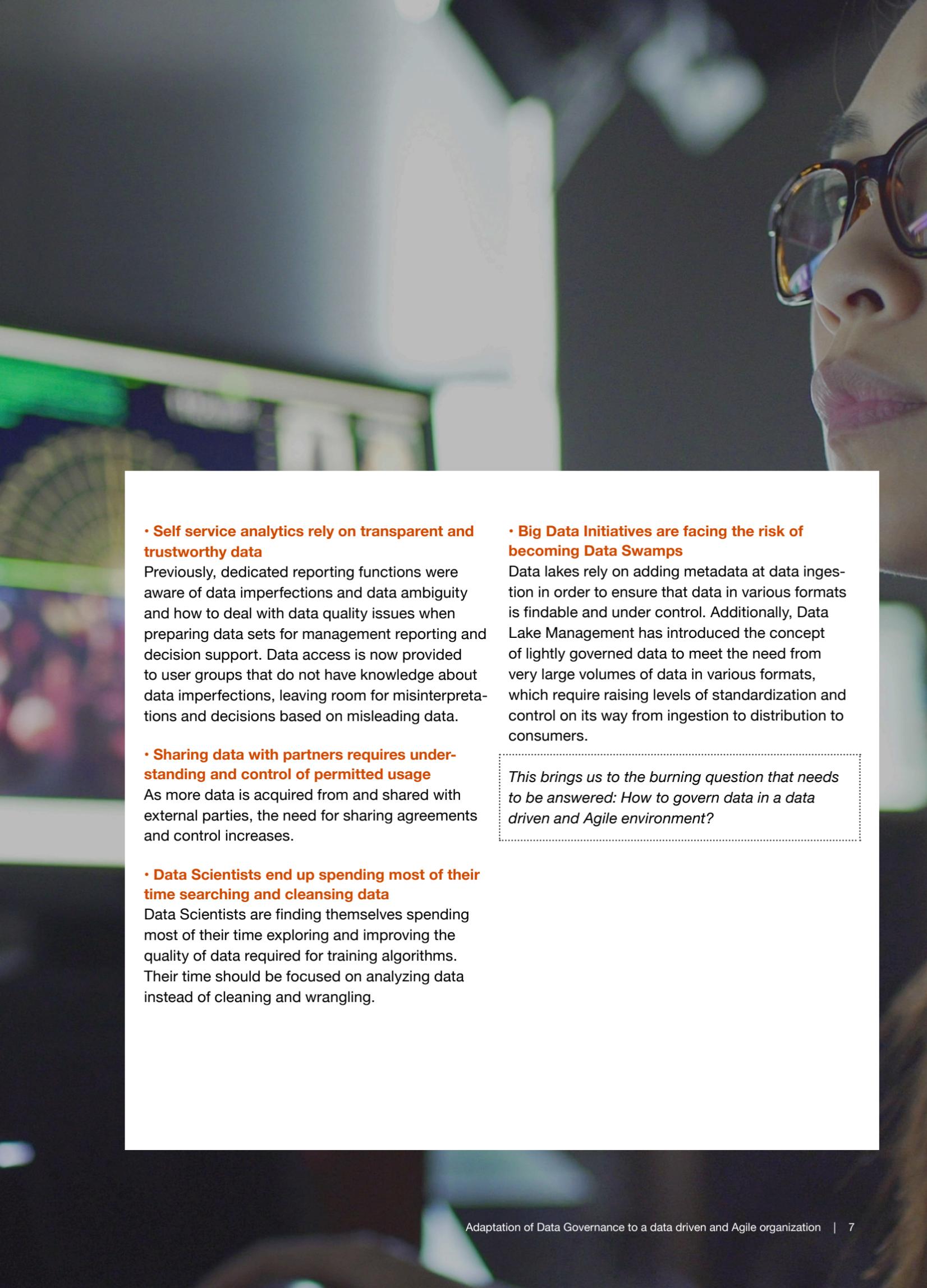
• Self service analytics rely on transparent and trustworthy data

Previously, dedicated reporting functions were aware of data imperfections and data ambiguity and how to deal with data quality issues when preparing data sets for management reporting and decision support. Data access is now provided to user groups that do not have knowledge about data imperfections, leaving room for misinterpretations and decisions based on misleading data.

• Sharing data with partners requires understanding and control of permitted usage

As more data is acquired from and shared with external parties, the need for sharing agreements and control increases.

• Data Scientists end up spending most of their time searching and cleansing data

Data Scientists are finding themselves spending most of their time exploring and improving the quality of data required for training algorithms. Their time should be focused on analyzing data instead of cleaning and wrangling.

• Big Data Initiatives are facing the risk of becoming Data Swamps

Data lakes rely on adding metadata at data ingestion in order to ensure that data in various formats is findable and under control. Additionally, Data Lake Management has introduced the concept of lightly governed data to meet the need from very large volumes of data in various formats, which require raising levels of standardization and control on its way from ingestion to distribution to consumers.

This brings us to the burning question that needs to be answered: How to govern data in a data driven and Agile environment?

Prerequisites for Data Governance to fit in a data driven and Agile landscape

Through our work with clients who are on their way to become data driven and Agile, we have been able to identify three success factors that distinguish the leaders from the peers that are lagging behind.

1. Ability to prioritize Data & Analytics initiatives based on strategic value
2. Data Governance implementation is performed in an Agile manner
3. High organizational adoption of the required change

Considering these success factors from an Agile Data Governance perspective, we can see that a set of key prerequisites need to be established.

1. Ability to prioritize data & analytics initiatives based on strategic value

- Visibility on how Data & Analytics supports business objectives (e.g. a Data & Analytics Strategy)
- Leadership focus on Data Governance
- Performance and value metrics

2. Data Governance implementation is performed in an Agile manner

- Acceptance that data standards and Data Governance can be defined incrementally using an Agile approach
- Data Governance roles are involved in Agile development teams
- Data Governance development is handled in a separate backlog

3. High organizational adoption of the required change

- Leadership buy in, ownership and recognition
- Communication and training
- Non-invasive structures building on what is present
- Support function providing expert support and training
- Sufficient staff time to spend on Data Governance activities
- User friendly and effective tools

In order to make these prerequisites for success visible and to provide guidance, we have translated our findings into seven guiding principles. These principles can guide organizations with the intention to adapt Data Governance to support a data driven, Agile and collaborative organizational culture.

Seven guiding principles for successful Agile Data Governance

1. **Data Governance prioritization must be aligned with business objectives**
Data Governance prioritizations must be aligned with business objectives in order to be applied where business value is maximized and to provide rationale for efforts.

2. **Agile Data Governance balances offensive and defensive use of data**
Data Governance practices must balance offensive (supporting insight generation and innovation) and defensive (compliance to legal, ethical and risk minimizing limits) use of data defined in Data Governance policy.

3. **Implementation of a Minimum Viable Capability (MVC) which proceeds in accordance to the Data & Analytics Strategy**
In order to focus the implementation on prioritized areas and to minimize efforts on developing back-office functions that do not support the prioritized need, an MVC approach is suggested. A scalable core is established that can be applied and scaled depending on the implementation area and strategic prioritization.

4. **Agile Data Governance is based on the existing governance structures through a non-invasive philosophy**
Traditional Data Governance focuses on detailed procedures while Agile Data Governance focuses on guidelines and capability to support the organization. The operating model and governance structure should as far as possible be based on existing governance structures and roles (non-invasive). Data ownership and data stewardship are pragmatically allocated to function owner, process owner or data specialist. Roles are assigned to staff based on personal fit, motivation and ability to allocate time.

5. **Data Governance and Agile development-teams have mutual respect and collaborate**
Mutual understanding and respect needs to be established between the data governance and the Agile development teams, based on a mutual understanding of policies and objectives that govern each area. Data Governance need to embrace the Agile concept of incremental development, and the Agile Development teams need to understand that data usage ultimately must comply with the Data Governance policy.
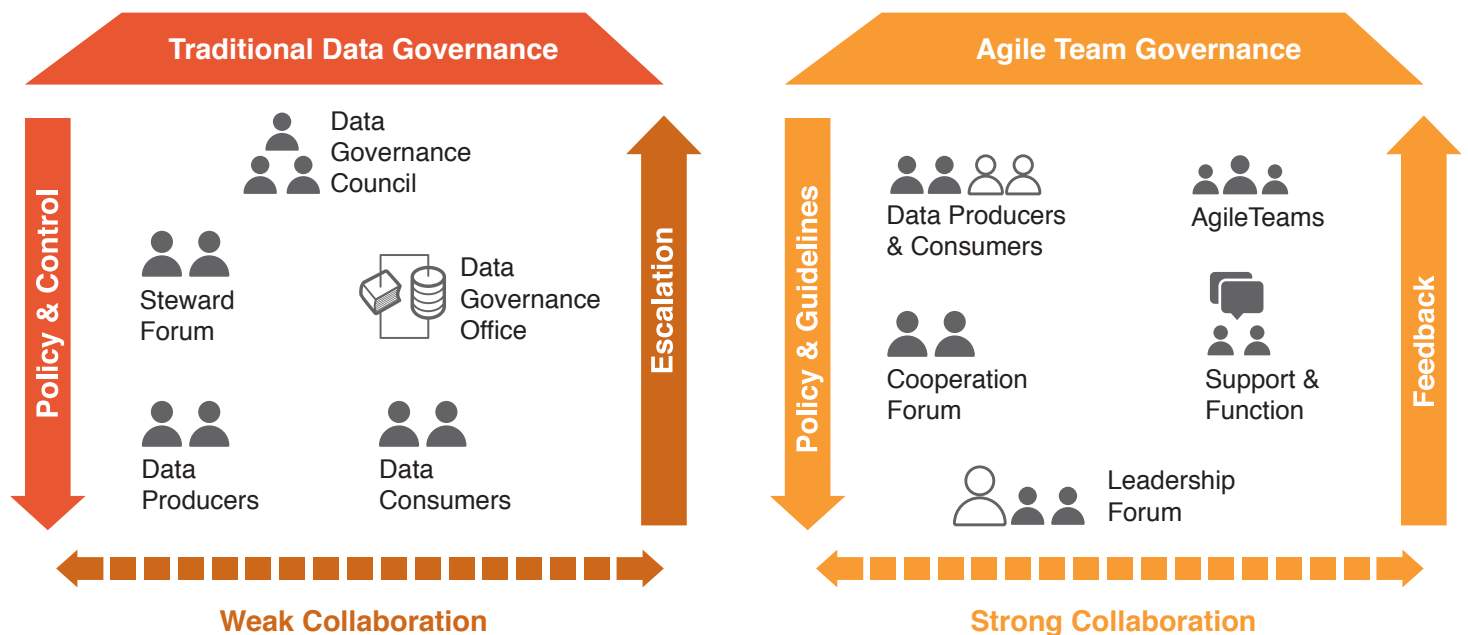
6. **Data Governance and management roles are provided with the right conditions and recognition for taking on responsibility**
Business expertise, training, time, support and tools must be provided to people involved in data governance and management. The leadership must recognize the Data Governance efforts in order to assure sustainable results. Collaboration requires forums, processes, effective tools and a culture that is supportive.

7. **Leadership actively communicates the criticality of data and Data Governance**
Leadership understanding of data and Data Governance needs to be mature and they must actively communicate its importance to the rest of the organization. Leadership sponsors Data Governance and stakeholders are provided sufficient resources and mandate.

Applying the seven guiding principles shifts the focus from compliance and top-down control to a focus on enabling organizations to leverage and optimize the use of data to achieve business objectives within regulatory and ethical constraints.

Traditional Data Governance

Policy & Control

- Data Governance Council
- Steward Forum
- Data Governance Office
- Data Producers
- Data Consumers

Escalation

Weak Collaboration

Agile Team Governance

Policy & Guidelines

- Data Producers & Consumers
- AgileTeams
- Cooperation Forum
- Support & Function
- Leadership Forum

Feedback

Strong Collaboration

Characteristics:

- Defensive enforcement of policies for compliance purposes
- Waterfall domain-wise implementation
- Detailed procedures to be interpreted by affected personnel
- Central function focusing on policy enforcement and production of procedures
- Responsibility without sufficient resources
- Weak collaboration and knowledge sharing between consumers and producers
- Manual procedures - single point systems support
- Low leadership acknowledgement for performing Data Governance tasks

- Balance between defensive and offensive use of data
- Agile implementation based on strategic prioritization
- Policy and guidelines implemented in systems easily available for end users
- Central function providing guidance, support and training
- Responsibility and resources aligned
- Institutionalized collaboration among user, producer and developer community
- Streamlined processes with embedded systems support
- Data governance tasks acknowledged and prioritized by leadership

A scalable Minimum Viable Capability is established which is then applied on a pilot use case

A strategic and Agile approach to Data Governance does not mean that a traditional structure is unnecessary, only that the design and implementation approach needs to be redesigned.
We recommend that the structure follows three basic steps that need to be customized according to organizational situation.

1.Build common understanding of the value of Data Governance for prioritized use case and roadmap to deliver required capability
Business value of Data Governance and Data Management is visualized by deriving requirements from business objectives.

The vision, guiding principles and roadmap for improving Data Governance maturity to the level of ambition is defined and agreed. Sequence is maturity assessment, vision, principles, target state, gap as-is versus target state, business case and high level roadmap to close gap. This creates a common understanding of the concept and value of Data Governance, status of present capabilities and an initial understanding of what and how to establish an MVC. It is a step that requires extensive involvement and support from leadership.

2.Build and launch a Minimum Viable Capability (MVC) that is scalable
An MVC consists of the structures and capabilities that are required to create a scalable Data Governance capability that can be applied on selected first use case and set of data in scope. The MVC is then continuously developed as it takes on new use cases and data sets.

MVC Components

Operating Model
The roles and forums that will interact with the defined responsibilities and mandate for decision making, including processes and training members. As data volume and diversification increase the operating model need to adapt.

Support Function
Expert Data Governance competence that will support implementation and operations.

Data Governance Guidelines
Guidelines (based on policies) outlining purpose and method for execution of Data Governance within the organization guiding day-to-day operations without pointing out detailed procedures.

Tools and Technology
Tools providing support to collaboration and distributed exercise of Data Governance and Data Management activities. Initially simple tools will be sufficient, when volumes and complexity increase scalable tools are required.

3.Integrating Data Governance into Agile development

Key prerequisite for implementing the principle "Data shall be treated as an asset" is to make Data Governance procedures a natural and value adding activity in development and in day-to-day procedures. This prerequisite is supported by adding Data Governance artifacts into the "definition of done" in all development following the Agile approach. By applying this approach, organizations will be able to grow data assets that are governed, shareable, reuseable in an Agile manner.

Initially an MVC is established that comprises the basic structures and capabilities required to provide Data Governance for the first data and analytics use cases. In the next iterations capability is developed to address requirements derived from data strategy and product backlog. Maturity of Data Governance will then evolve directed by Data Strategy and integrated with Agile development.
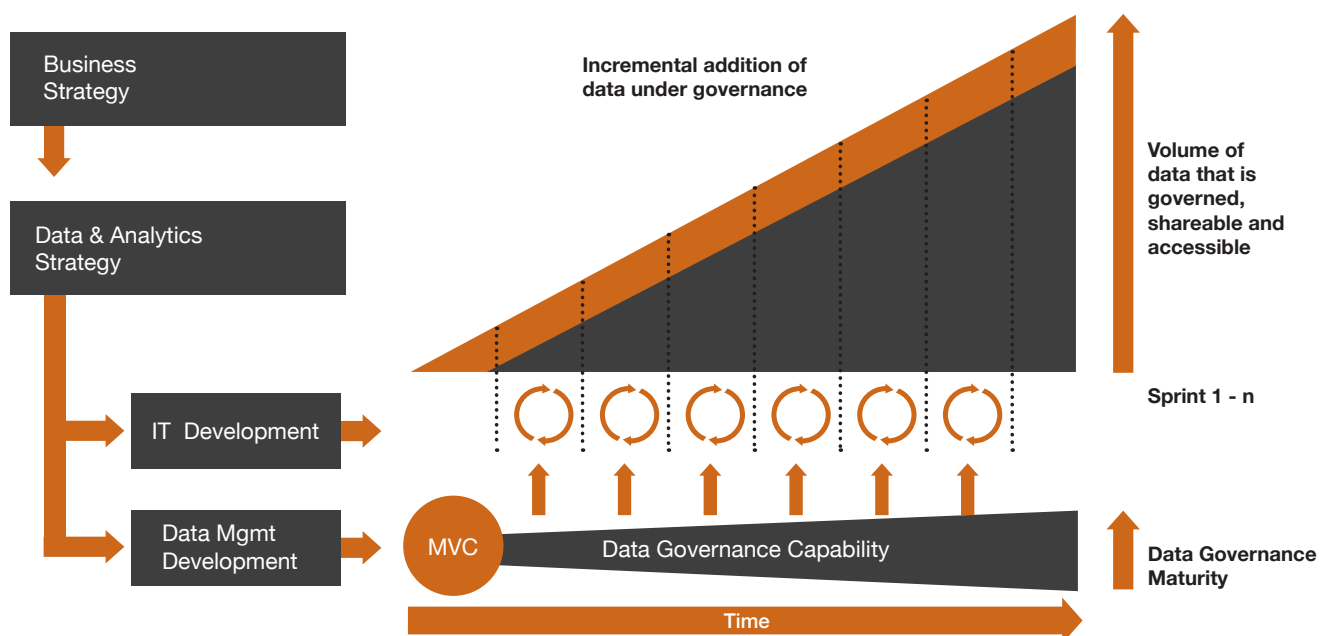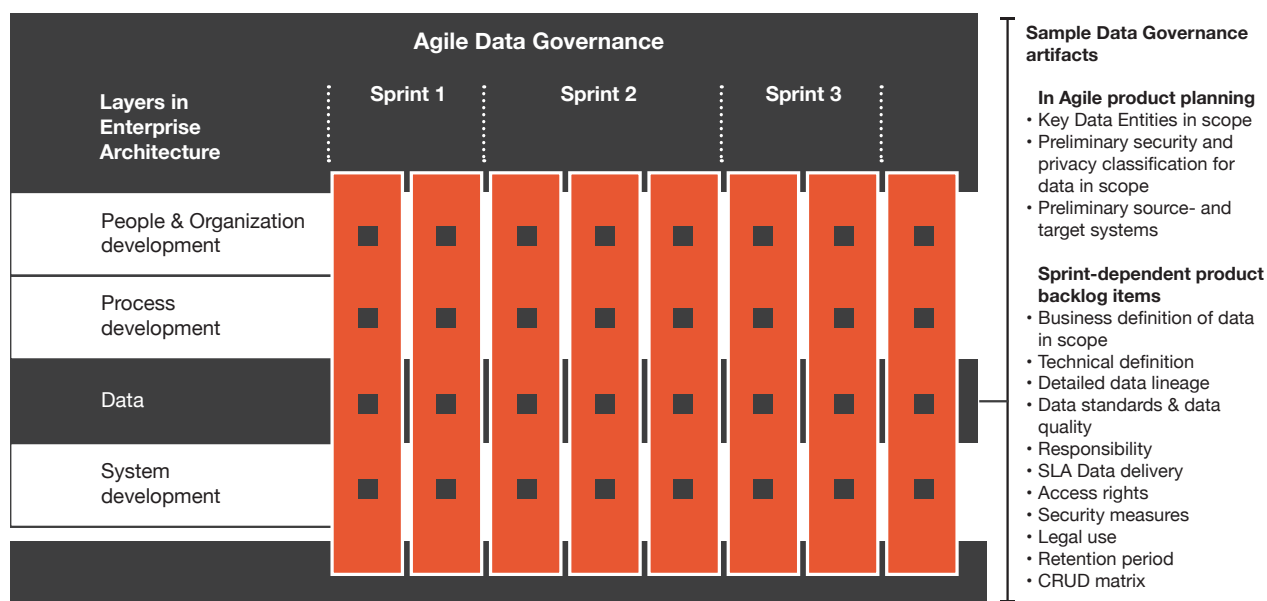

An in-depth walkthrough of suggested approach to integrate Data Governance into Agile development is presented in the following section.

Integrating Data Governance with Agile development

In Agile, the development plan is defined and maintained in the form of the product backlog consisting of product backlog items. All product backlog items have components belonging to one or more of the Enterprise Architecture layers: people & organization, processes, data and systems.

Data Governance artifacts are defined as part of the definition of done at the initial and the ongoing product planning.

Agile Data Governance

Layers in Enterprise Architecture	Sprint 1	Sprint 2	Sprint 3
People & Organization development			
Process development			
Data			
System development			

Sample Data Governance artifacts

In Agile product planning
- Key Data Entities in scope
- Preliminary security and privacy classification for data in scope
- Preliminary source- and target systems

Sprint-dependent product backlog items
- Business definition of data in scope
- Technical definition
- Detailed data lineage
- Data standards & data quality
- Responsibility
- SLA Data delivery
- Access rights
- Security measures
- Legal use
- Retention period
- CRUD matrix

Traditionally, stakeholders representing all architectural layers except data have provided product requirements translated into the product backlog. Absence of a Data Governance perspective and a capability adapted to support Agile development has left the data layer to be handled ad-hoc.

Agile Data Governance integrates Agile with Data Governance and creates forms for collaboration between Agile development teams and Data Governance organization.

In order to get an initial understanding of what to be considered during product development, the Agile product planning need to point out the following:

- What data entities are in scope?
- Who are the producers and the consumers?
- What Data Governance issues to anticipate?
- What are the preliminary security and privacy classifications of the data?
- What are the source and target systems?

When proceeding into detailed definition of product backlog items, Data Governance Support representative provides guidance on what Data Governance artifacts that should be part of definition of done, depending on the scope of the product backlog item. As insight evolves, the sprint planning can anticipate what and when Data Governance support is required.

Due to the nature of Data Governance, the Data Governance artifacts often require lead time in order to be agreed upon by affected stakeholders across the organization. Thus, product planning needs to take into account for the Data Govern-ance perspective early on.

In order to increase speed by minimizing lead time and to secure alignment in decision making, Data Governance issues can be distributed to a virtual cooperation forum. This forum consists of context dependent stakeholders, coordinated by the data steward appointed to the Agile development team. Suggested standards are shared for approval, adjustment and final approval with support of system tool (e.g. a data catalog tool). Decisions are only escalated if data standard or other data issue can not be agreed on, or if the issue is stipulated to be handled in a certain forum.

Agile Data Governance in Scrum

Based on experience, the table below indicates key Data Governance artifacts per Scrum event. Suggested Data Governance roles to be involved are indicated in order to assure that the Data Governance perspective is addressed during product development.

We would like to stress that artifacts and Data Governance effort are context dependent.
In order to keep the Agile development team lean and focused on delivering artifacts of value for the product in scope Data Governance artifacts need to be reconsidered at each sprint planning. The table below outlines sample Data Governance artifacts and roles to be included in each Srum event.

Scrum Event	Event Description	Scrum Artifact	Data Governance Artifacts (sample)	Agile Roles	Data Governance Roles involved
Product Planning	Management of the Product Backlog	• Product Backlog with Product Backlog Items	• Data Entities in scope • Preliminary security and privacy classification • Source and target systems	• Product Owner • Business developers • Requirements Analyst	• Data Governance Support • Enterprise Architect
Sprint Planning	Select the work (Backlog Items) to be performed in Sprint	• Sprint Backlog (set of Product Backlog items selected for the Sprint) • Tasks with a clear Definition of Done (DoD)	• Detailed definitions • Data Models • Final security classification • Data Lineage • Data quality standards • Retention periods • Data sharing agreement	• Development team	• Data Stewards • Security Specialist • (Data Modeller) • (Data Quality Expert)
Daily Scrum	Inspect how progress is trending toward completing the work in the Sprint Backlog	• Status of Sprint	• Depending on product backlog item(s)	• Development team	• Data Steward
Sprint Review	Held at the end of the Sprint to inspect the Increment and to adapt the Product Backlog if needed.	• Sprint Increment • Revised Product Backlog • Monitoring Progress Toward Goals	• Depending on Product backlog Item(s)	• Development team	• Data Steward • (Data Modeller) • (Data Quality Expert)
Sprint Retro-spective	Scrum Team inspect itself and create a plan for improve-ments to be enacted during the next Sprint	• A plan for improvements		• Development team	• Data Governance Support • Enterprise Architect

Client Case: Applying Agile Data Governance in a Big Data context

The following client example summarizes considerations and solutions for Agile Data Governance. Challenges and recommendations are described based on the different solution areas for a Big Data solution which aims to support automation, computer science and AI-based decision-making as well as self-service analysis.

Client's strategic investment in Big Data solutions initiated a need for Data Governance

PwC was hired to support a large Swedish client in building strategic Big Data capability to support advanced analytics and to automate decision-making. Development was directed by the client's data & analytics strategy. Data & Analytics Strategy use cases were defined and prioritized to be supported by a state of the art Data Lake technology platform. Proper Data Governance was identified as critical to assure correct handling and quality of massive amounts of personal data. This data must be handled according to stringent privacy and security policies while leaving room for insight generation and training of AI-algorithms.

A traditional approach to Data Governance was considered not to be effective

The current Data Governance practices were assessed to be unfit to support needs due to lack of established Data Governance structure and capabilities. Client's strong commitment to minimize bureaucracy and to delegate responsibility while being Agile, called for a different approach to Data Governance. PwC suggested an approach in accordance with the principles and guidelines for Agile Data Governance. Principles were anchored with client's key stakeholders in the data management and analytics community.

A minimum viable Data Governance capability was designed to accommodate data needs for the initial analytics use cases

An MVC was defined in parallel with definition of a comprehensive conceptual data model, identification of the critical data and the data sources for defined analytics use-cases.

The MVC operating model was designed to build upon the present operations and product governance structure (according to the non-invasive principle). In the present structure responsibility of data was not defined. Data ownership was assigned to the product area that had the highest data quality requirement for a given data domain.

Sample strategic, tactical and operational data decisions were defined and attributed to decisions forums in the operating model to guide where to escalate data issues and were to handle cross domain issues.

Data stewards were appointed based on knowledge, willingness and ability to take on responsibility. For the majority of data domains, this required one data steward from product area and one from the system side.

To support the organization in implementing Data Governance, a support function was defined consisting of Head of Data Governance, Information Modeller, Metadata Specialist, Data Quality Specialist and Security & Privacy Specialist. System support initially consisted of an interim metadata catalogue in Excel to later be replaced by a standard solution system. Client data management system suite already provided data quality support.

The MVC was implemented to govern the data required for the first analytics use case
During the SCRUM planning phase, it was not necessary to detail product data requirements, but there was a need to have an initial understanding of the data to be included for planning of the following phases. This includes key data entities in scope, preliminary privacy and security classification and preliminary source and target system.

During development, the Agile team explored available data to identify relevant data sets. When the solution design was finalized, the Agile team in collaboration with data stewards specified data formats, transformation rules and mapped information into the data catalogue. In addition, this group defined the data quality requirements and data quality process for data in scope.

Lessons learned:
- Establish common understanding of Agile Data Governance framework to be applied in the Agile development.
- Collaboration between the data stewards and the Agile team is critical in order to assure that the most accurate and valuable data is provided to the analytic solution.
- The data stewards play an important role in supporting the process of adding new information to information models and establishing Data Governance for data in-scope.
- User friendly and effective tools are vital to facilitate information sharing between data stewards and the development teams.

- In the development teams, it is beneficial that some developers evolve to become power users and take responsibility for maintaining information in decided tools.

As always: Frequent alignment with sponsor throughout the establishment is crucial.

Contact

Alexander Fritsch
alexander.fritsch@pwc.com
+46 (0)729 97 26 02

Olof Olsson
olof.olsson@pwc.com
+46 (0)729 80 99 43

Joakim Lannersten
joakim.lannersten@pwc.com
+46 (0)729 80 99 13

Helena Marttinen
helena.marttinen@pwc.com
+46 (0)709 50 75 32

PwC Sweden´s Data & Analytics team will help you optimize your data assets. Use data analysis and data science to make faster and better decisions. Work more efficiently and save money. Find new sources of revenue. And capitalize on the untapped business intelligence you already own through the power of data & analytics tools.

PwC Sweden is the market leader within auditing, tax and advisory services, with 2,800 people at 34 offices throughout the country. Our purpose is to build trust in society and solve important problems. Our 45,000 clients are comprised of both global and Swedish companies and organizations of all sizes, and public sector entities. We provide guidance to our clients and assist them in achieving their goals, regardless of their stage of development.

PwC Sweden operates as a separate and independent legal entity. We are a member of the PwC global network which has 250,000 people in 158 countries. This gives us the possibility of sharing knowledge and experience on a global basis in developing sustainable solutions for our clients.