



A Minority Economically Disadvantaged Woman-Owned Hubzone-certified Small Business

Use of Master Data Management and Data Governance to Improve Data Management

FEDCASIC – April 12, 2017

Manish Khera, VP, 703-407-7051, mkhera@harmonia.com

Randall Daly, Solutions Data Architect, 913-832-8566, rdaly@harmonia.com

Our Credentials

Harmonia has 150+ personnel who have experience in databases, data warehouse, business intelligence, reporting, dashboards, analytics, etc. We have multiple projects in which we are conducting Master Data Management and Data Governance.

US Department of Agriculture Farm Service Agency Database Management Office

- 1 Enterprise Data Warehouse (~1 PB when completed)
- 2,200+ databases
- 1.1+ trillion records
- 15,000+ ETL processes moving and consolidating 20+ billion records nightly
- 800+ reports utilized by 13,000+ users in 3,000+ county offices
- Master Data
 - 14.1+ million Customer Master data records reduced to 9.7+ million records

US Department of Education Office of Federal Student Aid

- 1 Enterprise Data Warehouse (72 TB)
- 30+ databases
- 900+ million records (\$1.3+ trillion loan portfolio)
- Master Data
 - 40+ million Student Loan records reduced to ????

Topic Overview

Given that;

- 80% of all data generated in 2008 was unstructured or big data. The 80% in 2008 increased to 95%+ by 2014.

www.dataversity.net/the-growth-of-unstructured-data-what-are-we-going-to-do-with-all-those-zettabytes/

- Data generated within the last 5 years is greater than all data generated in human history previously. (44x increase between 2016 and 2005). We are at 23 zettabytes. Total amount of human speech ever spoken is estimated to be 40 zettabytes.

wikibon.org/blog/big-data-statistics/

- Estimated cost of poor data to the U.S. economy \$3.6+ trillion in 2015 (\$18.04 trillion was the US GDP for 2015) – a 15% loss in productivity

www-new.insightsquared.com/2012/01/7-facts-about-data-quality-infographic/

Questions:

How will we manage all this data from disparate sources?

How will we ensure data quality to ensure confidence in its accuracy?

Tenets of Data Management

Data

- is a mission-critical corporate asset
- is the only asset with infinite re-usability
- needs to be managed with the same rigor and discipline as all other valued assets
- quality measurement is foundational to managing it effectively

Data Governance

- is an evergreen program, not a project
- is intended to establish and execute ongoing data management capabilities
- is typically an enterprise level program with C-suite support and top-down direction
- must be implemented from within a data management service organization at the client and attempt to influence outward

Data Quality

- is an ongoing enterprise capability , not a one-time integration effort
- will involve determining the metrics employed to measure data quality
- will monitor and report data quality only - sub-standard data will not be corrected within the Enterprise Data Warehouse; it must be corrected at source

“Rules of the Road”

Data

- will be limited to the source systems providing data to the enterprise (in our case the Enterprise Data Warehouse at the USDA Farm Service Agency)

Data Governance

- will be limited to the scope of authority of the client
- will be limited to the context of Data Quality Management
- all other data management functions are out of scope
- willing owners and stewards can be recruited/assigned by the client

Data Quality

- no corrective action will be expected for any of the sub-standard data identified

Critical Success Factors (PPTD)

People

- Actively engaged and supportive Executive Data Stewards (owners)
- Empowered Data Stewardship teams

Processes

- EDW project development incorporating data quality activities & tasks
- Recurring Data Governance sessions
- Quality focused business process modifications

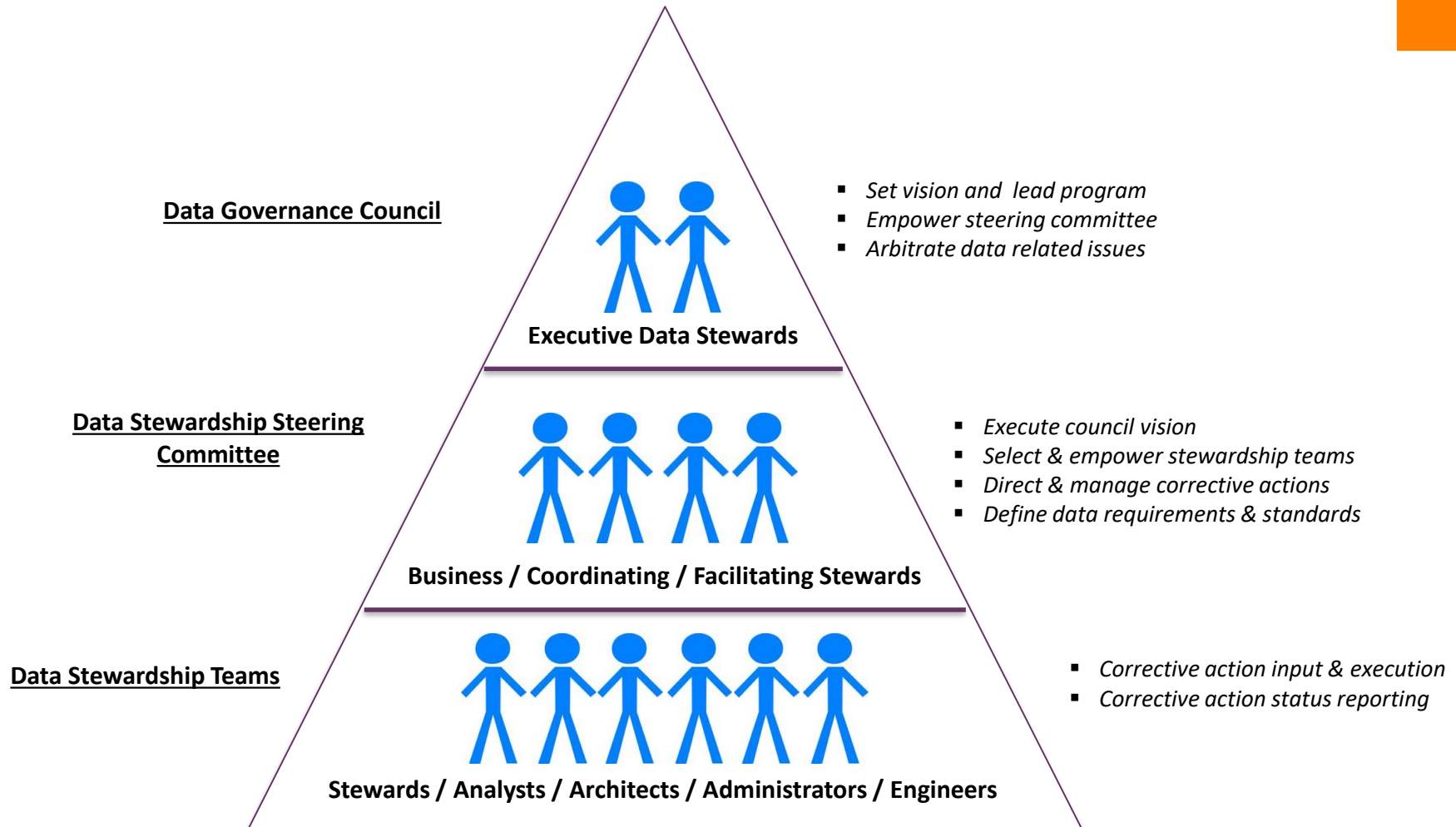
Technology

- Quality dashboards and reports
- Data profiling tools
- Quality focused technology modifications

Data

- Meaningful data quality standards and metrics

Data Governance Organization Framework (People)



Data Governance Organization Bodies

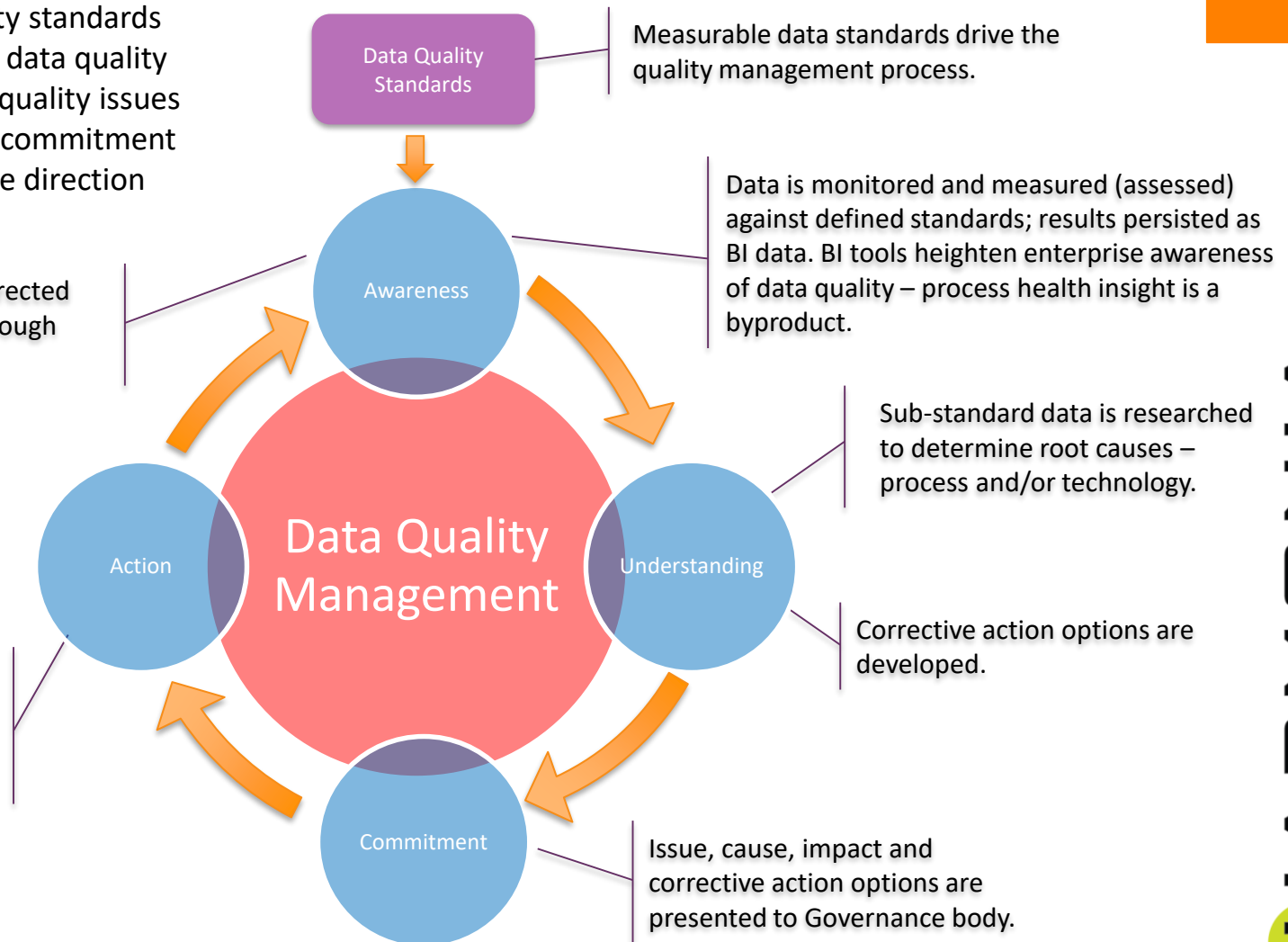
Organization	Description
Data Management Service Organization	One or more units of data management professionals responsible for data management within the IT organization. This is the DBMO within the FSA.
Data Governance Council	The primary and highest authority organization for data governance within the enterprise.
Data Stewardship Steering Committee	One or more cross-functional groups of coordinating data stewards responsible for support and oversight of initiatives launched by the Data Governance Council.
Data Stewardship Teams	One or more virtual teams consisting of stewards and technical disciplines collaborating on governance initiatives – typically within a specific subject area, led by a coordinating data steward in partnership with a data architect and a data stewardship facilitator.

Continuous Quality Management Cycle (Process)

1. Define data quality standards
2. Become aware of data quality
3. Understand data quality issues
4. Seek governance commitment
5. Act on governance direction
6. Repeat...

Cycle begins anew. Corrected data re-enters EDW through delta processing.

Governance approved corrective actions are executed. Data is corrected at source.

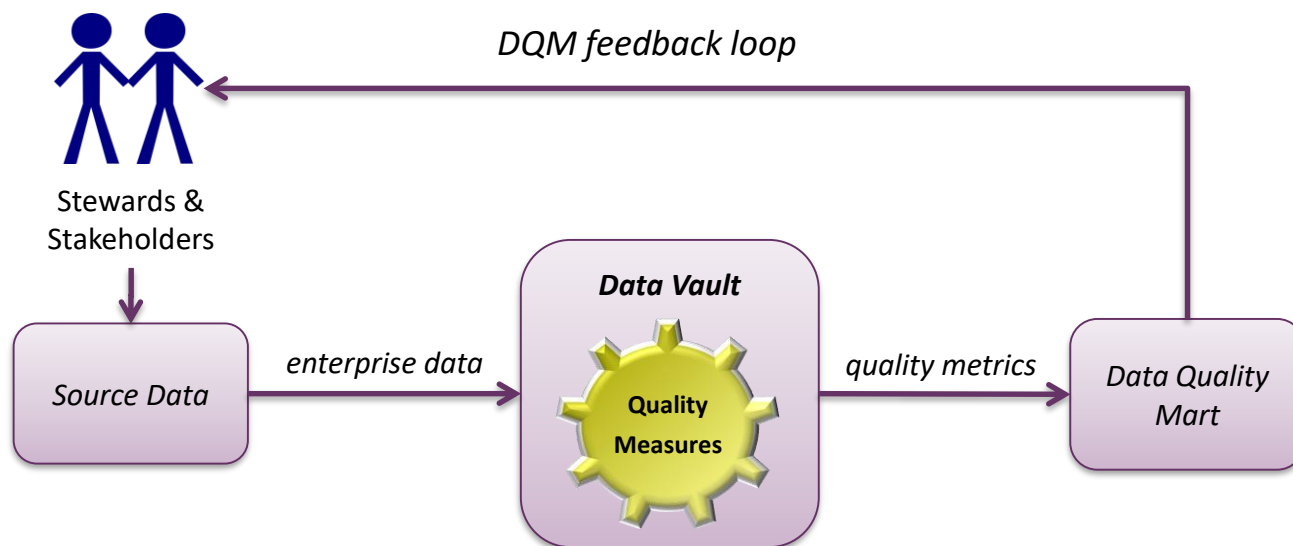


Awareness - Assess Data Quality

Management begins with awareness. Data stewards must achieve and maintain awareness of the quality of data they are responsible for.

A key contributor to this awareness is the ongoing monitoring and reporting implemented in EDW processes.

- Source data is assessed using established standards as it is processed through the EDW
- Assessment results (metrics) are persisted as a BI subject area and trended over time
- Data Quality reports and dashboards are provided to stewards and stakeholders (feedback loop)



Data quality metrics can also be used in confidence scoring the data for decision-maker consideration.

Utilize (Technology)

Capability	Description
Data Profiling	The capability to measure data against established quality rules. The size, type and complexity of data will drive the capability needed – from hand crafted SQL (simple) to sophisticated data profiling solutions (complex).
MDM Platform	A multi-domain master and reference data management solution that provides rule based entity resolution and data survivorship capabilities.
Issue Management	Substandard data, once identified needs documented and tracked, similar to a 'bug' found during software testing.
Metadata BI	The capability to measure, record and report on data quality metrics – Business Intelligence applied to data quality.
Data Modeling	The capability to document data and data structures at the Conceptual, Logical and Physical levels.
Process Modeling	The capability to document the processes, roles and responsibilities related to Reference Data and Data Quality management.
Data Glossary	The capability to define and communicate a common understanding of enterprise data and terms.

Understanding - Define Data Quality Issue (Data)

Data Governance members need a full and complete understanding of data issues – cause, business impact and the implications of proposed remediation.

Define Problem

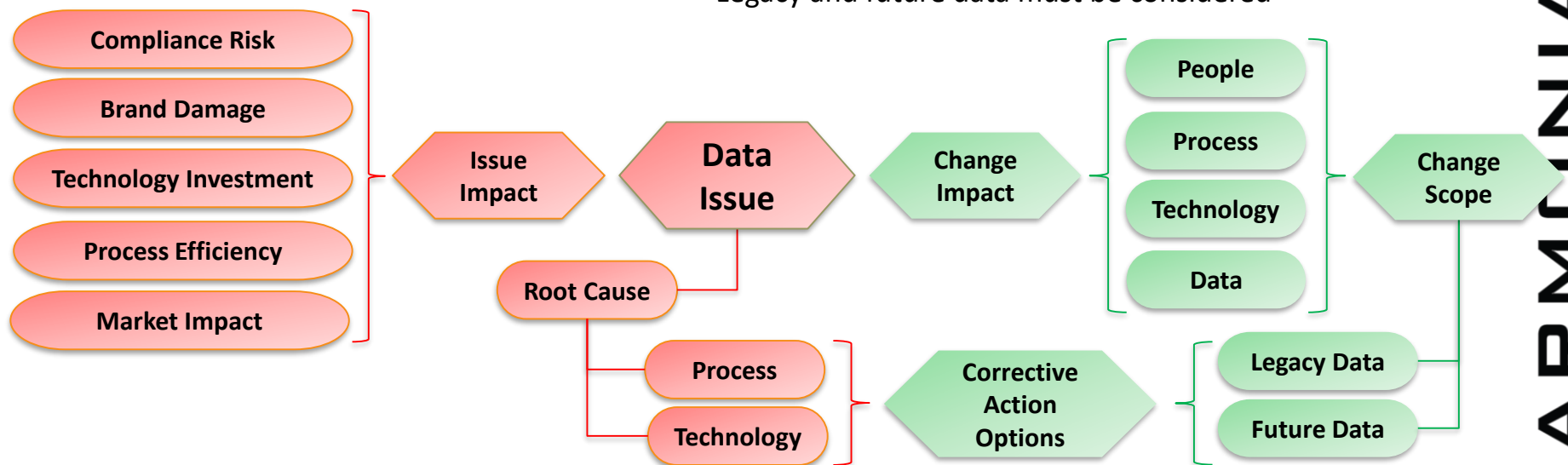
Substandard data detected:

- Root cause(s) determined
- Real and potential impacts identified

Provide Options

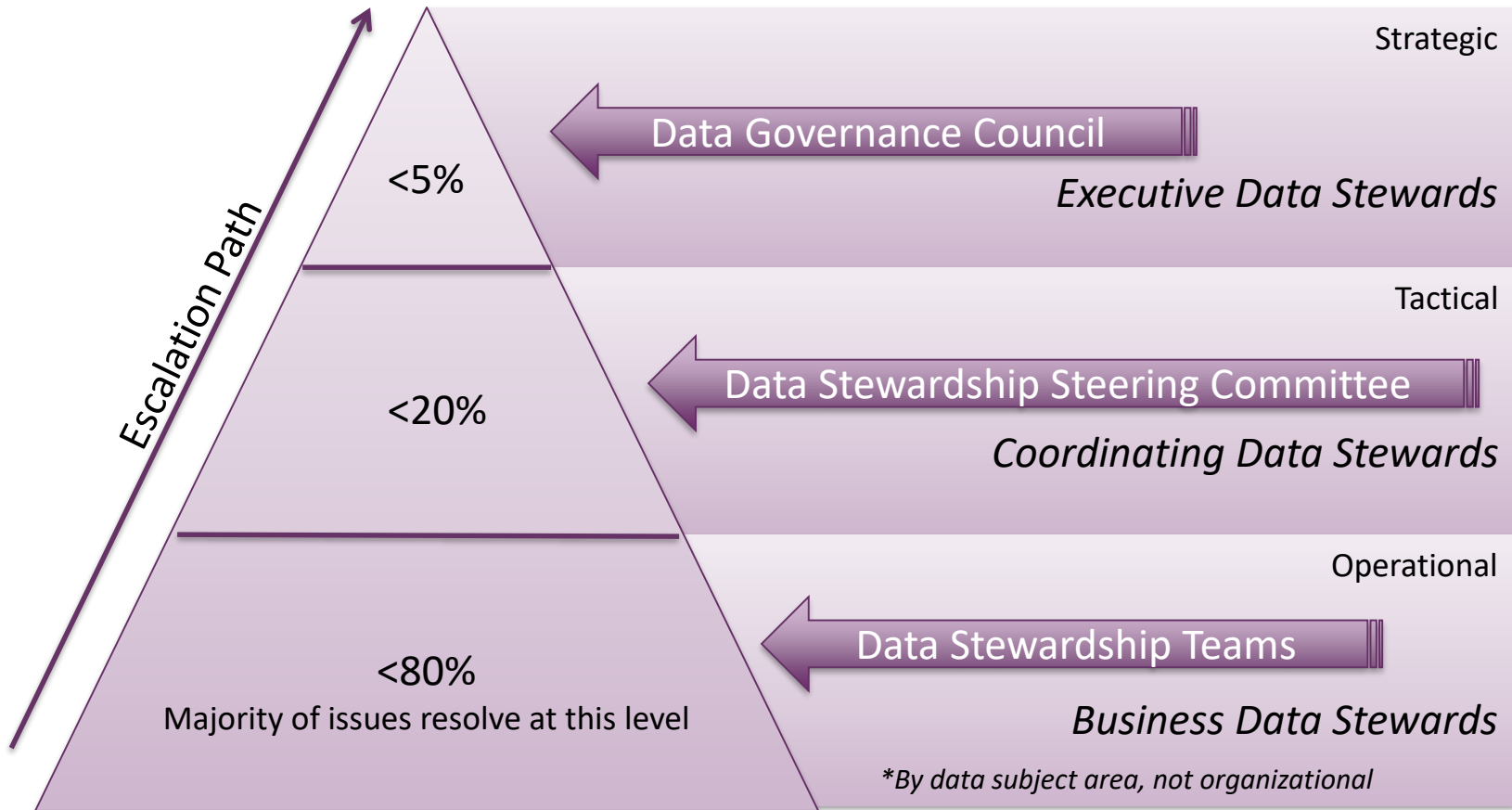
Corrective action options developed and presented to governance:

- Impact of recommended change(s) clearly understood and articulated
- Scope of recommended change must be reasonable
- Legacy and future data must be considered



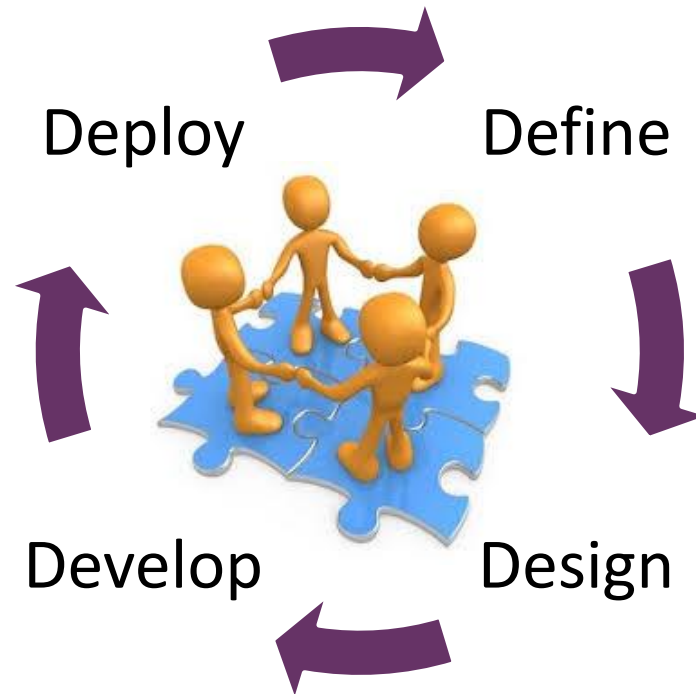
Commitment – Data Issue Resolution

Data issues are presented to the appropriate Data Governance body for resolution.



Action – Address Data Quality Issue

Data Stewardship teams develop and implement the corrective action approved by the Data Governance organization.



A virtual team consisting of business data stewards and the technical disciplines required to address the data issue