EDISON
building the data science profession
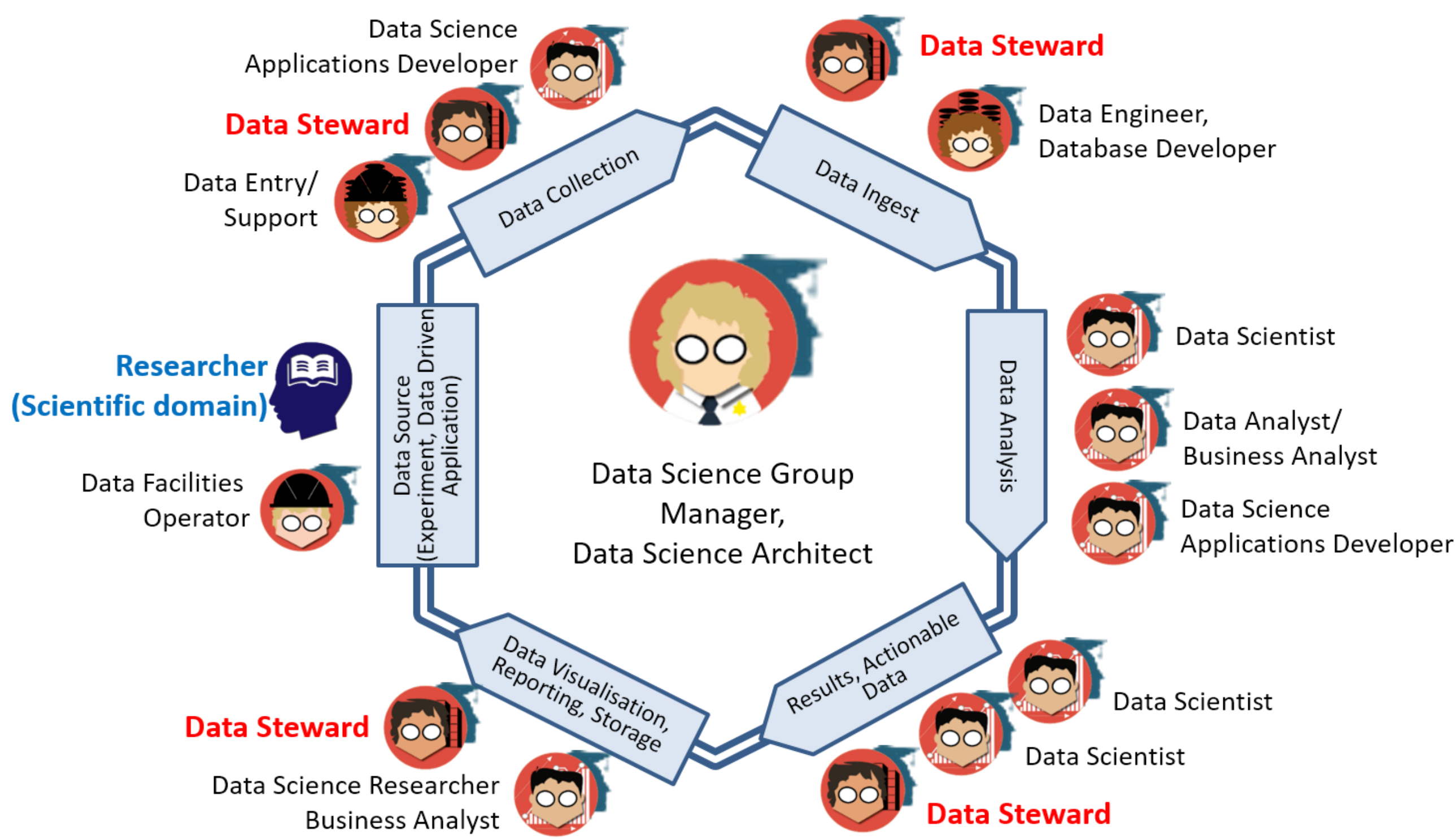
AI powered Research Data Management

How Artificial Intelligence and Machine Learning can Improve Data Management

Research Data Management Lifecycle and Associated Roles

Data Science Applications Developer

Data Steward

Data Entry/ Support

Researcher (Scientific domain)

Data Facilities Operator

Data Collection — Data Ingest

Data Source (Experiment, Data Driven Application)

Data Science Group Manager, Data Science Architect

Data Analysis

Data Visualisation, Reporting, Storage — Results, Actionable Data

Data Steward

Data Engineer, Database Developer

Data Scientist

Data Analyst/ Business Analyst

Data Science Applications Developer

Data Scientist

Data Scientist

Data Steward

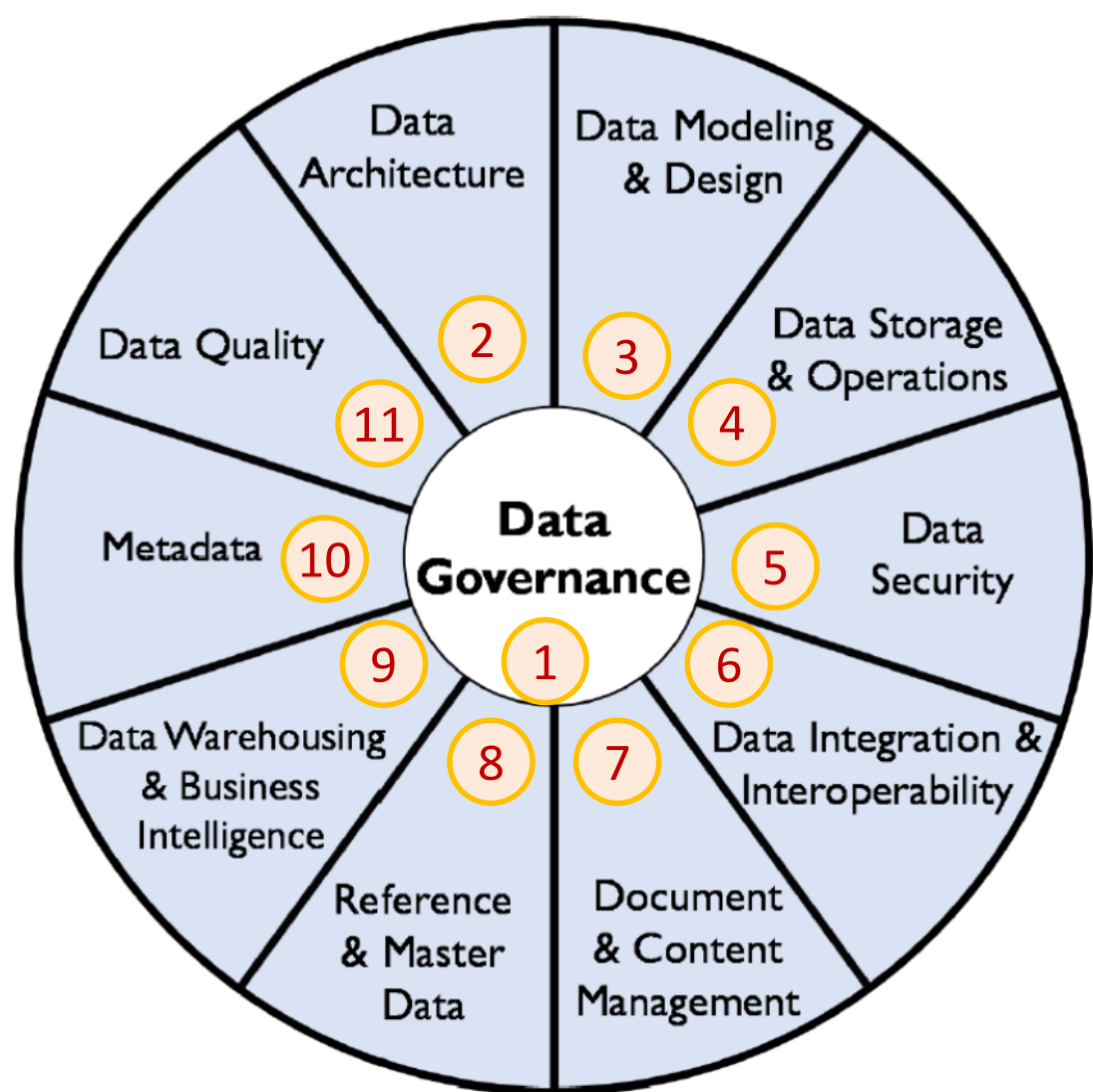**Data Steward**

Data Science Researcher Business Analyst

AI & ML assistive tools can increase quality and precision of Data Management by all actors/roles in the research data lifecycle

Research data management areas and processes that potentially benefit from AI&ML technologies
- Assisted by AI & ML tools (NLP/language recognition, text analysis, data mining, classifications, statistics, etc.)
  - Data entry (structured/unstructured) and initial classification
  - Metadata entry and management
  - Data quality assessment
- Powered by ML methods and tools (classification, cluster analysis, CNN, deep learning, SVM, PCA/ICA, dimensionality reduction, etc)
  - Storage & search oprtimisation, provenance, rapid access for complex analysis
  - Data pooling and integration for customary data analysis, including using Open Data and social media data in blended data analysis

Future data management assistive tools will benefit from human language interface that is already well supported and widely used with cloud powered applications.
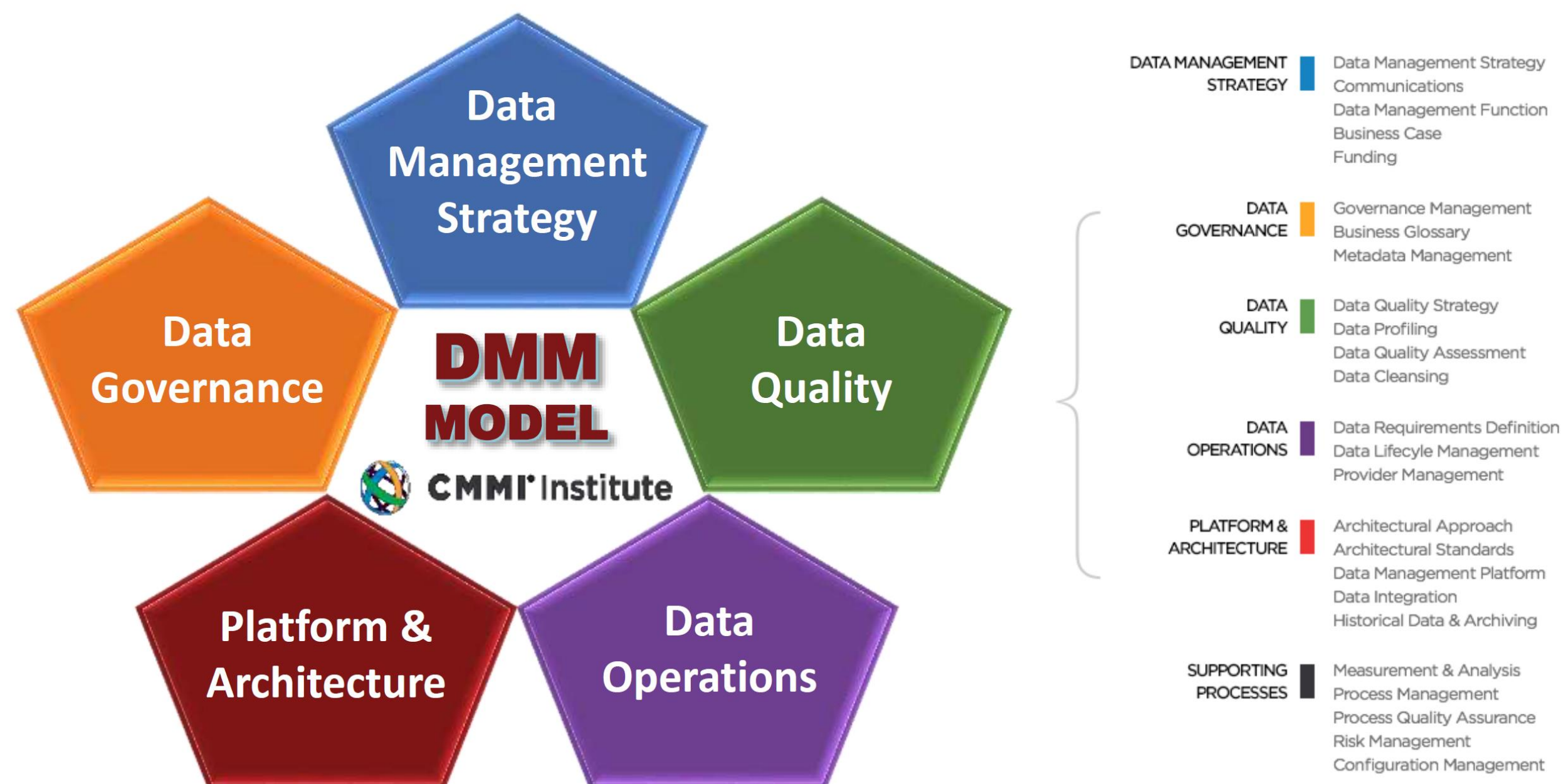
DAMA Data Management Knowledge Areas

Data Architecture

Data Modeling & Design

Data Quality

Data Storage & Operations

Metadata

Data Security

Data Governance

Data Warehousing & Business Intelligence

Data Integration & Interoperability

Reference & Master Data

Document & Content Management

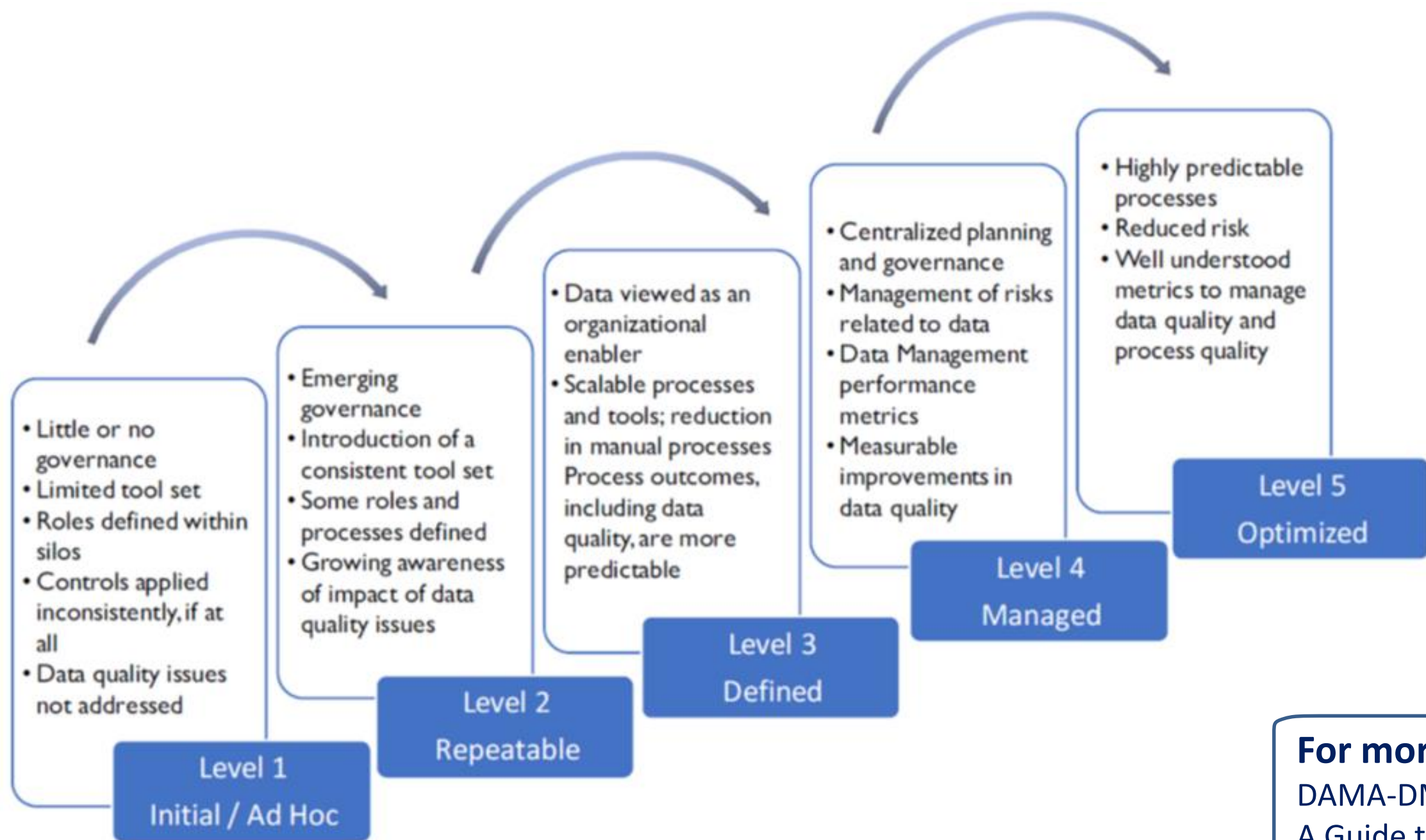The DAMA Wheel defines the Data Management Knowledge Areas.
- Data Governance is placed at the center of data management activities, since governance is required for consistency within and balance between the functions.
- The other Knowledge Areas are all necessary parts of a mature data management function:
  - Big Data and Data Science Analytics
  - Data Management Maturity Assessment
  - Data Management Organization and Roles Data Management and Organizational Change Management
  - Data Handling Ethics

DAMA Data Management Maturity Levels

- Little or no governance
- Limited tool set
- Roles defined within silos
- Controls applied inconsistently, if at all
- Data quality issues not addressed

Level 1 Initial / Ad Hoc

- Emerging governance
- Introduction of a consistent tool set
- Some roles and processes defined
- Growing awareness of impact of data quality issues

Level 2 Repeatable

- Data viewed as an organizational enabler
- Scalable processes and tools; reduction in manual processes Process outcomes, including data quality, are more predictable

Level 3 Defined

- Centralized planning and governance
- Management of risks related to data
- Data Management performance metrics
- Measurable improvements in data quality

Level 4 Managed

- Highly predictable processes
- Reduced risk
- Well understood metrics to manage data quality and process quality

Level 5 Optimized

CMMI Data Maturity Model (CMMI DMM)

Data Management Strategy

Data Governance

DMM MODEL CMMI Institute

Data Quality

Platform & Architecture

Data Operations

DATA MANAGEMENT STRATEGY — Data Management Strategy, Communications, Data Management Function, Business Case, Funding

DATA GOVERNANCE — Governance Management, Business Glossary, Metadata Management

DATA QUALITY — Data Quality Strategy, Data Profiling, Data Quality Assessment, Data Cleansing

DATA OPERATIONS — Data Requirements Definition, Data Lifecycle Management, Provider Management

PLATFORM & ARCHITECTURE — Architectural Approach, Architectural Standards, Data Management Platform, Data Integration, Historical Data & Archiving

SUPPORTING PROCESSES — Measurement & Analysis, Process Management, Process Quality Assurance, Risk Management, Configuration Management

CMMI DMM Maturity Level Assessment

Data Management Strategy, Communications, Data Management Function, Business Case, Program Funding, Governance Management, Business Glossary, Metadata Management, Data Quality Strategy, Data Profiling, Data Quality Assessment, Data Requirements, Data Cleansing, Data Lifecycle Management, Provider Management, Architectural Standards, Architectural Approach, Data Integration, Data Management Platform, Historical Data, Retention and Archiving, Measurement and Analysis, Process Management, Process Quality Assurance, Risk Management, Configuration Management

■ Data Strategy ■ Data Governance ■ Data Quality ■ Data Operations ■ Data Platform ■ Supporting Processes

CMMI DMM Maturity Levels:
1. Performed: Processes are defined ad-hoc, primarily at the project level.
2. Managed: Processes are planned and executed according to policy
3. Defined: Set of standard processes is employed and consistently followed
4. Measured: Process metrics have been defined and used for data management
5. Optimised: Process performance is optimised based on measured metrics

For more information refer to Data Management industry standards
DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition), DAMA International, July 2017. ISBN: 9781634622479
A Guide to the Business Analysis Body of Knowledge (BABOK Guide) [online] http://www.iiba.org/babok-guide.aspx
Data Maturity Model (DMM), CMMI Institute, 2018 [online] https://cmminstitute.com/data-management-maturity
IT Capability Maturity Framework (IT-CMF), The Body of Knowledge Guide, 2nd Edition, Van Haren Publishing, 2016.

Data Management is a part of the EDISON Data Science Framework (EDSF)
https://github.com/EDISONcommunity/EDSF
Contact: Yuri Demchenko <y.demchenko@uva.nl>

EOSC

UvA UNIVERSITEIT VAN AMSTERDAM

UNIVERSITY OF Southampton