

Field Report: Collibra Data Governance Center 4.0

Monday, April 15, 2013

To automatically subscribe/unsubscribe, enter your e-mail address on the form at this [link](#).

You may also unsubscribe by sending a blank e-mail to <mailto:MDM-Alert@tcdii.com> with the subject: UNSUBSCRIBE.

Why is “Data Governance” So Important?

MDM provides a trusted, consistent view of key information assets across the enterprise – ranging from “customers”, “products”, and “suppliers” to “locations” and more. In large enterprises, MDM is becoming a business transformation strategy as the cornerstone of every critical business process and business decision. Data is an important asset to most companies and leveraging that data properly can result in operational and IT cost savings as well as drive business growth. Furthermore, governing strategic data assets is foundational to a service-oriented architecture (SOA), which in turn facilitates business process management (BPM).

Clearly, MDM is an enticing proposition for many enterprises but to achieve these results a proper “Data Governance strategy” must be in place. When deploying MDM, a proper Data Governance discipline should consider the business drivers, project scope, roles and people filling each role, policies and procedures, data quality, inheritability, social norms, and the business operating model. Data Governance of master data is more than a single product or process, rather, it is an ecosystem of products, processes, people, and information. Based on recognition of issues at hand, an improving economy, and increasing regulatory requirements, businesses are now recognizing the opportunity to take a more strategic view of enterprise Data Governance. Additionally, the vendor community (including Collibra, Kalido, IBM and SAP) are delivering the vanguard of Data Governance frameworks expressly integrated with MDM hubs. Clearly, MDM and Data Governance are codependent/interdependent. During 2013-15, savvy enterprises must invest upfront in Data Governance to realize MDM sustainability and ROI.

Data Governance is a major IT initiative being undertaken by a large number of market-leading global 5000 enterprises. Both as an IT discipline and an integrated set of technology solutions, MDM continues to evolve at a rapid pace. Research analysts at the MDM Institute annually produce a set of ten strategic planning assumptions (milestones) to help global 5000 enterprises focus efforts for their own large-scale, mission-critical MDM projects. For product/platform strategy purposes, we provide a set of evaluation criteria that provide a framework of feature/functions we see in increasingly wide practice at IT organizations tasked with evaluating charting and executing their enterprise's Data Governance programs:

The majority of this Field Report on Collibra Data Governance Center 4.0 therefore represents our analyst opinion buttressed by the in-depth reviews, evaluations and (often) hands-on proof-of-concepts executed by the membership of the MDM Institute's Advisory Council.

1. Methodology
2. Data exploration/profiling
3. Data model, policy model & business glossary management
4. Rules/policy management
5. Decision rights management
6. MDM hub integration
7. Enterprise application integration
8. Integrated metrics
9. Multi-level, role-based security
10. E2E data lifecycle support

In summary, Collibra Data Governance Center 4.0 is an integrated product suite to make Data Governance practical and process-driven: implementing Data Governance operating models/processes, increase and sustain end-user adoption and report on compliance, and manage/measure maturity and ROI metrics. DGC 4.0 is an excellent choice for enterprises looking to embark on their Data Governance journey.

Collibra Data Governance Center Evolution

Collibra originally started as a spinoff from the Free University of Brussels, giving it a sound basis in semantic technology. Over the past five years, Collibra has established its presence as a respected and innovative Data Governance software platform. With the launch of its Data Governance Center 4.0 solution, strong 100%+ year-over-year revenue results, and a quickly expanding customer base, Collibra promises to be a market leader in the emerging field of Data Governance software solutions.

Collibra Data Governance Center 4.0 is arguably one of the first-to-market integrated and active Data Governance applications

-- via its governance capabilities for creation, maintenance and replication of master data "upfront" (upstream) of MDM hub deployment.

The focus of DGC is delivering out-of-the-box configurable flexibility via a web-based, collaborative Data Governance platform. Concurrently, the vendor has partnered with several thought leadership consulting firms in Data Governance (e.g., Cognizant Technology Solutions and Information Asset LLC) to deliver both integration with market-leading MDM hubs as well pre-populated business metrics for specific industries and horizontal corporate functions.

Figure 1 - Collibra Data Governance Center 4.0 Family

- **Business Semantics Glossary**
- **Reference Data Accelerator**
- **Data Stewardship Manager**

Source: The MDM Institute

Overview of Collibra Data Governance Center

Collibra Data Governance Center (DGC) is a business friendly, collaborative Data Governance platform. Based on their business role, users can provide input, reviews, comments, proposals, etc. To keep this manageable, DGC has a built-in workflow engine (with both out-of-the-box preconfigured and configurable workflows). This way the platform helps the enterprise to reach out to the right stakeholders at the right time (and make sure it is easy to follow up on any open items).

As a contemporary Data Governance software solution, DGC provides an open architecture and API set that is based on Java, JavaScript, [REST](#) (representational state transfer), and relational DBMS. The overall architecture is classical three-tier (client UI, application server, database server) and is well-behaved in terms of industry standards as DGC is built on and uses open standards such as [BPMN](#) 2.0 (business process model and notation), [SBVR](#) (semantics of business vocabulary and business rules), XML, ...

The DGC platform is comprised of three interlocked subsystems: Business Semantics Glossary, Reference Data Accelerator, and Data Stewardship Manager.

Business Semantics Glossary -- Different terms take on different meanings depending on the context. DGC's Business Semantics Glossary is combined with a data directory to provide answers to questions such as: Which applications are involved as source and targets? Who is responsible for them? Which entities and attribute do they contain? What are our target master data entities and attributes? Who should be involved in the mapping specification, and which agreements are relevant?

Reference Data Accelerator -- DGC's Reference Data Accelerator handles all types of reference data equally efficiently, making it easily accessible and manageable in a business-friendly user interface. Adding new types of reference data, different relations or updated hierarchical representations is performed easily and quickly via the web UI.

Data Stewardship Manager -- One of the single biggest challenges in Data Governance is sustained adoption, DGC drives end-user adoption through business-friendly end-user interfaces, extensive data publication and consumption possibilities, and integration with productivity tools and mobile devices (email approvals, iPad app, portal integration, "search everywhere",...). Collectively this is known as DGC's Data Stewardship Manager. Thus accessible data assets become true enterprise data assets as Data Stewardship Manager supports a structured, process-oriented data management strategy via:

- Automating and structuring the stewardship activities by replacing unstructured spreadsheets and e-mails
- Delivering processes to formalize data quality issue management, data policy creation and compliance, and data sharing agreements
- Providing both manual as well as automated data quality resolution processes
- Assigning artifacts (business terms, policies, rules, issues and reference data) to a data steward via dashboard so that such tasks can be easily managed and monitored

As part of the interactions with its Customer Advisory Council, the MDM Institute captures and promotes models such "top 10 evaluation criteria" for key MDM-related subsystems. During 2013-14 and as part of the background research for the much more comprehensive "**Master Data Governance: Market Review & Forecast for 2013-15**" report, more than thirty Global 5000 size enterprises shared their software evaluation processes and contributed commentary and supporting details for a set of "top 10" evaluation criteria for master data governance solutions. These evaluation criteria (figure 1) are discussed in more detail in the above referenced market study. The majority of this Field Report in turn takes these "top 10" evaluation criteria as a framework to discuss and understand the capabilities of the Collibra Data Governance Center 4.0 solution.

Summary Evaluation - Top 10 Evaluation Criteria

1. Methodology - Collibra's Data Governance methodology is process-enabled such that either IT organizations, or consultancies can easily set up and kick-off organization structures, process and metrics in an "accelerated" mode. Information Asset LLC is one such consultancy that specializes in helping organizations build out their Data Governance programs. Information Asset has created a number of templates in the Collibra environment including: a sample data governance charter, roles, responsibilities, RACIs, data quality metrics, business terms, policies and reference data sets. For the upfront MDM strategy development, both Collibra and its consultancy partners (Cognizant, Information Asset, et al) provide workshops as an introduction to identify data deficiencies and readiness issues when in Data Governance start-up mode. DGC's configurable operating model (which includes "configurable semantics" enable the extension of the metamodel (types, attributes, relations, ...) according to specific needs) as well as:

- Organizational setup -- e.g. Data Governance council and working groups; these can be organized as centralized, decentralized or federated
- Roles and responsibilities
- Workflows and processes

DGC's metamodel templates enable an organization to set up enterprise-level as well as specialized templates for different assets, domains and communities. The actual deliverables as part of the product provides for a "10 day quick start" which includes: software installation setup and configuration; knowledge transfer; and, coaching. In short, "ready to go" Data Governance.

**Figure 2 -
Overview of Collibra DGC 4.0**

STRENGTHS

1. **Web-based, collaborative Data Governance platform**
2. **Includes all key DG functions-- Business Semantics Glossary, Reference Data Accelerator & Data Stewardship Manager**
3. **Configurable predefined, operating model**
4. **Embedded business metrics**
5. **Supported by major Data Governance consultancies**

CAVEATS

1. **Modest proof points for integration with IBM MDM & Informatica MDM**
2. **Modest market traction to date--** Aspen Re, Barry Callebaut, COLT Technology Services, Eandis, ENTSO-E (European Network of Transmission System Operators for Electricity), Flemish Government, HSBC, ING, Janssen Research, Johnson & Johnson, Kyivstar GSM, NetApp, Spectrum Health, Standard & Poor's, VDAB
3. **Cloud strategy needs refinement**

Source: The MDM Institute

2. Data Exploration/Profiling - DGC integrates with various market-leading data quality tools (e.g., Trillium TS Discovery and Informatica Data Quality (IDQ) with IBM QualityServer planned for 2H2013]). The results of the profiling can be fed into DGC (e.g., candidate reference data, as input to approval process, thresholds and failing data rows, ...). DGC is a business user-focused, collaboration platform that is used by business stakeholders and data stewards to define business policies, business definitions, business rules, data quality metrics etc. and link that to the actual results of data quality measurements. Based on the governance operating model: roles, responsibilities and a configuration workflow engine, the platform can easily involve and coordinate all necessary stakeholders to review and approve rules. By using DGC with a DQ platform the organization can easily report data quality outputs (thresholds, failing rows, metrics, ...) to the owners in the different parts of the organization with the relevant roles and responsibilities. This includes the ability for organizational roll-up (e.g., comparing Sales vs. Finance across different lines of business). Based on these quality thresholds, DGC will automatically create data issues from failing business rules. Its business-facing, collaborative environment allows business stakeholders and data stewards to become maximally involved with minimal effort in resolving these data issues. DGC will intelligently route issues to the correct stakeholders based on the chosen governance operational model; and different issue resolution processes can be easily configured.

**Figure 3 -
Data Governance Platform
Evaluation Criteria**

1. *Methodology*
2. *Data exploration/profiling*
3. *Data model, policy model & business glossary management*
4. *Rules/policy management*
5. *Decision rights management*
6. *MDM hub integration*
7. *Enterprise application integration*
8. *Multi-level, role-based security*
9. *Integrated metrics (business, technical, & philosophical)*
10. *E2E data lifecycle support*

Source: The MDM Institute

3. Data Model, Policy Model & Business Glossary Management - The fundamental architecture of DGC is to first gain consensus about business vocabularies/semantics. As noted by leading consultants, an important first step in every Data Governance initiative is the identification and prioritization of critical data assets: what are the business goals and which critical business elements and attributes are related to those goals? One has to agree on the meaning of those key business elements and critical attributes, as they mean different things depending upon their context. The underlying metadata of DGC provides a unified business glossary environment for business intelligence and other knowledge workers to share their efforts at the business semantics level. To further unify all components, DGC includes a Global Search capability and can link to other BPM or workflow systems.

4. Rules/Policy Management - Unlike other non-integrated, non-active data governance solutions, DGC has a robust business rules management capability via the Data Stewardship Manager subsystem that is integrated with and shares metadata and policies with the target MDM hub application whose master data it governs. Initially the list of partners includes Cognizant for IBM MDM Hub and RealDolmen for Informatica MDM hub. Additionally, DGC's business rules can be used to drive or extend the governance via its "developed once, used many" repository which can be used by knowledgeable business end-users. This approach to "policies" provides guidance and control for how data should be handled (e.g., data sharing, security, approval, etc.). The organization can also set up detailed templates (e.g., description, exception scenario, and frequency) to collaboratively create and manage policies-- and then break them down into more specific business and data quality rules to link policies to other existing assets (e.g., business terms, metrics and data elements) to establish traceability. This provides practical management of data stewards and their more advanced assets: policy management, business rules, data quality issues, data sharing agreements, and out-of-the-box governance workflows.

5. Decision Rights Management - Support for governance councils' and committees' decision-making processes is built into the workflows that are provided out-of-the-box for LOB- and corporate-function specific roles -- whether it be material master data manager or other such roles. In short, the built-in workflows can be used to automate the operational aspects of Data Governance, and can be tailored to fit the organizational-specific process an organization may have in place (or evolve) to support arbitration, adjudication, escalation, and all the normal "decisioning" processes associated with data stewards and their councils/committees. Issue management in DGC addresses data issues which can arise anywhere: misunderstood reports, bad data entry, policy changes, misalignment, etc. When it comes to issue management and to actually resolving an individual issue, DGC enables a data steward to tap into a pre-defined decision rights hierarchy for certain cases (e.g., cross functional issue where bad data is actually bad because of missing source rules/process).

- Issue management uses the built-in decision rights management within DGC. The entire decision rights management is quite extensive and can be set up to match an organization's needs (including "federated" organizational models) via DGC's configurable roles/responsibilities (as part of the operating model) where a role represents a grouping of rights and responsibilities
- Fully configurable escalation, voting, decision making, and notification processes
- Assignment of roles at various hierarchy levels to determine scope of decision rights (i.e., "Sales" vs. Finance, "Product Domain" vs "Customer Domain", or "USA" versus "US" reference data value)
- Fully configurable hierarchies

6. MDM Hub Integration - DGC currently provides for bi-directional integration with a limited number, albeit, market-leading MDM hubs. It is a centralized solution which can be deployed as a hub or can be installed within the same instance as your operational MDM hub solution. DGC is most often deployed as the central creation hub/engine and therefore it "is" integrated with the system of record hub unlike the vast majority of Data Governance solutions currently marketed which do not integrate with any MDM hub at all. Additionally, it can be deployed in-line as a staging consolidation hub if desired. DGC also includes provisioning capabilities, e.g., "snapshots" can easily be delivered into MDM hubs through SQL or web services interface. Such connection to MDM hubs are also easily established in the coordination workflow (e.g., publish/sync to MDM web service of a new set of reference data immediately after approval by data steward).

7. Enterprise Application Integration - DGC provides an import/export end-user wizard for unstructured (TXT, DOC, PPT, etc.), semi structured (CSV, XLS) and structured (XML, XMI, XSD, RDF/OWL) I/O. Its governance engine provides other enterprise application integration (EAI) via Java and Web Services APIs. Also included are DGC's API (web services & REST) which can be used for EAI at multiple levels. Web access to DGC's objects and processes enable OEMs, consulting partners and IT organizations to further customer the Data Governance processes as might be required. DGC also supports creation and maintenance of reference data as might be expected of RDM (reference data management) capability found in a full function MDM platform. Specifically, it supports the creation of reference data "snapshots" which are files that can be exported to an operational data quality/ETL environment such as Informatica PowerCenter, IBM InfoSphere DataStage or Ab Initio.

8. Integrated Metrics (Business, Technical & Philosophical) - In addition, metrics are mappable to defined SLAs to trigger reminder or supervisory emails as well as other processes. Most importantly, in terms of closed-loop Data Governance processes, once such metrics are implemented upstream in DGC, they can also be measured downstream. Indeed, a lot of the need for technical governance metrics is removed as the business rules are managed much tighter, i.e., controls are much more upfront than downstream. If certain governance rules are too complex to define and not desired to be rigidly managed upfront, then DGC's Data Stewardship Manager can periodically run the technical rules if desired. Philosophical metrics such as "Data Governance readiness" or "maturity level assessments" are typically deliverables from organizations such as Cognizant, EMC Consulting, Information Asset, and other consultancies. For example, scorecards to see how well an organization is performing (e.g., one might be good at data quality, but have poor process quality).

9. Multi-Level, Role-Based Security - Within DGC, security is set up out-of-the-box as multi-level and role-based where "roles" are groupings of rights and responsibilities. These roles are fully configurable: i.e., organizations can easily set up their own roles (e.g., data custodian vs. data steward) at any time, including a detailed configuration of what the role is allowed/not allowed to do (permissions), as well as what the role should/should not do (e.g., responsibilities in context of a workflow). By assigning a role to an individual person or group of people at a certain level in the hierarchy (integration with identity management systems via LDAP is available), an organization can further specify scope of responsibility per the below scenarios:

- "Sales" level data steward - can do everything in Sales (and all domains/assets owned by Sales)
- "Customer Domain" level data steward (which may be owned by Sales organization) - can do everything in this domain for all the assets which are part of this domain
- "Individual asset" level data steward (e.g., "Customer") - can only perform Data Governance functions pertaining to this asset

10. E2E Data Lifecycle Support - DGC provides systemic end-to-end, full lifecycle of information as an asset, i.e., onboarding (new account creation), promotion, retirement, and redaction. As an E2E (end-to-end) lifecycle governor, it provides for master data to be managed via workflow and staging areas so any changed master data will be fully audited, and checked to see whether it should be published. Specifically, a "change request" in DGC is the mechanism to control the information lifecycle. All the change requests are stored in DGC and can be reviewed as a historical record of the governance process. Moreover, documents can be attached at time of creation and across the audit trail with such attachments.

Competitive Outlook

Prior to the release of Collibra's Data Governance Center 4.0, organizations that desired enterprise-strength data quality as part of MDM infrastructure depended upon systems integrators (SIs) and consultancies for such critical capabilities. As a result, major SIs such as Accenture, Cognizant, Deloitte, Tata, and Wipro developed very significant skills and pools of staff to develop custom "Data Governance" frameworks to solve these MDM data quality challenges.

With the formal delivery of the mega vendor's own Data Governance capabilities (IBM, SAP especially), and the increasing uptake in the market, these same SIs and consultancies will need to rapidly remake themselves as "value add" contributors to the increasing install base of MDM hub sites. During 2013, we expect the market to modestly embrace the mega vendors' Data Governance capabilities (due to cost, complexity and lack of overall integration) which will enable the SIs and consultancies to continue selling and deliver their custom solutions, while concurrently building out their own mega vendor Data Governance-to-MDM-hub credentials and capabilities. We project that certain smaller consultancies such as EMC Consulting, Information Asset, etc. will totally embrace the DGC platform to gain market momentum, while other much larger SI firms such as Cognizant Technology Solutions will add Collibra to their portfolio of Data Governance accelerators as they look to exit the custom Data Governance framework business (with its associated internal software development costs).

Heterogeneous Data Governance software solution providers such as Kalido and SAS/DataFlux will continue to deliver Data Governance solutions that directly integrate with (and innately require) their own Kalido MDM and DataFlux qMDM respectively and as a result are not direct competitors.

BOTTOM LINE

For enterprises approaching governance of master data, Collibra Data Governance Center 4.0 can provide an integrated, active solution with out-of-the-box configurability to support multiple organizational governance styles. This web-based, collaborative platform promises ongoing sustainability and lower TCO relative to the alternatives of custom frameworks or mega vendor-proffered Data Governance solutions.

Specifically, DGC leverages existing investments in process-oriented and semantics-driven Data Governance along with assets from select major Data Governance experts. While DGC is a relatively new product (compared to other solutions such as Kalido Data Governance Director and Informatica Data Director), its foundation in business semantics glossary and workflow rather than solely business process management (BPM) or data stewardship consoles provides a significant edge in delivering sustainable Data Governance. This governance-in-a-box approach will enable DGC users to quickly ramp up a Data Governance environment that is tailored to the business community's processes and semantics-- thus insuring ongoing success and sustainability with minimal IT support required.

A challenge for Collibra marketing is that the company needs to break through the perception that a small software firm can provide solutions for large North American or EMEA enterprises. On the other hand, it helps that distinguished system integrators and Data Governance specialist consultancies invest in the company via their selection of Collibra DGC 4.0 as platform of choice to deliver their own Data Governance solutions.

Coming to market during 2013-14 are Data Governance solutions characterized by pro-active integration of the governance function with MDM hubs. Collibra's DGC is in the vanguard of such products and is an excellent choice for enterprises looking to leverage a Data Governance-process oriented solution that includes integration with several marketing leading MDM hubs.

In summary, Collibra Data Governance Center 4.0 is an integrated product suite to make governance of master data practical and process-driven. It provides flexibility to implement commonplace Data Governance operating models to increase and sustain end-user adoption, to report on compliance, and to manage maturity and ROI metrics. DGC 4.0 is an excellent choice for enterprises looking to embark on their Data Governance journey.

See you at the next annual MDM & Data Governance Summit in your hemisphere.



Aaron Zornes

Chief Research Officer

www.the-MDM-Institute.com

Independent, Authoritative, & Relevant

****SAVE THE DATES****

More MDM programs get their successful start at MDM & Data Governance Summits than anywhere else

- [MDM & Data Governance Summit Europe 2013](#) -- Radisson Portman BLU - London ▪ April 15-17, 2013
- [MDM & Data Governance Summit Asia-Pacific 2013](#) – Doltone House, Darling Island Wharf – Sydney ▪ May 21-22, 2013
- [MDM & Data Governance Summit Tokyo 2013](#) – Belle Salle Iidabashi First – Tokyo ▪ June 11-12, 2013
- [MDM & Data Governance Summit Canada 2013](#) – The Carlu - Toronto ▪ June 26-27, 2013
- [MDM & Data Governance Summit San Francisco 2013](#) – Hyatt Fisherman's Wharf – San Francisco ▪ July 17-18, 2013
- [MDM & Data Governance Summit New York 2013](#) – Marriott Marquis Times Square – NYC ▪ October 20-22, 2013
- [MDM & Data Governance Summit Singapore 2013](#) – Marina Bay Sands Resort – Singapore ▪ October 2013

About the MDM Institute

The MDM Institute is the world's leading research and advisory consultancy exclusively focused on master data management. As chief research officer, Aaron Zornes delivers the technology-related insight necessary for its clients to make the right decisions in their use of master data management (MDM), customer data integration (CDI), reference data management (RDM) and data governance solutions to achieve their customer-centric business goals. The MDM Institute provides authoritative, independent and relevant consulting advice to senior IT leaders in corporations and government agencies, to business leaders in high-tech enterprises and professional services firms, and to technology investors. The MDM Institute delivers its research and advice to more than 60,000 clients in 10,500 distinct enterprises via Twitter, Linked In, Xing, Google+ and email newsletters. Additionally, each year more than 2,000 paid delegates attend its MDM & Data Governance Summit conference series held in London, New York City, San Francisco, Singapore, Sydney, Tokyo and Toronto (now in its seventh year). Founded in 2004, the MDM Institute is headquartered in San Francisco and has clients primarily in North America, Europe and Asia-Pacific. For more information, visit <http://www.the-mdm-institute.com>.

For additional info on this topic or other MDM Institute offerings, please contact info@the-MDM-Institute.com.

TO UNSUBSCRIBE

To automatically subscribe/unsubscribe, enter your e-mail address on the form at [this link](#).

You may also unsubscribe by sending a blank e-mail to MDM-Alert@tcdii.com with the subject: UNSUBSCRIBE.

© 2013 The MDM Institute. MDM Alert™ is a registered trademark of the MDM Institute.