Data Management activities
Centralize, decentralize or best-of-both-worlds?

Ard van der Staaij MSc RE and drs. Jeroen Tegelaar

There is a high need of standardization for better data quality within organizations. Most resulting from internal initiatives, like process optimization or requirements deriving from law and regulations. Standardization doesn't have to mean that an organization needs to centralize all activities to achieve control. There are different ways to manage the activities around data within an organization. This article gives an introduction on the topic, including some practical examples.

A. van der Staaij MSc RE
is a manager at KPMG Advisory N.V.
Risk Consulting, Enterprise Data
Management.
vanderstaaij.ard@kpmg.nl

Drs. J.A.C. Tegelaar
is a senior manager at KPMG IT Advisory
and one of the core members of the
Enterprise Data Management team.
tegelaar.jeroen@kpmg.nl

Introduction

The absolute net value of data management activities within companies has yet to reach its peak. Departments have an increasing need for good quality data for reasons of analysis, compliance, growth and efficiency. Maintaining their critical data is for that reason alone the most important activity in the era of digitalization we live in. Therefore, companies should think hard how they want to position and implement these activities for the long term and to consolidate it within a professional data management office.

The evolvement of the Data Management Office

There are several trends visible in the positioning of the data management office over time. But before we continue, we first need to clarify what the definition of a data management office is. We define a data management office as 'a delegated function of data management activities from the designated data owner'. For example, data maintenance activities and the management of the data lifecycle. The trends identified in figure 1 follow the evolution of the digital developments within organizations:

1. *Data management within the IT function*: back in time when organizations first started using automated systems (pre-ERP) as administrative tooling, data was just a record 'doing nothing'. When issues occurred, this was seen as an IT issue. Obviously, ownership of data was within IT.

2. *A first step towards the business*: starting to realize that data is the fuel of many primary business processes, organizations positioned data management activities in the back office of the business. These activities were primarily administrative and operational and did not yet concern the cleansing and/or enrichment of data.

3. *Standardization and centralization*: next, organizations started realizing the need for standardization and even centralization of data management activities for different reasons. For example, an ERP implementation, process integration activities; such as optimizing the supply chain and legislation and regulations (for example: SolvencyII, BCBS239, but also data protection regulations). Data collected in 'secondary systems' such as file shares was still left most of the time to the whims of the individual, with an incidental records manager as the proverbial exception to the rule.

4. *Hybrid data management organization*: currently many companies are struggling with the question of how far should they go with the centralization of the management of data. A centralized data management office often means long lead times, sometimes losing connection with the business. But a decentralized data management office (often within business units), means inconsistencies, less overall view, no specific knowledge of the business and the impact on the IT infrastructure.
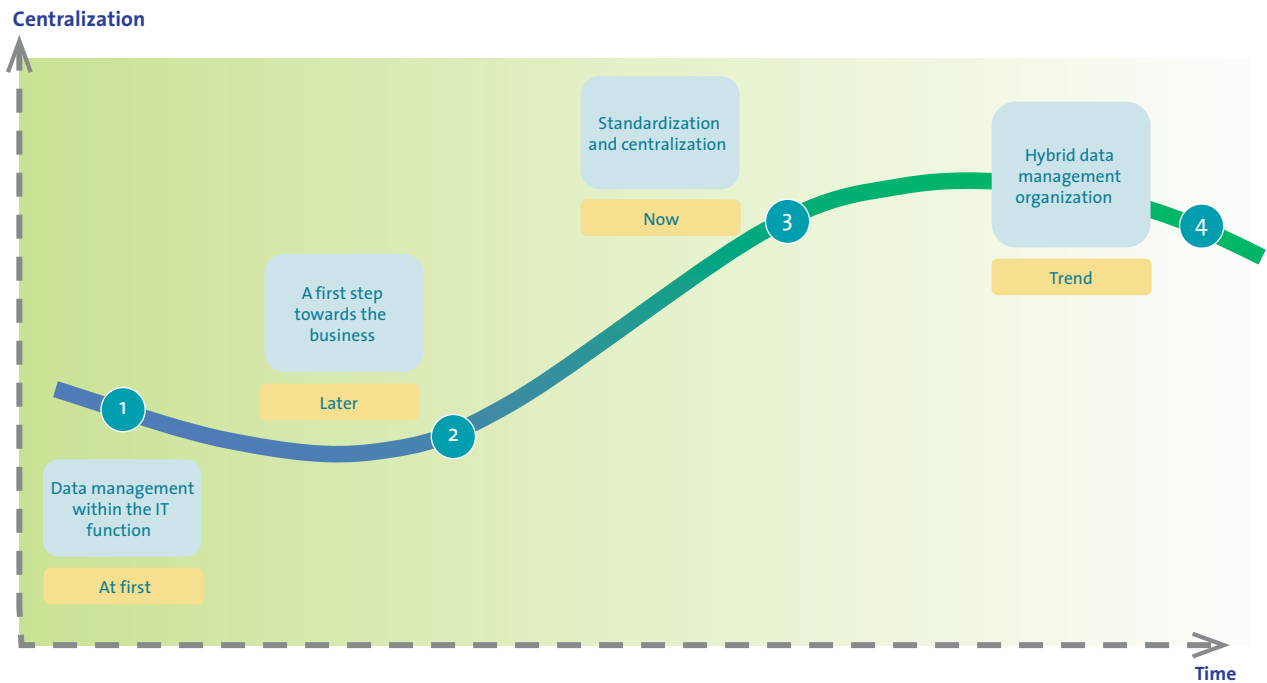
Centralization

Standardization
and centralization

Now

Hybrid data
management
organization

Trend

A first step
towards the
business

Later

Data management
within the IT
function

At first

Time

Figure 1. **The evolvement of the Data Management Office.**

In other words, there is no right or wrong in centralizing or decentralizing data management activities. Since, we are discussing the positioning of the data management office, there first needs to be a common understanding of two statements. These statements are a pre-condition for successful data management:
• Data governance needs to be organized top down at a strategic level ([Unen12]). A pre-condition for centralizing or delegating data management is a well-organized data governance structure. Data governance facilitates data ownership and then makes it possible to successfully centralize or delegate data management activities.
• Data should always be owned by the business ([Jonk11]). Independent from where the data management function will be positioned, the business is always accountable for their data. Shared service centers or data management offices therefore always have a delegated accountability to maintain the data which is primarily owned by the business (for example: a G/L account is owned by Finance, but can be maintained in a shared service center outside of the finance department).

The above statements will not be further discussed here, but are seen as a condition sine que non for organizing the data management activities.

In the next paragraphs, we will first identify the most important types of data management offices and the most important influencing factors before we elaborate on the positioning of these offices. We will cluster this in a tangible tool (see figure 4) which can be used to facilitate the discussion around centralizing data maintenance activities. This is made tangible with an example of product or material data.

Types of positioning the Data Management Office

As presented in the statements above, data governance needs to be organized top down at a strategic level. The underlying fact of most data management models is that data governance should be centralized as a precondition for an efficient data management office. In the end, dele-

We define a Data Management Office as 'a delegated function of Data Management activities from the designated data owner'

gated operational activities are in need of guidance. A simple example; delegating the data maintenance activities around material master data are in need of guidance. It is not only administrating data, but also validating and enriching the data based on standard definitions and business rules. Three operational models can be recognized, which will be elaborated hereafter.

Centralized

Data will be maintained centrally. The data management office handles all data requests centrally, with no intervention from other stakeholders. This is often a delegated function from a data owner in the business.

Hybrid

Data is maintained partly centrally and partly within other organizational departments. But, the complete data maintenance process end-to-end is managed centrally.

Decentralized

In a decentralized environment all data management activities are managed decentrally. There is no overall data management office managing the process.

Factors of influence: data dimensions, level of automation and available expertise

The choice of the best data management model is dependent on multiple influencing factors: data dimension, the level of automation and available expertise. In choosing the position for your data management office, the first thing to do is identifying the data dimensions that need to be maintained. For example, is the data maintained and used by more than one business function? Secondly, you should focus on the question which level of automation is required and for what processes? Finally you should answer the question, what is the level of expertise within the business related to data management? Can they maintain the quality of data at a decentral level?

Data dimensions

In the context of the decision to maintain data centralized or decentralized the data is categorized into four dimensions: corporate data, shared data, process data and local data. The key reason is that data which is maintained or used from an overarching perspective is potentially relevant for centralized data management activities.
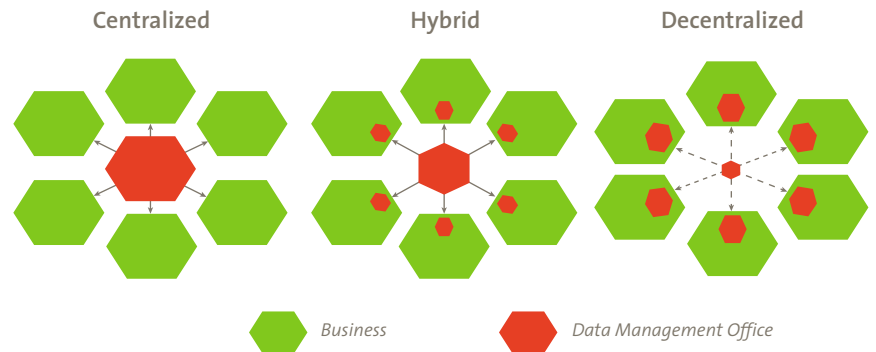

Figure 2. **Types of positioning the Data Management Office.**

Corporate data

External verifiable data is defined as corporate data. There is only one version of the truth which is valid for the entire organization. Corporate data is often externally supplied, for example acquired by a data supplier, such as GS1 where organization can acquire a set of GTIN (EAN) to uniquely identify their products. Also, this data can be delivered via a business relation (client, supplier, etc.). Since there can only be one version of the truth in the entire organization, the data management activities should be organized in a manner that for example the risk of inconsistency in the data and the use of outdated data is mitigated. Governance and even maintenance should be conducted preferably centrally.

Shared data

Like corporate data, there is only one version of the truth throughout the entire organization. Shared data is valid for the entire organization and used for multiple purposes within multiple processes, Such as the name/description of a product. This description should be unique throughout the entire organization. The difference is that this data is not externally verifiable. The risks are the same as with corporate data, but it takes more effort to control the data quality, since the data is not externally verifiable. Governance and maintenance should be managed preferably in the same manner as corporate data.

Process data

A variant of shared data is so called process data. This is data which is used over different disciplines but within one business process. A good example is that of a material or product which has multiple units of measures. One for commercial purposes (define planogram), one for logistical purposes (delivery on pallet or roll container), etc. Independent of the place of maintenance the governance

Data Dimensions	Example	Governance
Corporate data	DUNS number, GTIN, Bank Account Number, VAT number	Central
Shared data	Billing Address, Name, Description, Contracts, Reports	Central
Process data	Pricing, Logistical Unit/Transportation Unit (Pallet, Roll container, etc.)	Hybrid
Local data	Marketing data (only used by one department), pictures department day-off	Decentrally

Table 1. **Examples of data dimensions.**

should be managed centrally, but operational activities could be deployed decentrally.

Local data

Local data is only relevant for one discipline, within one process, one department for one purpose. If this data is incorrect it will affect only that specific department and not a complete process or even the entire organization. The risk is therefore relatively low, which means there is no direct need for central governance.

Level of automation and level of expertise

There are multiple factors to define whether data should be maintained centrally or decentrally. In the end these factors determine whether data can or should be maintained in the business (at source) or in a centralized department, which is purely focusing on data management activities.

As data is owned by the business, the data should preferably be maintained close to the source: the business ([Unen12]). In conclusion, the statement: 'Data should be maintained in the business, unless…' is used to define whether this is true or data should be maintained in a centralized data management office. Two important influencing factors are related to the level of automation and the level of expertise in an organization, related to data management activities.

The level of automation defines whether activities can be decentralized

The data management architecture within an organization is highly relevant to decide on the centralization of data management activities or not. As stated in [Unen12] there are three types of data management architectures. This example relates to Master Data Management (MDM), a specific field of expertise within Enterprise Data Management (EDM).

The first model refers to a consolidated approach, where data is maintained in several applications and consolidated in one MDM environment. Purely for the facilitation of consolidated reporting. In the second model, harmonized architecture, the operational maintenance is still in several applications. The difference is that data is centralized and governed in the back-end (MDM) solution, with push functionality to the feeding applications to have the golden record in place. At last we recognize a centralized MDM solution in the architecture, which is feeding all relevant applications using this data. In this last situation all data maintenance activities are centralized in one system.

A data maintenance process can be managed procedurally or it can be automated (or both). This is especially relevant when data ownership (source of data) is divided across multiple stakeholders. A valid and monitored working procedure or workflow needs to be in place to manage the

The key reason is that data which is maintained or used from an overarching perspective is potentially relevant for centralized Data Management activities
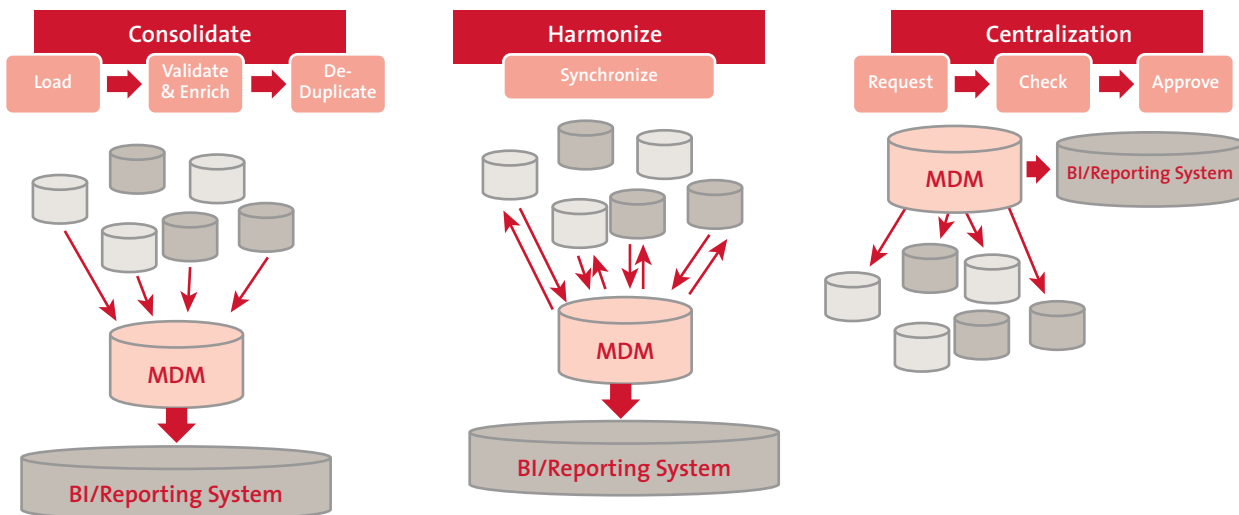
Figure 3. Representation of three types of MDM architectures ([Unen12]).

result: complete, accurate and timely data. Therefore, the more data maintenance activities are centralized and automated in a single system (at source) the more activities can take place in the business (decentralized). The main reason is that specific tooling can provide workflow functionality (request, check, update, approve and distribute) to manage the process over multiple stakeholders. A lot of reasons to centralize activities are not relevant in this situation, because extensive workflow functionalities can mitigate a lot of risks, for example:

• *Efficiency*: SLA/timings can be implemented in each workflow step;
• *Quality*: data quality checks can be implemented in each workflow step;
• *Responsibility*: workflow makes the process visible and easier to manage and employees gain a greater insight into how their activities affect other activities in other process steps.

If activities are decentralized in different systems, a strict process management procedure needs to be in place. The relevancy to centralize these activities is more relevant in this situation. This is because efficiency, quality and responsibility is harder to manage decentralized.

Level of expertise in the business

Of course not everything can be automated. Knowledge and expertise to perform data management activities in complex environments cannot be automated completely. Therefore each data management activity should be checked to see whether knowledge and expertise to perform this activity can be transferred to for example back-office functions. The employees managing the data should understand the relationships between data dimensions and attributes to be able to perform an impact analysis on how data will be affected when a new data attribute

The more data maintenance activities are centralized and automated in a single system, the more activities can take place in the business (decentralized)

Knowledge and expertise to perform Data Management activities in complex environments cannot be automated completely

is introduced or an old one is changed. They should be able to perform control measures on data quality and act when people start deviating their way of working from the desired one.

Decision tree Data Management Office

A decision tree can be used to decide whether data should be maintained in the business (decentralized) or in a data management office (centralized). All elements discussed above are integrated in this decision tree. The decision tree should be used separately for each data object which should be maintained.

The starting statement is as follows: 'Data should be maintained in the business, if…':

1. *the data dimension is local data*: local data can be maintained in the business. There is no risk of incorrect data which is used over multiple disciplines, departments or processes. Local data is often very specific with no need to delegate or centralize the data management activities of this data dimension. Corporate data, shared data and process data can potentially be considered for central data management in a data management office.
2. *knowledge and expertise cannot be transferred (level of expertise)*: if the knowledge and expertise needed for data management activities is very complex (in business expertise) and therefore it is not possible to transfer this knowledge to a delegate, these activities should be preferably in the business (at source).
3. *there is a high level of automation*: a high level of automation facilitates data management over multiple disciplines, departments or processes. Within one single source of truth and management, workflow manage-
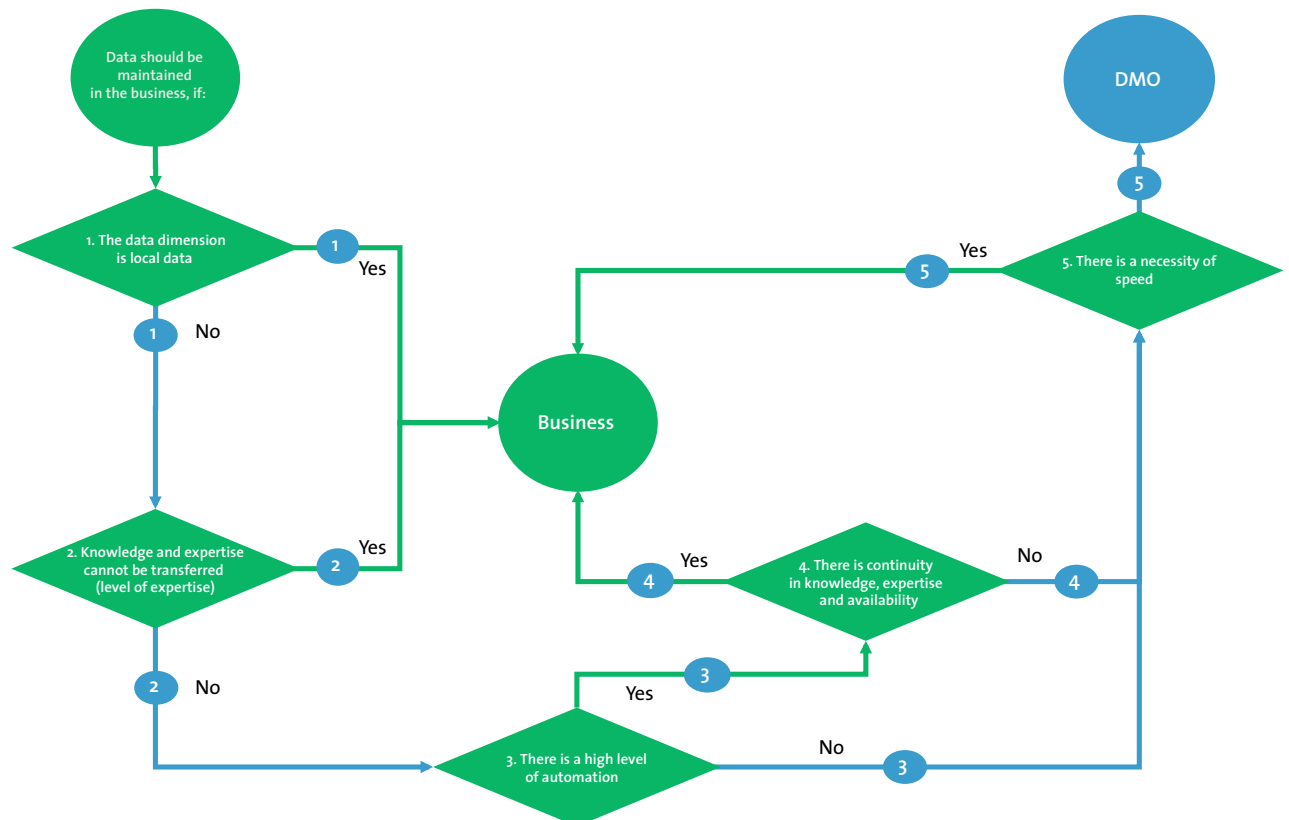

Figure 4. **Decision tree Data Management Office.**

ment (and related metadata) can facilitate the data management activities over multiple stakeholders.

4. *there is continuity in knowledge, expertise and availability*: continuity of data management activities is critical. Data management is no longer just data entry. There is specific knowledge and expertise needed to perform these tasks. For example the continuity of the personnel performing these tasks is relevant in defining if tasks can be done at the business or should be centralized. For example a sales assistant enters their own customers in a system. The job rotation in these functions is high and after one or two years the sales assistant is promoted. Tasks, knowledge and expertise is lost. To have sustainable knowledge and expertise available, continuity is very important in data management. This could therefore be a reason to centralize activities to a dedicated team, where rotation is lower and focus and knowledge management can more easily be integrated in their daily routines.

5. *there is a necessity for speed*: all decisions for centralized data management have been made. Centralization means delegation, which means an extra process step. What if the data needs to be directly available for use? For example a new customer calling a sales assistant requesting an order. The necessity of speed in the process is therefore a relevant influencing factor in deciding to centralize activities or not.

The outcome of using the decision tree is not directly set in stone. This decision tree is a tool to think about the reasoning for centralizing activities or not. The tool can be used for each data attribute (field), set of attributes or at an entity level.

Applying a practical example: product or material data

The management of data about a product or material is often complex in organizations. An example is given to make the use of the decision tree slightly more tangible. But first a quick introduction on some key benefits in the example organization and its product or material data management process.

1. *Receive external data*: different external data sources are the fundament of the reflection of the product or material in the system. Data could be derived from suppliers or for example external data pools. This is often corporate data such as GLN/EAN, with in addition shared data as, for example, the description of the product and unit of measurement (weights, measures).

2. *Enrich with internal data for primary processes (e.g. relevant for ERP)*: each discipline within the organization, relevant to a product or material is owner of several extra attributes, mostly specific for their process: process data. Examples here are pricing conditions (commercial department) and logistical unit/transportation unit (logistics).

3. *Enrich with internal data for additional purposes (e.g. e-commerce)*: next to data which is used for the traditional transactional ERP processes, other activities within the

Influencing factor	Relevance for product or material	Result
1. If the data dimension is local data.	All product or material data mentioned in the example above is not local data. Therefore all data is used for multiple purposes.	DMO
2. If knowledge and expertise cannot be transferred (level of expertise).	The knowledge of a complete (end-to-end) product or material is transferrable to delegates. If this case mentioned actuarial data (financial risk calculations) there might be an issue here, because of the complex matter.	DMO
3. If there is a high level of automation.	As stated in the example there is a high level of automation. Workflow functionality manages the data management process over various disciplines, where most administrative tasks are automated.	Business
4. There is continuity in knowledge, expertise and availability.	There is a low fluctuation in business function. When knowledge and expertise is steady in business positions, there is a basis for continuity.	Business
5. There is a necessity of speed.	Not applicable in this case.	N/A

Table 2. **Examples of using the decision tree on product or material data.**

company may need far more data for a product or material. Examples here are mainly omni-channel activities, for example, online. Additionally commercial texts, food & beverage information and other digital assets need to be maintained.

Additional context

The data architecture in this example is highly automated and centralized. The external data sources are interfaced into data management tooling, where advanced workflow functionalities manage the enrichment and validation of the data. Further, this organization is a mature business entity with very few fluctuations in business functions.

Since there are multiple data sources and a complex data management process the question arises: should we maintain the data centrally or can we manage to let each discipline maintain their own part of the data? In table 2 some examples show how to use the decision tree.

Conclusion

It is again important to state that there is no right or wrong when positioning a data management office. Before initiating a data management office it is important to identify the critical data and the level of expertise in the organization. In practice there are many cases where organizations forgot the first and acted on the latter, with the result an office full of people who were still formally performing data activities for their previous divisions while doing nothing on data management for the organization as a whole. We have also seen organizations doing the first while forgetting the latter, with the result a highly efficient design and target operating model for data management but with no commitment from departments and no commitment on dedicated or capable resources.

Secondly, data governance remains the critical precondition for success. Without ownership of data people will always give preference to local interests, instead of the interest of the organization as a whole. And it should be enforced from the top down, to ensure standardization, consistency and action on the right scale.

Finally, automation is instrumental in achieving your goals, but IT tooling is just one of the instruments you should use. Alone it can be a solution for the short term, but it will become a problem for the long term. Without the other organizational measures the IT department will again own the data and that is something the business will

not accept anymore. The level of automation when used properly will facilitate the movement of data management closer to the business, which is the actual owner and source of the data. Knowing your data is owning your data, and your office is the key.

References

[Jonk11] R.A. Jonker, F.T. Kooistra, D. Cepariu, J. van Etten and S. Swartjes, *Effective Master Data Management*, Compact 2011/0.

[Unen12] C.W.J.A. van Unen, A.S.M. de Goeij, S. Swartjes and A.J. van der Staaij, *Master Data Management: Do's & Don'ts*, Compact 2012/2.

About the authors

A.J. van der Staaij MSc RE is a manager at KPMG Advisory N.V. Risk Consulting, Enterprise Data Management and has over 7 years of experience within enterprise data management consultancy. He is specialized in master data management, but has a broad background in multiple other data management disciplines, like data governance, data quality management and information management. He is primarily focused on the organizational and process component within enterprise data management and delivered solutions to clients within a wide range of industries.

Drs. J.A.C. Tegelaar is a senior manager at KPMG Advisory N.V. Risk Consulting, Enterprise Data Management and has over 15 years of industry and advisory services experience. He has a background in Data Governance, Data Quality, Data Migration, Data Privacy and Enterprise Content Management. A lot of his engagements are about supporting companies in identifying and translating regulatory requirements on data to practical action for global business.