



The Four Primary Approaches to Data Quality Management

SUMMARY

Data Quality Management is a complex and varied set of approaches and tools. The large number of these approaches and tools can be confusing for both Data Management professionals and business people alike, and for all who attempt to maintain or improve the quality of data used by their organisations. This paper helps you find your way through this complex landscape by:

- *Providing a simple explanation of the main approaches*
- *Highlighting the main features of each approach*
- *Outlining the main strengths and shortcomings of each approach*

AUTHORS

This paper is the first output of the DAMA UK Working Group on “Approaches to Data Quality Management”.

Chris Maynard – Transforming Information

Zoey Husband – Ford Credit UK

Nigel Turner – Trillium Software

Neil Harvey – Diversey, part of Sealed Air

INTRODUCTION

While researching for this paper we have investigated approaches to maintaining and improving the quality of information derived from the electronic data in an enterprise. We have chosen to call these approaches, “Data Quality Management” (DQM). In our thinking, Data Quality Management is a subset of Data Management.

Distilling the published methodologies for Data Quality Management reveals four recurring themes. We call these the Primary Approaches to DQM. (It is possible that they apply to other areas of Data Management, such as Data Security, but we have not yet investigated them.) Most methodologies emphasise ONE of these primary approaches, some combine two, and a few recognise all four.

Each approach has its merits. Each has its shortcomings. We will now consider each in turn.

The Four Primary Approaches to Data Quality Management

PEOPLE-BASED APPROACHES

Changing people's attitudes and actions towards Data Quality Management through training, education and community building

A people-based approach assumes that the main cause of poor Information Quality in organisations is people. People can and often do create, amend and delete information and data wrongly. Often people do not understand the value of the data assets they work with and so fail to preserve, enhance or use those assets. So we focus on the people within the organisation, improving their awareness and skills in Data Quality Management. This will in turn address the basic shortcoming of Data Management, that of human ownership. Data and information are not there to generate work for computers. They are there to help people make decisions.

People-based approaches encourage people to communicate and work together to solve their information problems. They tend to regard Data Quality Management as a cultural and business change challenge - business transformation is the key driver. IT systems and process are just enablers. Outcomes of this approach include some interesting "stewardship", "community of practice" and "bodies of knowledge" (BOK) methodologies which, unlike the other approaches, provide a clear role for a Human Resource function in Data Quality Management.

Typical Environments

This sort of approach is often used in Research and Development departments of large organisations, particularly if they are geographically dispersed. These approaches often work well in groups of professionals working with a high degree of autonomy such as medics and lawyers. In contrast it is also effective in very small organisations, where just getting people to do things consistently can make the biggest difference.

Strengths

People-centric approaches result in strong and sustainable Stewardship and Governance models. IT Tools come and go, processes can be re-written. These approaches appreciate that people's behaviour, skills and attitudes are important in managing data and information. They recognise the importance of human error as a cause of poor data quality. They can be very cost effective and very motivating.

A key strength of a people-centric approach is it can produce results very quickly in areas where data quality is down to human error or lack of understanding. In communities of practice it may require little more than an evangelical figure and a few willing disciples as a driver.

Shortcomings

Frequent shortcomings are sustainability and penetration. There is only so much you can do in changing behaviour; it requires willingness on behalf of both the "educator" and the "trainee".

While quick fixes are possible, complete success requires building a true culture change. This usually takes a long time. It requires long-term engagement from people and is therefore easily hampered by political issues or poor relationships.

The Four Primary Approaches to Data Quality Management

PROCESS-BASED APPROACHES

Delivering improved information by improving business processes

Process-driven approaches see DQM as first and foremost a business issue and not an IT problem.

Failures in data quality are seen primarily as the result of business process failures. Poor data quality management and data quality are symptoms of poor process design and execution. Improving processes can therefore enhance data and its management. In addition these approaches often lay emphasis on getting to the root cause of the particular information quality problem and then in designing, developing and deploying processes that remove or mitigate the cause.

Typical Environments

These approaches are used in enterprises which rely on process discipline rather than individual skill. They work well in process-driven enterprises such as manufacturing, call-centres, large retail operations, and in organisations with a high staff turn-over.

Strengths

Process-based approaches are particularly attractive to business people as they tend to intuitively understand the value of process so they are often the easiest “sell”.

They are a very good way of involving operational people in Data Quality Management and of engaging management. The methodologies used encourage getting to the root cause of each problem, as they drive data correction at source, removing or reducing the need to use costly IT methods to transform the data.

If the correct Business Process Management tools are chosen to match specific situations and business environments, these approaches can be very effective. They can be quick to deploy and provide clearly defined statistics to demonstrate any improvements.

If the processes are developed with controls that enable them to identify errors early, they can ensure consistency.

Shortcomings

Not all Data Quality Management issues come from process failings. If this is not used in conjunction with technology approaches the control can be manually intensive and resource-hungry to implement, especially in larger organisations.

In large organisations process change can be expensive and resource-heavy and once set up, it tends to require ongoing control and assurance.

Process approaches tend to encounter difficulties in high-skill enterprises or when there is no repeated process, or no culture of process thinking.

The Four Primary Approaches to Data Quality Management

INFORMATION-BASED APPROACHES

Improving the nature of the information itself, usually by focusing on structure and meaning

Information and data need good design and wide standardisation. The focus is on designing data well so that a) it meets the enterprise's requirements, b) it can be easily re-processed by different systems and c) the information it represents can be productively shared between many people.

A failure to understand the significance of the information approach to Data Quality Management can lead to a range of baffling problems in spite of our best intentions with people, process and technology. High among these are problems of meaning such as the proverbial attempt to "compare apples with oranges". Incompatible data meaning can make automated interchange between systems *effectively impossible*. Even a well-implemented, single shared system can be quickly undermined by inconsistent reference data.

Terminology and practice among information-based approaches are diverse, but there is usually an emphasis on structure, meaning and standardisation. Attention to Structure (e.g. Entity-Relationship Modelling, Data Modelling, Normal Form, Business Rules) builds a foundation for data that is compatible, complete, stable, flexible, concise and fit for purpose. Attention to Meaning (e.g. Semantics, Taxonomy, Classification, Ontology, Meta-Data, Business Definitions) helps ensure that data is comparable, consistent and intuitive. Standardisation (of Codes, Names, Definitions and Structures) helps ensure that data is comparable, compatible and consistent across the enterprise.

Typical Environments

These approaches are characteristic of many Data Management specialists. They thrive in large organisations – large enough that standardisation cannot be achieved informally. They are useful to those which need a high degree of control or precision in their information – government departments, financial institutions, military and pharmaceuticals.

Strengths

At best these approaches help to bridge the divide between IT and non-IT functions. They provide a language and tools to lead both sides to a deep understanding of their enterprise. They help non-IT people to be precise about the information they need, making it easier for IT to meet those needs. Once well embedded they allow new functionality to be built-on quickly. They provide an excellent foundation for widespread data and information sharing.

Shortcomings

However, they require special skills. Getting good design used as standard across the enterprise can require a high level of discipline. They can become an "ivory tower" exercise with no practical relevance to the enterprise at large, especially if attempted with no regard to the other three primary approaches. Some of these Information approaches are difficult to apply when most applications are bought rather than built in house.

The Four Primary Approaches to Data Quality Management

TECHNOLOGY-BASED APPROACHES

Bringing the power of technology to bear explicitly on information quality

Technology-based approaches recognise that the data explosion experienced globally over the last ten years or so has been generated by advances in technology. These include the rise of the internet, mobile access to data, digital storage and social networking. The volume of electronic data now held by both organisations and individuals worldwide is such that it cannot be managed without the use of specialist data quality technologies.

There is no doubt that technology has a critical role to play in Data Quality Management. In response to this a plethora of established and emerging technologies have emerged to help organisations store, manage, exploit and access electronic data. Some key technology areas include databases and data warehouses, business intelligence applications, master data management (MDM) tools, data quality improvement software, data mining and analytic applications, metadata repositories, extract, transform and load (ETL) toolsets and data profiling tools.

Typical Environments

Examples of technology-based approaches can be found in IT consultancies, financial departments of large organisations and government departments.

Strengths

These approaches recognise that IT has a significant role to play in Data Quality Management and its improvement. They appreciate that specialist Data Quality Management tools can make a significant impact, especially in larger enterprises where the sheer quantity of electronic data demands automated Data Quality Management

Shortcomings

The implication of this approach in many organisations is that Data Quality Management is primarily the responsibility of the IT department. Improving Data Quality Management and its exploitation is seen as a key function of IT – delivering the right information at the right time to the right place at the right price.

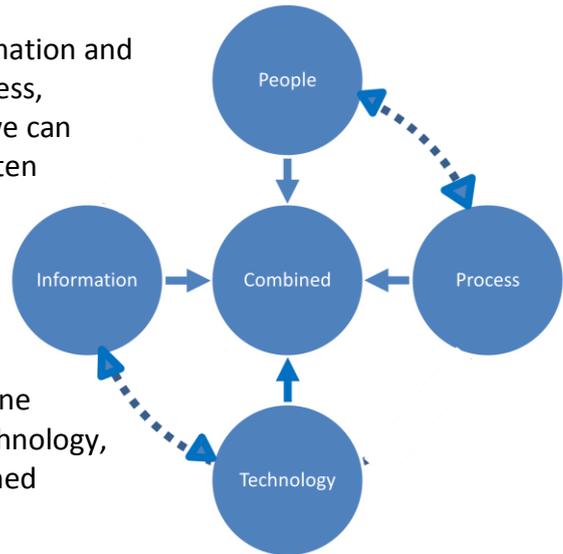
However, Data Quality Management is more than a technology issue. Many studies have shown that an organisation's ability to capture, store and exploit data for commercial advantage is heavily dependent on its business processes and the way its people create and change data. Many data quality improvement initiatives fail to meet their stated objectives because they start with the misperception that technology alone will deliver the required benefits. Buyers often fall for the hype of technology vendors. These projects often turn into costly failures, because applying new technology to problems can be an expensive investment. Sometimes focusing on people, process or information can deliver a better return on investment. In summary, improving technology can be a necessary part of improvement, but in itself is unlikely to give the enterprise the data quality it needs.

The Four Primary Approaches to Data Quality Management

COMBINED APPROACHES

Delivering superior information quality by combining more than one primary approach

The four Primary Approaches are People, Process, Information and Technology. Given that two pairs (Information and Process, People and Technology) – are often seen as opposites, we can map the four approaches like this. Two pairs that are often combined are linked by the dotted arrows.



If the four primary approaches have their strengths and shortcomings then it comes as no surprise to find that some methodologies are combinations of more than one primary approach. We found some that combine People and Process, some combine Information and Technology, and a few combine all four in what we might call Combined Approaches.

The existence of these approaches suggests that the Data Quality Management challenge is a holistic one, necessarily encompassing all four primary approaches. Only this can bring about enduring and comprehensive improvement.

Overemphasising one perspective can often negatively impact others. For instance introducing a new Customer Relationship Management system without changing business processes or retraining people can make the problem worse and not better. Trying to optimise one approach without properly assessing its wider ramifications can have the effect of sub-optimising the whole.

Typical environments

The more complex the environment the more likely it is that you will see such a model. Examples might be a global organisation with different divisions or regions in different states of maturity, a telecommunications company, a global bank or a large retailer.

Strengths

Data Quality Management is a complex challenge. Tackling that challenge with a combination of people, process, information and technology increases the chance of success. Progress in one dimension supports or stimulates progress in another, improving the chance of making and sustaining overall improvements in quality. The combined approach recognises this, approaches the problem accordingly, and so is more likely to find fundamental solutions to the root causes. It also forces business and IT to work together to address the problems.

Shortcomings

In practice a combined approach means that although individuals or groups can be given primary responsibility for leading the Data Quality Management change, they need to involve all affected stakeholders. Data Quality Managers are thus faced with an organisational transformation challenge, which is typically beyond their authority and ability. Combined approaches are rarely quick fixes. These approaches can be complex and time consuming and many people feel uncomfortable dealing with the required level of complexity.

The Four Primary Approaches to Data Quality Management

SO WHAT DO I AIM FOR?

Use whichever primary approach suits your organisation and your skills, but take time to learn about the others and always bear in mind that if you have problems getting one approach to be effective, there may be an answer waiting in another.

If you are a beginner in Data Quality Management, don't be afraid to get stuck in to just one of the primary approaches. Your route in can start with any of them. But if you want to develop a career in this area, then, over time, you should work to get experience of each of the others. This will maximise your value to your enterprise.

ACKNOWLEDGEMENTS

Throughout this paper we have avoided making reference to specific authors. All methodologies have their merits and shortcomings.

However, we found in our research two authors who have come independently to somewhat similar conclusions as ourselves. It is therefore right to acknowledge them here. John McKean in his book "Information Masters: Secrets of the Customer Race" writes of "a balanced investment across the seven information competency determinants of people skills, processes, organization structure, culture, leadership, technology and information itself." If, as we believe, his "people skills, organization structure, culture and leadership" map into our primary approach of "people" then we recognise the same four categories.

Dan Power in an article called "A 4-D Approach to Data Governance" writes that "working across the four dimensions of people, process, technology and information is critical to your data governance success." (Information Management Magazine, May/June 2011.)

AND FINALLY

The path to good Data Quality Management and Data Quality Management can be a long and lonely one. So keep in touch with other professionals. Remember it is a marathon and not a sprint. High quality and sustainable solutions take time to build.

Keep up to date with new ideas. This is a growing area of interest but as yet there is no Utopian solution. There is and will be a lot of good advice out there.

-o0o-

Chris Maynard
Nigel Turner
Zoey Husband
Neil Harvey

The Four Primary Approaches to Data Quality Management

STRENGTHS AND SHORTCOMINGS OF THE FOUR PRIMARY DATA QUALITY MANAGEMENT APPROACHES

Approach	Strengths	Shortcomings	When to consider?
PEOPLE	<ul style="list-style-type: none"> • Appreciates that people’s behaviour, skills and attitudes are important in managing data and information • Recognises the importance of human error as a cause of poor data quality • Improving Data Quality Management can ultimately only happen if people recognise its importance • Can be very cost effective and very motivating 	<ul style="list-style-type: none"> • Culture change usually takes a long time so not good for problems that require more urgent solutions • Absolutely requires engagement from people • Easily hampered by political issues or poor relationships 	<ul style="list-style-type: none"> • When people’s behaviour and motivation is a key issue • When enterprise is loosely controlled with high personal autonomy • High-skill enterprises • Small rather than large enterprises
PROCESS	<ul style="list-style-type: none"> • Recognises that good processes create good data, bad processes create bad data, so the process must be fixed to improve the data • Encourages getting to the root causes of quality problems • Business people tend to understand processes intuitively so encourages leadership and involvement, • A good way of involving operational people in Data Quality Management so encourages wide participation 	<ul style="list-style-type: none"> • Tends to be heavily manually intensive, so resource hungry to implement, especially in larger organisations • Once set up, this tends to require ongoing control • In large organisations, processes are difficult to change • May encounter difficulties in high-skill enterprises or when there is no repeated process 	<ul style="list-style-type: none"> • High staff turnover or process-driven enterprises • When driving Data Quality Management from outside IT or when the IT department cannot be involved • Can be an “easier sell” to the enterprise • Aligns well with process-driven enterprises e.g. manufacturing, call-centres, retail
INFORMATION	<ul style="list-style-type: none"> • Focuses on the data itself which is, after all, our goal • Starting with the information can bring insights of value to the business • An important contributor to understanding the problem – even when the solution is to be found with another approach • Recognises that business and IT need to work in partnership to resolve problems • Data has more permanence than process and the high level decision-makers in the enterprise often recognise the value • Encourages and supports “fact-based” decision making 	<ul style="list-style-type: none"> • Can be seen as inward looking, limited to Data Quality Management professionals so difficult to engage and prove its value to others • Takes insufficient account of business drivers and needs so difficult to prioritise activities • Sometimes improved data structures can be almost impossible to implement in practice 	<ul style="list-style-type: none"> • Soon • “Greenfield” situation • Mergers and acquisitions • Major integration exercises (of systems of organisations) • Where you don’t know about the company or the information space • Very useful in integrating systems or enterprises
TECHNOLOGY	<ul style="list-style-type: none"> • Recognises that IT has a significant role to play in Data Quality Management and its improvement • Specialist Data Quality Management tools can make a significant impact, especially in larger enterprises • The exponential growth of electronic data demands new tools to help automate Data Quality Management 	<ul style="list-style-type: none"> • Tends to reinforce the view that Data Quality Management is an IT and not a enterprise challenge • May be used to “hide” data problems and increase the problems longer term • Can be ridiculously expensive with uncertainty of return on investment • New technology can be very disruptive to an enterprise 	<ul style="list-style-type: none"> • When you have a clear set of requirements matched to a clear technology solution • When IT is driving Data Quality Management with little support from the rest of the enterprise • When large volumes of data are involved • In highly automated enterprises • When improvements need no human intervention