GDPR MEETS DATA MANAGEMENT

DATA CROSSROADS

Table of contents

Introduction

The General Data Protection Regulation (GDPR) is a new EU legislation. It lays down rules related to protection of personal data and free movement of personal data.

Only lazy business people are not talking about GDPR at the moment. The clock is ticking, getting closer to the date of 25th may 2018, and it is getting more and more exciting. There are countless interpretations and explanations for the meaning of GDPR and many companies have already started internal programs preparing for the new requirements. They realize that they need to search for both short- and long-term solutions. A short-term solution assumes finding means, often temporary, to be compliant by 25th May 2018. The long-term solution means significant changes in your business practices and technology and embedding the solution into daily operations in order to ensure 'data protection by design'[1].

In any case, at this point you need analyze the following, deeply and thoroughly:

- how GDPR affects your business;
- what strategy is best for you;
- which actions are necessary.

There are various stakeholders in every business, who are concerned about the compliancy with the Regulation. Data management professionals are only a part of this group, but as long as GDPR is about personal data, they are in the front line. This paper will give you an insight into the relationship between GDPR and data management. It provides you with a guideline, and main steps a company needs to undertake in order to improve or develop data management capabilities.

The **main concerns**

WHEN DOES GDPR CONCERN YOUR BUSINESS?

For a start, you need to make a quick evaluation whether your company has to comply with GDPR. There are several simple criteria[2] you need to take into account:

YOUR INVOLVEMENT WITH THE EU

Is your company operating within the EU? Or if your company is outside of the EU, does it it offer goods and services to **individuals in the EU**?

YOUR INVOLVEMENT WITH INDIVIDUALS AND PERSONAL DATA

What kind of groups of **individuals** are you dealing with? Think about private customers, employees, related parties to corporate customers.

What kind of personal data does your company handle? The Regulation gives a rather clear definition on what 'personal data' really is: 'information relating to an identified or identifiable natural person ('data subject').'[1] Also, you can find the extended definition in our Glossary on p.20.

The next step in your analysis is to evaluate whether the personal data also includes **special categories**. The answer will influence your data processing practices.

YOUR DATA-PROCESSING ACTIVITIES

What is your **role** in data processing?

What are you doing with the **information** you obtain?

If you determine the purpose and means of the processing of personal data, your company takes responsibilities of **Controller.**[1]

If you only participate in data processing on behalf of Controller, you are subject to duties of **Processor.**[1]

Both roles have their own legal obligations. You will proceed depending on your role in the data processing.

YOUR DATA-PROCESSING CAPABILITIES

What are your current personal data **processing capabilities**?

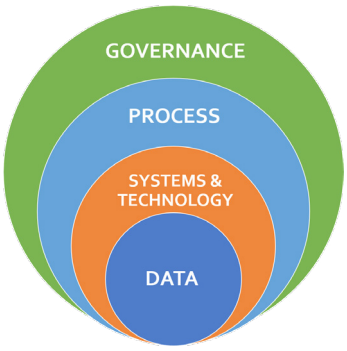
Which **gaps** have you discovered in comparison to to GDPR requirements?

Regardless of the role, the Regulation drives you to the re-assessment of your personal data processing capabilities and their optimization.

HOW DOES GDPR INFLUENCE DATA MANAGEMENT CAPABILITIES?

Various companies define and proceed with their business capabilities in different ways. One of the definitions stipulates that 'a business capability defines the organization's capacity to successfully perform a unique business activity'[4].

Usually you describe a business capability in terms of data, systems and technology, process, and governance.[7]



In order to be GDPR-compliant, you need to address the challenges in the areas of:

- data management;
- systems and technologies;
- business processes;
- governance.

As the main subject of GDPR is personal data and its processing, your first ambition now is to take a closer look at this data management area. It means defining what:

- data management means in the context of GDPR;
- requirements of GDPR affect data management;
- actions you need to execute in these areas;
- your deliverables will be.

DATA MANAGEMENT ANALYSIS

For analysis, you could use *DAMA-DMBOK 2*[3].According to this source 'data management is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles' [3].

DM-BOK 2 identifies 11 knowledge areas and presents it in the form of DAMA Wheel.[3]
The DAMA Wheel gives you an overview on data management ingredients. You need to identify for yourself, which functions you need to reach your current goal. In this book we hope to assist you with choosing the right components and putting them in correct order.

REQUIREMENTS FOR THE PERSONAL DATA YOU HOLD

GDPR applies most of all to personal data. Personal data relates to an identified (or identifiable) natural person – the 'data subject'[1]. Personal data can directly or indirectly identify a data subject. There are several categories of data subjects your company might deal with. Think about private clients, employees, prospects, listed contact, related parties of your corporate clients (i.e. ultimate beneficial owners, directors, legal representatives). Some personal data has to be recognized as of a special category.

From the data management point of view, you need to record the list of data elements and data sets, in order to:

- align the definitions of these data elements across your company;
- specify which data belong to the special categories;
- lay down relationships between data subject categories and the personal data elements.

When working with special data, you have to identify and embed the conditions into your data model / data storage, under which you will still be able to proceed with the processing of special data.

REQUIREMENTS

Art. 4.1
'Personal data means any information relating to an identified or identifiable person'

Art. 4.1
'Natural person...can be identified, directly or indirectly..'

Art. 9.1, 9.2
'Processing of [special categories] of special data shall be prohibited [unless one of the conditions] applies..'

ACTIONS

1. Define which personal data (sets) you are dealing with, can be characterized as personal data according to GDPR definitions.
2. Identify which of your data can be considered as personal data of special categories.
3. Make an analysis on direct and indirect data.
4. Define the terms and definitions of personal data and align them with other parties within your company. Create a business glossary.
5. Control and adjust the security and privacy classifications applied for personal data.
6. Map the main categories of data subjects to personal data elements.
7. Identify the main categories of data subjects which are relevant to your company.
8. Identify the conditions for proceeding with personal data of the special categories.
9. Adjust conceptual, logical, and physical models, as well as data glossary and metadata repository to be able to safeguard and maintain the information regarding personal data (incl. special categories), and the main categories of data subjects.

DELIVERABLES

1. Business glossary.
2. Conceptual and logical data model.
3. Physical data model.
4. Data dictionary.
5. Metadata repository.
6. Data privacy and confidentiality standards.

PRINCIPLES

GDPR outlines the main principles any organization has to comply with, in order to protect personal data during its processing. According to the Regulation you have to be able to demonstrate how you comply with the principles, at any time. 'The GDPR applies to both automated personal data and to manual filing systems where personal data are accessible according to specific criteria.'[2]

The definition of data processing in the Regulation overlaps with the definition of data processing, as well as data lifecycle[3,4]. DAMA-DMBOK 2 stipulates that 'data not only has a lifecycle, it also has a lineage'[3]. It means that in order to demonstrate your compliance with the Regulation you need a documented data lineage / flow. Data flow 'maps the movement of data to business processes, locations, business roles, and to technical components'[3]. Accuracy being a part of the requirement, forces you to think about the data quality framework. Such an approach drastically extends the role of data management professionals in the organization's task to become GDPR compliant.

Your documentation of data flow will include purposes of data processing as well as analysis of the whole data flow to ensure 'minimization' of data processing.

REQUIREMENTS

Art. 5.1 (a), 6
'Lawfulness, fairness and transparency'

Art. 5.1 (b)
'Purpose limitation'

Art.5.1 (c)
'Data minimization'

Art. 5.1 (d)
'Accuracy'

Art. 5.1 (e)
'Storage limitation'

Art. 5.1 (f)
'Integrity and confidentiality'

Art. 5.1 (d)
'Accountability'

ACTIONS

1. For each personal data element/set you need to identify the source, the relevant purpose of processing and the users. Update your metadata repository with this information.
2. To link the source of the data with its purpose of use, you can use data lineage as a tool. As a minimum, you can document data lineage in descriptive form[7,8].
The following elements should be taken into account:

- business process and roles;
- application landscape;
- data users (external and internal);
- a format to deliver the information in;
- controls and security embodied into processes.

For data transfer to third countries, you need to have a description of how you safeguard this data on its way to its recipients.
3. In order to keep data accurate, you need to build in data quality checks for the accuracy of personal data. It is advisable to build quality checks in the process chains, as well as on separate data element levels.
4. Per each data element, you need to check regulatory requirements for retention periods to ensure 'storage limitation'. The regulation provides some exceptions for data storage limitation (i.e. archiving purposes in the public interests etc.). You will also need to map different purposes of usage

per data elements with different retention period requirements.
5. Definition of security and privacy classification aspect of the data processing are also in your 'to-do' list in order to safeguard against unauthorised or unlawful processing.
6. Accountability of business roles in the process should be described as a data lineage element.

DELIVERABLES

1. Metadata repository.
2. (Descriptive) data lineage and metadata lineage
3. DQ controls built-in processes and on a level of separate data elements.
4. Data storage procedures.
5. Security classification.
6. Audit procedures for processing.

INDIVIDUALS' RIGHTS

GDPR stipulates some rights for individuals. Let us take the closer look into these rights from data management point of view. The biggest challenge here is separating data management accountabilities from IT and operations.

- The company must be able to provide information (notices) to a data subject about how their data is being processed, as this information is also a part of metadata and data lineage. You need not only to document and to maintain this information, you are obliged to provide this information within a limited period of time. So to put it simply: the descriptive way of maintaining data lineage might not be good enough.
- A data subject has the right to have their data erased or restricted to processing. Assume that you need to clean up several databases that comprise of a processing chain. Yes, this is IT-area, but in order to know what to erase, you need to have an up-to-date metadata repository, which is data management territory.
- Data portability means that 'data (can be) transmitted directly from one controller to another, when technically feasible'[1]. Yes again: a technical challenge. But before being able to do that, you need to know what data you need to transfer and where this data is located. Data managers are welcome to answer these questions.
- Automated decision making and profiling will be either a part, or the goal of data processing. Business rules that are a part of data processing are also a subject for data management. Business rules as techniques are used in different functional areas of data management as metadata, DWH and BI, data integration and interoperability.

REQUIREMENTS

Art. 12-14
The right to be informed

Art. 15
The right of access

Art. 16
The right to rectification

Art. 17
The right to erasure ('the right to be forgotten')

Art. 18
The right to restrict processing

Art. 20
The right to data portability

Art. 21
The right to object

Art. 22
Rights related to automated decision making and profiling

ACTIONS

1. The above mentioned are parts of data lineage and metadata repository maintenance. Should your company have a complicated application landscape with a high number of applications, you should think from moving from descriptive data lineage to an automated solution.

DELIVERABLES

1. Metadata repository.
2. Data lineage.

SECURITY

GDPR embraces two regulations concerning the security of data. First of all, your company has to ensure secure processing. You are supposed to 'implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk'[1]. Secondly, in case of breach, you have to communicate it 'without undue delay'[1] to data subjects, as well as to supervisory authority.

Again, this is to some extent a grey area in terms of responsibilities of IT, operations, data management, and security. Each company distributes responsibilities according to their own organizational structure. If you check the definition of 'data security' in DAMA-DMBOK 2, you will see that a lot of tasks are still related to the data management function. The main goal for security from the data management perspective is to 'protect information assets in alignment with privacy and confidentiality regulations...'[3].

REQUIREMENTS

Art. 32.1
'Implement appropriate technical and organizational measures to ensure a level of security'

Art. 32.1 (a)
The pseudonymisation and encryption of personal data

Art. 32.1 (b)
'Ongoing confidentiality, integrity, availability and resilience of processing systems and services'

Art. 32.1 (c)
'The ability to restore the availability and access to personal data

Art. 33.1
'Without undue delay...notify the personal data breach'

ACTIONS

1. For many companies, security is an existing business function that needs its own methodology and tools. Extra attention should be paid to the personal data elements subject to the new regulation. Large companies often encounter a similar issue: the levels of security and privacy have been defined on an application level, not for separate data elements. Thus security audit has to take place, focusing on the identified personal data elements.
2. An ongoing discussion is the protection of keys used to pseudonymized and encrypt data. Combination of the keys and the pseudonymized or encrypted personal data can be considered as indirect data for GDPR purposes. So, safeguarding of the keys is an additional focal point for data management.
3. The requirements might lead you to a necessary revision of data security architecture and building additional security access controls.

DELIVERABLES

1. Revised security and privacy classification.
2. Revised data security architecture.

ACCOUNTABILITY OF CONTROLLERS AND PROCESSORS

As mentioned above, your company could recognize itself either in the role of Controller or Processor. Accountabilities for these roles slightly differ, although the requirements regarding data processing are quite similar.

Some of these accountabilities (i.e. implementation of data protection policies) might fall outside the data management function. Some of them (i.e. maintaining of processing activities) lie in the grey area. For example, recording of 'the categories of recipients to whom the personal data have been or will be disclosed'[1] is a part of metadata and data lineage capabilities.

The most significant requirement is building data protection by design. Different companies might have their own interpretation as for what it means. As (meta)data lineage seems to play an important role in the ability to comply with the regulation, 'data protection by design' can be consequently interpreted as '(meta)data lineage by design'. 'Data lineage by design' is known as the one of the greatest technical challenges.

Requirements for the implementation of technical and organizational measures call for participation of different business functions. Depending on organization, the distribution of tasks may vary.

REQUIREMENTS

Art. 24.1
'Implement appropriate technical and organizational measures. That processing performed in accordance with this Regulation'

Art. 24.2
'Implementation of appropriate data protection policies..'

Art. 25
Data protection by design and by default

Art. 26.1
'maintain a record of processing activities'

ACTIONS

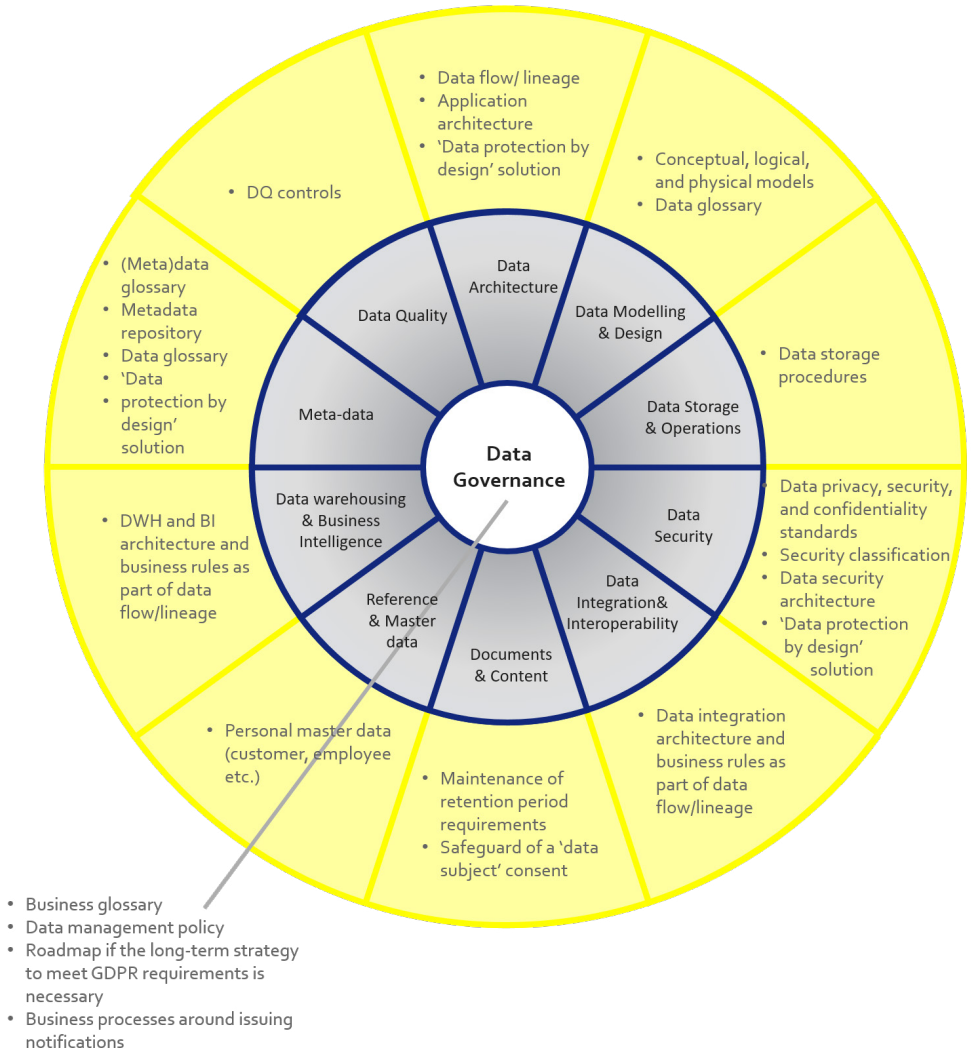
1. For data management, the most important part is paying attention to documentation of the list of personal data elements, purposes of data processing, groups of recipients, processes and applications involved in data processing.

DELIVERABLES

1. Adjusted data management policy.
2. Data management roadmap in the area of data lineage.

Data management **capabilities** and **deliverables** for **GDPR compliancy**

You made a brief analysis on dependencies between GDPR requirements and data management capabilities. In order to develop an action plan, first, you should briefly summarize the data management capabilities involved. Then you continue with the main deliverables. The DAMA Wheel can help with visualisation.



What you see might come as a shock: in order to meet the GDPR requirements, you need to implement data management on a full scale. Don't worry, there a few positive points, though:

- personal data is the only scope;
- it will be a good 'pilot' project for those companies that just started thinking about data management;
- you still can opt for a short-term solution, and in the process the long-term solution might become clearer.

The solution you choose for your company will depend on your company's size, maturity of data management and available resources.

Data management implementation: **where to start?**

You do not set up data management just for fun. You do it because you have certain needs, a certain goal will drive you to do it. Depending on this particular driver, you need to find data management components that will serve the purpose. You implement these components in a certain order. We offer you a **5-step approach** as one of the ways to do it.

You repeat the process as soon for each separate driver you have. You might change the order of steps, or make iterations, depending on your company's needs.

Choose the main driver.

The driver defines the set of data management components you will implement. Think about improving data quality, compliancy with regulations, implementing predictive or prescriptive analytics.

Step 1. Define who needs what data and why.

Within a company, data is being shared among different business functions. You need to find out the main stakeholders and discover their concerns related to the set of data.

Step 2. Agree on *who* will be doing *what*.

As data management is a shared responsibility, you need to define and divide relevant tasks and responsibilities.

Step 3. Structure your data-house.

This part depends greatly on your driver. For a specific driver, you need only a specific limited number of universal data management components. As soon as your driver changes, you need to add or to adjust the mix.

Step 4. Assess the gaps between what you have and what you need.

You company has already operating data management functions. First, you need to define what tasks you take into the scope of data management. Second, based on one of the existing maturity and readiness assessment techniques, you assess the level of maturity of your company. Third, you define the desired level. As a result, you have executed a gap analysis.

Step 5. Keep going.

Based on the gap analysis, you specify tasks to be done. Based on your current company culture and practices, you specify the way of achieving your goals.

A more detailed explanation of this approach can be found in *The Data Management Cookbook*.

Some final thoughts...

If you have reached this point, by now you must realize the role of data management in relation to GDPR compliancy.

Every company will find its own way. Hopefully, this book has provided you with ideas on how and where to start. All you need to do is to design a feasible approach to be able to build 'personal data protection by design'.

Feel free to contact Data Crossroads for free advice or more in-depth consulting or workshops.

Works cited

1. Regulation (EU) 2016/679 of the European parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

2. Overview of the General Data Protection Regulation (GDPR).
https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/

3. DAMA-DMBOK 2. Data Management Body of Knowledge, 2nd edition

4.The DAMA Dictionary of Data Management, 2nd edition 2011.

5. Open Group Standard. TOGAF Version 9.1. The Open Group
www.opengroup.org

6. BIZBOK Guide. Business Architecture Body of Knowledge.

7. BCBS 239: Data Lineage or Information Value Chain.
http://datacrossroads.nl/2017/03/16/bcbs-239-data-lineage-information-value-chain/

8. New vision on Data Lineage/ flow in DAMA-DMBOK2
http://datacrossroads.nl/2017/09/10/new-vision-on-data-lineage-flow-in-dama-dmbok-2/

IMAGES

Image 1, p.7: Capabilities model. Copyright © by Data Crossroads.

Image 2,3, p.7,14: The DAMA Wheel. DAMA International's Guide to The Data Management Body of Knowledge (DAMA-DMBOK), Second edition. Copyright © by Dama International

Image 4: The 6-step approach to data management implementation. Make your data work for you. Copyright © by Data Crossroads.

Glossary

Accuracy
the degree that data correctly represents 'real-life' entities. Accuracy is difficult to measure, unless an organization can reproduce data collections or manually conform accuracy of records.
DAMA-DMBOK 2, p.457

Controller
the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
General Data Protection Regulation, Art.4(7)

Data Architecture
identifying the data needs of the enterprise (regardless of structure), and designing and maintaining the master blueprints to meet those needs. Using master blueprints to guide data integration, control data assets, and align data investments with business strategy.
DAMA-DMBOK 2, p.100

Data Governance
the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets.[3]
DAMA-DMBOK 2, p.69

Data Integration and Interoperability
managing the movement and consolidation of data within and between applications and organizations.
DAMA-DMBOK 2, p.271

Data Modeling
the process of discovering, analyzing, and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model. This process is iterative and may include a conceptual, logical, and physical model.
DAMA-DMBOK 2, p.121

Data Security
definition, planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets.
DAMA-DMBOK 2, p.219

Data Storage and Operations
the design, implementation, and support of stored data to maximize its value.
DAMA-DMBOK 2, p. 170

Data Quality Management
the planning, implementation, and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers.
DAMA-DMBOK 2, p.451

Data Warehousing and Business Intelligence
planning, implementation, and control processes to provide decision support data and support knowledge workers engaged in reporting, query, and analysis.
DAMA-DMBOK 2, p.382

Document and Content Management
planning, implementation, and control activities for lifecycle management of data and information found in any form or medium.
DAMA-DMBOK 2, p.304

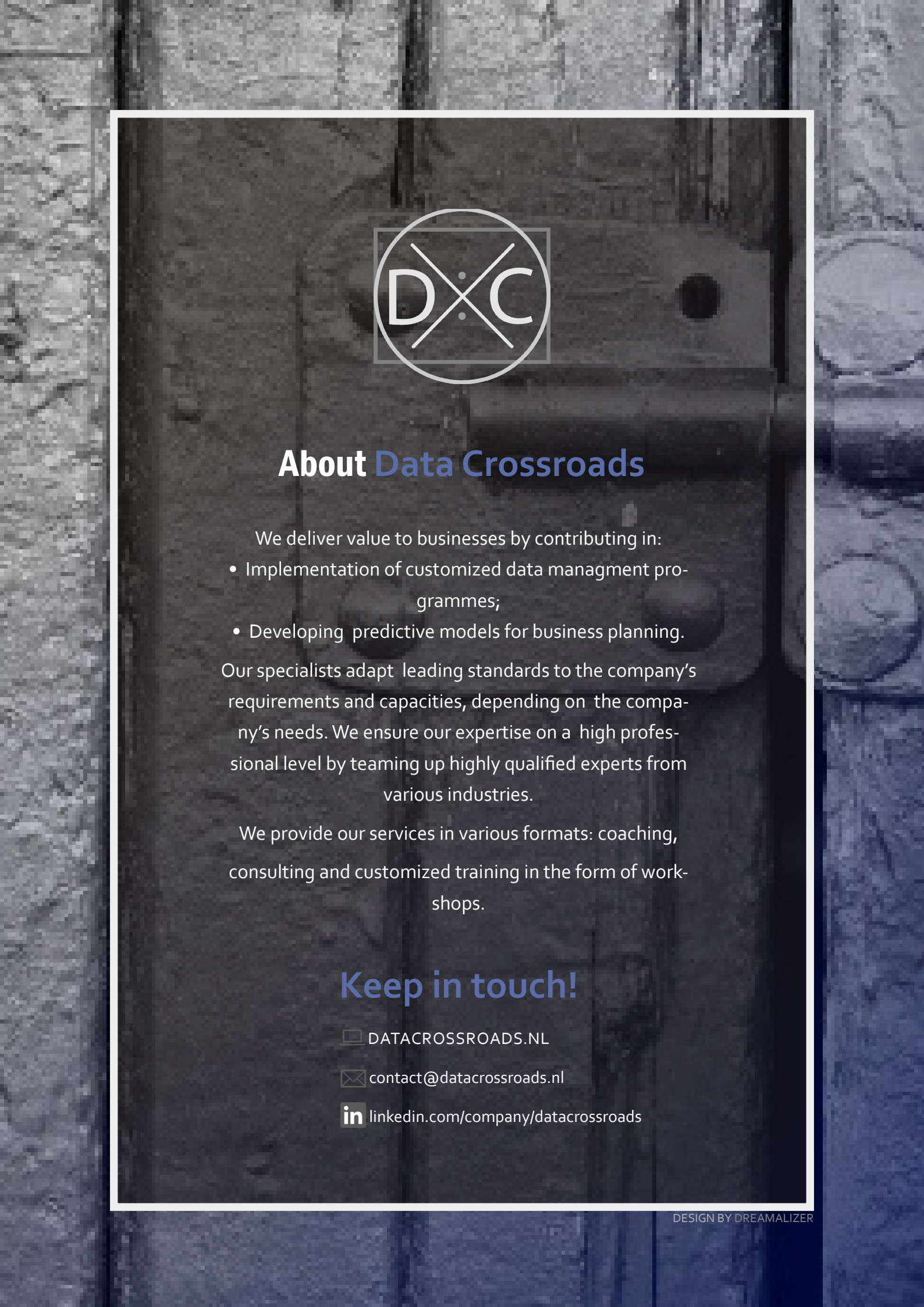
Metadata Management
planning, implementation, and control activities to enable access to high quality, integrated metadata.
DAMA-DMBOK 2, p.419

Personal Data
any information relating to an identified or identifiable natural person ; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
General Data Protection Regulation, Art.4(1)

Processing
any operations or set of operations which is performed on personal data or sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.[2]
General Data Protection Regulation, Art.4(2)

Processor
a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.
General Data Protection Regulation, Art.4(8)

Reference and master data
managing shared data to meet organizational goals, reduce risk associated with data redundancy, ensure higher quality, and reduce the costs of data integration.
DAMA-DMBOK 2, p.348



About Data Crossroads

We deliver value to businesses by contributing in:

- Implementation of customized data managment programmes;
- Developing predictive models for business planning.

Our specialists adapt leading standards to the company's requirements and capacities, depending on the company's needs. We ensure our expertise on a high professional level by teaming up highly qualified experts from various industries.

We provide our services in various formats: coaching, consulting and customized training in the form of workshops.

Keep in touch!

DATACROSSROADS.NL

contact@datacrossroads.nl

linkedin.com/company/datacrossroads