

Squaring the circle:

Using a data governance framework to support data quality



Introduction

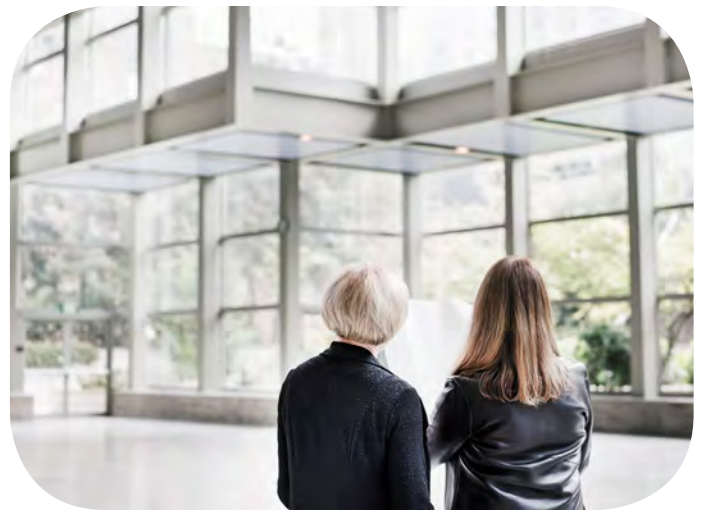
Most organizations wish for better quality data which makes it surprising just how many of them are still relying on short term, tactical approaches to managing their data quality. Nicola Askham, The Data Governance Coach shares her advice on how implementing a data governance framework can significantly increase the long-term success and sustainability of your data quality initiatives.

While data governance may be a relatively young data management discipline (Wikipedia cites it as an emerging discipline), organizations have wanted and sought better data quality data for a long time. So long in fact that it makes you wonder why they have not identified a lasting solution!

So what is the problem? For many years companies have sought to improve the quality of their data by applying quick fixes. These usually take the form of cleansing the data after it has been captured, often in a separate system, such as a data warehouse. This has led to well-developed approaches to analyzing datasets in order to identify deficiencies and good techniques for cleansing and enriching the data. However, despite the maturity of these techniques, they need to be constantly and frequently repeated to maintain lasting quality. Now you may be asking why is that? Why does the data not stay “fixed”? Well quite simply, it is due to a lack of governance

or control over the data. In order to maintain a consistently good level of data quality, it must be proactively managed, ensuring that it is captured accurately and deterioration in its quality is prevented.

This is where a structured data governance framework makes a huge difference. Data governance comprises a range of activities undertaken by the business, with its primary purpose being to improve the quality of the data. However there is still confusion about how data governance and data quality relate to each other. The aim of this paper is to clarify that relationship and illustrate how a data governance framework is used to support and embed data quality activities within your organization.



Definitions

Although data governance professionals are generally very focused on data definitions for their organization, there are still differing definitions of what data governance actually is and its scope.

Because of this I'd like to clarify what is meant by the terminology used throughout this white paper:

Data governance	The exercise of authority, control, and shared decision making (planning, monitoring and enforcement) over the management of data assets.*
Data governance framework	An organization's approach to managing the quality of its data i.e. what they are going to do, who is going to do it and how they intend to achieve it.
Data profiling	An approach to data quality analysis, using statistics to show patterns of usage and patterns of contents.*
Data dictionary	Any place where business and/or technical terms and definitions are stored.*
Data quality	The degree to which data is accurate, complete, timely and consistent with all requirements and business rules.*
Data quality management	The application of practices to improve data quality.*
Data glossary	Similar to a Data Dictionary, but enriched with further information about the data e.g. its relationship with other data items, its data owner and where that data is used.
Master data management	Processes that control management of master data values to enable consistent use.*
Reference data	Any data used to organize or categorize other data. Usually consists of codes and definitions.*

*The DAMA Dictionary of Data Management

The relationship between data governance and data quality

I often describe the relationship between data governance and data quality as symbiotic. After all the relationship is based on a mutual interdependence between the two data management disciplines. Having spent many years focusing on data governance and data quality, it is quite clear that you would not want to do one without the other.

However, as mentioned earlier, this is sadly often not the case. It is rare for a company to decide to commence a data governance initiative for no other reason than because it is “best practice”. The purpose of having a data governance framework is to manage and improve data quality. So why would you go to the effort of defining and implementing a framework, if you were not going to use it to monitor and improve the quality of your data?

But for many years organizations have embarked on data quality initiatives without implementing a data governance framework to support them. The result being that many data quality improvement projects are, at best, tactical fixes which have only short-term results. At worst, they fail to have any significant impact from the start with the exception of the financial costs invested in the exercise.

The reason for this is that without data governance often roles and responsibilities are not set up and agreed, nor are the organizational processes needed to proactively manage data quality. For example setting up Data Quality reporting without a data governance framework in place begs the question to whom would you report? In some instances the initial recipient of such reports may be interested in the results and may even take action to improve the quality of the data as a result. But if the task has not been allocated as a definite role responsibility, how can you guarantee this will continue when their priorities change? Will they remember to tell their successor about the reports if they change roles? On many occasions I have

come across data quality reports disappearing into a black hole because the person who originally requested them has moved on rendering a valuable insight into the health of an organization's data as useless, time consuming and costly.

It is only through using a data governance framework, where data quality and data governance become entwined, that sustainable benefits begin to appear.

Having all three components of the data governance framework is key to the success of your data quality initiatives. For example, having a policy defining the roles and responsibilities without set processes will result in an inconsistent approach at best.

Having a policy and processes defined without roles and responsibilities will result in no-one taking ownership for the processes and therefore nothing happens!

What is a data governance framework?

A data governance framework is a structure put in place allowing organizations to proactively manage data quality. There is no one size fits all approach. For it to work, it must fit with the organization's culture, structure and practices. Having said that, the basic components of policy, processes and roles & responsibilities will be found in every framework.

Policy

Having a policy stating that your company will "do" data governance is a key part of the framework. Without it there will be no authority to change the way your organization thinks about and manages its data.

In a few instances progress without having a policy may be made when data governance is a key focus, however as future business priorities change it is likely that your initiative will slow or at worse be abandoned. Having a policy as part of your organization's risk management approach gives you the authority, and your stakeholders the requirement, to implement data governance.

Processes

These are the processes (and the deliverables arising out of those processes) which need to be followed in order to manage your data quality. They can include things such as data quality reporting and data quality issue management.

Having your processes defined and documented facilitates a consistent and repeatable approach to managing data across your organization.

Roles and responsibilities

The third vital component is to define who is responsible for data quality in your organization and the scope and extent of their responsibilities.

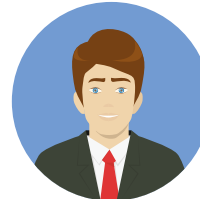
Typical organizational roles include:



Data Owner
accountable for the quality of a defined dataset(s)



Data Steward
responsible for the quality of a defined dataset(s)



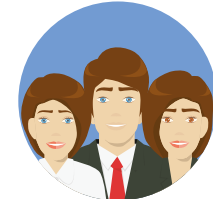
Data Producer
responsible for creating/capturing data according to the Data Consumers' requirements



Data Consumer
responsible for defining what makes the data they use of sufficient quality to use



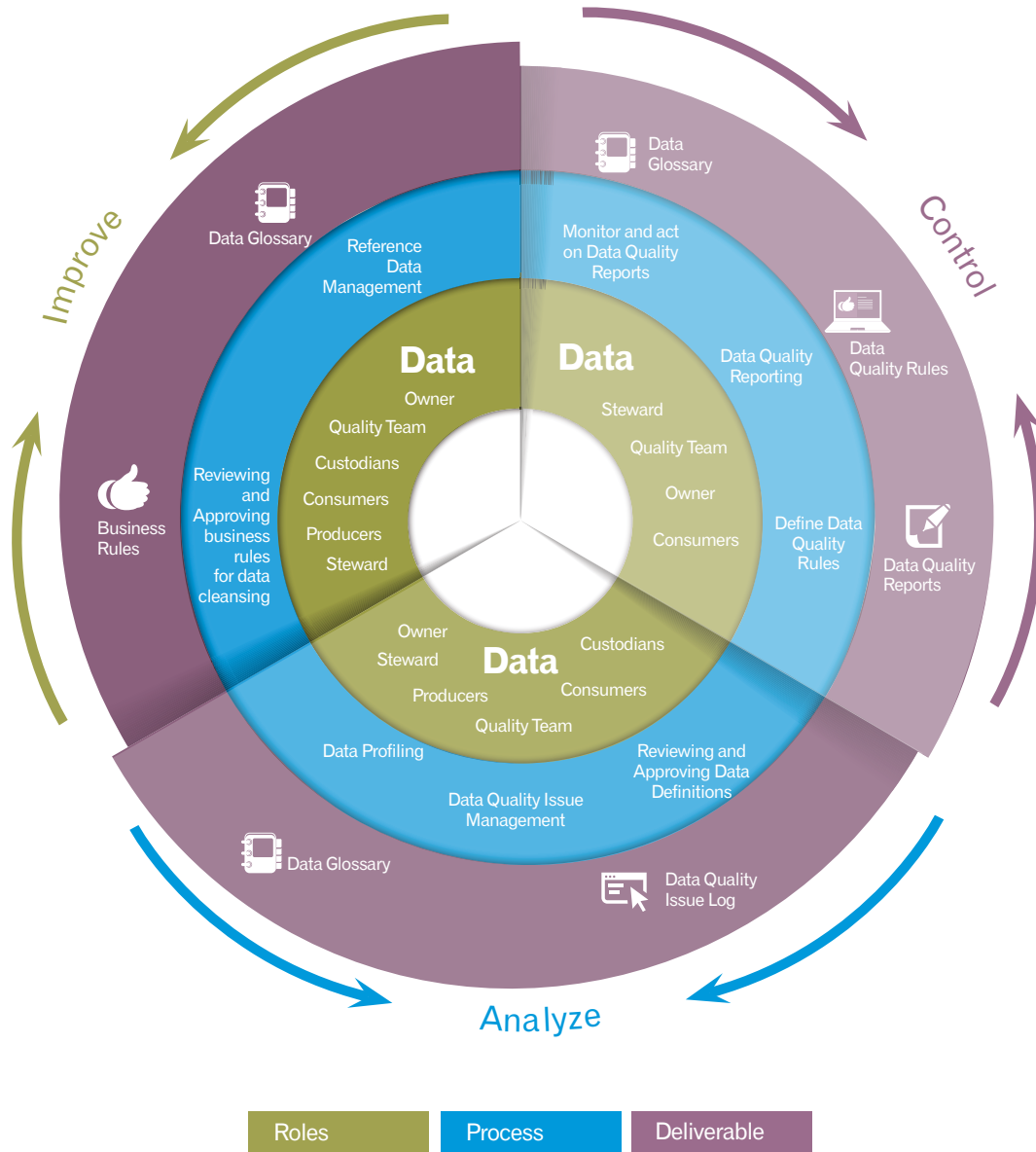
Data Custodian
responsible for maintaining the data on IT systems, in accordance with business requirements



Data Quality Team
responsible for undertaking and supporting data quality activities across the organization

How does a data governance framework relate to data quality?

As this is often an area of confusion I've found it useful to demonstrate the relationship between a data governance framework and data quality using a diagram:

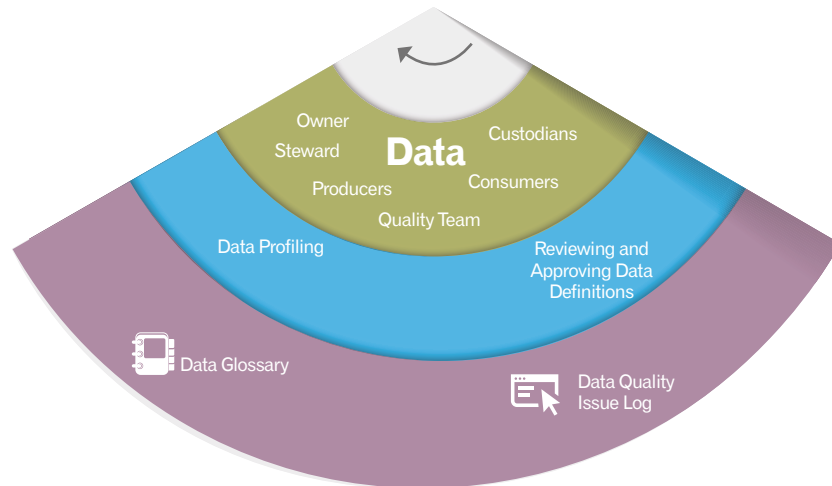


In the next few sections of this white paper, I address each segment of the diagram explaining how a data governance framework supports data quality activities. Specifically I will address the type of processes, the people responsible for undertaking them and deliverables resulting from those processes. Keeping in mind that no one size fits all, the examples that follow are for illustrating how it could work without advocating a definitive solution for

your organisation. For example, you may choose different processes or allocate responsibilities to different roles.

It is worth mentioning that creating a data policy is a one-off exercise completed early on in your data governance initiative. It mandates the activities that follow and for this reason it is not referred to below:

Analyze



1. Process - Data profiling

While data profiling itself can be regarded as a pure data quality activity, the insight which it gives often has wider benefits to an organization if those results are shared with the correct stakeholders.

Data profiling done in isolation (usually either to investigate data quality issues or as part of data project) provides results that are used in isolation. How much more powerful and impactful would it be if the results of the data profiling were shared with data owners and data stewards who in turn can add business expertise and insight to the results?

Roles

The Data Quality Team is responsible for profiling the data and sharing the results with the appropriate stakeholders.

The Data Steward(s) and **Data Owner(s)** are responsible for reviewing the profiling results and providing business expertise and insight to the Data Quality Team. The Data Quality Team is then able to further refine the profiling results.

Deliverables

Data Profiling Report

2. Process - Reviewing and approving data definitions

Conversely, data profiling results can be an extremely useful input when working on data definitions.

Sadly while data dictionaries are usually created when a new system is developed, often they are seen as project deliverables and subsequently become lost to the business. To truly understand and manage your data, it must be defined and those definitions held where they are readily accessible by business users i.e. in data dictionary or data glossary. However, it is useful to be aware that different users of the same data may have different beliefs as to the importance, relevance and meaning of that data. It is therefore very important that when creating a data dictionary or data glossary that you have a process to firstly draft and then review and approve those definitions.

The good news is data profiling can provide a valuable input into this process as it illustrates the values currently used for each data item. This can be extremely useful when agreeing on a data definition.

Roles

Data Quality Team - responsible for managing and coordinating the process amongst the various stakeholders and providing the data profiling input.

Data Steward - responsible for reviewing the data profiling inputs and drafting the definition(s).

Data Consumers - provide input to the definition(s).

Data Producers - provide input to the definition(s). As the people capturing the data in the first place, they may have valuable insight into its definition(s) and uses within your organization.

Data Owner - approves the data definition(s).

Deliverables

Data Profiling Report

Data Glossary

3. Process - Data quality issue management

If the primary purpose of implementing a data governance framework is to improve the quality of your data, then having a process to investigate and act on data quality issues is a vital process.

It should be noted that this process cuts across both the Analyze and Improve segments of the diagram. As it commences with the identification and analysis of the issue, I have included the detail of the process in the Analyze segment.

Roles

Data Quality Team - responsible for managing, coordinating and reporting on the process. The Data Quality Team will be responsible for maintaining the Data Quality Issue Log, supporting or undertaking investigations to establish the extent and cause of issues. This team may also be involved in any remedial activities such as data cleansing, education and system fixes.

Data Consumers - responsible for highlighting issues with the data they use and reporting these issues to the Data Quality Team.

Data Stewards - responsible for working with the Data Quality Team to investigate and resolve data quality issues.

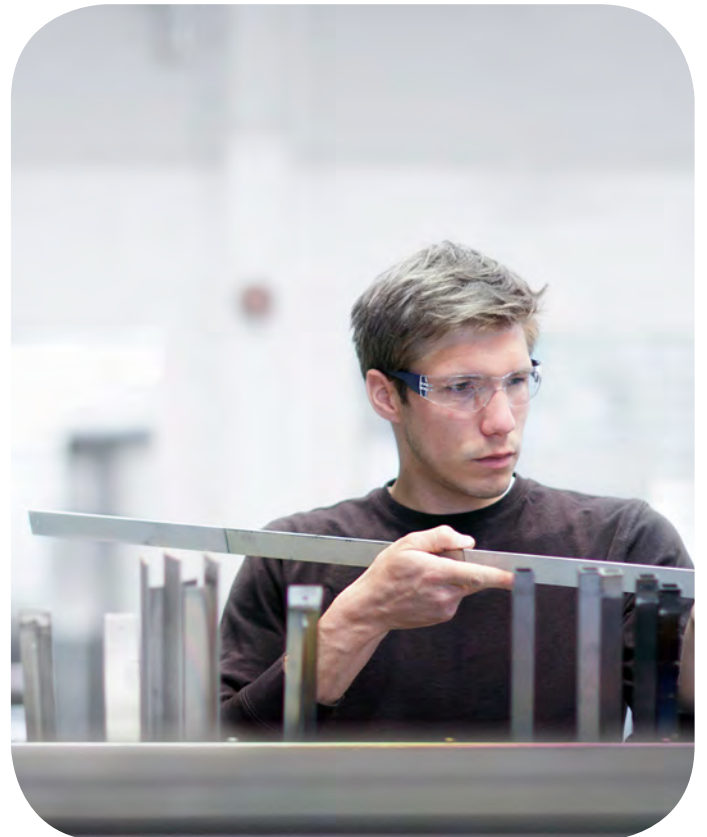
Data Producers - responsible for working with the Data Quality Team to investigate and resolve data quality issues.

Data Owner - responsible for approving any remedial action.

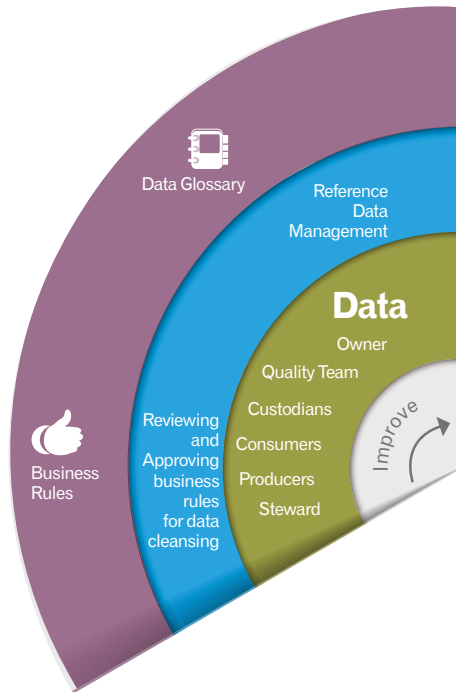
Data Custodian - responsible for implementing any system fixes required as part of remedial activities.

Deliverables

Data Quality Issue Log



Improve



Once your data has been analyzed, either as a result of issues being highlighted or from proactive data profiling, the next logical action is to improve that data. So let's look at a couple of process examples which support that action:

1. Process - Reviewing and approving business rules for data cleansing

Having discovered a requirement to cleanse or fix a dataset, it will be necessary to get input from business stakeholders in order to agree the rules by which the data will be cleansed. Of course, with your data governance roles and responsibilities in place identifying the correct people to involve here will now be quick and easy. It is useful to have a repository (a Data Glossary) to document those rules ready for use in future data cleansing activities or even for data quality monitoring and reporting. Once again the results of any data profiling are a useful input into this process.

Roles

Data Quality Team - responsible for managing and coordinating the process amongst the various stakeholders and providing the data profiling input.

Data Steward - responsible for reviewing the data profiling inputs and drafting the business rules.

Data Consumers - responsible for reviewing and giving input into the proposed business rules.

Data Owner - approves the business rules for use.

Deliverables

Business rules

2. Process - Reference data management

Reference Data Management is a highly focused implementation of Master Data Management, controlling the definitions of the reference data used within your organization. Without a data governance framework in place, reference data management is hard to achieve.

For example, if IT is asked to add a new country code, how do they know if the requester is authorized to make that change?

Roles

Data Quality Team – responsible for identifying the Data Owner of a reference data item and ensuring that any changes to reference data items and their definitions are updated (if already recorded), or captured, in the Data Glossary.

Data Owner/Data Steward responsible for reviewing and agreeing any proposed change to reference data codes and definitions.

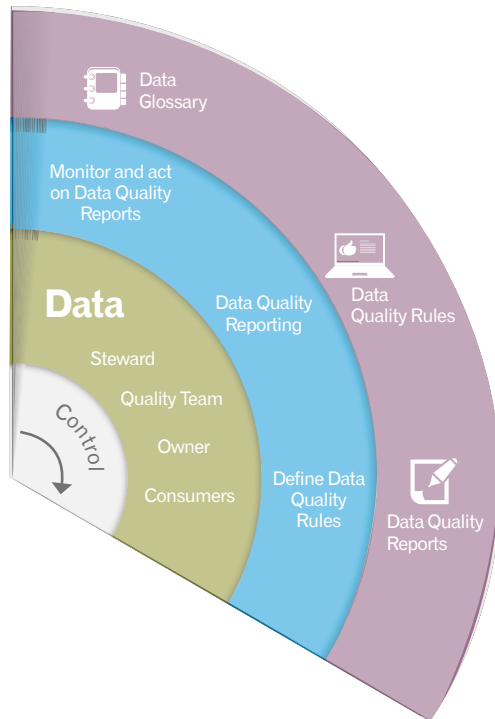
Data Consumers - responsible for requesting new reference data items or any changes to reference data items.

Data Custodian - responsible for updating any systems that use that reference data item to accept the new or amended code/value.

Deliverables

Data Glossary

Control



While having a data governance framework can add greater and more sustainable benefits when you are looking to analyze and improve your data, its contribution is greater when it actually comes to controlling your data. You are unlikely to have much success (even in the short-term) in controlling data without a data governance framework. After all, data governance is about putting in place those controls to proactively manage your data. To illustrate this let's look at some of the processes which fall in this segment:

1. Process - Define data quality rules

This process is similar to that of reviewing and approving business rules for data cleansing. However whereas the data cleansing process is reactive the process of creating data quality rules is proactive. This proactive process allows you the ability to report on the status of your data quality at any point in time.

Roles

Data Quality Team - responsible for managing and coordinating the process amongst the various stakeholders and recording the Data Quality Rules for future use (possibly in the Data Glossary or Data Quality reporting tool).

Data Consumers - responsible for defining what makes the data good enough for them to use i.e. their data quality requirements

Data Steward - responsible for reviewing input from all Data Consumers and drafting the Data Quality Rules

Data Owner - responsible for approving the Data Quality Rules

Deliverables

Data Quality Rules

2. Process - Data quality reporting

Only after the data quality rules are defined are you able to instigate a process for reporting on how the data measures up against those rules.

Role

Data Quality Team - responsible for producing and distributing data quality reports to appropriate stakeholders

Deliverables

Data Quality Reports

3. Process - Monitor and act on data quality reports

Obviously just producing and distributing data quality reports may not have the desired effect. While some people may take it upon themselves to review and act upon these reports, they may do so in an inconsistent manner, while others may not take action at all! You can resolve this situation by having a process which clarifies what action needs to be taken and by whom.

Roles

Data Owner - ultimately accountable for ensuring that the data quality reports are monitored and acted upon.

Data Steward - responsible for monitoring and acting upon data quality reports.

Data Quality Team - responsible for monitoring and escalating when substandard or declining levels of data quality are not acted upon.

Deliverables

There are unlikely to be specific deliverables from this process as any action arising will result in another process being triggered e.g. Data Cleansing or Data Quality Issue Management.

Summary

I hope these examples have clarified:

- The relationship and interdependence of data governance and data quality
- How designing and implementing the right data governance framework for your organization supports your data quality activities and helps you embed them into business as usual
- And finally why a structured framework is vital to ensure the long-term success and sustainability of these data quality activities

Analyze. Improve. Control. A solid data governance framework can support your entire data quality program.

[Learn more](#)

About the author

Nicola Askham, The Data Governance Coach, is an independent data management consultant. Her experience in coaching both regulatory and non-regulatory organizations to design and implement full data governance



frameworks, is unique within the Data Governance field. The coaching approach enables organizations to self-manage the process beyond initial implementation.

Nicola's coaching and Data Governance workshops, including Solvency II, ensures your data governance framework is embedded as an integral part of your business as usual policy. The benefit for you is that once the framework is in place your organization will be confident, competent and compliant.



Experian
53 State Street
Boston, MA 02109
T: 1 888 727 8822
www.edq.com

© 2016 Experian Information Solutions, Inc. • All rights reserved
Experian and the Experian marks used herein are trademarks or registered trademarks of Experian Information Solutions, Inc. Other product and company names mentioned herein are the property of their respective owners.