Syndicate Data Incorporation into Business Intelligence

Amjad Riaz
Syndicate Data Incorporation into Business Intelligence

Submitted by Amjad Riaz to the University of Skövde as a dissertation towards the degree of M.Sc. by examination and dissertation in the School of Humanities and Informatics.

23 November 2011

I hereby certify that all material in this dissertation which is not my own work has been identified and that no work is included for which a degree has already been conferred on me.

Signature: _______________________________________________
Syndicate Data Incorporation into Business Intelligence

Amjad Riaz

Abstract

Organisations today are working in an increasingly competitive environment where business success rests on the ability to make high quality decisions. Consequently, comprehensive knowledge about the organisation itself as well as the market in which it operates is required. Data warehouse (DW)-based business intelligence (BI) solutions can fulfil this need by integrating data from internal and external sources to provide useful insights that will assist organisational key-players in their decision-making. The specific incorporation of syndicate data (which is a type of external data) is particularly important because it enriches data content and maximises its full value. Although previous research strongly indicates that supplementing internal data with SD enhances the decision capabilities of an organisation and gives it a competitive edge in the marketplace, the literature on SD incorporation itself is of a very limited scope. Therefore, the aim of this work is to explore current practices in SD incorporation into DW-based BI solutions. A questionnaire study on the identification, acquisition, integration and usage of SD was conducted with BI consultants. Besides confirming that SD integration into DW-based BI solutions is common, the results also provide insights on how to identify SD suppliers, different data acquisition approaches, data distribution methods, integration approaches, types of SD, and SD application areas that are being used. Propositions for future work, which will extend the findings accounted for in this work, are also included.

Key words: business intelligence, data warehouse, external data incorporation, syndicate data incorporation, syndicate data.
Acknowledgements

Whatever insights it may offer, this dissertation would not be complete if it did not include my deepest gratitude to all those who have made this research possible.

First and foremost, I would like to thank my supervisor Ph.D. Mattias Strand for his guidance and boundless support throughout this project. It is an honour to work with such a dedicated mentor and I am grateful for the sound advice you’ve given from all the way across the North Sea. Special thanks also go to Ph.D. Mikael Berndtsson for his time and feedback regarding this work and for providing assistance in numerous ways.

Most importantly, I am indebted to my family for their love, encouragement and motivation, especially when I needed it most. I dedicate my dissertation to my parents, Riaz Ahmed and Tahseen Riaz, who have taught me the value of a lifelong pursuit of knowledge. I am where I am today because of you. Finally, to my wife Zahra Amjad who has been endlessly patient and an inspiration, thank you for believing in me.
# Table of Contents

1. Introduction .................................................................................................................. 1  
   1.1. Problem Area .......................................................................................................... 2  
   1.2. Aim & Objectives .................................................................................................... 3  
2. Frame of Reference ........................................................................................................ 5  
   2.1. Business Intelligence ............................................................................................... 5  
   2.2. Data Warehouse ...................................................................................................... 7  
   2.2.1. Comparing DW with Operational Systems ....................................................... 8  
   2.3. External and Syndicate Data ................................................................................... 8  
   2.3.1. Defining External and Syndicate Data ............................................................... 8  
   2.3.2. The External Data Incorporation Process .......................................................... 9  
   2.3.3. Types of External Data ...................................................................................... 10  
   2.3.4. External Data Sources ....................................................................................... 11  
   2.3.5. External Data Integration Approaches ............................................................. 12  
   2.3.6. External Data Usage .......................................................................................... 12  
3. Research Approach ....................................................................................................... 14  
   3.1. Research Method .................................................................................................... 14  
   3.2. Research Process .................................................................................................... 18  
   3.2.1. Selecting the Tool .............................................................................................. 18  
   3.2.2. Sample Selection ............................................................................................... 20  
   3.2.3. Ethics, Privacy, and Security Concerns .............................................................. 21  
   3.2.4. Developing Questions for the Online-Questionnaire ....................................... 21  
   3.2.5. Online-questionnaire Implementation ............................................................... 22  
   3.2.6. Data Analysis ..................................................................................................... 24  
4. Analysis and Results ..................................................................................................... 25  
   4.1. Background ............................................................................................................ 27  
   4.2. Main Questions ....................................................................................................... 32  
   4.2.1. Identification ...................................................................................................... 32  
   4.2.2. Acquisition ........................................................................................................ 36  
   4.2.3. Integration .......................................................................................................... 40  
   4.2.4. Usage ................................................................................................................ 42
4.3. Concluding Questions.................................................................48
4.3.1. The Current trend in Syndicate Data Incorporation ...............48
4.3.2. Perceived Importance of Syndicate Data in Generating Insights....48
4.3.3. Future trend of Syndicate Data Incorporation............................51
5. Conclusions.................................................................................55
6. Discussions....................................................................................57
  6.1. Reflection on the Research Approach .........................................57
  6.2. Aim fulfilment ............................................................................58
  6.3. Putting the results in wider context.............................................60
  6.4. Future Work................................................................................61
7. References.....................................................................................63
Appendix 1: Invitation.........................................................................66
Appendix 2 Online-questionnaire User Interface.................................67
List of Figures

Figure 1: A best practice BI environment .......................................................... 6
Figure 2: External data incorporation process .................................................. 10
Figure 3: The four-step processes of research method selection ..................... 14
Figure 4: Research process ............................................................................. 18
Figure 5: Basic structure of the online-questionnaire ....................................... 23
Figure 6: Distribution of respondents by country ............................................. 28
Figure 7: Distribution of respondents by working experience ......................... 29
Figure 8: Distribution of respondents by business role ................................... 30
Figure 9: Distribution of respondents by current Industry .............................. 31
Figure 10: Industry distribution by previous BI projects .................................. 32
Figure 11: Commonness of SDSs identification approaches ............................ 34
Figure 12: Commonness of syndicate data acquisition approaches ................. 36
Figure 13: Commonness of syndicate data distribution methods ..................... 38
Figure 14: Commonness of syndicate data integration approaches ................ 41
Figure 15: Commonness of syndicate data categories ..................................... 43
Figure 16: Commonness of different syndicates data application areas .......... 46
Figure 17: Commonness of syndicate data incorporation ............................... 48
Figure 18: Syndicate data’s importance (with SD experience) ....................... 49
Figure 19: Syndicate data’s importance (with no SD experience) ................. 50
Figure 20: Perceived importance of syndicate data ........................................ 51
Figure 21: Future trend of SD incorporation (with SD experience) ................. 52
Figure 22: Future trend of SD incorporation (with no SD experience) .......... 53
Figure 23: Expected future trend of syndicate data incorporation ................... 53
1. Introduction

Organizations are working in an environment driven by “hyper-competition” (Ding, 2009). In such an environment, business success is directly related to high quality decisions made by management (Carlsson and El Sawy, 2008). In a highly competitive environment, organizations need to focus on smarter and more efficient use of information (Zeljko, 2007). The fundamental element in the initiation of any decision making is to cater for capabilities to supply the right information, to the right person, at the right time (Solomon and Gray, 2008; Bogza and Zaharie, 2008; Ding, 2009).

Business intelligence (BI) provides such capabilities to organizational key-players and thereby assists them in driving operations in an effective and efficient manner (Zeljko, 2007; Wixom and Watson, 2010). In addition, Wixom and Watson (2010) claim that BI is an essential prerequisite for the success of organizations, when competing in a highly competitive marketplace. According to them, there is no universally accepted definition of BI and, in general, they define it as (Wixom and Watson, 2010, p.14):

“Business intelligence (BI) is a broad category of technologies, applications, and processes for gathering, storing, accessing, and analysing data to help its users make better decisions”

Moreover, for being able to stay informed regarding vital aspects, such as competitors and customers, data is the key element of any business intelligence solution (Inmon, 1996; Inmon, 2005; Strand and Carlsson, 2008; Bogza and Zaharie, 2008; Ponniah, 2010). BI consists of two main activities, getting data in and getting data out. The former is traditionally called data warehousing (DW) (Watson and Wixom, 2007) and DWs are a vital cornerstone in any comprehensive BI environment (Solomon and Gray, 2008) and Wixom and Watson (2010). For getting data in, data warehousing involve integrating data from many sources into a centralized repository, as a means of achieving an enterprise-wide, integrated view of vital business dimensions, such as customers, suppliers, and products (Inmon, 1996). In literature, there are numerous definitions of a DW. However, in this work, the following, widely adopted definition by (Inmon, 1996, p.33) was chosen:

“A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management’s decision making process.”
The sources feeding the DW may reside internally, as well as, externally to the organization (Watson and Wixom, 2007). External data may provide insights into an industry’s trends and empowers management to compare their organization’s performance against other organizations. In contrast to external data, internal data only provides a picture based on present and past activities of the organization which is not enough to compete in a highly competitive marketplace (Pirttimäki, 2004; Ponniah, 2010). In alignment, although the internal data often withholds the volume majority in implemented DWs, external data is attracting more and more interest. Moreover, most external data is incorporated from suppliers specialized in collecting, compiling and selling data (Strand, 2005). Kimball refers to these companies as syndicate data suppliers (SDSs). Consequently, Strand (2005) entitles this data as syndicate data and defines it in the following manner (Strand et al., 2006, p.982):

Business data (and its associated metadata), purchased from an organization specialized in collecting, compiling, and selling data, and targeted towards the strategic and/or the tactic decision making processes of the incorporating organization.

Why is then external data in general and syndicate data in particular exposed to an increased interest? There are several explanations to be found in literature but for this work, the following quotation from Inmon (1996, p.191) was used to illustrate the importance of data originating from outside the organizations: “external data plays an important role in the understanding of information over time, though it does not say anything directly it says everything about the universe that the company must work and compete in.”

1.1. Problem Area

Current literature indicates that there is a lack of research in the area of business intelligence in general and advocates a need to further address BI-related research questions with a wider variety of strategies (Jourdan et al., 2008). Wixom and Watson (2010) explicitly express that BI is an underexplored area. Although much is known about conceptual foundation and frameworks, there is a lot more for business intelligence academics to explore. In other words, the area of business intelligence deserves more attention (Jourdan et al., 2008).
In addition, many organizations have already realized the importance of “competing on analytics” and external data in general and syndicate data in particular is one way for an organization to increase their competitive edge (Strand and Carlsson, 2008). Moreover, if organizations want to get the full value of their data content, they need to use relevant syndicate data in conjunction with their internal data (Strand and Carlsson, 2008). Furthermore, the growth in the implementation of DW-based BI-solutions has increased significantly and DW developers now showing an increased interest in syndicate data (Strand and Carlsson, 2008). In alignment, Ponniah (2010, p.60) concludes that “syndicate data in now becoming big business”.

Still, although there are many indications in existing literature that external data (which syndicate data is a sub-type of (Strand et al., 2003)) is a very important ingredient for decision making in BI (Pirttimäki, 2004; Watson and Wixom, 2007; Bogza and Zaharie, 2008), when trying to go into details, regarding e.g. types of data, suppliers of the data and applications, literature is of limited scope. For instance, although Strand et al. (2003) categorized types and sources of external data, their study is focused on all types of external sources and related data within data warehousing, with only a limited number of respondents and in a delimiting Swedish setting. The work by Strand is a starting point, but does not provide the broad picture in the context of BI solutions.

Therefore, based on the above reasoning, one may argue that there is a need to acquire more knowledge about syndicate data incorporation into DW-based BI solutions.

1.2. Aim & Objectives

The aim of this work is: “to explore current practice of syndicate data incorporation into DW-based BI solutions”.

In order to fulfil the aim of this work, the following objectives need to be answered:

1- Which syndicate data supplier that are used and how are they identified?
2- How is the syndicate data acquired?
3- How is the syndicate data integrated into DW-based BI solutions?
4- How is the syndicate data used in DW-based BI-solutions?
The establishment of the objectives is grounded in the external data incorporation process (EDIP) introduced by Strand and Wangler (2004). According to them, the process for incorporating external data (which thus also applies to syndicate data) is comprised of the following four activities: identification, acquisition, integration and usage (further details regarding the EDIP can be found in Section 2.3.5). For this work, each objective is based on obtaining a deeper understanding about the different activities of the EDIP, in the context of DW-based BI solutions. The purpose of the objectives is elaborated upon below.

Answering Objective 1 supports the fulfilment of the aim by providing extended knowledge and a frame of reference regarding current approaches for identifying SDSs. Answering Objective 1 will also extend current knowledge regarding which SDSs that are actually used. Answering the second objective, which deals with the general principles of how syndicate data is acquired, intends to clarify, and at the same time extend current knowledge regarding; how companies obtain the data, which distribution technologies they apply and the frequency of the acquisition. Answering Objective 3 supports the fulfilment of the aim by gaining a general idea of current syndicate data integration approaches and will result in a deeper understanding regarding different approaches. Finally, answering the forth objective will provide knowledge on the different types of syndicate data that are used and the areas in which this data is typically utilised. Thereby, it contributes to the fulfilment of the aim by increasing the knowledge regarding current practice of syndicate data usage.
2. Frame of Reference

2.1. Business Intelligence

In order to understand what business intelligence entails, it is essential to establish what this concept actually means. However, obtaining a universally accepted definition of business intelligence is by no means straightforward since existing literature offers several differing definitions.

According to Solomon and Gray (2008, p.175), BI refers to a system that “combines data gathering, data storage, and knowledge management with analysis to provide input to the decision process”. Zeljko (2007) on the other hand, mentions the use of technology to collect and effectively use information to improve business effectiveness. Yet other definitions claim that business intelligence is the process of extracting, transforming, managing and analysing business data (Niu et al., 2009) or getting the right information to the right people at the right time (Ding, 2009). However, of all the views put forward, we have chosen to adopt the definition provided by Wixom and Watson (2010, p.14) who state that:

“Business intelligence (BI) is a broad category of technologies, applications, and processes for gathering, storing, accessing, and analysing data to help its users make better decisions.”

Our primary reason for adopting this definition of BI is the comprehensive and broad scope it offers. Wixom and Watson’s (2010) definition is applicable in most cases as it highlights the key characteristics of business intelligence; namely, gathering data from source systems, storing the data, accessing the data and analysing the data using business intelligence technologies and applications. Furthermore, this definition emphasizes the core purpose of the BI environment, which is assisting business users in their decision making.

As mentioned previously, there are many variations in business intelligence environments. In fact, many organisations may have different environments. For example, some may not have any dependent data marts for (DW), while others may not have DW at all. (Wixom and Watson, 2010).

Given that DW is a cornerstone in BI solutions (Solomon and Gray, 2008; Wixom and Watson, 2010) and our work focuses specifically on DW-based BI
solutions, the generic and widely applicable BI environment presented in Figure 1 (adopted from (Wixom and Watson, 2010, p.15)) is well aligned with this study and contributing with an extensive context to the research aim at hand.

Figure 1: A best practice BI environment

Below, the different components in Figure 1 are described

On the right hand side of Figure 1, there is a representation of the many types of users and applications accessing the available data. There are many different producers who analyse the data and create content for information consumers. The potential applications that could be used for this include (in an excerpt); SQL queries, drillable reports, OLAP, EIS, dashboard, scorecard, alerts and data mining or predictive analysis.

On the left hand side of Figure 1, we see the variety of data sources from which data is generated; for example, Point-of-Sale (POS), OLTP and ERP systems, as well as external sources. Organisations extract data from these sources, transform the data and then load it into the DW. Traditionally this process is called ETL (extract transform and load) but due to additional ways of integrating data, the common term is now Data Integration.
Finally, three important components are shown at the bottom of Figure 1, namely, Metadata, Data Quality and Governance. Technology is involved in all of them to some extent but all three components are actually about people and processes. Metadata supports the IT people who obtain the data and the users who retrieve the data, while Data Quality refers to the fact that it is very important for the data to be of a high quality so that it can be used on a long term basis. The Governance component is responsible for ensuring that the BI meets organisational goals, and consists of people, committees and processes.

The centre of Figure 1, i.e. the DW will be more extensively describes in the up-coming section, due to its critical role for this work.

2.2. Data Warehouse

According to Wixom and Watson (2010) and Solomon and Gray (2008) the data warehouse is a cornerstone in a generic and comprehensive BI environment. As such, it is important to fully grasp the concept of a data warehouse and to define this term. There are many varying definitions in existing literature. However, it is not in our interests to assess the strengths and weaknesses of each of them here. Instead, we will rely on the most widely accepted definition, which was introduced by William H. Inmon, one of the pioneers of DW.

According to Inmon (1996, p.33):

“A data warehouse is a subject oriented, integrated, non-volatile, time variant collection of data in support of management’s decisions.”

These key characteristics can be described as follows (Inmon, 1996):

Subject Oriented means that the data provides information on a particular business area such as customers or employees, instead of also including information on other areas such as the on-going operations of the organisation.

Integrated refers to the way data is gathered from multiple internal and external sources and then integrated into the warehouse. In addition, the data must be consistent in format, naming and other aspects, and may also be detailed, summarized or aggregated.
Non-volatile means that once data is integrated into the data warehouse, it should not change.

Time-variant means that a data warehouse maintains both historic and recent data for a longer period of time. In contrast, operational databases maintain only the most current and up to date data values.

2.2.1. Comparing DW with Operational Systems

One of the motives for adopting the DW definition proposed by Inmon (1996) was that it clearly expresses the support for management, as opposed to operational systems supporting day-to-day tasks. For clarification, this section accounts for the key differences between operational systems and DWs.

Devlin (1997, p.14) defines operational systems as systems which “run the business in real time, based on up-to-the-second data, and are primarily designed to rapidly and efficiently handle large numbers of simple read/write transactions.” According to Ponniah (2010), operational systems are online transactional processing systems (OLTP), which are used to run day-to-day company operations and to support core business processes. Operational systems store recent information about a single entity such as an order, invoice or customer. On the other hand, a DW contains both historic and recent data (Inmon, 1996). DW does not generate fresh data but processes large amounts of data in order to generate insights for strategic and tactical decision making by the management (Ponniah, 2010).

Another difference is that operational system data is updated constantly but in DW, data is being accessed constantly (Inmon, 2005). Furthermore, a DW holds data for long periods of time compared to an operational system. The time horizon to hold data for DW is typically 5-10 years whereas operational data is held for 60-90 days (Inmon, 1996; Inmon, 2005).

2.3. External and Syndicate Data

2.3.1. Defining External and Syndicate Data

Generally, there are two main ways in which the externality of the data can be understood. Firstly, it could mean that the data is acquired from outside the
organisation. Secondly, from a database viewpoint, external may refer to the fact that the data is stored and maintained outside a particular database, such as the local or distributed database (Strand, 2005). As mentioned in section 1.1, our research concentrates on data that can provide an insight into an industry’s trends and empower management to compare the organisation’s performance with other organisations of the same type. Therefore, in the context of externality which we have adopted, the data is understood as being acquired from outside the organisations.

Strand et al. (2003) claim that finding one common definition of external data is difficult. Devlin (1997, p.135) defines external data as: “Business data (and its associated metadata) originating from one business that may be used as part of either the operational or the informational processes of another business”. Meanwhile, Kimball (1996) refers to external data as data purchased from a specialized data supplier in the form of syndicate data (a subtype of external data (Strand et al., 2003). Finally, Strand et al. (2006, p.982) suggests that:

“Syndicate data is business data (and its associated metadata), purchased from an organization specialized in collecting, compiling, and selling data, and targeted towards the strategic and/or the tactic decision making processes of the incorporating organization.”

Strand et al. (2003) argue that Devlin’s (1997) definition is more general and may include sources such as the Internet and business partners. On the other hand, Kimball conceptualizes external data but does not define it (Strand et al., 2006). We have adopted Strand et al.’s (2006) suggested definition for this work. The reason for choosing this definition is that it is more specialized and separates syndicate data from external data according to its characteristics. Furthermore, conceptually, this definition is more contextualized with respect to syndicate data (Strand et al., 2003) and focuses on tactical and strategic decision making.

2.3.2. The External Data Incorporation Process

As stated in section 2.1, the data warehouse is a cornerstone in a generic and comprehensive BI environment (Wixom and Watson, 2010; Solomon and Gray, 2008). However, as we mentioned in section 1.2, there is a lack of knowledge about data incorporation into BI solutions.
In this section, we will briefly describe and summarize the external data incorporation process presented by Strand and Wangler (2004). This process is illustrated in Figure 2 (adopted from (Strand and Wangler, 2004)).

The external data incorporation process can be divided into four main activities:
1- Identification
2- Acquisition
3- Integration
4- Usage

*Identification* refers to the activity of locating and assessing sources within the market from which data can be obtained. Some common sources are listed in section 2.3.4.

*Acquisition* is the process of acquiring data from external sources and injecting data into internal systems.

*Integration* activity involves the data integration process, data modelling with respect to the underlying schemas, ensuring data quality and data storage.

*Usage* activity includes data interpretation, determining the purpose of the data as well as its appropriate usage in the decision making process and conceptually mapping the data with internal data.

### 2.3.3. Types of External Data

In this section, we present the different types of external data that may be incorporated into a DW-based BI solution. According to a study conducted by Strand et al. (2004), there are two particularly common types of external data
that are frequently incorporated into DW – industrial code and address information.

*Industrial code* is used to classify organisations and services. For example, the SNI (Svensk Näringsgrenindelning) index is used to identify the industry an organisation is involved in. In this classification system, the industry may also be divided further into different levels according to industrial sector.

*Address Information* is another important type of external data which helps organisations to keep track of their customers, partners and other participants in the marketplace. Without incorporating external data, it would be too difficult to ensure that such important information is always up to date. The usage area of address information is explained in section 2.3.6.

### 2.3.4. External Data Sources

Data can be acquired from various external sources (Devlin, 1997). However, the existing literature is not very comprehensive and does not provide enough detail about them. Nonetheless, there are some common external data sources which have been identified by Strand et al. (2003) based on the Swedish market. Below, we describe and summarise the different categories which these sources fall into.

*Statistical institutes* offer information about the commercial, industrial, social, economic and general activities as well as condition of the general population.

*Syndicate data suppliers* provide economic data about companies, which helps businesses to reduce credit risk and also helps them to find profitable customers and manage vendor efficiency.

*Industry organisation* provides data about specific industry sectors.

*County councils and municipalities* provide demographic information about the population of a particular area e.g. electoral register etc.

*The Internet* contains a massive amount of information and may be used to obtain information about competitors e.g. offers, prices etc.
Business Partners provide data from other companies or suppliers and also help businesses to find out about the overall performance of the environment.

Bi-product data suppliers are companies which create huge amounts of data as a bi-product of their core business and which sell this data to other organisations as a way of using the data to cover their own costs.

2.3.5. External Data Integration Approaches

According to Strand et al. (2004a), there are four integration approaches and the decision of which integration approach to choose is fully dependent on the nature and purpose of the external data being integrated into the DW environment (Strand et al., 2004a). These approaches are explained briefly below.

Star-schema dimension integration: In this approach, external data is stored in a separate dimension in the data warehouse. This requirement arises when an organisation does not want to mix external data with internal data because of the poor data quality.

Dimension attribute integration: In this approach, the external data is contained in different attributes. In such cases, the external data is stored and presented in a dimension that is based on the internal data.

Attribute value integration: In this approach, external data is mixed with the internal data on an attribute value level.

Spread-sheet integration: In this approach, external data is stored in a reference table and mixed in a spread-sheet manner.

2.3.6. External Data Usage

In this section, we will explain the different uses of external data. As stated in previous sections, there are different types of external data sources. However, organisations may not enjoy the full advantage of having varied data from these different sources unless the data is used in a productive way. Here, we present a few different usage areas based on the studies conducted by Strand et al. (2004b).
Update Addresses: The most common usage of external data involves updating customer addresses (Strand et al., 2003). According to Strand et al., (2004b) this usage area is very important because when a company sends out any information to their customers, it is vital that this information reaches the correct person.

Increase the precision of marketing campaigns: Another important usage area of external data is in increasing the precision of marketing campaigns. Marketing plays a key role in the success of any business and sometimes involves campaigns targeted at a specific segment of the population who are likely to become a company’s customers. External data, for example a list of home owners, car owners etc., can help determine how the organisation can be most effectively promoted to the intended audience.

Segmentation of Stakeholders: In this usage area, external data such as credit information helps organisations to create segmentations of their customers and other stakeholders.

Identify Customer Trends: Companies may acquire data about customer trends through customer transactions made within the company. Many companies also outsource their marketing and customer services to other business partners and data provided by these business partners can provide a more complete view of customers’ buying trends.

Determine Geographical Coverage: This involves acquiring population data from SDSs for comparison with internal data. The comparison provides opportunities for companies to better plan their business policy on a regional basis. Strand et al. (2004a, p.4) give the example of a newspaper where, “external address data give the number of households in the particular area of interest. That numbers are then compared to the customers derived from internal data. The quotient gives the newspaper’s percentile coverage of households in the area.”

Support star-schema dimension design: This plays an important role in the success of DW and thus is important as a usage area. In a star-schema dimension design, dimensions are divided into increasingly refined and specialized categories which are then used when designing the aggregation levels of star-schema dimensions related to companies or customers.
3. Research Approach

The key decision for all research lies in selecting an appropriate and systematic method which will support the fulfilment of the aim.

To ensure that this selection of research method was conducted systematically, the four-step process presented by Berndtsson et al. (2008) and illustrated in Figure 3 (adopted from Berndtsson et al., 2008, p.68) was followed. The figure also indicates which sections of the report that correspond to each step of the process.

![Diagram of four-step process]

Figure 3: The four-step processes of research method selection

3.1. Research Method

In every field of science, progress depends heavily on the contributions made by research (Marczyk et al., 2005). According to Marczyk et al. (2005) the purpose of research is to answer questions and acquire new knowledge. However, Berndtsson et al. (2008) argue that “research” is a semantically overloaded term and is sometimes used in a very general sense. As such, it is important for us to understand that in an academic context.
“Research is used to refer to the activity of a diligent and systematic inquiry or investigation in an area, with the objective of discovering or revising facts, theories, applications etc.” (Berndtsson et al., 2008, p.10).

According to Dawson (2002), research methodology is a philosophy or general principle which guides the research. In addition, as indicated by Berndtsson et al, (2008), there is a lack of consensus among authors as to whether data gathering should be referred to as “methods” or “techniques”. For the sake of consistency and to avoid any ambiguity in the terminology used to mean data gathering, we will use the term “research method” in our work.

The choice of which research method to use depends on how the problem is specified in the research question, since this would point towards the most appropriate research methodology for the chosen subject (Dawson, 2002). Dawson (2002) also suggests that it is not wise to compare the strengths and weaknesses of each methodology (Dawson, 2002). Our aim (refer to section 1.2) is to conduct an exploratory study on SD incorporation into DW-based BI solutions, as a means of broadening current knowledge. In addition, the domain coverage will also be different, compared to the work conducted by Strand and associated. Dawson (2002) refers to this as domain triangulation.

In research, there are numerous methods available for data gathering including literature analysis, interviews, case studies, surveys, implementation and questionnaires (Dawson, 2002; Berndtsson et al., 2008). A number of these research methods may be applicable for this work. However, the purpose of the study and the chosen research methodology can help us identify the most appropriate research method (Dawson, 2002). As Dawson (2002) points out, considering budget and time constraints is also important. Thus the research methods which fulfil these criteria and have the potential to answer our research problem can be narrowed down to interview or questionnaires.

In order to fulfil the aim stated in section 1.2, we decided that questionnaires would be the most appropriate research method for our study. Still, even if deciding upon questionnaire as the method of choice, different types of questionnaires are possible to apply. With the advent of the internet, as a new medium of communication, online-questionnaires are getting increasingly popular (Singh et al., 2009). As a result, there is a growing preference amongst researchers for using web-based questionnaires instead of traditional paper–based ones (Singh et al., 2009). Online-questionnaires have many advantages
over paper-based questionnaires in terms of cost, speed, appearance, flexibility, functionality, and usability (Dillman, 2000; Kwak and Radler, 2002; Bandilla et al., 2003; cited by Lumsden, 2006; Singh et al., 2009). Online-questionnaires are also more interactive compared to paper-based ones as they are able to include pop-up instructions, error messages and links (Lumsden, 2006; Singh et al., 2009).

The primary reason for selecting online-questionnaires instead of conducting interviews is the reasonably large number of respondents that are anticipated. As we are aiming for an exploratory study, as stated in section 1.2, a reasonably large number of respondents will be required (the argument for this is presented in section 1.1), and a questionnaire would allow data collection to be performed in a more effective and systematic way. Furthermore, an online-questionnaire facilitates the collection of information from a large number of respondents simultaneously, thus enabling data collection in less time and in a cost effective manner. It means when respondents submit their response at the end of the online-questionnaire, a response transcript is ready in the form required for the data analysis phase. This is helpful as it enables us to avoid the extra time and effort it takes to transcribe each response and send it back to respondents for verification, which is a necessary part of using the interview research method.

Another consideration in method selection is respondent and response management. Online-questionnaires use a database to store data which is very convenient for managing, analysing and quantifying results; whereas interviews involve a great deal of manual work that may leave room for error while formulating, analysing and quantifying results. In addition, according to Williams and Scandura (2000), applying different research methods over time using domain triangulation provides more insight on a given subject area. For this work, it means that besides triangulating the domain, compared to the work conducted by Strand and associates; also the research method becomes triangulated, i.e. going from interviews to questionnaires.

However, online-questionnaires have been criticised, due to the following reasons; coverage, non-response, measurement and sampling errors (Lumsden (2006). Below, these errors are briefly presented.

Coverage errors result from “not allowing all members of the survey population to have an equal or non-zero chance of being sampled for
participation in a survey” (Dillman, 2000, p.9; cited by Lumsden, 2006). With online-questionnaires, there is a digital division as they only reach respondents who have access to the internet (Lumsden, 2006; Singh et al., 2009). However, it is important to note that the issue of coverage does not just apply to online-questionnaires but affects traditional survey methods as well (Lumsden, 2006). Still, Lumsden (2006) argues that the increasing trend in using computers and the internet has minimized the issue of coverage with respect to online-questionnaires.

Non-response errors occur when respondents either fail to respond or abandon the questionnaire before completing it (Lumsden, 2006). The main reason for respondents abandoning or failing to complete a questionnaire is that they tend to underestimate the time required to complete it. To avoid this problem, many features can be incorporated into the questionnaire to provide some means of estimating completion time (Crawford et al., 2001; cited by Lumsden, 2006).

Sampling Errors can arise for online-questionnaires when only a small, and therefore unrepresentative, portion of the targeted group of respondents reply due to the rest of the sample population being unintentionally excluded due to factors such as differences in connection speeds, bandwidth limitations, browser configurations, monitors, or hardware. It is important to take technical user requirements into consideration when designing an online-questionnaire (Lumsden, 2006).

Measurement Errors may occur when questions are worded in a manner that fails to elicit meaningful or relevant answers from respondents. Besides making it difficult to obtain interpretable data, poorly structured questions can confuse and frustrate respondents when they are filling in the questionnaire (Dillman, 2000; cited by Lumsden, 2006).

Based on the arguments presented above, an online-questionnaire will be used. Foremost due to the opportunities it caters for addressing respondents which are required to hold a rather detailed competence, including BI, as well as DWs and SD. If conducting a traditional questionnaire, the sampling of the respondents would be really hard and finding them an utmost challenging task.
3.2. Research Process

When designing and implementing an online questionnaire, it is important to have a list of systematic and logical steps that are to be completed (Lumsden, 2006). In order to provide readers with a clearer understanding of the research process, Figure 4 was included. Figure 4 also intends to guide the reader with relation to the upcoming parts of this section.

![Figure 4: Research process](image)

In the sections that follow, we will briefly discuss the decisions that were made and explain the rationales behind our choices.

3.2.1. Selecting the Tool

As already decided that research method would use an online questionnaire, the next important decision point involved finding out whether it would be better to build custom online questionnaire application or to buy one off the shelf. As expected, there are advantages and disadvantages to both choices. In the case of
making custom online-questionnaire application, the main advantage is that it provides more control over the application and thus would be able to tailor its features according to online-questionnaire requirements. Furthermore, we would be able to program the application to generate results in the exact format we would later be required to present our findings in. In our particular case, the disadvantage of making our own online questionnaire application is that it would require more time and resources than the scope of our project allows. Meanwhile, the benefit of buying a ready-made application is that such applications are generally more reliable, having been developed and tested by experts and proven effective for the many other organisations already using them.

We initially decided to build our own online-questionnaire application as we have all the necessary skills required. However, after gathering all the application requirements and initiating the application analysis phase, we found that it would not be possible to complete all the online-questionnaire development phases of designing, building, testing and deployment within the given time scale. As such, it was more feasible for us to purchase a ready-made online-questionnaire application.

With this in mind, the next important task was to find a reliable online questionnaire application that would satisfy our questionnaire requirements. As above, we highlighted the issues that need to be addressed when choosing an online questionnaire solution (see Figure 4). These acted as the requirements upon which we based our choice when looking for the most suitable application. We carried out a target search on the internet and found a number of companies which provide online-questionnaire tools, such as SurveyMonkey, SurveyGizmo, FreeOnlineSurveys, Kwik Surveys and so on. After careful comparison of the different applications in terms of features, such as; reliability, designing facilities, data security, backup and launching facility, SurveyGizmo was chosen.

While we do not intend to discuss all the differences between the various solutions providers mentioned above, we will list the key advantages of using SurveyGizmo (SG). SG provides an easy way to create online surveys/questionnaires with no experience needed at all. It allows 26 basic question types with customizable options for research, insights, tracking and marketing purposes. Additionally, SG is capable of handling advanced requirements, enabling us to design a highly sophisticated questionnaire. For
example, one of the requirements for Question 6 in our online-questionnaire (see Appendix 2), where respondents must indicate whether they ‘agree’ or ‘disagree’ with the key definitions, is that if the respondent disagrees with any of the definitions, the application should take the respondent directly to the ‘Thank you’ page (see Section 3.2.5). As SG provides the facility to create logic, branching and jumping etc., we were able to meet this requirement. SG also provides features which allow the data to be analysed and presented in a comprehensive way. For example, it supports question-by-question analysis, cumulative reports and so on. Furthermore, it provides the option to download the resulting data as a file e.g. a CSV file. This feature was very useful for us as we intended to create custom reports by transferring the data into a database and running a custom query(s) to analyse the data.

Publishing our questionnaire was also quite easy with SG as it provides almost all the standard ways of doing so, including through search engine indexing, sharing the URL, email campaigns and social networking. We found this very useful and helpful during the online-questionnaire publishing phase.

Besides all these features, one of the most attractive advantages for us was that SG offers free accounts with full, unrestricted features for student research projects.

### 3.2.2. Sample Selection

To provide the relevant knowledge and understanding, a respondent’s role becomes more vital. Therefore, we chose to target business intelligence (BI) consultants in our study. The key characteristics of BI consultants are often involved in many different projects and therefore have broader and more knowledge on SD incorporation into DW-based BI solution and understanding of the current market. These characteristics are instrumental in gaining significant insights into current practice in the marketplace and help to answer our research question in a satisfactory fashion. However, as BI is a rather broad knowledge domain, we also targeted individuals in roles such as BI Business Analyst, BI Business Modeller, BI IT Designer, BI Production Support, BI Manager, Data Steward, Head of Information Management, IT specialist etc.

There are many ways to solicit responses online and respondents can be contacted via emails, online forums or web sites (Singh et al., 2009). We started by contacting prospective respondents via one of the professional
forums on LinkedIn We first got a subscription to the BI Intelligence groups and then invited the other members in these groups to participate in our study. The same procedure was also used on other social networking websites like Facebook and other open BI forums.

3.2.3. Ethics, Privacy, and Security Concerns

Ethics, privacy and security concerns are very important in online-questionnaire research. Online-questionnaire raise several kinds of privacy breach issues including those that are physical, informational, psychological or interactional (Singh et al., 2009).

Keeping in mind these concerns and to reassure our respondents, before administering the questionnaire, we comprehensively described how potential ethical, privacy and security concerns in our study would be addressed. We also let the respondents know how we would handle all the data from their responses in order to help them feel more comfortable about taking part in the online-questionnaire. Our statement on how ethical, privacy and security concerns will be addressed is given in Appendix 2.

From a data security point of view, SG is responsible for keeping the data from each online-questionnaire safe, as their online data protection policy guarantees. However, the online questionnaire data was also downloaded as a Comma-Separated Value (CSV) file and imported it into the database system, in order to run custom database queries for data analysis, thus removing the data from SG’s care. Where this occurred, as stated in the privacy policy (see Appendix 1, Appendix 2), we ensured that only those involved with conducting the project would be able to access the responses obtained.

3.2.4. Developing Questions for the Online-Questionnaire

In the previous sections, we described the key decision points in designing the online-questionnaire. Once these were dealt with, the next step was to develop the actual questions for our study (see Appendix 2). At this stage, we were careful to bear in mind that these questions should facilitate the collection of the required material or data such that the aims and objectives of our study would be satisfied.
There were also several other considerations that were taken into account during this process. Firstly, we considered the fact that the questions should be drafted in a way that makes their context clear to respondents. To achieve this, we included an introduction section in our online questionnaire (see Appendix 2) which explained the context of the study as well as the definitions of key terms that would be used in the questions asked. Secondly, we tried to ensure that the questions were worded in a manner that allowed them to be quickly and easily understood as well as responded to. There were number of iterations in drafting questions and this resulted in a number of changes including adjusting the sequence of the questions, rewording them, adding as well as removing some questions and finally dividing them into the three groups of: Background, Main Questions and Concluding Questions. The purpose of each groups are presented in next section (3.2.5). However, in short, the questions and their related response fields, including the fixed alternatives, are based on findings in existing literature – particularly the work conducted by Strand and associates. This is in alignment with the argument presented in 3.1, where the need for a domain and method-triangulating approach of their work was expressed and motivated.

### 3.2.5. Online-questionnaire Implementation

The online questionnaire consists of five web pages in total (Appendix 2) and its basic structure is shown in Figure 5. In this section we briefly explain the purpose of each page and describe how we applied question logic, jumping and branching features, as described in section 3.2.1, to different pages and questions.

**Introduction:** This web page describes the purpose of the online questionnaire as well as the aims and objectives of our study. It also provides respondents with information on the potential ethical, privacy and security concerns involved.

**Background:** The aim of having background questions is to obtain more details about the respondent’s company, country, industry, and experience. All of the questions are optional except for Question 6, which asks the respondent whether he/she would ‘agree’ or ‘disagree’ with the important definitions that are used in our study. This question must be answered in order to proceed to the next page. As it is vital to assure that the respondents share the same definitions chosen for this work regarding BI, DW and SD. Furthermore, if the
respondent ‘disagrees’ with any of the definitions presented, the application will jump to the “Thank you” page. The same kind of logic is applied for Question 11 which asks the respondent about external data incorporation into DW-based BI solutions. If the respondent has never used external data in the projects which he/she has participated in, then there is no need to present main questions as they are all based on external data. Thus, the application takes the user directly to the concluding questions page.

![Figure 5: Basic structure of the online-questionnaire](image)

**Main Questions:** This web page presents the main questions which ask the respondent about different types of SD, SD sources, acquisition approaches, integration methods and finally, the usage of SD. As mentioned, these are the main questions in our study and the resulting responses determine whether the aims and objectives of our study are met successfully (see Appendix 2).

**Concluding Questions:** The fourth page of the online questionnaire presents the concluding questions which ask about the future trends in the incorporation of SD, the need for SD incorporation within DW-based BI solutions and also for any additional comments the respondent may have.

**Thank you:** The final page of the online-questionnaire thanks the respondent for taking part and presents a message of appreciation as well as the ‘Submit’ button.
To minimize the rate of incompletion, we include an estimate of the time needed to fill the questionnaire in the invitation letter sent to respondents (see Appendix 1). This allows respondents to better plan their availability to fill the questionnaire. Additionally, we have incorporated a questionnaire progress bar feature which helps the user to see the number of questions that have been completed and to estimate the completion of the questionnaire in terms of percentage. (Lumsden, 2006)

3.2.6. Data Analysis

For a statistical analysis of each variable, we have used several ways to calculate the average. By average, we mean a single value which gives a general representation of the entire data set in place of the random distribution of a list of values. As we have used different measures of the average including the mode, median, arithmetic mean and rating scale, we will now briefly go through each of these types of statistics.

Mode – the most common value in a data set
Median – the value in the middle of a distribution of values that have been arranged in ascending order.
Arithmetic Mean – Sum of all value divided by the total number of values
4. Analysis and Results

In this section, the results of the online-questionnaire are presented. A total of 95 responses were received, out of which 50 are complete, 5 are partial, and 40 are blank. A complete response is one where the respondent has viewed all the pages of the online-questionnaire (for reading purposes, the terms on-line questionnaire and questionnaire will hereafter be used interchangeably), answered at least the compulsory questions on the background page (Question 6, see Appendix 2) and submitted the questionnaire. In contrast, if the user exited the online-questionnaire at any stage without submitting the questionnaire (see Appendix 2), The response is considered as partial. There are also 40 responses which are blank. Blank means that the respondents have read through the introduction page of the questionnaire, but did not proceed any further, i.e. the respondents have not given any responses to the actual questions in the questionnaire.

For obvious reasons, blank and partial responses has not been included as part of the material being analysed. The summary of the responses are found in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>No of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed</td>
<td>50</td>
</tr>
<tr>
<td>Partial</td>
<td>5</td>
</tr>
<tr>
<td>Blank</td>
<td>40</td>
</tr>
<tr>
<td><strong>Total Individual viewer of the online-questionnaire</strong></td>
<td><strong>95</strong></td>
</tr>
</tbody>
</table>

For being able to fulfil the aim of this work, it is crucial that the underlying empirical data obtained is from respondents who, besides contributing with a complete response, also share our understanding/interpretation of a “data warehouse” (Section 2.2), “business intelligence” (Section 2.1) and “syndicate data” (Section 2.3.1). It was also considered important that the respondents have an own previous experience of syndicate data incorporation into DW-based BI solutions. An analysis of the complete responses resulted in four groups of responses, following the dimensions agreeing on definitions (agree/disagree) and previous experience of syndicate data incorporation into DW-based BI-solutions (yes/no).
The first group is composed of respondents who agreed with our chosen definitions and also have their own experience of syndicate data incorporation into DW-based BI-solutions. From this group, data from all three question’s categories (background, main questions and concluding questions) was included in the analysis.

The second group consists of respondents who agreed with all three definitions but lacked experience with incorporating syndicate data into DW-based BI solutions. From this group, only data from the third question category, “concluding questions”, was analyzed. This data was compared with the data obtained from the first group, in order to contrast aspects such as current importance of and future trends in syndicate data incorporation into DW-based BI solutions. The reason we did not discard this data is that it would be interesting to compare the viewpoints of those who have relevant SD incorporation experience with those who do not. This would allow us to ascertain whether there are any general differences in opinion between the two groups regarding the future of incorporating syndicate data into DW-based BI solutions. In short, the comparison was intended to broaden the understanding of syndicate data incorporation into DW-based BI-solutions.

In the third group are the respondents who disagreed with one or more of the definitions listed in the online-questionnaire. Since they did not agree with the definitions of the vital concepts that were on which this study is based, their responses have not been included in the empirical material that was analyzed. The fourth group is composed of respondents who not only disagreed with the definitions but also had no prior experience with SD. Based on the same reasons for discarding the data obtained from the third group, these responses have also been excluded. The distributions of the respondents over the four groups are summarized in Table 2.

Table 2: Completed Response Breakdown

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Definitions</th>
<th>Experience with Syndicate Data</th>
<th>No of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agree</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>Agree</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>Disagree</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Disagree</td>
<td>No</td>
<td>0</td>
</tr>
</tbody>
</table>
An important point to note about the online-questionnaire is that the respondent is only required to answer Question 6 (see Appendix 2) in order to proceed to the next page (where the main questions are located), which is important for study (see section 3.2.4 for more detail). All other questions are merely optional. This means that respondents could skip any questions they wished to. According to Singh et al., (2009), it is problematic and unethical for respondents not to be given the choice to skip questions they do not want to answer. We therefore provided the option of answering certain questions with “N/A” (not applicable) where appropriate. Providing the option to answer a question with “N/A” always poses a challenge when it comes to interpreting the resulting quantitative data. After all, “N/A” may mean either of the following:

1) The respondent simply did not want to answer a particular sub-question or failed to choose any of the possible answers provided;
2) The respondent did not find any of the other alternatives applicable and thus chose “N/A” as a valid and reasonable answer in itself.

For added clarification, the number of respondents choosing the answer “N/A” for each question in the main section of the questionnaire are indicated accordingly, so that the total number of responses always sums up to 30. This enables us to clearly see the proportion of respondents who answered “N/A”.

As a final comment, it is worth mentioning that the estimated time for answering the complete questionnaire was 28 minutes, which is considered as a fairly long time. However, this was unavoidable given the number of questions the respondents were asked. This estimation was obtained using a SG feature that displays an approximation of the accumulated answering-time as questions are added.

4.1. Background

Distribution of respondents by geographic region

Below, the geographical distribution of the respondents is illustrated in Figure 6. As can be seen, most of the respondents (30%) are residing in the UK. 17% of the respondents are from the United States of America (USA), 23% are from various countries across Europe and the remaining 30% of respondents are from 6 different regions throughout the world.
Although this study was intended to be internationally representative, the results show that it was dominantly participated by respondents from two countries, the UK and the USA. This was initially surprising because invitations to answer the questionnaire were sent using open BI professional forums around the world and more respondents from other regions were expected. Nonetheless, there could be many reasons this geographical distribution of respondents was obtained. For example, there may be less BI consultants from other parts of the world participating in professional and social groups. Alternatively, there could have been a language barrier, as the online-questionnaire was only presented in English.

Another possibility is that the geographical distribution of respondents mirrors the usage of SD around the world. Although this is merely speculation at the present time, SD incorporation into DW-based BI solutions might be more prevalent in the UK and USA compared to other countries. As a result, these countries would yield far more individuals who have an interest in SD incorporation or qualify as respondents. In contrast, BI consultants from countries where SD incorporation is relatively uncommon may have felt that they lacked the relevant knowledge and therefore did not even bother to answer the online-questionnaire.
Distribution of respondents by working experience

The amount of experience the respondents had with respect to working with DW-based BI solutions is varied. However, most respondents reported having 4-5 years’ of experience (33%). A fair amount of the respondents (27%) were also highly experienced in this field, having worked with DW-based BI solutions for more than 10 years. The full distribution of the working experience of the respondents is shown in Figure 7.

![Pie chart showing distribution of respondents by working experience](image)

**Figure 7: Distribution of respondents by working experience**

On the whole, these results are natural and as per expectation. As we mentioned in Section 3.2.2, we mainly aimed to target experienced BI personnel and this also seems to have had an impact on the resulting set of respondents. Highly experienced BI experts may have also been more willing to respond to the questionnaire due to a greater level of confidence in their ability to contribute to the study in terms of knowledge. As can be seen, only a handful of people with less than 1 year of BI experience responded. Others who are just starting to gain expertise in this area may have felt that they have no valuable insights to offer and so did not fill in the questionnaire.

**Distribution of respondents by business role**

The majority of respondents were BI consultants, constituting 40% of the total number of persons who took part in this study. Other prominent roles were BI
business Analyst (20%) and BI Manager (17%). Only 11% of the respondents did not specify their role (i.e. N/A). The role distribution of the respondents is illustrated in Figure 8. As mentioned in section 3.2.2, only respondents holding certain BI roles would be able to provide the specialist knowledge and understanding relevant to the current study. Thus, when recruiting respondents, we specifically targeted BI consultants. The fact that the results are showing approximately half of the respondents to be BI consultants is therefore very natural and as per our expectation.

Figure 8: Distribution of respondents by business role

Distribution of respondents by current Industry

Consulting (23%) appeared as the most common main industry of the respondents; followed by Information Technology (20%), Banking (10%) and Retailing (7%). However, the results showed a large range of industries being represented in our sample, as illustrated in Figure 9.

This distribution generally reflects a pattern that would be expected in the overall BI industry. After all, most BI experts work within consultancy firms and engage with clients across several industries. Additionally, the high proportion of respondents who work in Information Technology (IT) can be explained by the link between IT, which involves the gathering, processing and
distribution of data, with BI. Additionally, the implementation of BI systems always relies heavily on IT.

![Distribution of respondents by current Industry](image)

**Figure 9: Distribution of respondents by current Industry**

**Distribution of respondents’ previous industry experience**

As mentioned (and illustrated in Figure 9), respondents most commonly listed their current industry as consulting. This corresponds with the role of “BI consultant” being held by a large proportion of them (role distribution shown in Figure 8). In addition to their main industries, we asked respondents to tell us the industries in which they have conducted BI projects. Banking (50%) appeared to be the most common sector for this, followed by Insurance (27%) and Automotive (23%). All others sectors which were specified are reported in Figure 10.

Banking, Insurance and Automotive are all sectors which involve a high degree of risk and speculation. Poorly informed judgements within these industries can
lead to massive financial losses. It is thus logical that the application of BI is particularly common in these areas.

![Figure 10: Industry distribution by previous BI projects](image)

4.2. Main Questions

4.2.1. Identification

Identifying sources of Syndicate Data

The respondents named 28 SDSs, with the most common ones being Experian (40%), followed by AC Nielsen (30%) and Equifax (30%). In the following subsections, some details regarding these three, most common SDSs, are given. The descriptions are not intended to be complete, but they were included to give some details regarding the variety of business conducted by these suppliers.
Experian is one of the largest global information service companies and operates in 80 countries around the world. It helps businesses to manage credit risk, fraud prevention and target marketing, covering a variety of sectors including financial services, retail, home shopping, telecommunications, utilities, media, insurance, automotive, leisure, charity and property. In dealing with credit risk, Experian also compiles credit profile reports and credit ratings for individuals as well as businesses. A credit profile provides a substantial amount of information about an individual or business including current and previous addresses, credit account details, credit account histories, and so on. This may be used by organisations such as banks or other businesses to make a decision about credit lending or business partnership opportunities. Such data could be acquired on demand or via subscription and is available for both individuals and businesses.

Equifax is one of the largest credit reporting companies in the world and operates in 14 different countries across Northern America and Europe. Like Experian, it supplies credit intelligence to individuals as well as businesses and mainly provides data related to credit profiles and credit rating.

In contrast, the services offered by AC Nielsen allow businesses to measure their market performance, gain insights into consumer behaviour, analyse market dynamics, resolve marketing and sales issues, and capture growth opportunities. This array of services is mainly based on providing market data that assists businesses in exploiting opportunities in the market.

As Pirttimäki (2004) states, it is very important for organisations to recognise and utilise external information in order to understand the whereabouts in their business environment. To cater for this understanding, they also require reliable data from trusted suppliers. Thus, one would assume that the sources preferred by BI experts are those which they have found to be most accurate and dependable. The data sources that were most commonly referred to by the respondents in our study were large scale companies operating in many countries around the world and with an above millions client-base. While not explicitly stated, it seems that BI experts find these well-known companies which specialise in collecting and compiling data – in other words, established SDSs – to be the most reliable sources of SD. This is in accordance with the findings reported by Strand and Wabler (2004) claiming that data acquired from SDSs is more reliable and of a higher quality, compared to other sources
of external data, such as the Internet, bi-product data suppliers, and governmental agencies.

**Approaches to identify syndicate data suppliers**

In order to find out which approaches organisations most typically use to identify SDSs for DW-based BI solutions in practice, we presented respondents with four different possible approaches, based on the work by Strand et al., (2004b) and asked them to indicate how common each one is in current practice. Our findings are presented in Figure 11.

![Figure 11: Commonness of SDSs identification approaches](image)

The first approach that used to identify a SDS is through utilising previous personal experience of the consultant. 37% of respondents reported this approach as “very common”, while 30% of respondents said that it was “common” and 7% of respondents considered it “rare”. Given a rating scale of 1 to 4 where 1 = never used and 4 = very common, a mean rating of how common this approach is was calculated and we found that, on average, it was
regarded as “very common”. The popularity of this approach could be attributed to the fact that the respondents in our sample have an average of 4-5 years of experience working with DW-based BI solutions (see Section 4.1). They would thus have plenty of previous experience of data suppliers to draw on and even built up some own preferences.

The second approach, which in short described as existing collaborations, 33% of respondents rated this as “very common”, 27% as “common” and 20% as “rare”. The average rating for this approach confirms it as common.

Meanwhile, the third approach for identifying SDSs was defined as “Data suppliers heard of (companies heard of)”, meaning that data suppliers are identified based on their reputation. Most respondents rated this approach as “common” (37%) but a fair amount considered it to be “rare” (26%). Unlike the first two approaches, only 17% of respondents rated this approach as “very common”. The average rating for this approach showed that, as a whole, it is considered “rare”. The fourth, and last, approach organisations could use is to conduct an “Active search to identify suppliers”. This does not a very popular option, with the majority of respondents rating it as “rare” (37%), 30% as “common” and hardly any respondents stating that it is “very common” (10%). The mean rating for this approach also showed it to be “rare”.

If we compare the overall rating for each of the four approaches above, it is clear that the first approach is the one most commonly used to identify external sources. This makes sense given that most of our respondents are BI consultants (section 4.1) and thus have a significant amount of experience working with data suppliers on DW-based BI projects (section 3.2.2). After all, as Strand et al. (2004b) states, consultant experience is one of the most common approaches to identifying reliable SDSs.

SDSs identification is comparatively more difficult than identifying sources of internal data because external data is located outside an organisation’s boundaries. It is difficult not only to identify knowledge of the available external sources, but also to verify the quality of the data (Strand et al., 2003). Data validity and data reliability thus become important constraints in the SDSs identification activity. This was reflected in our results which showed that organisations prefer to rely on identification approaches where the validity and reliability of sources have been verified through first-hand experience (Approaches 1 and 2) rather than those where sources are only known through
what has been heard (Approach 3, average rating = rare) or where there is completely no prior personal knowledge about the sources (Approach 4, average rating = rare).

Our findings thus allow us to conclude that while there are many approaches to identifying SDSs. BI consultants or experts in similar roles can play an integral part in the identification activity. Any previous experience of working with different SDSs (whether by the consultant or within the organisation itself) is also valuable as it ensures knowledge about the data’s quality and reliability.

4.2.2. Acquisition

Contractual settings for data acquisition

Generally speaking, there are, based on the work by Strand et al., (2004b) two main data acquisition approaches that organisations may use when syndicate data is to be incorporated: on-demand and via subscription contracts. The commonness of these approaches is illustrated in Figure 12.

![Figure 12: Commonness of syndicate data acquisition approaches](image)

Acquiring data “via subscription”, was considered to be a “very common” approach by 44% of the respondents, which is a relatively large proportion of the sample, compared to those who thought that it was “common” (23%) and
those who considered it to be “rare” (13%). Indeed, given a rating scale of 1 to 4 where 1 = never used and 4 = very common, a mean rating of how common this approach is showed it to be, on average, “very common”. Although “on-demand” data acquisition was also found to be used fairly often, it seemed to be a slightly less favoured method in comparison. Just 20% of respondents agreed that it is “very common”, while 37% said that it is “common” and 17% considered it “rare”. 3% of respondents even stated that they had “never used” the approach of acquiring data on-demand. On average, on-demand data acquisition was rated as “common”.

This is in alignment with Strand et al.’s (2004a) findings which revealed that the more common approach between the two contractual settings is to acquire data via subscription. Still, it would seem natural to assume that on-demand data acquisition would be more popular because it is usually the cheaper option. However, when making a choice between these two acquisition approaches, one has to consider more than the surface costs of SD. To elaborate, medium-scale companies might not need a large data set and therefore, opting for a subscription contract would probably cost them more than is worthwhile, thus they may be advised to acquire the data on-demand. However, large scale companies usually require a large data set and need to keep data fresh (i.e. frequent up-dates and refresh intervals). Therefore, it would be more efficient and economical for them to sign a subscription contract.

**Methods of data distribution**

Syndicate data may be distributed using different methods. Based on the work by Strand et al., (2004a), the respondents were presented to four different methods. In order to collect new knowledge, the respondents were also given the opportunity to include methods not covered by Strand et al., (2004a). A summary of the resulting responses is given in Figure 13.

Out of the four alternatives we presented, FTP technology and DVD-ROMs or CD-ROMs is most widely used by all respondents. Practically all respondents reported FTP technology to be either a “very common” (37%) or “common” (58%) distribution method, with only 5% rating it as “rare”. For DVD-ROMs or CD-ROMs, only 5% reported them as “very common”. Nonetheless, they were still considered to be “common” by a significant proportion of respondents (74%). Despite the general popularity of this medium of data
distribution, 21% stated that they had “rare” this method to distribute data. A possible reason for this is that DVD-ROMs, and more so CD-ROMs, are slowly being phased out as forms of portable data storage. The usage of email attachments amongst the respondents seemed to be more mixed, with 12% regarding this approach as “very common”, 35% perceiving it to be “common”, 35% stating that it is “rare” and finally, 18% reporting it to be a method that they “never used”. The popularity of web hotels was similarly divided but was clearly less used than email attachments. Only 36% reported it to be “common” and 28% stated that it was “rare”. 36% indicated that this method is “never used”.

![Commonness of syndicate data distribution methods](image)

**Figure 13: Commonness of syndicate data distribution methods**

As mentioned above, the respondents were also given the opportunity to include other methods, not covered in existing literature. Two other methods were mentioned: “Application API” (referred to by only 1 respondent who rated it as “very common”) and “Web Services” (63% rated as “very common” and 38% rated as “common”).
According to the average rating system, the most common (i.e. “very common”) method of data distribution appeared to be via “web services”, which is a method suggested by several respondents using the “other” answer option. In comparison, using FTP technology or a DVD-ROM/CD-ROM both received an average rating that showed them to be “common”; and using an email attachment or a web-hotel to distribute data was overall considered to be “rare”.

Although the average ratings suggest that web services are the most common way to distribute data, this result was calculated from responses given by a very small number of respondents who only made up a small proportion of our entire sample. As such, it is likely that these ratings are based on a disproportionately skewed sample and cannot be considered as a statistically reliable conclusion. Thus, it is important to expand our data analysis to gain more insight and a clear answer to the question of which syndicate data distribution methods are most commonly used. However, it is first necessary to understand what the term “Web Services” actually means.

According to Bosworth (2001), “Web service’ describes specific functionality, value delivered via Internet protocols, for the purpose of providing a mechanism for another service or application”. In terms of how web services can be used to distribute SD, one scenario is to make a company’s processes available to the appropriate partners and vendors using the Internet (Bosworth, 2001).

Since some of the respondents mentioned web services as “very common”, it is reasonable to assume that the SDSs use service-oriented architectures which they use as a base-line for delivering syndicate data through web services. We argue that it is the advancements in network computing and the Internet that has helped web services to become an increasingly important method of distributing SD. Furthermore, they are a cost effective, reliable and a quick way of distributing data to corporations.

Going back to the issue of data analysis, it would be constructive to examine the actual number of participants who indicated how common each distribution method is. For FTP technology, which can be described as a means of distributing data through the Internet in a secure way, 7 respondents provided a rating of “very common” and 11 respondents rated this method as “common”. In contrast, only 5 respondents rated data distribution using web services as
“very common” and 3 respondents provided a rating of “common”. Based purely on a direct comparison of the number of respondents providing a rating of “very common” or “common” for each distribution method, FTP technology is the most common way to distribute data to corporations. This result is in alignment with the existing research by Strand et al. (2004a) which states that FTP technology is the most common distribution approach of SD. Still, since Strand et al., (2004a) already indicated that XML is an existing method, it is not unlikely that the web services will increase its commonness even further in the future.

4.2.3. Integration

Data Integration Approaches

After the data has been acquired and transformed according to given rules, the next stage is to integrate it into a repository component of a DW-based BI solution (often, the data is directly integrated in some way or another into, or aligned with, the data storage component of the DW). Depending on how the data will be used, a number of different data integration approaches may be applied. Figure 14 illustrates the outcome of the online-questionnaire with respect to the four types of data integration approaches (adopted by Strand et al., (2004b)) and how common they are.

The first approach, star-schema dimension integration, includes external data that is stored as a separate dimension in a star-schema (see section 2.3.5). The results show that close to half of the respondents (37%) regarded this as a “very common” approach, while most of the remaining respondents (33%) stated that it was “common”. Only a minute minority (7%) indicated it as “rare”. Interestingly, none of the respondents indicated that this approach as “never used”, which may be taken as evidence for the fact that this approach is largely accepted. The second approach, dimension attribute integration, involves separate attributes which contain the external data and are stored in a dimension that is mostly based on internal data. This approach was also much regarded as “very common” (40%) or “common” (30%), with only 7% respondent rating it as “rare”. The third approach, attribute value integration, is one in which internal and external data is stored and mixed under one attribute. Once more, most respondents thought that this approach was “very common” (40%), or “common” (30%), with just 3% respondent rating it as “rare” and another 3% reporting that it is “never used”. Conversely, the fourth, and last,
approach, spread-sheet integration, was largely regarded as “rare” (32%) or “never used” (13%). In this approach, data is not actually integrated into the DW but is instead stored in a spread-sheet manner and mainly used for comparison with the internal data.

![Graph showing commonness of syndicate data integration approaches](image)

**Figure 14: Commonness of syndicate data integration approaches**

Using the means calculations previously explained (section 3.2.6); the first three approaches were all found to be, on the whole, “very common”. As a matter of fact, all of them had almost identical average rating values. This suggests that all three approaches are roughly equally common when integrating data into DW-based BI solutions. This verifies the work by Strand et al. (2004a) who believed that:

“External data may be acquired and integrated in many different ways. Probably, if this study was conducted in a few years’ time, there would probably be more than one organization that applied more than one integration approach” (Strand et al., 2004a, p.512).
Our results are the actualisation of Strand et al.’s (2004a) speculations. As can be seen, the first three data integration approaches accounted for are very commonly used. In terms of the rarity of the fourth, and last, data integration approach, even the results of this work confirm the results presented by Strand et al. (2004a) already in their study, a similarly small percentage of the respondents reported this approach as being used. Still, in the perspective of advances in analytics and BI, one may speculate that the spread-sheet approach will increase or may arise under several, diverging shapes.

4.2.4. Usage

From a syndicate data usage perspective, this section accounts for the following aspects. First, with respect to different categories of SD, 9 broad categories, adopted from (Strand et al., 2004b) were included. The categories included; sales and marketing data, credit data, competitive data, industry data, economic data, demographic data, commodity data, psychometric data and meteorological data and asked them to rate its commonness. In addition, the respondents were also given the opportunity to report any other category which they have encountered. According to the Results (as illustrated in Figure 15), for “Sales and Marketing Data”, in this category data involves that leads to take initiatives of prospects or established customers, 50% rated as “very common” and 23% rated as “common”. “Credit Data” information in this category used to assess the risk about landing money to the customer, 44% rated as “very common”, 17% as “common”, 13% as “rare” and only 3% rated as “never used”. “Competitive Data” helps businesses to get broader view of the current market landscape and also helps decision makers to make decision effectively based on competitive data, 30% rated as “very common”, 33% as “common”, 7% rated as “rare”. “Industry Data”, category includes all types of data that generates from an industry which does not belong to your own industry, 34% rated as “very common”, 20% as “common” and 13% as “rare”. “Economic Data”, this category may refers to syndicate data types, which provide information about the past and present of the economy and are in the form of time series, 40% rated as “very common”, followed by 23% as “common” and only 7% as “rare”. “Demographic data“, data in this category belong to human population and represent geographical location, identification or describing populations, 43% as “very common” and 27% rated as “common”. “Commodity Data”, data in this category belongs to demand and supply of goods, 17% rated as “very common”, 33% rated as “common”, 10% as “rare” and only 3% rated as “never used”. “Psychometric Data”, in this category data
is belong to psychological measurement includes measurement of knowledge, abilities, attitude personality traits and educational measurements, 3% reported as “very common”, 7% as “common”, 7% rated as “rare” and 40% rated as “never used”. “Meteorological Data”, includes data which determine the impact of weather and climatology on local or regional level, 4% respondents rated as “very common”, 23% rated as “common”, 37% “rare” and only 3% “never used”.

![Commonness of syndicate data categories](image)

Figure 15: Commonness of syndicate data categories

According to our statistical data analysis, “Sales and Marketing data” category appeared most common among the 9, followed by “Credit Data”, “Competitive Data”, “Industry Data”, “Economic Data” and “Demographic Data” appeared as “common”. Only “Meteorological Data” category appeared as “common” and “Psychometric Data” category appeared as “never used”.

43
Specific Types of Syndicate Data Used:

In order to determine what kind of SD being incorporated into DW-Based BI solution, the respondents were asked to give examples of SD used in their current or previous DW-based BI projects. This query was presented as an open-ended question accompanied by empty text boxes (see Appendix 2) which allowed respondents to give answers based on their experience. The respondents reported using a vast variety of SD types (35 in total) but the most common appeared to be credit ratings, customer addresses, sales and marketing data, and economic data. As mentioned, other less universal examples of SD were also given but these will be described later on in this section.

Many respondents indicated that credit ratings are often incorporated into DW-based BI solutions. A credit rating is basically an evaluation of how likely it would be for an individual or a corporation to default on their debts. This is determined from the said individual or corporation’s past record of acquiring credit and making adequate repayments along with their asset availability and extent of liabilities. Such information helps companies to make strategic decisions when establishing the feasibility of starting business partnerships or trading with other worthwhile companies. Credit ratings also help in assessing an individual’s financial position before money is e.g. lent to him or her.

In the work carried out by Strand et al. (2003), it was found that credit rating is an important type of SD. Indeed, the findings presented above show that credit data, a category of SD which includes credit ratings, is one of the most common type of SD used. This would be expected, considering that banking is the industry in which respondents most commonly reported having conducted DW-based BI solutions (see Section 4.1). Given that the financial sector deals extensively with lending to consumers and forming partnerships with other businesses, it is natural that many of the respondents mentioned credit ratings as a key example of SD that has been incorporated into their projects. Analysis of these results leads to the conclusion that credit ratings play an integral role in terms of SD usage in informing strategic decision making.

From the responses given in the online-questionnaire, customer addresses appear to be another common example of SD used by BI consultants. This information assists organizations to; update address information, locate customers, and analyse the geographic distribution of their clients, vendors and other relevant market players. Customer addresses are also used in marketing
initiatives; for example, where promotional material and advertising is sent to prospective or established customers. It is necessary to remember that the type of data referred to when using this term is not restricted to mailing addresses and post codes but also comprises other contact details such as telephone and fax numbers. All of these are used to collect information about possible new customers as well as emerging companies. As was explained, this data is mainly used in direct market campaigns and thus must be kept updated and accurate, as a means to avoid “bad will” and costly returns due to notifications or payments that have been sent to the wrong place or person (Strand et al., 2004a). It is worth pointing out that customer addresses fall under the sales and marketing category of SD which is also regarded as one of the most commonly used categories of SD as reported by BI consultants. However, the importance and frequent usage of this particular example of SD has already been reported in existing literature; most notably, in Strand and Wangler’s (2004a) interview study on external data incorporation into DWs.

Many respondents also gave the examples of economic data or economic statistics when listing the types of SD that have been incorporated into their BI projects. These are usually quantitative measures of an economy over a time series; for example, the consumption and income levels for a set of households in a country. Such data is then used to aggregate sectors and industries into one economy. From our viewpoint, it is important for an organization to integrate economic data into DW-based BI solutions because it allows those in charge to determine how current economic factors may affect their companies. This in turn helps ensure that the strategic decisions made will lead to the best outcome.

**Application Areas for Syndicate Data:**

As we discussed in the previous section, there are many different types of SD which are incorporated into DW-based BI solution and each of these can be applied in various ways, as a means of supporting decision-making and analytical activities. As a starting-point, the respondents were exposed to the six application areas originally reported by Strand et al. (2004b). Following the previous conduct of the questionnaire, the respondents were also asked to indicate how common each application area is. Several empty text boxes marked “others” were also provided (see Appendix 2) so that the respondents could indicate the other application areas which they may have encountered in current practice. This was included as a means to extend current knowledge on
syndicate data application, when incorporated into DW-based BI solutions. Furthermore, a rating of commonness for any additional application areas listed was also requested. The resulting findings are illustrated in Figure 16.

![Figure 16: Commonness of different syndicate data application areas](image)

Out of the six alternatives we presented, updating addresses was considered by respondents to be one of the most common SD application areas. Almost half of those asked (47%) stated that such usage is “very common” and 13% of respondents agreed with the rating of “common”. Only 7% of respondents reported it to be “rare”. Similarly, SD seems to be often used for increasing the precision of marketing campaigns, with 47% of respondents reporting this as “very common” and 17% believing that it is “common”. Once more, only 7% of respondents felt that it was “rare” as an application area. Although, applying SD in the segmentation of stakeholders was also fairly common, it appeared to be practised slightly less frequently. In comparison to the first two alternatives, fewer respondents (37%) rated this usage area as “very common” and the tendency to regard it as “common” (20%) or “rare” (10%) was higher.

Incorporating SD into DW-based BI solutions for the purpose of identifying customer trends seemed to be a popular practice, with almost all respondents reporting this to be either “very common” (40%) or “common” (30%). The same applies to the application area of determining geographical coverage.
which was rated as “very common” by 33% of respondents and “common” by another 30%. Meanwhile, using SD to support a star-schema dimension design stood out as more atypical. Just 17% of respondents indicated that this was “very common” and not many more rated it as “common” (13%). Besides being labelled as “rare” by 20% of respondents, some even believed that it was “never used” (7%).

According to the average rating system that has been used throughout our statistical data analysis, the SD usage areas of updating addresses, identifying customer trends and increasing the precision of marketing are all “very common”. In general, these results support the existing literature on SD which highlights updating customer addresses and increasing the precision of marketing campaigns as major usage areas (Strand et al., 2004b). However, there is one notable difference. In their study, Strand et al. (2004b) reported that only 25% of respondents mentioned identifying customer trends as a usage area and thus concluded that it is not “very common”. In contrast, our statistics reveal that this no longer holds and identifying customer trends has clearly emerged as an important application area for SD. Given the increasing level of competition in the market, it is enormously beneficial for leading companies to constantly stay one step ahead and to remain savvy about customer behaviour. For example, having up-to-date data about customer buying trends and total household incomes can aid an organization in working out the demand and supply in a particular region and thus adjust its output to obtain maximum growth. Knowing about customer trends could also help companies to launch new products and services in ways that are more likely to appeal to consumers.

Many companies across all industries have succeeded in improving their business competitiveness by implementing customer relationship management (CRM) initiatives. This is evident from their increased returns and ability to decrease operational expenses. Managing customer relationships (which often involves the use of customer addresses) is a powerful and efficient strategy for maximising customer satisfaction and loyalty (Chen and Popovich, 2003).

As a whole, using SD both in the segmentation of stakeholders and to determine geographical coverage appears to be “common”. This shows good alignment with Strand et al.’s (2004b) study results. Unfortunately, none of the respondents took the opportunity to mention other application areas for SD.
4.3. Concluding Questions

4.3.1. The Current trend in Syndicate Data Incorporation

According to most of the respondents (57%), SD incorporation occurs frequently. In fact, 20% of the respondents indicated SD as “always” being incorporated into their DW-based BI projects, while 24% stressed that this practice is “very common” and 13% stated that it is “common”. In contrast, only 23% of respondents reported that it is “rather uncommon” (Figure 17). These results lead to the conclusion that SD incorporation into DW-based BI solutions is widely used to enrich data contents, which enables companies to derive the full value of their data contents and improves their ability to exploit opportunities in the market. The existing literature is in support of these findings and SD has repeatedly been found to be a very important ingredient for decision making in BI (Pirttimäki, 2004; Watson and Wixom, 2007; Bogza and Zaharie, 2008).

![Figure 17: Commonness of syndicate data incorporation](image)

4.3.2. Perceived Importance of Syndicate Data in Generating Insights

In order to account for how prior experience of SD incorporation into DW-based BI solutions may influence perceptions regarding its importance in
generating insights, the responses obtained were divided into two major groups (refer to Section 4, Table 2 for more details). The data from the first group (respondents who have experience of SD incorporation) revealed a greater tendency to view syndicate data as very important in generating insights using DW-based BI solutions. As seen in Figure 18, 53% of the BI consultants considered SD to be “very important” and 20% stated that it was “important”. Only 10% of them rated it as “less important”. A notable finding was that there were no respondents at all who indicated that SD was completely unimportant for generating insights. The mean rating, given by the respondents in the group representing an own experience, shows that, SD incorporation is “very important”.

![Pie chart showing syndicate data’s importance with SD experience](image)

**Figure 18: Syndicate data’s importance (with SD experience)**

In contrast, while most respondents who did not have experience with SD incorporation still regarded it as “important” when generating insights, less of them reported it as “very important” (26%). Instead, 56% of these respondents rated it as “important” and 6% thought of it as “less important”. A small proportion of respondents (6%) were “unsure” of its significance, which can probably be attributed to them having an insufficient amount of knowledge to make a definitive judgement about the importance of SD incorporation. As with the first group, it was noted that none of these respondents believed SD
incorporation to be totally unimportant. The perception regarding the role of SD in generating insights is illustrated in Figure 19.

![Figure 19: Syndicate data’s importance (with no SD experience)](image)

A statistical analysis carried out to determine the average rating given by respondents from this group found that, collectively, they perceived SD incorporation into DW-based BI solutions as “important”. The data acquired from this group is very interesting because it suggests that, even without first-hand experience of SD incorporation, BI consultants understand its value in enriching data contents. This means that the complete lack of SD incorporation in the DW-based BI projects these respondents have participated in should not be taken as an indication of a disregard for SD itself. There could be many other reasons they have no experience with SD incorporation. For instance, they may simply not have had any opportunities to apply SD or the companies they are working with might not have had the budget for it. Alternatively, such a project may be in the pipeline for them but is not yet at a stage of implementation.

Before drawing any conclusion about the overall importance of SD in generating insights, it would be beneficial to combine the results from both groups of respondents (Figure 20). The combined results reveal a general consensus that SD incorporation is important. Almost half of the sample (44%) considered it to be “very important” and another 32% rated it as “important”. A
very small proportion of the sample (9%) considered it to be “less important” and only 2% of respondents were “unsure”.

![Figure 20: Perceived importance of syndicate data](image)

In summary, regardless of prior relevant experience or not, both groups share the view that SD is important and helpful in generating insights using DW-based BI solutions. This conclusion is very much in line with the existing literature which states that external data (which syndicate data is a sub-type of [Strand et al., 2003]) is a very important ingredient for decision making in BI (Pirttimäki, 2004; Watson and Wixom, 2007; Bogza and Zaharie, 2008). As Pirttimäki (2004) reminds us, any information or expertise that enhances an organisation’s ability to make effective decisions is beneficial and should not be ignored. Instead, effort should be made to acquire and take full advantage of it. Therefore, internal data on its own is insufficient for understanding the implications of the current state of the market. Additional insights are needed for more accurate business forecasting, improving strategic decision making, detecting novel business opportunities and sustaining a competitive edge.

### 4.3.3. Future trend of Syndicate Data Incorporation

As confirmed in previous section, SD is commonly employed in DW-based BI solutions and is regarded as important for generating insights (Section 4.3.2). In this section, the expected future trend for syndicate data incorporation is explored. As before (see Section 4.6.2), separate analyses of data from two the
different respondent groups (Section 4, Table 2: Group 1 and 2) were conducted to gain a better understanding of how prior experience of SD incorporation might influence respondents’ opinions.

Amongst the respondents with SD incorporation experience, 30% agreed that incorporating SD into DW-based BI projects would “drastically increase” in the future. A slightly bigger proportion (33%) predicted that it would simply “increase” and hardly any respondents (3%) thought that it would “decrease”. A small percentage of respondents (3%) were “unsure” what would occur. Based on a rating scale of 1 to 5 where 1 = unsure and 5 = drastically increase, a calculated mean of the answers given by these respondents indicated that, on average, SD incorporation is expected to “drastically increase”. A more detailed breakdown of the responses is provided in Figure 21.

![Pie chart showing future trend of SD incorporation](image)

**Figure 21: Future trend of SD incorporation (with SD experience)**

Interestingly, almost all respondents with no SD incorporation experience, (81%) thought that SD incorporation into DW-based BI projects would “increase”. Although none of the respondents in this group believed that SD incorporation would “decrease” in any way, 13% stated that they were “unsure” what would occur in the future with regards to this practice. The collective expectation for the future trend of SD incorporation is that it would
“increase”. The full range of responses that were given is illustrated in Figure 22.

Figure 22: Future trend of SD incorporation (with no SD experience)

For a broader picture of the expected future trend of SD incorporation, we combined the results from both groups as illustrated in Figure 23.

Figure 23: Expected future trend of syndicate data incorporation
Almost all the respondents (70%) expected SD incorporation into DW-based BI solutions to “increase” in one way or another, with 20% even going so far as to say that it would “drastically increase”. As illustrated in Figure 23 above, just 2% predicted that the use of SD incorporation would fall. A data analysis using our average rating system showed that, as a whole, SD incorporation is expected to increase. This clearly suggests that the market for syndicate data will continue to grow and that the majority of organizations expect to keep using it. This is in alignment with the standpoints expressed by Ponniah (2010), who explicitly mentioned that the growth in the implementation of DWs has increased significantly and that DW developers now have access to many other types of syndicate data that were
5. Conclusions

This chapter summarises the general conclusions that can be derived from the results presented in the previous chapter.

The results of this study show that, in general, companies pay a lot of attention to the accuracy and reliability of the SD they wish to obtain. Wherever possible, data is sourced from reliable and specialised SDSs. In the data analysis, Experian, Equifax and A.C. Nielsen emerged as the most commonly used SDSs. All three of these corporations are globally renowned commercial data suppliers and specialise in collecting, compiling and selling SD. In terms of how such SDSs may be identified, two approaches are being used commonly in the current market: 1) the consultants are familiar with a particular supplier through personal experience or participation in previous projects, and 2) the organization is already working with a supplier in some way and thus that SDS becomes “inherited” for BI purposes. This suggests that companies in the current market mainly rely on some form of previous experience (either their consultants’ personal experience or knowledge “inherited” from within the organization) when identifying possible SDSs.

Syndicate data acquisition involves two main issues: 1) contractual settings (which determines the frequency with which SD should be acquired), and 2) distribution method (which determines the manner in which SD is transferred). With regard to contractual settings, it was found that both subscription-based and on-demand data acquisition are widely used. However, companies are more likely to sign a subscription contract than acquire data on demand. As with most business decisions, cost is the key determinant of which approach a company will choose. This is in turn affected by the size of the data set required and how up-to-date the data must be. Acquiring SD via subscriptions is more suitable for companies which need to refresh their data very frequently. It also meets the requirements of having a large data set and thus is generally more suited to large scale companies. In contrast, an on-demand contractual setting is more suitable for companies which do not require frequent up-dates for their data and which are satisfied with a relatively small data set. As a result, small to medium sized companies usually prefer on-demand data acquisition for SD.

With regard to how SD is distributed during the acquisition activity, evidence from this study indicates that FTP technology – which is a means of
Conclusions

distributing data through the Internet securely – is very common. Web services are also increasingly becoming an important data distribution method because they are a very cost effective, reliable and quick way of distributing data to companies. The growing popularity of this SD distribution method is one of the more significant findings to emerge from the present study as it is a completely new contribution to the existing literature on SD acquisition.

After the data has been acquired and transformed, the next step is to integrate it into a DW-based BI solution. Out of the four SD integration approaches that were named in Strand et al.’s (2004a) study, three were found to be very common in the current market. These are: star-schema dimension integration, dimension attribute integration and attribute value integration.

The relevance of the above finding is clearly supported by Strand et al.’s (2004a), prediction that “Probably, if this [their] study was conducted in a few years time, there would probably be more than one organization that applied more than one integration approach”. Evidently, there is no longer one particular approach of integrating SD which can be claimed to be the most common. The growing amount of SDSs and different types of SD available means that companies must now be willing to apply several approaches simultaneously, depending on which is most suitable for the data being dealt with. Nonetheless, the present research shows that the fourth approach, spreadsheet integration, is nearly fully phased out; mainly due to advancements in BI-tools in general and internet technology in particular.

Once SD has been integrated it can be used, and at this point, the business decision maker comes into play. In order to understand how SD is used in DW-based BI solutions, it is vital to recognise which types of SD that are being incorporated. The results reported here indicate that there are at least 35 different types of SD being used. However, the main examples of SD utilised on a regular basis are credit ratings, customer addresses, sales and marketing data and economic data. The prevalent use of these specific types of SD ties in fairly well with the areas of application for SD that were found to be most common; namely, updating addresses, identifying customer trends, and increasing the precision of marketing campaigns.
6. Discussions

6.1. Reflection on the Research Approach

Carrying out this study has provided valuable experience in terms of the planning and implementation needed for achieving the aim of this work.

Overall, the implementation of the research process proceeded according to plan. The only weakness we encountered with our chosen research method was the lack of face-to-face interaction with respondents, which produced a possibility of the questions and their contexts being misinterpreted. Although we provided some contact details in the Introduction section of our online questionnaire so that respondents could seek clarification if needed, an improvement would have been to also include other means of communication such as live chat or other forms of social media.

In terms of selecting an online-questionnaire software tool, SurveyGizmo (SG) proved to be a good choice as it provided almost everything we needed to implement our online-questionnaire. While SG also had many helpful data analysis tools with which to analyse responses, these were ultimately not used because they only produced a general data analysis report that lacked the level of depth required in this study. Instead, we exported all the gathered data from SG to our database and performed data analysis.

In regard to data collection, the online-questionnaire was kept live for 16 weeks, which in fact was 4 weeks longer than originally intended. This is the only part of the process that deviated from the estimations made in our project plan, and can be attributed to the way we recruited our respondents. We initially attempted contacting our target sample by posting a general invitation on websites like LinkedIn, Facebook and other BI forums, but this failed to elicit many responses. Thus, we decided to contact each potential respondent with a personalized invitation, which was very effective and enabled us to achieve our targets but was more time consuming.

The number of partial and blank responses that were obtained was relatively high compared to that of complete responses. Presumably, this was due to the length of time required to complete the entire questionnaire. To ensure that respondents were able to allocate the time needed, we took the precaution of informing them approximately how long the questionnaire would take. This
estimate was based on both our own time assessment and that made by SG. Even so, one of the respondents explicitly complained that it took longer than expected. In view of this, while we acknowledge that it is very difficult to estimate exactly how long a questionnaire containing a combination of open and close-ended questions would take, it may still be worthwhile to conduct a more intensive analysis to produce a truly accurate estimation.

On the whole, the responses given were very insightful and helped develop a deeper understanding regarding SD incorporation into DW-based BI solutions. Thereby, the choice of method, with the underlying intention of method triangulating the work conducted by Strand and associates, must be considered as proper. Only one portion of the results – those regarding SD application areas – were somewhat disappointing. We expected respondents to introduce new SD application areas and were surprised to find that none were reported. While we were unable to find a definite reason for this, a plausible explanation is that respondents were simply feeling tired by the time they reached this question, which is stated at the end of the main questions section. At this point, the SG tool estimates (see. previous explanation of this tools feature in section 3.2.1) that the respondents have spent 24 minutes with the questionnaire.

None of the respondents voiced any concerns about ethical issues or the privacy policy which was clearly stated in the Introduction section of our online-questionnaire as well as in the invitations we posted. This is a very good sign and indicates that the respondents were satisfied with our data handling methods.

### 6.2. Aim fulfilment

In order to fulfil the aim of this work, “to explore the current practice of syndicate data incorporation into DW-based BI solutions”, the following research objectives were established:

1- Which syndicate data supplier that are used and how are they identified?
2- How is the syndicate data acquired?
3- How is the syndicate data integrated into DW-based BI solutions?
4- How is the syndicate data used in DW-based BI-solutions?

In the following section, discussions on the answering of each of the research objectives are included.
**Objective 1 - Which syndicate data suppliers that are used and how they are identified?**

This work contributes with new knowledge on the specific SDSs that are being used in the market and highlights the strategies preferred by organisations when identifying which SDSs to use. In addition, it provides a frame of reference for consumers who are unsure on how to approach SDSs. Based on the above, we argue that Objective 1 has been achieved, since it contributes with new knowledge and thereby extends related work.

**Objective 2 - How is the syndicate data acquired?**

This study contributes by confirming the existing knowledge on contractual settings for data acquisition and provides a frame of reference for organisations that wish to choose a data acquisition approach that best suits their syndicate data incorporation requirements. With regard to data distribution, evidence from this study confirms existing knowledge on data distribution methods. In addition, one of the most significant findings to emerge from this study is that web services appear to be an increasingly popular SD distribution method. To summarise, this work offers new knowledge on how SD is acquired and also confirms the results of Strand’s research. Therefore, it can be argued that Objective 2 has been accomplished.

**Objective 3 - How is the syndicate data integrated into DW-based BI solutions?**

This work contributes with new knowledge on the way current advancements have influenced how SD is integrated into DW-based BI solutions and provides an up-to-date comparison with the results obtained by Strand et al. (2004a). Based on the conclusions and the results presented in the previous chapter, we argue that Objective 3 has been met.

**Objective 4 - How is the syndicate data used in DW-based BI-solutions?**

The results of this work indicate that there are many types of SD available and highlight the most common types that are used in current practice. Furthermore, it provides vital information on how SD is used. Even though these results are mostly confirmatory and do not really add much knowledge
Discussions

about application areas that were previously unaccounted for, they do corroborate previous work. Therefore, it can be argued that Objective 4 has been answered.

6.3. Putting the results in wider context

The contribution of this work is relevant to researchers as well as practitioners and increases the current knowledge regarding SD incorporation into DW-based BI solutions. It does so by confirming and adding to the available information on the identification, acquisition, integration and usage of SD. In addition, it extends current knowledge by introducing details (e.g. on suppliers and acquisition approaches) previously unaccounted for in published literature. Furthermore, the findings of this work are a significant contribution to the scientific community because they enable the identification of future research directions (as accounted for in Section 6.4).

To be more specific, the findings can be used by companies who already practice SD incorporation to work out what other types of SD would enhance their decision capabilities and what the application areas of these types of SD are, thus maximising the full value of the data content. Furthermore, companies could compare the capabilities of their existing SDSs with those reported here, which would ultimately help them work out the best SDS for providing quality data that can enrich the data content of their DW-based BI solutions. Companies which are planning to explore SD incorporation for the first time can also benefit. Our findings provide a lot of information regarding the process of incorporating SD into DW-based BI solutions, since it is based on insights provided by actual BI practitioners with ample experience in SD incorporation within the current market. The results also provide support to SDSs who are in the process of enhancing their data capabilities as it offers useful information on current trends in SD consumption; for example, the types of SD which are most in demand, the types of SD distribution technology competing SDSs are using, and so on.

Most of the results obtained in this study are very well aligned with the existing literature and should be considered relevant to the subject. The only issue is the extent to which they can be generalized. Our initial intention was to include an internationally representative sample in our study. Unfortunately, this was not possible and around 50% of the respondents were either from the United Kingdom or the United States of America. Still, one may argue that 30
complete responses, from experienced consultants with an internalized knowledge from relevant projects, constitutes a good ground for being able to state that the results are rather generic and creates a valuable extension to current literature.

Furthermore, the results are also intended to inspire and guide companies to make further use of their DW-based BI-solutions as a means of increasing their competitive edge. In this perspective we have found that SD is definitely an opportunity path to follow. As Ponniah (2010, p.60) states: “Syndicate data is becoming big business”.

Finally, whenever one embarks on a journey related to the investigation of BI usage in general, and adding on data to large DWs in particular, the implications of the resulting findings need to be considered from an information ethics point of view. Given that the sole purpose of these systems is to integrate huge amounts of data related to companies or private persons, it is inevitable for ethical concerns to arise. The results of this work show a number of application areas where SD may contribute by “completing” the picture businesses have of their customers, including through providing additional information regarding sales and marketing patterns or credit ratings. Both these data types may be used to support decisions regarding customer segmentation or to hinder individuals from becoming customers. Although these decisions are perfectly sound from a profitability point of view and these applications areas are part of the actual reason SD so important, one may ask, “At which point will we have crawled too deep under the skin of our (e.g.) customers?” There is no easy answer to this question but it must be considered, especially since the results of this work show that SD incorporation will increase and that it is considered as very important by more than half of the respondents in this study.

6.4. Future Work

As with all research, we hope that this work contributes new knowledge that will be used by others and that it will have an impact among practitioners. Still, research is an on-going process, where bits and pieces are tied together to form something bigger. As such, keeping in mind the motive for this work, we argue that there is a need for further research within this area. Although the results of this work reveals some new details and confirms current research, there is still a lot to be done. Therefore, we suggest some ideas for future work.
Firstly, since we previously commented upon the rather brief and undetailed answers related to applications of SD, it would be valuable to conduct further studies which are delimited to the usage activity of the incorporation process. In addition, extending current knowledge regarding SD usage is vital for creating an in-depth understanding on how SD may contribute in different decision-making situations and to verify the actual impact SD has on decisions being made. Therefore, further studies aimed at giving details regarding SD usage are needed.

Secondly, the aim of this work was to explore current incorporation of SD into DW-based BI solutions. As in previous work, we triangulated the method used for extracting new knowledge. Yet, there are many details left to be uncovered, and to do so, further methods may need to be applied. For example, in-depth case-studies at experienced companies would probably contribute many useful insights, especially since such studies would uncover how SD is used in day-to-day activities. Case studies would also broaden the coverage of respondents and other materials since such studies would be targeted towards the actual users of SD, rather than the consultants and others we targeted in this work, in order to achieve a more general understanding.
7. References


Appendix 1: Invitation

Dear Respondent

We are carrying out a study into “Syndicate Data Incorporation into Business Intelligence (BI)”. We are inviting business intelligence consultants (or people with a similar level of experience) to complete our study and we would like you to be one of our participants. Your help will be highly appreciated.

Your responses will remain completely confidential and secure. Only authors and people related to the project related will be able to see the transcripts. Once we have finished this study, all responses will be presented in a summary format and all primary data will be destroyed.

If you have any questions please send an email to: a10amjri@student.his.se if you would like to participate, please follow the link below:

http://edusurveygizmo.com/s3/491128/syndicate-Data-Incorporation-into-Business-Intelligence-BI

We hope that you will take few minutes to complete an online-questionnaire.

Thank you in advance for your valuable feedback.

Kind Regards,

Amjad Riaz
Appendix 2 Online-questionnaire User Interface

| Welcome to online-questionnaire. |

**Introduction:**

The purpose of this online-questionnaire is to explore how external data from specialized data suppliers (hereafter referred to as syndicate data) are used in data warehouse (DW)-based BI-solutions. To succeed with the aims of the research, obtaining your knowledge and expertise is crucial. The information collected will be used to identify the types of syndicate data and their suppliers. Furthermore, the purpose of the online-questionnaire is to find out which types of analysis and decisions are supported by the syndicate data. In addition, the results of the work will be used to assist organizations in improving their competitive edge and help them to make more out of current investments in DWs and BI-solutions, as well as current and future investments in external data.

The results of this study will be presented in a master thesis, written under the supervision of the University of Skövde, Sweden. If you want to receive an e-copy of the thesis you may state so on the next page and an e-copy will be sent to you immediately after the completion of the thesis (presentation planned for mid-June 2011).

**Privacy Policy:**

Your responses will be completely confidential. Only authors and project related people will be able to see the transcripts. Once we have finished the study, all responses will be presented in summary format. Your comments might be quoted but in an unidentifiable manner.

**Online-questionnaire Support:**

If you have any questions you can email allamp@student.hs.se with your questions. We will respond as promptly as possible.

We hope you will take few minutes and contribute with your expertise.

Disclaimer of Liability | Contact us

Next
Syndicate Data Incorporation into Business Intelligence (BI)

Background

1. Your Full Name

2. Your company

3. In which country are you personally located?

4. Do you want to receive a copy of the report?
   If 'yes' then you must provide your email address.
   ○ Yes
   ○ No

6. There are many definitions of BI, DWHs and syndicate data. For comparative reasons, it is therefore vital that you adhere to the following definitions of the concepts introduced at the Introduction page.

<table>
<thead>
<tr>
<th>Term</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Intelligence (BI) is a broad category of technologies, applications, and processes for gathering, storing, accessing, and analyzing data to help its users make better decisions. (Winsem &amp; Watson, 2016, p. 14).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data warehouse: A data warehouse is a subject-oriented, integrated, Time Variant, non-volatile collection of data in support of management's decision making process. (Kimball, 1996, p. 23).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syndicate Data: Syndicate data is business data (and its associated metadata) purchased from an organization specialized in collecting, compiling, and selling data, and targeted towards the strategic and/or the tactic decision making processes of the incorporating organization. (Brand et al., 2005, p. 95).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why? (only if you disagree).

7. How many year(s) have you been working with BI solutions?

8. Which role best describe your current position?

<table>
<thead>
<tr>
<th>Role</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>If your role does not exist please specify 'other'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Which is your main industry?

<table>
<thead>
<tr>
<th>Industry</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>If your industry does not exist please specify 'other'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
10. In which industries have you conducted BI projects where syndicate data has been incorporated into DW-based BI solutions?
Select all that apply
- Aerospace and Defence
- Agriculture
- Automotive
- Banking
- Chemical/Petroleum
- Construction and Real Estate
- Consulting
- Consumer Goods
- Education
- Energy/Natural Resources
- Entertainment/ Media/Publishing
- Government/Public Sector
- Healthcare
- Information Technology
- Insurance
- Investment Management
- Logistics and Distribution
- Manufacturing
- Pharmaceuticals
- Professional Services
- Retailing
- Telecommunications
- Transportation/Travel/Tourism

11. Has syndicate data been incorporated into any DW-based BI project, in which you have participated?
- Yes
- No

How common is it?
- Please Select -
12. Which syndicate data suppliers have been used in the projects you have participated in? Please give the company names

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

13. Suppliers of syndicate data may be identified in various ways. Please indicate which approaches you have experienced and how common they are?

<table>
<thead>
<tr>
<th></th>
<th>Very common</th>
<th>Common</th>
<th>Rare</th>
<th>Never used</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>The consultant is familiar with some supplier(s) from previous projects (Personal experience AND Participation in other projects)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The organization is already working with the supplier (Inherited)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data suppliers heard of (Companies heard of)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active search for identifying suppliers.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other

Comments

14. The data acquired from syndicate data suppliers may follow different contractual settings. Please indicate how common each approach is?

<table>
<thead>
<tr>
<th></th>
<th>Very common</th>
<th>Common</th>
<th>Rare</th>
<th>Never used</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Via subscription contracts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On demand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other

Comments
15. The syndicate data may be acquired into organization by applying different data acquisition approaches. Please indicate how common each approach is?

<table>
<thead>
<tr>
<th>Approach</th>
<th>Very common</th>
<th>Common</th>
<th>Rare</th>
<th>Never used</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Transfer Protocol (FTP) technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-ROM or CD-ROM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email attachment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments

16. The syndicate data may be integrated into organizations by applying different data integration approaches. Please indicate how common each approach is?

<table>
<thead>
<tr>
<th>Approach</th>
<th>Very common</th>
<th>Common</th>
<th>Rare</th>
<th>Never used</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star-schema dimension integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension attribute integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute value integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spreadsheet integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments

17. Syndicate data may be clustered into the following categories. Please indicate how commonly they are used?

<table>
<thead>
<tr>
<th>Category</th>
<th>Very common</th>
<th>Common</th>
<th>Rare</th>
<th>Never used</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales and marketing data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitive data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commodity data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychometric data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meteorological data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments
18. What syndicate data have been incorporated into BI projects that you have been involved in? Please give examples of specific data

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

19. Below you find common application areas of syndicate data. Please indicate how common they are and feel free to introduce other applications not mentioned in the default alternatives.

<table>
<thead>
<tr>
<th>Application Area</th>
<th>Very common</th>
<th>Common</th>
<th>Rare</th>
<th>Never used</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update addresses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increase the precision of marketing campaigns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation of stakeholders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify customer trends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine geographical coverage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support star schema dimension design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other 1 (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other 2 (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other 3 (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other 4 (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other 5 (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other 6 (Please specify below)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other 1:   
Other 2:   
Other 3:   
Other 4:   
Other 5:   
Other 6:   

Comments:

...
Syndicate Data Incorporation into Business Intelligence (BI)

Concluding Questions

20. How important do you consider syndicate data to be in generating insights using DW-based BI solutions?

- Very important
- Important
- Less important
- Not important
- Unsure

21. Syndicate data incorporation into DW-based BI solutions will

- Drastically increase
- Increase
- Decrease
- Drastically decrease
- Unsure

Why?

22. What do you consider about the future trends, regarding the incorporation of syndicate data into DW-based BI solution?

23. Is there anything you would like to add?

Thank you for completing our online questionnaire of “External Data Incorporation into Business Intelligence (BI)”. Your inputs will be very valuable and play a key role to build the report.