Data Discovery—Right on Time

How Automation Underpins Data Discovery

AUGUST 2017

A White Paper by
Dr. Barry Devlin, 9sight Consulting
barry@9sight.com

A Discovery Hub® is an architectural construct designed to underpin good data quality and governance for business intelligence (BI) systems based on data discovery tools. It addresses the challenges that arise when data discovery tools, such as Qlik or PowerBI, are successfully adopted and drive business people to forage for source data wherever they can find it.

This paper proposes three occasions during the roll-out of BI when it is appropriate to adopt a Discovery Hub. Experience shows that it most often occurs after data discovery has become valuable to the business and when IT is struggling to maintain data quality in the face of business data collection from disparate sources.

Four use cases show how this approach, combined with TimeXtender Discovery Hub®, applies across different industries and business sizes, offering significant business benefits and improved data provision and use.

Contents

IN COLLABORATION WITH TIMEXTENDER

www.TimeXtender.com

W hether you are a small, medium or large organization, you may have many challenges in the emerging digitalized world. Tracking and managing your performance should certainly *not* be among them.

Modern self-service, *data discovery* tools offer business users a considerably easier and more productive environment to identify opportunities and analyze problems than previous business intelligence (BI) tools. The freedom to explore data without having to rely on a small and overstretched IT team to build and customize reports allows true innovation in decision making.

With freedom comes responsibility—the obligation to ensure that the data underpinning discovery is fit for purpose. That such data is correct, comprehensive and timely. With freedom to play with data, business users are tempted to gather it from every corner of the enterprise. Sometimes that's fine. But for core business information—the record of the true, legally binding performance of the business—strict control is needed to comply with regulatory bodies and, in some cases, to protect the CEO from jail! For other types of internally sourced data, consistency across business functions is highly desirable to reduce internal conflict or improve customer satisfaction. And with the increasing use of externally sourced information, management of such sources and consistent alignment with internal data is vital.

A Discovery Hub® is an architectural structure that provides the required level of governance and agility for data used in self-service BI to address the above responsibilities. The Discovery Hub structure consists of four layers:

- **Operational Data Exchange (ODX):** a store where IT provides **access** to all data of likely value and with minimal cleansing or transformation from the sources of interest

- **Modern Data Warehouse (MDW):** contains **trusted**, cleansed, and consolidated core business information, preferably in a loosely normalized, usage neutral structure, to support the common data needs of data discovery and self-service environments

- **Semantic Layer:** modeling capabilities to define and deliver department- or purpose-specific data structures and function with superior **ease of use**, based on common business terms and standard definitions

- **Front End:** a wide range of tools, including self-service BI tools such as Qlik, Power BI, Tableau, Excel that offer **visualization** or reporting of business data; statistical tools that provide predictive **analytics**; and emerging machine learning tools that augment human decision making
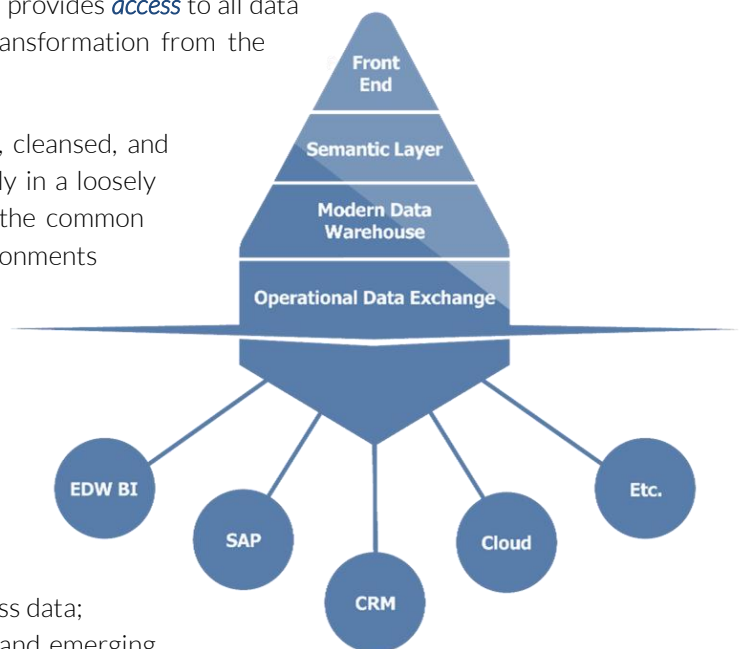

Figure 1: Discovery Hub architecture

Designing. building, populating and maintaining a governed and agile Discovery Hub is a project requiring both technical and business skills, as well as clearly defined responsibilities across business and IT departments. In particular, the Operational Data Exchange

is a job for IT, whereas the other layers are typically modelled by business users with do-main knowledge (such as BI staff). The job can also be outsourced to an experienced ser-vices company. In either case, you need them to be fast, efficient and value for money. Traditional ETL approaches tend to be too complex or expensive in these circumstances, and often address only the process between the operational sources and the ODX.

Automating the entire process—from source systems to front end and from design through to ongoing support—is central to speed and efficiency in delivering and updating a Discovery Hub. We thus require *Data Discovery Automation (DDA)* tooling, which ex-tends the existing concept of data warehouse automation to span the four layers of the Discovery Hub and, in particular, to support the data discovery tools used directly by business people. TimeXtender's DiscoveryHub® software has been designed with these aspects in mind.

So, what would justify investment in the Data Discovery Automation tools and processes to populate and manage a Discovery Hub? Of course, the business case starts with better and faster data-driven decision making and action enabled by data discovery. However, the actual justification for the Discovery Hub emerges only as the business realizes that the agility provided by the data discovery must be backed by responsive governance of data sourcing in order to assure the quality of the decisions taken. This is the topic of the next section.

Practical Indicators for a Discovery Hub

S preadsheets! I only mention them here because they have long illustrated the challenge of getting the right balance between data-savvy business users and those responsible for data quality—often, and incorrectly, assigned to IT.

Since the first spreadsheets appeared—way back in 1979 with VisiCalc on the Apple][—business users have had the ability to explore data to their heart's content. That's won-derful and it remains so in data discovery. However, these liberated business users also collect data for their spreadsheets from every imaginable source. That's less wonderful if you are concerned with data quality, reliability, and compliance.

Addressing data quality with a warehouse

The most common solution to this problem has been to implement a data warehouse as the certified data source for all such free-wheeling business users, whether they are us-ing spreadsheets, BI or data discovery tools. The data warehouse assures data quality, but often also introduces complexity and delays, first in its initial delivery and later in its ongoing maintenance via traditional ETL tools. Data Warehouse Automation addresses such agility problems, but on its own may not be enough. The combination of a Discovery Hub and DDA takes this focus on quality and agility to the next level, setting the stage for the next generation of business intelligence.

Consider, for example, a situation where a BI or discovery tool such as Qlik has first im-proved business user productivity. This is often followed by the realization that there ex-ists a data quality problem as users begin to get data from multiple, uncertified sources. IT is called in and builds a data warehouse as the mandated source for data. In some cases,

the initial implementation goes well and problems only arise when user needs or source systems change. In other cases, the initial implementation is too slow and cumbersome[*]. In either case, the initial and welcome improvement in BI productivity is poisoned by well-meaning attempts to solve data quality issues. IT and the warehouse are blamed.

This situation is quite common and often leads to stalled data discovery projects with consequent loss of business value, confidence, or interest. The solution is to move to a Discovery Hub approach supported by Data Discovery Automation.

Addressing data quality with a Discovery Hub

The design of the Discovery Hub is specifically adapted to meet the needs of both quality and agility of a data discovery environment. The preferred, loosely normalized, wide-table model in the MDW layer is general enough to support agile data discovery and lightly enough modeled to combine data quality and agile maintenance as data needs change. Even in cases when the MDW is modeled in traditional more normalized or dimensional formats, the discipline instilled by the ODX and Semantic layers can aid agility and encourage a gradual migration to the preferred MDW model.

These benefits apply to three distinct scenarios where a Discovery Hub may apply:

1. *Early roll out with initial data discovery project*

   Implementing a Discovery Hub at the earliest stage of a data discovery roll out has both benefits and challenges. In addition to the clear data quality goal, the availability of the Discovery Hub from the get-go accustoms business users to first seeking the required data there, reducing their need or habit to look elsewhere. This implies that the Discovery Hub must be populated with a superset of the initial data needs of the business users. Deciding the scope and content of the additional data is the challenge: casting the net too wide leads to the type of delays associated with the normalized data warehouse approach, while an overly narrow focus leaves users searching elsewhere too often. A careful but "outside the box" interaction between business and IT—often the role of a BI Center of Competence (BICC)[†]—when defining use cases and their possible future data scope will be essential.

2. *Data discovery maturity leads to data quality and compliance issues*

   This is perhaps the easiest time to introduce a Discovery Hub. A broad set of data requirements—both met and unmet—are already known, so the scope of the data

The most common time to begin implementing a Discovery Hub is after data discovery has been widely adopted by the business and data quality and consistency issues arise from uncontrolled data sourcing.

---

[*] Either of these issues can arise, irrespective of the data warehouse approach taken. A subject-oriented, hub-and-spoke ("Inmon") architecture can lead to a model that is too broadly scoped for a data discovery environment and thus take too long to deliver or upgrade due to its complexity. A dimensional ("Kimball") approach may result in a model that is too narrow and focused on a particular business need and query approach, which reduces flexibility and drives unnecessary rework in the data discovery environment. In essence, both approaches may inhibit agility unless appropriate methods and structures surround them.

[†] See the blog by Mikkel Kvist, Head of BICC at MT Højgaard for valuable insight on how a BICC can help: http://blog.timextender.com/four-signs-your-bi-architecture-might-need-automation

model is well-defined. Users are supportive because they are already aware of data quality issues. The importance of a speedy delivery cannot be underestimated here. Users are familiar with sourcing data for themselves and will be unlikely to tolerate any significant development delays as a result of turning to IT. Ideally, IT should be expecting the emergence of data quality problems (they are widespread!) and implementing the first stages of a Discovery Hub in anticipation of these needs. Yes, this does require IT to take the initiative to invest in a Data Discovery Automation tool before having a confirmed business need!

3. *Classic EDW stalls on data quality*

The existing enterprise data warehouse (EDW) solution will most likely form the starting point for designing the Discovery Hub, unless it has fallen far behind current data discovery needs (in which case, refer to bullet 2 above). If the existing structure is normalized rather than dimensional, the level of redesign will be less.

The goal is to move quickly to a loosely normalized, wide-table design that data discovery users can easily understand and adopt as their preferred data source going forward. To speed the transition, all viable aspects of the existing warehouse data model should be examined and reused in a simplified form. In some cases, it may even be possible to use some or all of the existing EDW as a source for the Discovery Hub. In such a transition, a suitable DDA tool is vital, not only at the beginning of this move but, more importantly, to enable easier and faster reaction to changes in business needs and data sources in the future.

Discovery Hub use cases

The proof of the pudding is in the eating, it is said. So, let's take a look at four real-life Discovery Hub use cases across the three Discovery Hub initiation scenarios—early roll out, data discovery maturity, and classic EDW stall.

The first two use cases relate to the situation when data discovery maturity leads to data quality issues, illustrating that this driver can apply to very different organization sizes and structures. The first example is of a charity, Direct Relief, that operates worldwide, but has a very small, centralized management structure. The second example, Nedap, is also a global organization, but has a more de-centralized management structure. Both introduced a data discovery tool to empower decision makers, and found that data quality and governance issues quickly came to the fore. The third example, ABAX, a fast-growing European telematics company, illustrates the benefits of rolling out a Discovery Hub with the initial data discovery project. The Nordisk Film use case, shows how a Discovery Hub can be a solution when an existing enterprise data warehouse stalls.

Relieving pain and poverty in the world with good data

Charity Navigator's 2016 list of *"10 of the Best Charities Everyone's Heard Of"* put Direct Relief in joint top spot with a perfect score, based on financial performance, transparency, and accountability. Data, analysis and reporting play a central role in ensuring that Direct Relief reached and maintains that score and, more importantly, achieves its vision of improving the health and lives of people affected by emergencies, catastrophes or poverty.

Although only a small organization of some 70 full-time, professional staff, Direct Relief depends implicitly on a wide variety of data from multiple sources to manage its work. These include information about people in need, usage of the products supplied, to whom which products are shipped, donations from medical and pharmaceutical companies, and so on. These are derived from internal operational sources like SAP, as well as a range of Excel spreadsheets and external data sources. For many years, the organization has driven its analysis and reporting needs using complex tools which required in-depth knowledge and were thus confined to a few individuals in the organization.

Direct Relief rolled out data discovery tool, QlikView, as an improved solution to their analysis and reporting needs. Business users from all areas (CEO, warehouse supervisions, Program Managers, etc) now use Qlik on a daily basis and are delighted with the analysis and reporting function provided. As these users have become proficient in Qlik use, they demanded additional data from new sources. Therein lay the new challenge.

IT have long provided the basic data sets for analysis and reporting. They were used to searching for and finding the data they needed. However, it was a completely manual process that left them relying on memory of where they had found what data and/or where they had saved it. With the additional enthusiasm and demands of the business for new data, IT were in danger of becoming a bottleneck to the business use of data.

In essence, Direct Relief had gone from Qlik newbie to expert in months, but quickly reached the limits of the organization's ability to source data quickly and cleanly enough to feed their new environment. Modern tooling was required to achieve new levels of agility in data sourcing and TimeXtender's Discovery Hub® suite fit the bill. Data modeling—understanding the data sources and streamlining naming conventions—was a vital first step to the creation of a Discovery Hub where all common data was collected and made available within a single platform, directly accessible to all Qlik users.

The outcome was an immediate improvement in the speed and agility of data delivery from IT and faster, more reliable analysis and reporting. New information needs for donors or recipient organizations were satisfied more quickly and with better quality. Furthermore, the availability of a well-modeled and designed Discovery Hub allowed IT to take a step back and undertake ongoing examination and improvement of all available information using the metadata gathered and stored in TimeXtender's Discovery Hub®. With the foundational data about funding, donors, products, medical supplies in one place, transparency of information is facilitated and accountability demonstrated.

> **What we wanted to do was very complex but this made everything very easy.**
>
> *Dawn Long, Director of IT and Quality, Direct Relief*

Business users in the warehouse, shipping, and all operations now have direct and immediate access to all data required and can easily identify, for example, what was shipped in any given period, what supplies have arrived, outstanding requests, and so on. With data now well-organized in a Hub, easily sourced and maintained with Discovery Hub® software from TimeXtender, users throughout the organization can ask for the information they need and be assured that it either exists already in the Hub or can be rapidly added.

Looking forward, Direct Relief plan to train some departments to develop their own reports in QlikSense. The data model will still be maintained by IT in the Discovery Hub. This will help address one of the main issues previously seen: the CEO could ask the same question to 3 different people and get 3 different answers. Of course, this typical prob-

lem can't be eliminated entirely, because users can filter any way they like in the Qlik environment. However, having common formulae and standard definitions applied up front in the Discovery Hub provides a more consistent set of results to routine questions.

Ensuring accuracy and timeliness of business insights

Nedap N.V. is a long time global supplier of identification and security solutions, headquartered in the Netherlands and with some 750 staff in eleven offices worldwide. These solutions are also delivered across a range of industries, including Healthcare, Libraries, Retail, and Livestock Management in Agriculture, organized in nine business units.

Like every global enterprise with multiple lines of business, Nedap is faced with the challenge of reporting along different dimensions. Summarization along business line must deal with the differing accounting practices and systems in use in the various countries. Reporting by geography encounters product and marketing differences across business units. Such challenges are often assumed to be the preserve on large multinational corporations with staff and money to throw at the reconciliation issues. But Nedap cannot afford such luxuries. Technology must solve the problem and deliver both accuracy and timeliness of reports and answers.

Nedap's first step was to introduce data discovery tool, QlikView, to empower business people to create their own analyses and gain insight into the different dimensions of the business. Of course, the new-found ability of the business to easily ask any question they wished immediately created a challenge for IT. According to Application Manager, Andre Bleumink: *"to make a meaningful comparison of business units, we would have to load everything from all the different sources before creating a single, standardized database for each report. Eventually, we just couldn't keep up with it."*

To minimize disruption to current and successful systems of reporting and insight within the geographical regions, Nedap's approach was to build a global level system on top of those existing data warehouse solutions. These regional data warehouses already extracted data from the local ERP systems, reconciling it across business units, and creating typical, traditional data warehouse / data mart systems, which delivered departmental and cross-functional insights to management within the regions. The cleansed and reconciled regional data within these warehouses then became the sources for the Discovery Hub running at the global level. Reconciliation at the global level addressed issues of differing data standards across regions and allowed both full company-wide reporting as well as cross-country reporting by business divisions. Combined with the automation provided by TimeXtender, this allowed meaningful and timely performance comparisons between business units and between geographical regions.

> It used to take days for us to create accurate company-wide and cross-country business unit reports, but now it would be more like a few hours. The time saved really is that much.
>
> *André Bleumink,*
> *Application Manager,*
> *Nedap*

Business benefits began with time and cost savings for IT. With the above automation, all consolidation of data from different regions could be handled by just two people. With the extensive logic needed to refine and present appropriate business data encoded in the metadata repository of Discovery Hub® and scripts, data preparation was reduced from days to hours. By assembling all data from multiple sources into a single repository, the need to repeat the same processes multiple times has been reduced and, in some cases, eliminated. Furthermore, the layered structure of the Discovery Hub allowed IT to subset data according to user needs and relabel it with business-relevant names and

structures. The separate Semantic Layer also allowed business people the freedom to use other tools, such as Power BI or Excel, as they saw fit.

Needless to say, the improved accuracy and timeliness of the data was highly appreciated by the business! QlikView reports had already increased the appetite of the business for data with their ease-of-use and stylish presentation. The Discovery Hub provides the assurance to the business that the data is a true representation of the business. If there are problems with the data loads, IT are informed promptly and can take immediate action. With accurate and trustworthy data available from a single source, business uptake grows, decision making is improved and actions are taken more quickly.

From tracking vehicles to measuring KPIs

ABAX, based out of Larvik, Norway, focuses on GPS vehicle tracking solutions in both the business and private sectors, in the Nordic region, more broadly in Europe, and in China., ABAX employs 350 people, covering sales, services, marketing, administration, and development for over 20,000 customers with 200,000 subscriptions.

By late 2015, following a period of rapid growth, both organically and by acquisition, ABAX found that their management reporting approach, then a manual system based on cutting, pasting and sharing Excel spreadsheets, was becoming untenable. An initial proposal for a traditional data warehouse solution was considered too complex, requiring substantial manual maintenance that was beyond the resources of the small IT department. An automated and agile approach was required, capable of delivering time savings of 80-90% in development and maintenance. This led to the implementation of a solution using Discovery Hub® and TimeXtender together with local partner bWise.

The business foundation for these infrastructural decisions was the widespread adoption of a set of key performance indicators (KPIs) spanning the entire business. With a business built on long-term services contracts for tracking and logging asset location and business travel, it's vital that management understands the top measures of business success and can drill into them along multiple axes. Data discovery tool, Qlik, was chosen for its ease of use and flexibility. Business people had standard dashboards, of course, but must also be able adapt to changing circumstances. The ability to take advantage of new, evolving data sources—such as the Internet of Things—was a vital consideration, not just for the front-end tool, but also for the data sourcing.

From the beginning, ABAX recognized that such agility was heavily dependent on data quality and consistency. Ensuring this governance demanded a well-structured environment separating data loading from transformation and cleansing, and from subsequent delivery. Furthermore, given the company's history of acquisitions for growth, the ability to include new and novel data sources into its management information environment at short notice was particularly important. The Discovery Hub approach met these needs.

> The important thing is to get from data discoveries to innovative action... automating that process was key to our success.
>
> *Christian Hofling, Chief Finance Officer, ABAX*

Since implementing their Discovery Hub solution in 2016, ABAX has hired additional BI staff to take advantage of the data now available to gain market share and to meet its goal of becoming best in class in a fast-moving industry.

Through the lens of business diversity

Nordisk Film has been in the movie business for 111 years, which makes it the oldest continuously active film studio in the world. Today, the company does more than make movies and TV series. It has expanded into operating a cinema chain in Denmark and Norway, distributing Hollywood movies, as well as electronic games and consoles, and selling live event tickets. As part of the Egmont Group since 1992, with a 2016 turnover of over half a billion Euro and nearly 1000 staff, Nordisk Film has grown up. Managing the complete electronic entertainment value chain—development, production, marketing and distribution—brings new challenges of dealing with a diversity of business drivers and measures.

A data warehouse dating from 2007—built for a single data source and a limited set of business measures—was becoming unfit for purpose in this new diverse technical and business environment. Managing high volume sales, such as games and tickets, requires very different processes and measures than the traditional film production business, moving reporting and analysis from monthly to daily—and potentially to hourly—as the media business becomes increasingly digitized.

With a multiplicity of new sources at the back-end and growing business demand for timely and flexible analysis and reporting, Nordisk Film decided to replace the original data warehouse with TimeXtender's Discovery Hub® approach. This allowed, for example, business users to access and use a combination of hard accounting data and commercial and customer data from a variety of sources on a single platform to better understand the behavior of cinema and gaming customers and respond more quickly to their changing needs.

In the old data warehouse environment, the task of extracting and preparing data for was complex and cumbersome. In fact, a weekly visit by an external expert consultant was required to hand code specific report requests. With Discovery Hub® from TimeXtender, the vast majority of this type of work now handled by business users and the IT manager working together. More complete and accurate data, provided with speed and agility, now enables business users to become more expert in using analysis and visualization tools, and—more importantly—focus on making better informed business decisions.

> We can have an initial report up and running within 15 minutes, which gives us time to look at the result then develop and optimise the solution.
>
> *Mikkel Hansen, Systems Finance Manager, Nordisk Film*

Conclusions

Delivering a data discovery environment with the right data, on time, and maintaining and updating it consistently over time demands excellence on many fronts. The use cases described here emphasize four key lessons:

1. **Data discovery use drives business benefit**
   Choose best of breed discovery tools that offer ease of use, flexibility, and powerful visualization support to business people to gain maximum benefit from BI investment.

2. **Business benefit drives unmanaged data collection**
   Build or engage in-depth data management and governance skills to ensure and maintain data quality and promote confidence in and ongoing use of data provided.

3. *Unmanaged data collection frustrates business people*
   Use powerful data discovery automation tooling to build, manage, and continuously update and maintain the data supply chain to the data discovery environment with ease and elegance.

4. *Business people need simple, managed access to data*
   Build a Discovery Hub infrastructure as soon as politically feasible—which may only be when business discovers data quality issues.

As seen from these use cases, the combination of a data discovery tool, a Discovery Hub approach, and powerful automation software—such as Discovery Hub® from TimeXtender—offers significant business benefits and improved data provision and use across a wide variety of business sizes and industries. Organizations embarking on data discovery projects in support of business goals would be well advised to consider a Discovery Hub architecture in parallel with the BI roll-out to gain maximum business value and promote best data quality practices.

> A Discovery Hub approach, combined with automation tooling, offers maximum business value and promotes best data quality practices.

Dr. Barry Devlin is among the foremost authorities on business insight and one of the founders of data warehousing, having published the first architectural paper on the topic in 1988. With over 30 years of IT experience, including 20 years with IBM as a Distinguished Engineer, he is a widely respected analyst, consultant, lecturer and author of the seminal book, "Data Warehouse—from Architecture to Implementation" and numerous White Papers. His book, **"Business unIntelligence—Insight and Innovation Beyond Analytics and Big Data"** *(http://bit.ly/BunI-TP2) was published in October 2013.*

Barry is founder and principal of 9sight Consulting. He specializes in the human, organizational and technological implications of deep business insight solutions combining all aspects of internally and externally sourced information, analytics, and artificial intelligence. A regular contributor to Twitter (@BarryDevlin), TDWI, BACollaborative, and more, Barry is based in Cape Town, South Africa and operates worldwide.