

## PUTTING DATA IN BUSINESS CONTEXT

### *THE VALUE AND VIRTUE OF A BUSINESS DIRECTORY*

September 2014

A White Paper by

Dr. Barry Devlin, 9sight Consulting  
[barry@9sight.com](mailto:barry@9sight.com)

*Data without context is meaningless. It is also valueless. Without a well-understood business context, any derived information and subsequent decisions are open to multiple interpretations or, worse, misinterpretation. It is the context—and, by extension, a Business Directory that manages this context—that promotes the value and virtue of data.*

*In a world of rapidly growing big data from many sources, data virtualization is opening a treasure trove to business users. Exploration and experimentation is rife; opportunities for innovative analytics and decision making abound. However, without a process to move from exploration to production, many of the gains may be lost. The adaptive decision cycle and sandboxing, built on the Business Directory, offer a solution.*

*This paper explores the evolution of metadata to context-setting information and describes the Business Directory in which it is stored, managed and made available to business users. It proposes the adaptive decision cycle as a model to bridge from discovery and exploration to production and governance. Finally, it offers an emerging vision of sandboxing as an environment where data can be reunited with the context needed both for informed innovation and optimized production.*

Sponsored by:

Cisco

[www.cisco.com/go/datavirtualization](http://www.cisco.com/go/datavirtualization)



## CONTENTS

- 2 From metadata to context-setting information
- 5 Information usage in the analytic mind
- 6 Business Directory—providing the context to content for business users
- 8 An emerging vision—uniting content and context
- 10 Conclusions

Since the birth of civilization, philosophers have deliberated on the meaning of life. And since the earliest days of decision support, business users have struggled with the meaning of data. Today, as business intelligence is enhanced by business analytics, extended into real-time operations, and offered on smartphones and tablets to everyone from delivery people to C-level executives, the question of *meaning* has become pivotal. In fact, data meaning—in its definition and exposition—constitutes perhaps the most important, yet under-appreciated, aspect of IT in the modern business.

Sadly, this paper is silent on the meaning of life; the meaning of data and the semantics of information are, however, at its core. The meaning of data does, unfortunately, present problems almost as intractable as those of the more existential question. Even the written word itself—the foundation of all definitions of meaning—can be ambiguous. Try reading aloud this short sentence “*I didn’t say she stole my money*” seven times, placing emphasis in turn on each of its seven words and notice how the meaning changes<sup>1</sup>. Small wonder then that the business definition of *profit* can be contentious.

While fundamental issues in treating semantics in computing clearly remain, the urgency to provide some practical solutions increases daily. Data is becoming ever more central to business decision making, and is extending to people with little or no training in data use and interpretation. Increasingly, moreover, data is being combined from multiple sources. Take a now common type of visualization, in this instance from the New York Daily News<sup>2</sup>. The story covers violence and safety on the New York subway, written mainly in terms of incidents per 100,000 trips. The main graphic shows total number of crimes per station over a five year period, and is interactive to allow switching to incident rates. As Robert Kosara discusses in a recent blog<sup>3</sup>, a key question to consider here is which, if any, of the calculations in the graphic or, indeed, many that are impossible due to lack of data, actually measure safety. One might even argue that personal safety is a largely subjective feeling, rendering the entire report meaningless. Meaning may often be in the mind of the beholder.

*As diverse data from disparate sources is combined for analysis by ordinary business users, the provision of real, usable meaning becomes a necessity.*

The ongoing rush to data is this decade’s gold rush; it will deliver immense value to business. In particular, the combination of data from multiple sources—traditional and new, internal and external to the enterprise—is becoming a key competitive arena. However, wresting value from such diverse data demands novel thinking and innovative solutions to one central question: what does the data truly mean to its users? Some answers may lie in two simple, albeit challenging, words: *context* and *usage*.

---

## FROM METADATA TO CONTEXT-SETTING INFORMATION

---

*In common use almost every word has many shades of meaning, and therefore needs to be interpreted by the context.*  
Alfred Marshall

---

As far back as the mid-1980s, the first documented<sup>4</sup> data warehouse architecture, built (by myself and other colleagues in Ireland) for IBM internal use, included a business data directory (BDD). This was defined as storing and managing “*the descriptions of the data stored in the BDW [Business Data Warehouse], the information concerning stored procedures, and the descriptions of business processes*”. Information in the BDD included descriptions of tables, columns, procedures and parameters, values of coded fields, as well as data sources, ownership, derivation methods and update dates. As shown in Figure 1, the BDD was positioned as joined to the BDW itself, that is, stored with the business data in the relational environment.

By the mid-1990s, the information stored in the BDD was widely called metadata. However, because of the focus on populating the warehouse, the metadata collected in many projects was mainly technical in nature, emanating from ETL (extract, transform and load) tools. Experts talked of business metadata to reemphasize non-technical content, while David Marco even declared<sup>5</sup> that metadata must include “...all physical data (contained in software and other media) and knowledge (contained in employees and various media) from inside and outside an organization, including information about the physical data, technical and business processes, rules and constraints of the data, and structures of the data used by a corporation”. Declaring knowledge in employees’ heads to be metadata may be extreme, and may pose some difficulties to its extraction! But, it does emphasize the dilemma: what is metadata?

A return to basics shows that the fundamental purpose of metadata is to provide the *context* in which any piece of information is used or capable of use, and from which its meaning and boundaries can be derived by its user. In well-bounded and -defined applications, such as operational systems, IT has previously modeled the data and business users are familiar with its content and usage; the context is known, and metadata is thus of limited interest in such applications. In decision support, narrowly focused data marts can also thrive in the absence of metadata for the same reasons. Explicitly documented context only becomes vital when data from diverse sources is brought together or when such conjoined or unfamiliar data is made available to new users. This is why metadata first came to the fore with the emergence of data warehousing. And it is why data virtualization and external data further emphasize the need for context for their effective and correct use.

The principle and practice of metadata does, by now, carry some considerable baggage. Its breadth and focus of content depends on who you ask. It is often cited as the area of business intelligence with the least successful implementation. And it is most recently associated with governmental snooping as a class of “non-information” (such as telephone numbers and locations) that can be legally harvested and analyzed at will. Even the moniker *data*, rather than *information*, is unfortunate, given that data is usually seen as highly structured and largely numerical in nature. For all these reasons, I have suggested<sup>6</sup> that we replace the underperforming word *metadata* with a term that accurately reflects its use and structure: *context-setting information* or CSI if you long for another acronym.

### **THE BUSINESS VALUE OF CONTEXT-SETTING INFORMATION**

It’s obvious that data without context has no value. Is that 30 the profit in dollars on selling a bicycle or a summer temperature in degrees Celsius? As the amount of available CSI expands, the business value of the information clearly increases. How was the profit calculated? When and where did that temperature occur? However, adding ever more context may or may not be valuable. The source of the temperature information may give a clue to its likely accuracy, but knowing who defined the calculation of profit in the business may contribute little to a particular analysis. It is not at all obvious at which stage additional context becomes superfluous or even distracting to the business user. In fact, further thought will reveal that the added value of additional CSI depends on a variety of factors, such as the data source (internal vs. external, highly vs. loosely structured, purchased vs. built, single vs. multiple, etc.), type of analysis process (reporting, problem solving, discovery, exploration, etc.), legal weight of the results, and many more. Indeed, the level of personal knowledge of an individual business user can further determine the value of offering more or less CSI in a particular instance.

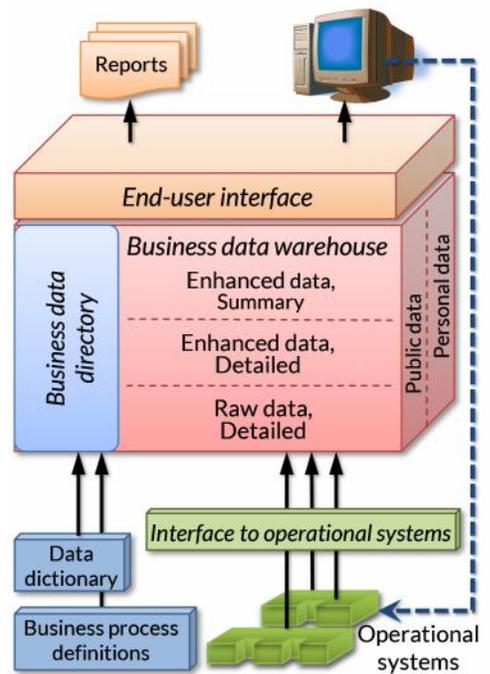


Figure 1: The business data directory, after Devlin & Murphy, 1988

Context-setting information should be considered as an integral part of the information resource of the business rather than something separate and different. Indeed, depending on the specific situation or business usage, the same information may be considered to be context-setting, reference, descriptive or some other type of information. For example, information about the sourcing of some external data set may be CSI for a business analyst, allowing determination of valid uses, but may be billing information for the accounting department. This shift in vantage point enables us to see that CSI comes from many sources beyond ETL or modeling processes. We can see that some CSI, such as column names in a relational table, is very stable, while other CSI, such as date and source of last update, may be quite volatile. We move from trying to define and distinguish business, technical and other types of metadata to seeing that all information use requires context and that any piece of information may set the context for the use of another. We recognize that some CSI is embedded in the system tables of a relational database, while other CSI exists in the documentation of a NoSQL-based system or, as Marco might point out, in the head of the Hadoop programmer.

With such a wide variety of sources, types and uses of context-setting information, it may be tempting to assume that no categorization is either possible or useful. In fact, CSI can be reduced to four broad categories of information needed by a business (or IT) person to determine what can be done with a particular *business matter*, be it information, process or even person<sup>7</sup>. These categories are:

*Context-setting information(CSI) describes the identity, provenance, currency and usage of all information used by the business.*

1. *Identity*: Name, description, meaning, structure, possible values and similar CSI are the most basic, independent attributes of a business matter and are the starting point for any use, from collection, creation or change to analysis.
2. *Provenance*: The source of the business matter in terms of physical location, ownership, guarantees of integrity, etc. that define (or allow definition of) limits and boundaries of acceptable use.
3. *Currency*: The timeliness of a business matter, relating to its absolute time of creation or use, and to its relative time of creation or use with respect to expectations, plans or other relevant events.
4. *Usage*: Any process that can be or has been applied the business matter, capable of altering its content, relevance, structure, etc. and thus influences its further use.

As such context-setting information is progressively (and it will not be done in a single iteration) collected, created and curated in the business, an asset of enormous value emerges. Although often referred to as a *Business Directory*, it is most likely a distributed and diversely structured set of information, of which perhaps a limited core set is separately instantiated and managed. Indeed, given the breadth and variety of the above CSI, any successful implementation will almost certainly limit the amount of unique or duplicate information held and use data virtualization techniques to access required information *in situ*. Interestingly, as data virtualization drives the need for CSI, in turn, CSI drives the need for virtualization.

Finally, recognizing the almost infinite variety of skills and interests of potential users, we see that access to and use of context-setting information must be highly adaptive to user behavior and needs in different situations. For example, a user in data discovery mode will need to explore the CSI in a loose and non-directed manner, while a user checking the validity and application of a query or report will probably follow a highly structured approach. A data scientist has very different needs in exploring or understanding information context when predicting customer churn based on Twitter sentiment and relationships than a C-level executive trying to ensure that a regulatory report is sufficiently accurate and complete to avoid going to jail.

## INFORMATION USAGE IN THE ANALYTIC MIND

*The ultimate authority must always rest with the individual's own reason and critical analysis.*  
Dalai Lama

Just as the old concept of metadata has become inadequate in the world of big data and analytics, so too has our prior model of how business users behave as they explore, analyze and base decisions on information. It's time to move our thinking about usage from the old opposing modes of BI- and spreadsheet-based decision making to an integrated approach that I call the *adaptive decision cycle*.

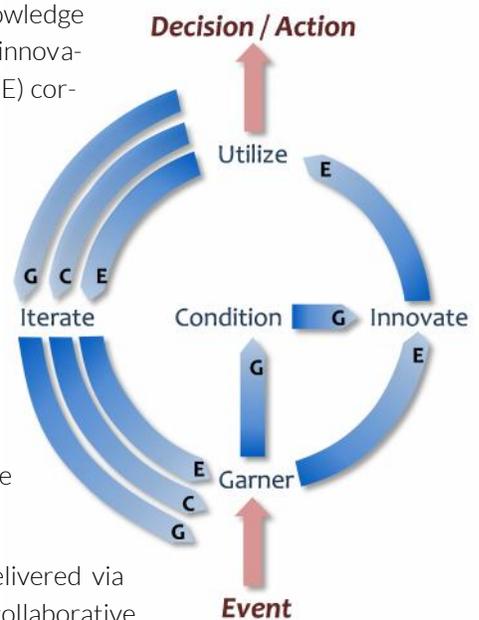
In traditional business intelligence, the flow of information from business events to required decisions is essentially unidirectional. Data is *garnered* from a number of sources and *conditioned*—cleansed, combined and enriched—by IT through a data warehouse or mart prior to being presented to the business user. This user explores and analyzes the received data, *utilizing* it to come to a decision. However, this last step is effective only if the data is complete. BI systems typically allow users to manipulate the data they contain, but often limit the user's ability to add data from additional sources. Instead, users must *iterate* back with IT through a development process to have any additional, required data garnered and conditioned. This model, which is good for report generation and dashboards, is termed *center-out decision making* because it depends on a central authority to ensure data quality for informational needs. However, the delays it introduces in innovative data analysis are unacceptable in many situations. This leads to the alternative *edge-on* model, where business users perform all aspects of data garnering, conditioning and iteration themselves, usually equipped with nothing more than a spreadsheet. Of course, real world BI typically consists of a mix of these two models, which satisfies neither data management nor decision agility needs.

*The adaptive decision cycle models the journey from innovative analytics, through business validation to optimized implementation in production.*

With the ongoing explosion of data sources, from mobile devices to social networking and other big data sources, which are subject to continuous change in structure and content, a new approach, combining the best of center-out and edge-on, is required. This new model, the adaptive decision cycle, addresses the needs of both data management and decision agility by defining three different roles, which corresponds to different paths through the cycle, as seen in Figure 2:

Figure 2:  
Adaptive decision cycle,  
after Devlin,  
2013

1. *Exploring*: usually an individual, working alone, uses personal skills, knowledge and incentives to garner, condition, utilize and iterate information in an innovative and unconstrained manner in pursuit of the business goal. This path (E) corresponds exactly to the current edge-on model.
2. *Cultivating*: usually a group of people of similar abilities and status, often in peer or immediate reporting relationships, validate and confirm (or otherwise) the approach of the explorer. This role provides for rapid, peer-reviewed iteration (C) of data requirements and analyses.
3. *Grounding*: a corporate or other central role, combining data steward and IT skills, responsible for ensuring accuracy of data sourcing, performance, and other data management aspects before any novel analysis is promoted to a wider use within the business. This path (G) through the cycle aligns with the center-out model of traditional BI.



This model can certainly be delivered in today's world with exploration delivered via sandboxes and similar approaches in the virtualized BI environment, while collaborative

and social networking tools support peer review and scoring in the cultivation and grounding roles. However, a more integrated environment is needed to support the different roles, manage the processes involved and maintain the underlying CSI required. Perhaps the biggest challenge is, as ever, the organizational change such an approach demands. Nonetheless, it is only through the adaptive decision cycle that personal or group innovation in analytics or decision support can be reviewed and corroborated for business validity and optimized and promoted to the production environment.

---

## BUSINESS DIRECTORY—PROVIDING THE CONTEXT TO CONTENT FOR BUSINESS USERS

---

*Business is not just doing deals... Finally, business is a cobweb of human relationships.*  
Ross Perot

---

As mentioned above, contextualizing content for business users is a key requirement in data virtualization. The reason is simple. In contrast to a traditional data mart or warehouse, where IT has already cleansed and contextualized the content for its proposed users, data virtualization offers users the nearly unlimited opportunity to easily acquire and aggregate data from one or more sources. This data may be relatively unfamiliar or totally new to these users. Its business meaning may be poorly understood or even unknown. The conditions under which the data can be combined may be unclear.

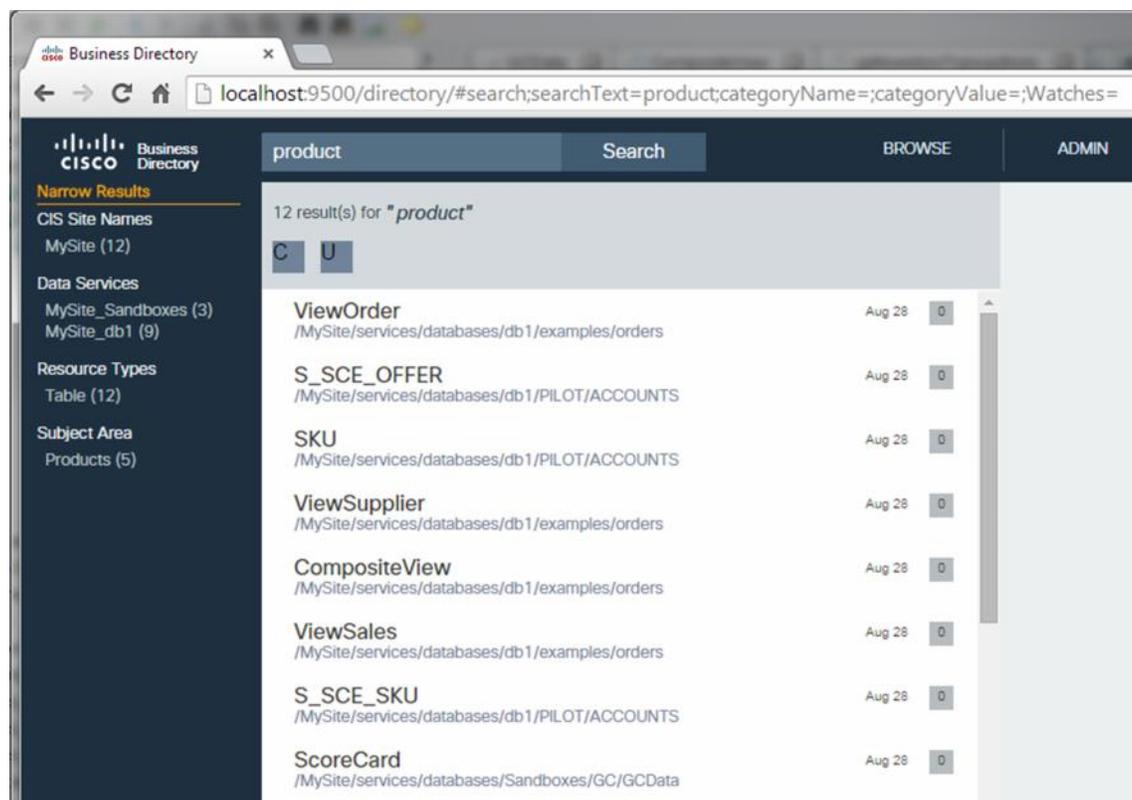
Today's business demands access to content from an expanding number and variety of data sources, from traditional databases to less structured forms, both internally and externally sourced. Therefore, there is an urgent requirement to provide context to business users, closely integrated with the available content itself. Cisco (formerly Composite Software) Data Virtualization, a leading vendor in this area with their Cisco Information Server product, has introduced a Business Directory to begin to address this need for context-setting information, firstly for business users, but also for the IT groups responsible for providing virtual data access.

Users commonly access the context-setting information via one of two mechanisms. First is *search*, a basic Google-like entry field that accepts free-form text and returns a list of matching items of business matter, such as views, tables, columns, etc., where the phrase searched for occurs in title, descriptions and so on. In contrast to this loosely exploratory mechanism, the second approach, *browsing*, involves a faceted search. Figure 3 shows a simplified view of the results of such browsing. Based on the type of drill-down commonly used in online retail websites, this approach provides a very structured view of the CSI in a manner that is also very familiar to business users.

The context-setting information displayed comes from three primary sources. The first is that normally available in the Information Server, created by the IT and business teams responsible for the original linking of sources, as well as that produced by the technical environment itself. This data includes view, table and column names, descriptions that exist in the source environments, lineage, dates and times of creation and update where available, and so on. An interesting feature here is the ability for business users to sign up for notifications of data source changes, linking users to the fast-paced tempo of modern business, particularly as it becomes increasingly linked to the mobile world. Further areas of personalized delivery currently include showing recent activity and creating favorites. Such *personalization* becomes increasingly important for ease of use as data volumes and variety increase.

The second source of this CSI is the data stewards and similar administrators who create and populate *categories* and custom properties. A category is any dimension for organizing data subject areas or

Figure 3:  
Cisco Business  
Directory  
appearance



topics that the business might find useful, from system-defined types (tables, views, files, etc.) to business-defined topics (department or business area where used, privacy level, etc.), as shown in the leftmost column of Figure 3. They provide business users an easy way to navigate data around particular subject areas. Custom properties are resource-level features defined by administrators, providing additional flexibility in the way that data is described and classified. One use could be to describe important governance aspects such as data reliance in particular situations, approvals for use, and so on of a particular set of data.

Third, business users can comment on any item described in the Business Directory. Such comments are made available in the CSI to all users and allow a simple, searchable, long-term trail of conversations among business users about any interesting aspect, change, etc. related to a particular item. At present, only a single-threaded, linear view is offered. This provides a simple collaborative environment to encourage business users to get involved in the ongoing evolution of business terminology and usage, as well as allowing IT to get advance information about emerging trends in usage or emerging data requirements.

Of particular interest here is to note that CSI has three levels of sourcing of increasing levels of informality: *provision*, governance and daily usage. At the first of these levels, CSI is very stable and standardized. This provides a firm foundation for meaning although, as happens in traditional data warehouses, there may be somewhat of an overemphasis on technical metadata. To ensure broadest use, it is important that all Information Server instances are registered to provide the widest base of available metadata. At the middle, governance, level, the formal aspects of business usage come to the fore. Including such CSI in the same environment as the basic metadata is particularly important, because it bridges a gaping chasm that exists in most organizations. *Security*, of course, plays a key role here; users can only see the CSI related to data to which they already have access. Finally, at the third and daily usage level, allowing business users to have their say democratizes the whole experience and offers a new possibility to make context a real part of the business conversation.

As a first version implementation, the Cisco Business Directory focuses firmly on putting in place a foundation for delivering context to business users. It draws primarily on the technical and descriptive metadata already collected within the Cisco Information Server. However, by addressing governance aspects and enabling user engagement, it avoids being locked into the old “technical metadata” cage. Areas for future focus certainly include additional access paths to the CSI, in particular directly from BI tools, further automation of governance-related aspects and expanded collaborative features.

*For users of data supplied through the Cisco Information Server, the Business Directory provides the context needed to maximize its use and value.*

So far, we’ve focused on how the context of data is exposed to business users and, to a lesser extent, where IT has a role in creating it and making it available. But, exposing context is only half of the journey. The real value emerges as this context is linked to the actual data content and embedded in the usage patterns of business users. The Business Directory provides information on how to access the virtualized data resources discovered by searching or browsing the CSI. And, of course, users can click on a table or view to directly upload the source data through the Information Server to a BI tool.

By exposing the CSI to users, the Business Directory encourages reuse of existing data and limits the ongoing reinvention of the analyst’s wheel. Business users simply take the data found and, subject to security constraints, upload them into any BI tool or spreadsheet they wish via the Information Server. However, the needs described in the adaptive decision cycle in Figure 2 above demand a more comprehensive and integrated solution.

---

## AN EMERGING VISION—UNITING CONTENT AND CONTEXT

---

*Information's pretty thin stuff unless mixed with experience.  
Clarence Day*

---

The vision is to support the full adaptive decision cycle, from initial exploration, through peer review cultivation, all the way to grounding, where analyses of wider interest can be promoted to production. The Business Directory function and CSI maintained there, both current and envisaged, are central to this process. Additional *sandboxing* function provides users with a structured, collaborative environment for data exploration and preparation, generating further CSI that, stored and used fully, supports the cultivation and grounding phases of the adaptive decision cycle.

Today, a common scenario involves analysts, often working in small teams, undertaking projects BI, analytics, visualization and data mining projects collecting and preparing corporate data sourced from Composite Information Server. Typically, 50%-80% of each project is spent in data preparation—identifying relevant data sets, discovering relationships among them, and formatting, integrating and transforming data to expose the most relevant information for consumption in the downstream analytic tool. Projects are often short-lived, and once the needed result is obtained, the assembled and integrated data may be of no further interest.

However, some projects may have wider, longer term application. And, even in those that don’t, the work may uncover interesting and useful context-setting information—previously unrecognized meanings, restrictions in use, relationships between datasets, etc. To support reuse of project data and CSI, a planned application on top of Composite Information Server will allow the creation of short-lived, locally-scoped workspaces—sandboxes—where analysts can collaborate on data preparation. A sandbox is an exploration area for all content, sourced both from the Information Server and

from private or local sources, wherein the context and process of use well-bounded and managed. The Business Directory is integrated throughout the process of preparing data, both enabling the analysis with current CSI and capturing data about sandbox work to further enrich the CSI.

The following description provides a medium to long term outlook of how sandboxing would support analysts to engage in exploration, cultivate data quality through collaboration in peer reviews, and facilitate grounding to production where appropriate. A possible work cycle is as follows:

1. A sandbox is provisioned for the project on a Cisco Information Server. A sandbox primarily collects project data sources and view definitions that transform that source data to expose information. All work done in the context of the sandbox will be related to this project. All analyst team members are given access to the sandbox.
2. Given the business question, an individual analyst explores available data sets in Business Directory, identifying the most relevant ones based on the available CSI. References to these data sources are added to the sandbox, which enables their use in building views. The solution may also need other data sources, such as departmental data marts, Excel workbooks, third-party data from services like Dun and Bradstreet, or CSV or XML files of relevance, etc. The analyst defines these as new data sources local to the sandbox in order to access these. They behave exactly like comparable data sources in the Information Server, but can only be referenced by views defined in the sandbox. Analysts can tag and annotate these sources within the scope of the sandbox, just as they would common data sources. As part of their exploration, analysts can profile all sandbox data source contents, to discover quality issues, potential relationships, novel uses and more.
3. Analysts define views on top of the data sources to integrate data from multiple sources and transform it to expose the information relevant to answering the question. These sandbox views are published via a Composite Information Server database dedicated to the sandbox. Access rights are granted to the analyst team who collaborate and, if necessary, iterate around these steps—moving from exploration to cultivation phase of the adaptive decision cycle.
4. With all data preparation done within the Information Server, technical metadata such as data lineage is shared and can be made available to the Business Directory. Analysts can explicitly use tagging and notes to provide richer contextual information to collaborators, which can be propagated to the Business Directory. Less obviously, analysts make implicit assertions of new CSI as they work, which can also enrich the Business Directory. For example:
  - Using a data source from the Business Directory in a sandbox asserts that it is a valuable data source. The Business Directory could spotlight most frequently used sources.
  - Using a “rogue” data mart as a source asserts that it exists and has value. The Business Directory could expose the existence of such useful but un-vetted sources to searchers.
  - Joining two data source tables strongly asserts a significant relationship between those two sources on those columns.
5. Many sandboxes will be retired at project end—the published Information Server database is removed. However, the view definitions, notes, tags, relationships, usage statistics, and so on can be archived and indexed. Some such CSI can be exposed in a separate, less-curated section of Business Directory. Others, like statistics about data source usage, may subtly inform Business Directory search and browsing results, steering users to the resources most valued by others.

*A sandbox environment offers users an area for exploration of all content, bounded by context and supporting the cycle of innovation to production and back.*

6. For projects that create views of long-term value, a more formal lifecycle exposes the views, along with any data sources used and CSI generated, to the IT or data governance function for formal grounding—further refinement and promotion to the curated part of the Business Directory. This phase may include social voting, building visualizations of an organization’s data assets, mapping emerging or diminishing areas of interest, identifying data and analyses to be optimized for promotion to production, and even charging or compensating organizational units based on their production or consumption of specific data assets.

This entire cycle occurs within the sandbox environment, providing full visibility into and control of how data assets are used, without impeding the agility of decision makers. And it enables full tracking and governance of such usage, for auditability, optimization and promotion to production when needed. With sandboxing built upon the Information Server and integrated with the Business Directory, individual and collaborative analyses provide context-setting information for the long-term benefit of the entire business that would otherwise be lost.

---

## CONCLUSIONS

---

*With so much information now online, it is exceptionally easy to simply dive in and drown.*  
Alfred Glossbrenner

---

In a modern, increasingly virtualized data environment, we need to move from overwhelming data availability to universal, optimized ease of use. This is enabled directly through context—the matrix in which every data element is embedded. And as data becomes ever more extensive and complex, and business users must interact with and gain value from ever wider and more novel swathes of information, context becomes the keystone to bridge from volume to value.

We are beginning a journey from disconnected decision making to agile, integrated insight. As business users understand the context of ever more data accessed *in situ*, they can begin to shift from a mindset of “quick, load up another spreadsheet” to one of “let’s move seamlessly from innovation to production”. This demands an integrated environment linking context and content and a framework to allow users to move from exploration, through validation, to production.

*The journey from disconnected decision making to agile, integrated insight begins with the context provided by the Business Directory.*

This vision of context and content drives new tools and techniques to manage and share context-setting information and to embed it into the adaptive decision cycle of content usage. The initial phase of this journey is to make data context visible and viable. This involves shifting thinking from metadata in its old technical sense to context-setting information that bridges the IT and business worlds. Business Directory takes this first step. Second, the concept of sandboxing provides a lightly managed and monitored environment in which users can innovate with content as needed. The context-setting information collected here includes technical metadata, actions performed on the data, and business-related information, including user comments and conversations. This CSI provides a foundation linking experimentation with the eventual move to production.

In essence, with data virtualization, we are moving from the traditional, center-out approach, which emphasizes business needs for quality and consistency, to a more free-flowing edge-on view, where timeliness and innovation are favored. However, both aspects are necessary: data quality does need to be controlled, just as analytic freedom needs to be preserved. The combination of Business Directory and the emerging sandboxing environment opens up a world of analytic innovation while providing a clear path to data quality and optimization of the production environment.

---

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 new book, "Business unIntelligence—Insight and Innovation Beyond Analytics and Big Data" (<http://bit.ly/BunI-Technics>) was published in October 2013.



Barry is founder and principal of 9sight Consulting. He specializes in the human, organizational and IT implications of deep business insight solutions that combine operational, informational and collaborative environments. A regular tweeter, @BarryDevlin, and contributor to [ITKnowledgeExchange](#) and [TDWI](#), Barry is based in Cape Town, South Africa and operates worldwide.

Brand and product names mentioned in this paper are trademarks or registered trademarks of Cisco and other companies.

---

<sup>1</sup> Anderson, Monica, "Semantics, Models and Model Free Methods", 2012, <http://videos.syntience.com/ai-meetups/smamfm.html>

<sup>2</sup> Ryley, Sarah and Donohue, Pete, New York Daily News, June 22, 2014, <http://www.nydailynews.com/new-york/nyc-crime/daily-news-analysis-reveals-crime-rankings-city-subway-system-article-1.1836918>

<sup>3</sup> Kosara, Robert, "Putting Data Into Context", July 23, 2014, <http://eagereyes.org/basics/putting-data-into-context>

<sup>4</sup> Devlin, Barry A. and Murphy, Paul T., "An architecture for a business and information System", IBM Systems Journal, Vol 27(1), (1988), <http://bit.ly/EBIS88>

<sup>5</sup> Marco, David, "Building and Managing the Meta Data Repository: A Full Lifecycle Guide", Wiley Computer Publishing, New York, 2000

<sup>6</sup> Devlin, Barry, "Business unIntelligence—Insight and Intuition Beyond Analytics and Big Data", Technics Publications, New Jersey, (2013), <http://bit.ly/BunI-Technics>

<sup>7</sup> Ibid. The logical, IDEAL architecture of Business unIntelligence proposes that everything of interest to a business, or business matter, can be described in terms of people, process and information.