

Playmarts: Agility with Control

Reconnecting Business Analysts to the Data Warehouse

December 2008

A White Paper by

Dr. Barry Devlin, 9sight Consulting
barry@9sight.com

One key group of potential users of Business Intelligence tools and Data Warehouse data has been seriously underserved for many years. These power business analysts resort to a mishmash of desktop productivity tools to support them in their search for innovative answers in the ever-growing jungles of enterprise data. IT departments trying to keep control of the corporate data asset through the data warehouse architecture denounce the “spreadmarts” thus created. Indeed, business analysts themselves also often wish that there was a better way.

This paper introduces a new category of data mart—the “playmart”—to address the needs of both business analysts and IT. In essence, a playmart provides agility for users through freeform, iterative exploration of data from multiple sources (playing with the data) in a safe, controlled and traceable environment (a mart). The goal, simply, is to provide users the freedom to innovate, while maintaining corporate auditability; to find unique business value in enterprise data and to ensure it is valid and repeatable when needed.

The playmart is more than a concept, however. A new tool—Lyza™—provides a significant first set of the functionality that the playmart requires.

Contents

**The World of
Business Analytics**

**The Joys and Sorrows of
Exploratory Analytics**

Introducing the Playmart

**Lyza: A Playmart
Implementation**

Conclusion

In collaboration with:

Lyzasoft, Inc.
www.lyzasoft.com



Joe Figger is a power business analyst¹ at Bravo Insurance Group Inc (BIGI). He's the one who comes up with the most unusual ways of looking at the company's vast store of financial, market and personal data, leading to some of the most innovative products BIGI has created.

Joe has negotiated access to lots of data over the years. Some of it is in the data warehouse; and some exists in a variety of systems and files Joe just found. He also stores a considerable amount of data on his PC. Joe runs some really large spreadsheets. He's an Excel wizard, combining data from multiple sources, with some of the most complex formulae you can imagine. He produces many substantial graphical and tabular PowerPoint presentations for senior management. Although he does resent how much repetitive and administrative work is involved.

Now, to handle ever larger spreadsheets, he has persuaded his manager to buy him the biggest, fastest PC available as he designs a new multi-spreadsheet analysis combining current sales and demographic data with historical information that will blow away the competition. And that's when things start to go badly wrong for Joe. Because IT has to sign off on that PC, Jill Joy, the manager of the Business Intelligence team, gets to hear about it. And given her battle to get the company data under control and her ongoing war against the spread of the spreadmarts, Jill vetoes the purchase.

Will Joe figure out how to get his new analysis done and save the company? Will Jill put a stop to his gallop and bring the company data under control? Or is there a way that both of them can win and save BIGI? Find out later in this paper...

The World of Business Analytics

Business analytics falls into four categories with differing goals, levels of end-user freedom and tooling. Although there is some overlap, these categories well describe the different roles of business analysts and how to support them. This paper focuses on one category—Exploratory Analytics. But first, we need to understand the other three and how IT typically supports them.

Query and Reporting: The oldest and simplest approach, reporting starts with skilled IT staff designing static reports, which are run on a regular basis and distributed to business managers. Reporting provides little or no flexibility to users. However, ad hoc query can be made available to users allowing more flexibility and deeper analysis. Based on a formal language such as SQL, querying requires awareness of data constraints and relationships; hence its restriction to power users.

Management Applications: Performance management, business planning and similar applications, often delivered as dashboards, are essentially automated versions of manual tasks previously based on reports and queries. While dashboards provide a limited level of analytical function, their main focus is to show specific data through an intuitive and attractive user interface to managers and executives, rather than to support power users such as business analysts.

On-Line Analytical Processing (OLAP): OLAP, in its various forms, is conceptually a very specific form of the relational model that supports slice, dice and drill-down analysis of data. Because these analyses are so common and useful, OLAP has become a key technology in business analytics.

These three types of analytics are closely tied to Business Intelligence (BI) and the Data Warehouse architecture first described in the mid- to late 1980s² as an approach to support all types of tactical and strategic decision making. The architecture recognized that decision makers often didn't know in advance exactly which data would be needed for any particular decision or how it would be used. Therefore, the goal was to make all business data available in a non-prescriptive manner.

Meeting that goal is made difficult by limitations in the existing IT environment. First is the diverse and distributed nature of the data sources. Second, syntactical and content errors and inconsistencies abound in the data. In addition, there are performance and security concerns with access to existing data sources. The resulting architecture is shown in figure 1. Its key characteristics are:

1. Data is extracted from the operational systems, cleansed, reconciled and stored in a normalized and historical database—the enterprise data warehouse—using relational technology

2. Subsets of this data, optimized for particular uses, are extracted from the warehouse into data marts for use by decision makers on a variety of technologies, often broadly relational in nature
3. Data flows unidirectionally through the environment in order to assure maximum control of data consistency and quality
4. Metadata, at all levels in the architecture, defines data provenance, meaning and manipulation

Despite some evolution in the intervening years, these characteristics remain largely intact as the foundation of most successful BI implementations today.

The first three categories of business analytics fit well in this architecture. Query and reporting tools, and the related management applications run directly off the data mart level or the data warehouse levels of the architecture. OLAP, with its relational model, also fits well into this architecture and has become the most common form of data mart. Most importantly, all three categories maintain data consistency and quality controls that have been applied as part of the warehouse design and population processes.

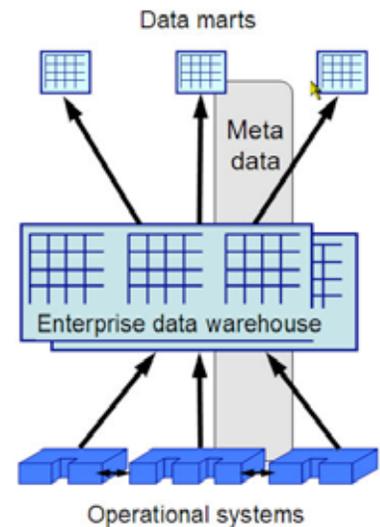


Figure 1:

Data Warehouse Architecture

The Joys and the Sorrows of Exploratory Analytics

Exploratory Analytics, the fourth type of business analytics, is the most poorly defined category but also gives the greatest potential for innovation and return on investment. Simply put, analysts ingeniously combine all available tools and data to allow them to explore any business issue in depth!

The most common form of Exploratory Analytics today is hand-crafted by business analysts. They gather data from multiple sources and integrate it manually, based on their deep knowledge of the actual business applications and data. They apply a variety of sorting, joining and querying techniques, often in an ad hoc manner, producing tabular or graphical reports. The ubiquitous tools used are desktop productivity tools such as spreadsheets, desktop databases and presentation tools.

(Another well-known exploratory approach, data mining, reverses the usual process by eliminating the need to formulate specific queries against the data, but is beyond the scope of this paper.)

The output of sophisticated exploratory analytics is of high commercial value. It provides rare insights into how the business actually works and innovative solutions to real management and operational issues. However, the question arises: why do analysts choose to perform such high value, complex work in personal productivity tools, particularly spreadsheets?³

- ❑ **Familiarity, low cost and ease of use:** spreadsheets are ubiquitous and used daily, and are often believed to provide the “only solution” to many immediate analysis requests
- ❑ **Flexibility:** with a wide variety of analytic functions, users can play with data as they need to
- ❑ **Ad hoc development and local control:** analyses are developed in a piecemeal fashion, growing as needs demand
- ❑ **Speed:** with such familiarity and low IT responsiveness, users can answer new questions rapidly

These “joys” of exploratory analytics can be summed up in a single word—agility. However, there exists considerable evidence⁴ of substantial “sorrows” that arise from the widespread use of spreadsheets and desktop tools in exploratory analytics:

- ❑ **Logic and data errors:** independent research repeatedly shows that many organizations continue to find major data and logic errors in existing and widely-used spreadsheets
- ❑ **Maintenance:** once built, the logic of the interlinked formulae can no longer be easily understood and unanticipated changes in source systems can cause data and calculation errors

- **Productivity:** refreshing spreadsheets with regularly changing data is tedious and transferring data between tools is manual and repetitive
- **Auditability:** traceability of data manipulation is difficult in a pure spreadsheet environment

Due to this lack of control, most data management experts and IT departments oppose the use of spreadsheets for business analytics. The spreadsheet-based approach does not easily fit in the data warehouse. Data is extracted from a variety of sources, not only the warehouse. Data quality is thus questionable, as is the integrity of joins between warehouse-certified and other data. Because spreadsheets are user-controlled, data can be distributed and manipulated in a non-auditable manner. In short, the data consistency and quality created by the warehouse can be, and often is, undone.

Wayne Eckerson, Director of Research, TDWI, coined the term “spreadmarts”⁵ to describe the proliferation of analysis data beyond the control of the data warehouse. He says that spreadsheets “proliferate like poisonous vines, slowly strangling organizations by depriving them of a single consistent set of information.” On the other hand, the innovation and business value derived from this agility of analysts using this environment cannot be simply thrown aside.

Is there a better way? Can we achieve a sufficient level of control of data quality and auditability and still provide enough agility to enable business analysts to be innovative and productive? If we revisit the choice of tooling, the answer is most certainly “yes”!

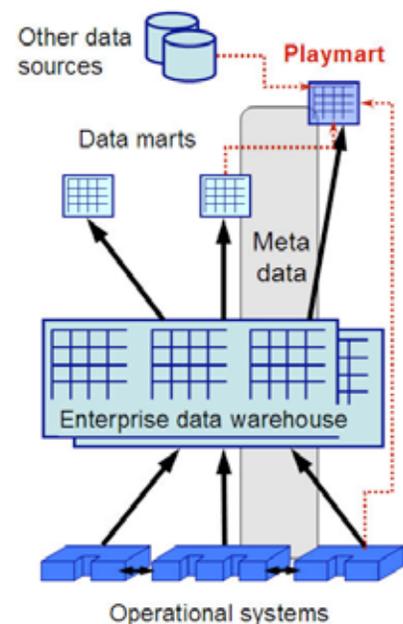
Introducing the Playmart

A “playmart” is a new form of data mart. The rationale is to provide a safe, bounded area where business analysts can play with data without undue constraint while maintaining data quality and governance. The playmart promotes agility with control through eight key characteristics, supporting agility, control or a combination of both:

1. **Comprehensive, integrated and user-friendly tooling (agility):** a full and well-integrated set of data gathering, analysis and presentation tools, suitable for business analysts, is available
2. **On-demand sourcing (agility):** data is loaded into the playmart as required by the business analyst, based on immediate needs. Data sources may be the data warehouse or marts, but also any other server-based or PC-based data sets
3. **Closed-loop cycle (agility):** the full analytics cycle from exploration to decision-making and back is supported, allowing the analyst to move seamlessly between these working modes over multiple rounds of business planning
4. **Collaborative working (agility):** all levels of collaboration from simple presentation sharing to full collaborative editing of data are supported
5. **Strong boundaries (control):** to maintain auditability, export of editable data from the playmart is strictly limited and controlled. Analysis outputs such as reports and presentations are exported only in a non-editable format and with supporting descriptive and security information
6. **Pervasive tracking (control):** all activities performed in the play mart are recorded and tracked
7. **Automation and production-enabling (both):** ad hoc analyses can be rerun with updated data on an ongoing basis in the playmart. Furthermore, the IT department can promote particular analyses to an external production environment and embed in the business flow on a formal basis and/or define preferred, pre-loaded sources of data for use in certain cases
8. **All-inclusive metadata (both):** a comprehensive set of metadata on all aspects of data quality and sourcing, usage and manipulation is the basis of

Figure 2:

The Playmart in the Data Warehouse Architecture



almost all of the above points. Once stored, it is available for use by business analysts on an ongoing basis

By providing this significant level of tracking and auditing of what data is used and how it has been combined, manipulated and analysed, the playmart supports the enterprise need to manage data quality and the drive towards a “single version of the truth”. From a regulatory point of view, it can be an important tool in tracking compliance under, for example, Sarbanes Oxley legislation.

For the power business analyst, the playmart provides all of the agility needed without the potential for data anarchy. The playmart allows iterative development of analysis approaches and enables the business analyst to bring data sets into the environment as and when needed. Because the playmart tracks what data is included in the analysis, data management concerns can be addressed and if there is sufficient demand, the data in question can be made available through the structure of the warehouse in the future.

For the business as a whole, the playmart unlocks the deep mastery of the real business process and data possessed only by business analysts. By enabling them to play safely with the data in a monitored environment, they can turn their mental image into an instantiated model that is amenable to repeated use and eventually promotion to production by IT.

Figure 2 positions the playmart in the data warehouse architecture. It is placed higher than standard data marts, reflecting its relative closeness to its users. The controlled, and thus preferred by IT, sourcing is shown by the black arrow from the enterprise data warehouse. The red, dashed arrows show additional sources of data that a business analyst can call upon. Like other data warehouse components, metadata about the playmart is recorded and stored where appropriate.

While one could envisage that many business analyst needs in the playmart could be met by the current personal productivity toolset, it is clear that the data management and auditability requirements could not be met so easily. A new class of tool is required that combines the best of both worlds.

Lyza: A Playmart Implementation

Lyza was introduced in September 2008 as a desktop data analysis solution⁶, running on Windows, Mac and now Linux. Its primary audience is business analysts, for whom it aims to provide a complete, comprehensive toolset within a bounded environment. The fact that the environment is bounded and the way that it is structured positions Lyza as a playmart. In this section, we describe the functionality provided and compare it to the eight key playmart characteristics above.

As can be checked out via the free 30-day trial⁷, Lyza provides a substantial level of synthesis, analysis and reporting in a single, consistent interface. According to business analysts, these tools meet many of their needs and can even give them new ideas about how they might be able to take innovative analytical approaches. These tools, in the order in which would typically be used, are:

- **Synthesize:** the data integration tools provide access to a wide variety of heterogeneous sources, from which data can be imported into the environment. Data from different sources can then be combined by visually connecting related columns.
- **Analyze:** the analysis tools use a drag and drop model to filter, sort and perform computations that leverage spreadsheet-like functions.
- **Publicize:** the charting and reporting tools provide output to presentations, dashboards, Adobe PDF files and more.

The data synthesis and analysis function provided by Lyza is closely based in name, form and approach on that provided by spreadsheets. The columnar model dictates that some typical commands, such as “look-up” functions, are not needed. The similarity to the spreadsheet world and breadth of coverage are key factors in encouraging business analysts to move to this new environment. Lyza also provides a largely seamless link from analysis to publicizing. This eliminates tedious and error-prone manual transfers of data between tools, a significant productivity gain for many analysts.

In contrast to the cell-based approach in spreadsheets, Lyza is column-based. All actions operate on entire columns automatically, removing the need to copy formulae down the columns as seen in spreadsheets. Furthermore, internally, Lyza is also columnar, which gives fast, scalable performance that can leverage parallel processing function in multi-core processors. There are no limits on the size of datasets, numbers of rows or columns that can be handled.

While there are gaps in the range of analysis and presentation function provided, Lyza clearly goes some considerable way to meeting the first playmart agility characteristic of providing *comprehensive, integrated and user-friendly tooling*.

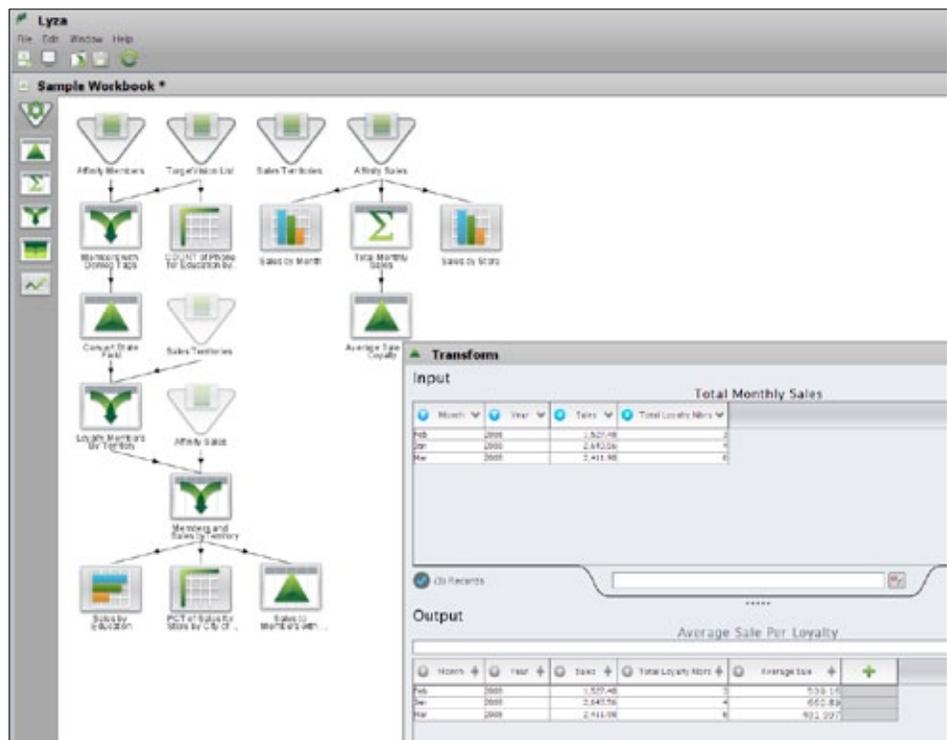


Figure 3:
Building an Analysis in Lyza

Looking to the second playmart characteristic, *on-demand sourcing*, Lyza recognizes that business analysts need the freedom to work with data from different sources at short notice and provides access in the synthesis function to flat/text files, Excel, Access, SQLServer, MySQL, Oracle, DB2, Sybase, PostgreSQL, and any other ODBC compliant databases. Lyza provides the ability to use these various sources as required by the analyst, without going through the bottleneck that IT is often seen as. This, together with the intuitive interface, can provide a significant level of agility for users.

Providing these tool sets in a single environment gives good support the third agility characteristic—*the closed-loop cycle*—for a simple but important reason: in contrast to prior approaches, Lyza recognizes that business analysts typically solve problems in an iterative fashion across these three areas.

For example, an analyst may first combine data from two sources and create some new data based on that. He sees that another source contains information that can be combined with the previous results to give more useful data. He graphs the output to check his reasoning and then goes back to the integration stage to bring in yet another data source and does some further analysis. Finally he creates a presentation for management based on the combined data. At a later stage, he may go back to the analysis in order to update the input data, perhaps slightly change the analysis based on management feedback and then create an updated presentation.

What this amounts to is an adaptive workflow, crossing data acquisition, combination, analysis and presentation, being built in real-time by the analyst. Lyza's interface is essentially a workflow builder, where data and tools are dragged and dropped onto a workbook and linked together. As each step is added, an appropriate window opens to allow the step to be detailed, as shown in figure 3.

Closing the loop from the presentation and decision-making activity back to the analysis phase will prove challenging in Lyza's current desktop-only incarnation. Similarly, support for *collaborative working*, the fourth agility characteristic and a key emerging need in decision making will require a server-based engine in Lyza for comprehensive support. However, Lyza's workflow-based model does support a somewhat better understanding of workbooks shared between co-workers than is the cases with spreadsheets.

From the above discussion, Lyza clearly stacks up fairly well in terms of the agility needs of the playmart. So, how does it rate for IT and data management? To what extent are the control and auditability characteristics supported?

First, Lyza provides an environment within which the business analyst's entire process can be fully met, potentially enabling the **closed boundaries** control characteristic. From a data management viewpoint, a major difficulty of the current personal productivity toolset is that once spreadsheets or presentations are distributed, the recipients can and do change the analyses and outputs to suit their own needs. While there may be no malfeasance involved, the loss of control and the subsequent debates over "which version is correct?" are costly and counter-productive.

Lyza has the features required—creating snapshots and timestamping analyses and presentations and restricting export to read-only formats. As long as the users stay within the Lyza environment, version control is considerably simpler than in the personal productivity toolset. However, Lyza also allows the export of datasets to external tools. While clearly needed in the product for migration and compatibility, this feature would need to be restricted at implementation time in a playmart. Such exits can never be closed entirely, but it will be vital going forward that Lyza provides a combination of top-class analysis and presentation features within the environment, while ensuring that exiting is both less easy and well-tracked.

Pervasive tracking, the second control characteristic of playmarts, follows in large measure from the workflow-based approach adopted by Lyza. The product maintains a record of all the steps performed in an analysis, from data sourcing right through to the creation of presentations in a secure, encrypted format. Processes and the results they produce can also be frozen and timestamped to provide a permanent record of the analysis of a particular period.

The workflow-based approach also confers another important benefit. Because Lyza knows the data sources and the processing performed on them, it can provide a means for the analyst to simply run the entire process again, but using updated data. This contributes to the **automation and production-enabling** characteristic of the playmart. However, in its present release, Lyza has focused on the individual needs of business analysts rather than on the process by which useful analyses can be brought to a more production-oriented level. In principle, the workflows contain sufficient information to enable translation to any preferred production toolset, but this is unproven. On the sourcing side, there is no way to pre-populate workbooks with preferred data via ETL (extract, transform and load) tooling.

Finally, workflows built by the analysts are, of necessity, described by an extensive set of metadata, stored by Lyza as XML. This store forms the basis for satisfying the final playmart characteristic—**all-inclusive meta-data**. This existing metadata is clearly used for tracking and auditing purposes and is made available to the business analysts through the user interface. Significant expansion will likely be needed in order to support production-enabling and additional automation and auditing functionality. Furthermore, as the current version of Lyza is entirely client-based, with the metadata residing there too, there is clearly a future requirement to move or duplicate this metadata on a server for security and manageability.

Conclusion

Let's return briefly to the story of Joe Figger and Jill Joy. Did they find a win-win solution?

With the architectural framework of the playmart, Jill can define a strategy that allows business analysts more freedom to innovate, while regaining control on data proliferation. She recognizes that spreadsheets per se are not the problem; it's just that business analysts haven't received the solution from IT to meet their needs. Although she was looking for ways to ban spreadsheets altogether, she now understands that analysts like Joe have their reasons for using them. And, anyway, prohibition has long proven to be a failure.

Joe, for his part, can begin experimenting with Lyza using a more integrated and intuitive toolset for his next big breakthrough. The workflow approach gives him a clearer picture of what he's doing and allows easier maintenance of his calculations and reuse of analysis steps. The columnar approach does require some

"I often have to do quick analysis of a client's storage environment in order to position product features in a timely and accurate manner. I recently received a write out of a log that monitors the size of every file system in the client's environment. It was only 40MB, but contained nearly 400,000 rows. I was able to quickly and easily drag the log into the Lyza desktop and was immediately able to view, scroll and browse the file, as well as perform the necessary data analysis required for the proposal—all in a matter of minutes. And if the client wants me to analyze another month's worth of data, I can simply drop it into my workflow that I built in Lyza. I am confident Lyza will continue to be an invaluable tool in the highly competitive storage market where I work."

Chris Thomas, Senior Account Technical Consultant, EMC.

readjustment after the cell-based world of spreadsheets, but it has proven its worth in simplifying development and speeding processing of large volumes of data.

More broadly, Lyza clearly has a level of agility in its function and user interface that attracts support and enthusiasm from business analysts, but the key to its future success as a playmart will depend on addressing the remaining control and auditability demands of IT for enterprise level acceptance. Key to this, Lyza must broaden its perspective from a pure desktop tool to provide a comprehensive server-side set of function for metadata storage and management, auditability and tracking, as well as support for IT to easily migrate analyst-developed workflows to the production environment as required.

In summary, Lyza, although still a version 1.0 product, meets quite a considerable portion of the basic requirements of a playmart. It provides sufficient analytic function to persuade business analysts to consider moving away from personal productivity tools, which is a necessity if IT is to rein in the proliferation of spreadmarts. From an IT viewpoint, it is well-architected with basic tracking and metadata function already present. Together, these considerations suggest that Lyza has the potential to address the full set of playmart characteristics in future development iterations.

Dr. Barry Devlin is among the foremost authorities on business insight and data warehousing. He is a widely respected consultant, lecturer and author of the seminal book, “Data Warehouse—from Architecture to Implementation”. Barry’s current interest extends to a fully integrated business, covering informational, operational and collaborative environments to offer an holistic experience of the business through IT. He is founder and principal of 9sight Consulting, specializing in the human, organizational and IT implications and design of deep business insight solutions.

About Lyzasoft Inc.

Founded in 2008, Lyzasoft, Inc. is a privately held company founded from business intelligence (BI) and data processing roots. Lyza provides a desktop software solution that enables end users to perform simple to complex data transformations and analyses without the reliance on expensive IT implementations. The solution enables you to present and communicate your findings as well, all within the Lyza operating environment. The toolkit was designed by analysts, for analysts.

Lyzasoft, Inc.
1675 Broadway, Suite 1300
Denver, CO 80202

www.lyzasoft.com

Brand and product names mentioned in this paper may be the trademarks or registered trademarks of their respective owners.

¹ The title “business analyst” has many diverse meanings. In this paper, we use it to refer to business users who perform relatively complex collection, analysis and interpretation of data from a wide variety of sources.

² “Data Warehouse—from Architecture to Implementation”, Barry Devlin, Addison-Wesley, (1997).
<http://www.amazon.com/Data-Warehouse-Implementation-Barry-Devlin/dp/0201964252>

³ “Strategies for Managing Spreadmarts: Migrating to a Managed BI Environment”, TDWI Best Practice Report, First Quarter 2008, www.tdwi.org/research/display.aspx?ID=8874.

⁴ See, for example, the European Spreadsheet Risks Interest Group, www.eusprig.org/index.htm

⁵ See “Taming Spreadsheet Jockeys”, TDWI Case Studies and Solutions, TDWI e-newsletter, July 2002.
<http://www.tdwi.org/display.asp?id=7167>

⁶ See www.lyzasoft.com.

⁷ See www.lyzasoft.com/try.php.