# IBM® InfoSphere™ Discovery

## Speed Time to Value for Information-Centric Projects

### Highlights:

- **Accelerate time to value for critical initiatives like data integration, governance and archiving**
- **Define the business objects (logical groupings of data) that serve as essential inputs into information-centric business projects**
- **Capture standard database relationships and discover the complex associations that are typically hidden from view**
- **Identify the complex transformation rules applied to populate a warehouse or operational data store**
- **Prototype and test source to target data mappings for application consolidation and master data management**

## Information-centric projects transform the business

In a value-conscious environment, IT managers are leading information-centric projects to transform the business and deliver cost savings. Data integration, warehousing and master data management (MDM) projects enable effective cross-promotions by providing greater insight about clients and prospects. Meanwhile, projects such as application consolidation and data archiving reduce operating costs and improve the performance of mission-critical CRM and ERP systems. Across all industries and geographies, data governance requirements demand continued focus.

For an on-time successful deployment of these critical initiatives, organizations must first possess a complete understanding of their existing data assets. For example, business managers must ask:

- Which high-value clients are eligible for the new sales promotion?
- What version of the data should we use for the ERP instance consolidation?
- Which of these service order records can we move into the archive?

- How were these customer purchase history figures calculated?
- Which applications contain confidential patient names and account numbers?

IT managers must survey a complex, distributed landscape in order to provide the answers. Why is the enterprise data environment so complicated? The reason at least, is simple. Most organizations execute their business processes across siloed, interrelated applications, databases and platforms. Without explicit information to describe how data elements are related across these systems, managers have limited insight into critical information assets.

## Data Discovery Provides Insight

Data discovery is the process of analyzing data values and data patterns to identify relationships. These disparate data elements are linked into logical units of information or 'business objects' such as customer, patient or invoice. A business object represents a group of related attributes (columns and tables) of data from one or more applications, databases or data sources. Discovery is also used to identify the transformation rules that have been applied to a source system

**Information Management**

**InfoSphere™ software**

when populating a target such as a data warehouse or an operational data store. Once accurately defined, these business objects and transformation rules provide the essential input into information-centric projects like data integration, MDM and archiving.

IBM InfoSphere Discovery™ provides market-leading capabilities to automate the identification and definition of data relationships across the complex, heterogeneous environments prevalent in IT organizations today. Covering every kind of data relationship, from simple to complex, IBM InfoSphere Discovery provides a 360° view of data assets.

Standard relationships such as primary or foreign key constraints are defined directly in the database catalog. However, there are more complicated data relationships which are not easily visible. For example, application defined relationships, are not contained in the database catalog but instead are enforced through the processing logic of the application itself. In other situations, complex business rules may have been applied to transform source data as it is moved into a target system.

IBM InfoSphere Discovery analyzes the data values and patterns across heterogeneous sources to capture hidden correlations and bring them clearly into view.

By applying heuristics and sophisticated algorithms, Discovery performs a full range of data analysis and transformation logic detection. It accommodates the widest range of enterprise data sources including relational databases and any structured data source that can be represented in a text file format such as a hierarchical database.

Without an automated process to identify data relationships and define business objects, organizations can spend months performing manual analysis, – with no guarantee of completeness or accuracy. IBM InfoSphere Discovery's automated capabilities accurately identifies relationships and defines business objects, speeding deployment of

information-centric projects by as much as ten times. Faster implementation means that IBM clients can exploit new revenue and cost containment opportunities, not available to the competition.

**IBM InfoSphere Discovery Facilitates Data Archiving Projects**
Industry analysts recognize the power of database archiving to cut storage costs and ensure optimal performance of critical business systems. As Forrester Research, Inc. notes, "Why store inactive data when such data is not required to be accessed by users immediately? Moving inactive data to another instance or archive system not only makes production databases more efficient, but it also lowers cost and improves manageability.[1]"
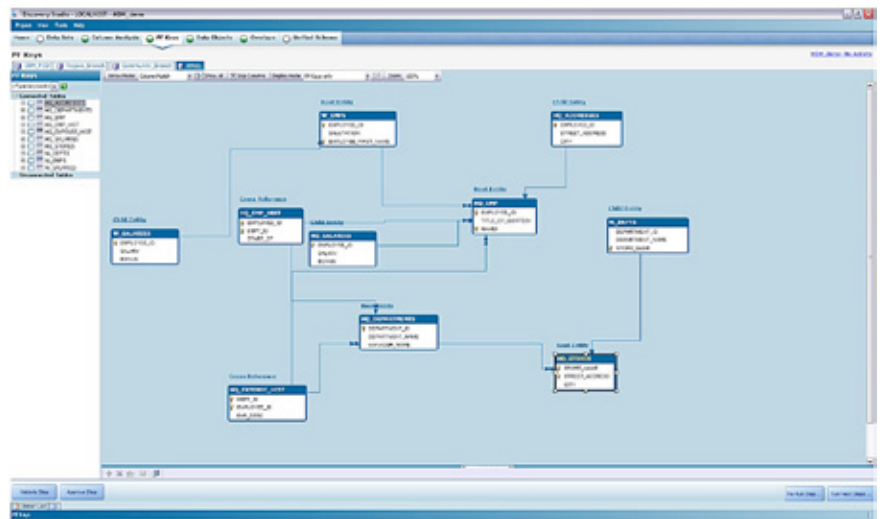


*Figure 1: IBM InfoSphere Discovery identifies the relationships that link disparate data elements*

A critical first step in any archiving project is the accurate representation of the business object to be archived. IBM InfoSphere Discovery complements the IBM Optim™ Data Growth Solution by capturing both database-defined and application-managed relationships that make up the business object to be archived. By using IBM InfoSphere Discovery together with the IBM Optim Data Growth Solution, organizations can quickly and easily implement best-practice archiving capabilities, reducing storage requirements and delivering substantial cost savings.

## Data Consolidation Projects are Implemented Accurately

Data consolidation, a major component of application migration, warehousing and MDM initiatives, is considered an especially challenging activity. Managers must interrogate multiple sources of data to determine how these systems were aggregated and transformed into an application target. Many source systems represent legacy technology, for which there is little documentation or domain expertise.

IBM InfoSphere Discovery automates the identification of complex, cross-system transformations and undocumented business rules. It provides advanced capabilities for discovering scalar transformations, conditional logic, aggregations, arithmetic equations, and other sophisticated business rules that would be nearly impossible to identify by manual inspection alone.

During a data consolidation project, the IT staff must also define new transformation rules accurately and precisely. These new rules will form the basis for mass movement of data between systems. Transformation errors can cause significant business impact and once the logic has been applied to the data, the process is extremely expensive to reverse.

Using IBM InfoSphere Discovery, IT can prototype and test new transformation rules for completeness, before any data is physically converted and moved. These IT validated rules can then be transferred to IBM InfoSphere FastTrack where Business Data Analysts can augment with additional documentation and business logic before generating IBM InfoSphere DataStage extraction, transformation and load (ETL) jobs. As a result of using IBM InfoSphere Discovery, organizations are able to minimize risk and deploy projects faster to support critical decisions.
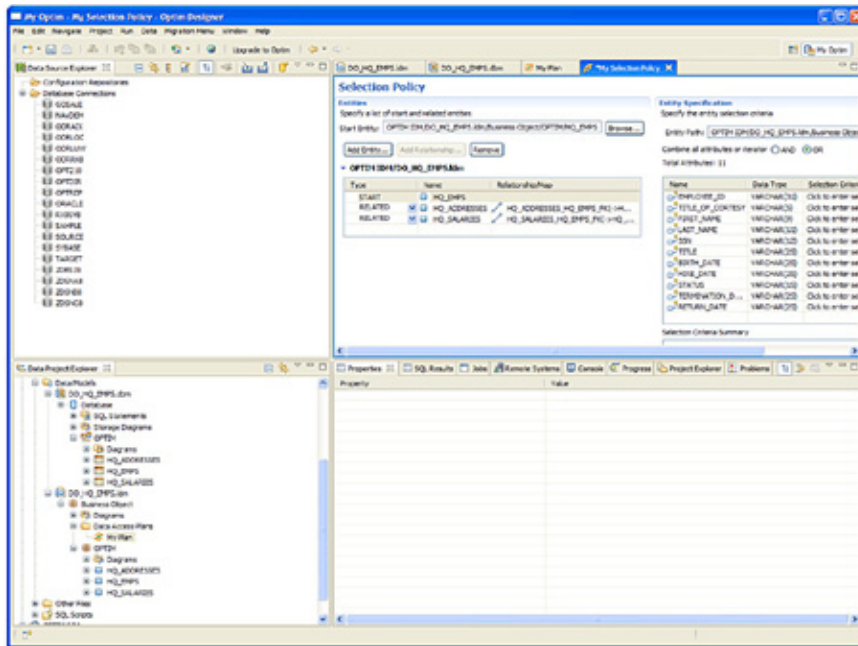


*Figure 2: IBM InfoSphere Discovery identifies the business objects to be archived using the IBM Optim™ Data Growth Solution*

**Success: IBM InfoSphere Discovery Delivers Business Value**

*Information-Centric Business Initiative: Application Consolidation*

A major international automotive manufacturer planned to consolidate numerous, legacy ERP systems from acquired companies. To promote consistency and improve visibility, the organization would also build master data hubs for Customer, Vendor and Materials entities.

*Challenge: 16 Legacy Systems; 31 Separate Data Repositories*

IT managers had inherited multiple, inconsistent data repositories across its portfolio of legacy ERP systems. There were 14 repositories for the Customer entity alone, with another 9 for Vendor and 6 for Materials. Data Analysts needed up to 7 days to manually analyze a single attribute within each repository. In addition the results could not be verified until data was moved to the new system. The organization deployed IBM InfoSphere Discovery to minimize the risk associated with manual relationship discovery and data mapping.

*Success: 90% Reduction in Work Effort; Trusted Information*

Automated cross-system analysis and data mapping across the 16 ERP systems resulted in a 90% reduction in work effort, greater visibility into potential data problems and lower project risk. Analysts can test relationships and business rules without moving a single row of data. Most importantly, business users are demonstrating increased trust in the consolidated system. They can now get answers to their questions quickly and reliably.

*Hidden Instances of Sensitive Data are Uncovered*

Some sensitive data is easy to find, for instance, telephone numbers in a column named 'tel_number'. However most application databases are more complex and sensitive data is sometimes compounded with other data elements or buried in text or comment fields. Data in one system may be moved and manipulated to be used in other places. Subject matter experts (SME) can sometimes offer insight, but only if they fully understand the system.
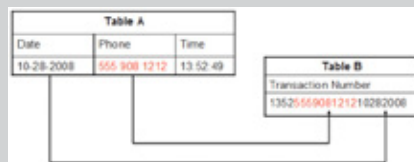


*Figure 3: Confidential information hidden in compound fields poses a privacy risk to the organization*

IBM InfoSphere Discovery enables organizations to identify all instances of confidential data across their environment – whether clearly visible or obscured from view. IBM InfoSphere Discovery works by examining data values across multiple sources to determine the complex rules and transformations that may hide sensitive content. It can locate confidential data items that are contained within larger fields, as shown in Figure 3, or that are separated across multiple columns.

IBM InfoSphere Discovery delivers automated capabilities that offer greater accuracy and reliability than manual analysis. When used in conjunction with the IBM Optim Data Privacy Solution, IBM InfoSphere Discovery provides the most effective enterprise-scale approach for locating and masking sensitive data across complex, heterogeneous environments.

*IBM InfoSphere Discovery Speeds Time to Value*

Information-centric projects form the basis of business evolution today. IBM InfoSphere Discovery drives project completion and faster time to value, helping organizations sustain a competitive edge. With advanced capabilities to identify related data, uncover hidden instances of sensitive information, and define business objects, IBM InfoSphere Discovery provides the fundamental input to business-critical initiatives like warehousing, governance and more. InfoSphere Discovery automates the complex process of discovery, providing broader coverage and greater accuracy which manual data inspection methods could not begin to offer.

*IBM® InfoSphere™ Discovery*

## About IBM InfoSphere Solutions

IBM InfoSphere products accelerate the delivery of trusted information for business optimization. Business optimization requires information that is accurate, complete, in context and actionable. Achieving this level of trusted business information requires transforming, reconciling and maintaining information, and delivering it in real time to the people, processes and applications that need it. IBM InfoSphere products stand alone in the market by offering the breadth of capabilities required for the end to end management and delivery of trusted business information to ensure About IBM InfoSphere Solutions

## About IBM Optim Integrated Data Management Solutions

IBM Optim Integrated Data Management Solutions offer proven, integrated capabilities to manage enterprise application data from

requirements to retirement. With IBM Optim, teams can share data artifacts (like models, policies and metadata) to align data management with business goals and improve collaboration. Today, organizations of all types leverage IBM Optim to improve performance, streamline database administration, speed application development, and enable effective governance. IBM Optim delivers better business outcomes, at lower cost, with less risk, while providing capabilities that scale across enterprise applications, databases and platforms.

## For more information

IBM InfoSphere Discovery complements the value of IBM Information Management solutions, including the IBM InfoSphere and IBM Optim product portfolios. To learn more, contact your IBM sales representative or visit the following Web sites:
ibm.com/software/data/ips
ibm.com/software/data/optim-solutions