



Linked Enterprise Data

Principles, uses and benefits





Linked Enterprise Data

At a glance

Business information systems have developed in incremental fashion. Each new operating need has generated an ad hoc application: ERP, CRM, ECM, directories, messaging, extranet and so on. IS development has been driven by applications and processes. Each new application has created another data silo, with the result that companies today are faced with a new challenge: how to manage and extract value from this disparate, isolated data.

Meanwhile, faster economic cycles, globalization and the Internet have ushered in a new playing field. In the face of intense competition and acceleration throughout the economy, companies must transform themselves in order to survive, grow or conquer. Pyramidal organizations are being replaced by networked operations: businesses are adopting a more flexible structure so as to respond more quickly and unleash their creativity. Operating decisions are being made at every level to improve performance. But in order to become more effective and drive the company strategy forward, employees must have access to contextual information appropriate to their responsibilities.

Linked Enterprise Data offers a radically different form of information access designed to fulfill that objective.

It creates a unified informational space that draws on all of the company's documents and data, whether structured or unstructured, and even external data captured from the Web.

Scalable and flexible by nature, Linked Enterprise Data creates the operational data needed in each functional unit, without modifying existing applications or disrupting the information system.

Linked Enterprise Data gives the unique opportunity to efficiently create new applications targeting specific business needs, such as mobile apps for sales or support, reusing and leveraging existing data available in your information system.

By providing each user with the business perspectives he or she needs, Antidot solutions open new opportunities for seeking and accessing information.

Linked Enterprise Data returns data management to the CIOs and restores autonomy for both functional units and users.

« *Through 2015, organizations integrating high value, diverse, new information types and sources into a coherent information management infrastructure will outperform their industry peers financially by more than 20%.* »

Gartner – september 2011

1. Data: the new IS frontier

Agility has become the key to success. To grow their business, companies must respond to the market by introducing new products and services at an increasingly rapid pace, so as not to be outdistanced by competitors. Economic, legal, social and environmental risks weigh ever more heavily and require an ongoing effort to adapt accordingly. The very perimeter of the company is constantly changing as a result of buyouts, mergers and restructurings.

At a time when everything is moving faster, **information poses a strategic challenge**: it is a fundamental resource and asset for any company, and the cornerstone of its business. Information has come to play a central role **both in decision-making and in implementing strategy**.

To respond to the market, companies need an agile information system.

Over time, as a result of operating pressures, the proliferation of tools and changing technology, the corporate information system (IS) has become fragmented. Whereas the database formed the heart of the system 20 years ago, today there are countless sources of data: ERP, customer relations systems, sales management, legal dossiers, internal directories, marketing ECM, e-mail servers, intranets, extranets, websites and so on. All of these sources are related to applications that were created in response to business needs and fed by day-to-day company operations and employee activities.

And yet this huge quantity of data, structured to varying degrees, is not being mined for the wealth and scope of its potential by either employees or information system applications. Why not? Primarily because the data is closed off in an array of partitioned, seal tight silos. Each source is accessible via a user interface or, in the best cases, via an API or specific Web Service.

The social Web, Open Data, mobile applications... are new tendencies that have emerged at a rapid pace. They generate needs, and hence put more strain on the

company IT strategy. But these major shifts also represent tremendous opportunities. To respond to those challenges and cope with market expectations, you need a way to reuse your data efficiently.

And there we see the true challenge posed by a modern IS: how can we create business applications, quickly and at low cost, that make use of existing data without altering the silos, in order to meet the needs of functional departments? How can we make the data in one application available to other applications without creating additional complexity, bearing in mind the constraints involved in preserving the information system's integrity, maintainability and scalability? How can we explicitly **link data from different silos** that is implicitly related? How can we **incorporate external data** to enhance our applications without disrupting the IS?

Structural modifications to existing silos are not a viable solution, and duplicating data merely adds to the complexity. Nonetheless, some solution must be found for making the data malleable and dynamic; our duty is to provide powerful business objects that are directly usable to those who need them. Control of information demands new tools for capturing, mining, searching and accessing data so that employees can successfully perform the tasks required of them.

Business Intelligence, Master Data Management, Service Oriented Architecture, Federated Search... various solutions have appeared over the past twenty years. They all attempt to tackle the problems raised by data proliferation. As discussed further, these tools undoubtedly provide operating benefits, but in most cases they entail a long and costly deployment process and make the system more complex. And none of them offers a genuine solution to the challenges of a scalable and agile IS.

We need to think about access to information within a unified space that receives data from every source in the company, a space from which it's possible to create new services.

**But how can we create this unified space?
By what process and using what technology?**

Are you able to meet the informational needs of your functional divisions based on the data you currently have in your applications? How quickly and at what cost?

How simply can you create a mobile app for the sales people leveraging products information coming from various sources ?

2. Linked Enterprise Data

Linked Enterprise Data is a revolutionary approach that meshes each and every enterprise data, wherever they come from, to create a global and unified information space from which new business information is created to solve operational needs.

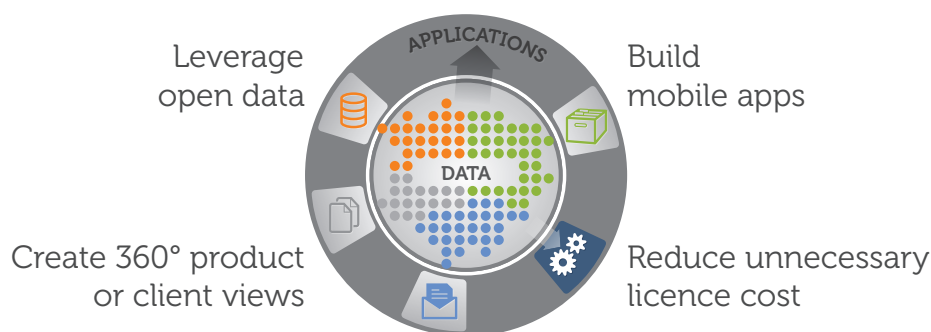
Linked Enterprise Data is different: it is a game changer. LED takes into account each and every type of enterprise data - structured records, documents or office files - where old-style approaches are limited to align relational databases or make applications communicate together. LED breaks down walls and interconnects the data itself at the most detailed level. By federating the content of heterogeneous silos, Linked Enterprise Data creates a unified and coherent warehouse that exposes and shares new knowledge objects tailored to decision-making and action required by business units.

To deliver this change in perspective, Linked Enterprise Data relies on the Semantic Web principles and technologies, created and standardized by W3C over the last 10 years. Also known as Web3.0, this revolution is transforming the Web from a documentary space reserved to human into a gigantic decentralized knowledge base that computers can use to solve complex problems by directly exchanging information. Linked Enterprise Data brings to the corporate IS those technologies that represent a comprehensive response to the challenge of creating an agile, high-performance and open information system: internal data sources are linked together and eventually consolidated with external data.

Linked Enterprise Data is scalable and flexible by nature. You set up LED without modifying existing applications or disrupting the IS. It is pragmatic: you can begin with a targeted project that will offer a quick response as well as a demonstrable return on investment. Then the informational graph is gradually enlarged, on a project-by-project basis, as you are asked to generate new applications, without undermining previous work.

By reading this white paper, you will understand how Linked Enterprise Data brings to your information system enormous benefits:

- **Break down walls between data** and offer an information hub in which the data itself is urbanized rather than the applications. Each application draws on data from throughout the IS by operating through this information hub.
- **Create links between existing data**, whether internal or external.
- **Reduce complexity** by offering a unified methodology for data exchange between applications. Whether the data is internal or drawn from the Web, LED opens up a range of possibilities for consolidating the full range of information available to you.
- **Deliver new applications** such as mobile apps: based on standard modern Web technologies, LED allows you to be reactive and even proactive to market demands.
- **Open, standardized, secure, long-term and effective** technological environment defined by the W3C.



The Linked Enterprise Data makes companies more agile by creating new ways to reuse and leverage the data at their disposal.

USE CASE

Antidot's
internal
search
engine

This real life case illustrates issues that companies face in terms of data management and how they try to improve their business efficiency through information.

Like any company in search of efficiency and performance, Antidot uses everyday a variety of tools that each responds to a given challenge or need:

- Sales and administrative management: CRM and ERP.
- Customer support: ticket management, customer back office, extranet.
- Production coordination and tracking: back office operations to manage our SaaS offering datacenter.
- Knowledge capitalization and dissemination: knowledge base, wiki, ECM.
- Software production by R&D: version managers (SVN, Git), administration and coordination of functional and technical upgrade requests, product bug tracking, automated test benches, production of documentation.
- Communication: e-mail and instant messaging.
- File servers.

Antidot employees, each having their own user profiles and separate objectives, all need access to the right information for their work:

- For administrative and management personnel: consolidated administrative reporting for a customer (project status, days consumed, cumulative invoicing and outstanding invoices), number of support tickets submitted and still open, traffic and service indicators for customers in SaaS mode, internal personnel involved, employee tracking, etc.
- For sales and pre-sales personnel: identification of contacts listed in the CRM; information on projects carried out with other companies from the same industry; technical modules typically proposed in similar projects; appropriate references and customer scenarios.
- Consultants and project managers have multiple needs:
 - » For managing and carrying out the project in accordance with sales commitments: sales proposal, purchase order, special administrative terms and conditions and special technical specifications for government contracts, functional specifications and scope, etc. For optimal access, these documents should be classified by genre, with each successive version aggregated with the final version.
 - » During the development and deployment phases: access to technical documentation, the knowledge base and guides to best practices based on appropriate product strategies.
- For customer support: summary administrative record; past and pending support tickets; any similar requests and incidents involving other customers; links with the internal knowledge base and technical documentation; list of internal staff for escalation procedure; identification of internal experts who have worked on similar projects, modules or functions; information on products and modules deployed.

Thus, what Antidot needs is to provide each employee with contextual informational resources that are consistent with identified use scenarios. We do have a state of the art unified enterprise search engine: Antidot Finder Suite, the award-winning full featured semantic search solution. But it falls far short of fulfilling these needs. According to the classical search engines paradigm, AFS indexes documents and structured data records. Therefore, it answers documents and records. But a list of documents as a reply to a query is not a valuable business information. This is why we must go further and create the knowledge objects that everyone needs

Linked Enterprise Data provides the solution

3. Understanding the new paradigm

The basis of the Web

Originally designed to serve as a universal document publication system, the Web has radically evolved over the past 15 years. With the advent of Web 2.0 and social networks, it has been transformed into a participatory and interactive ecosystem in which the user is no longer just a reader, but an integral participant.

Web has also had a technological impact, since companies have had to overhaul their information systems, just as software solutions have been redesigned to reflect the principles and technology standardized by the W3C. The IT village has been webified: a single tool, the browser; a single language for presenting data, HTML; and a single communication protocol, HTTP.

In 2012, Web pages and data in the tens of billions can be found online: statistics, timetables, recipes, books, user reviews, stock prices, music, blogs and wikis, personal photos and video... Nothing escapes the Web, and a significant

portion of all human knowledge and information is now available online. And yet all this information is only available to people, since the Web is literature. Computers are unable to consume and leverage this information in order to resolve complex problems.

Web 3.0, also known as the Semantic Web, is the latest iteration of the Web, in which computers can publish, capture, process and exchange information automatically and unambiguously.

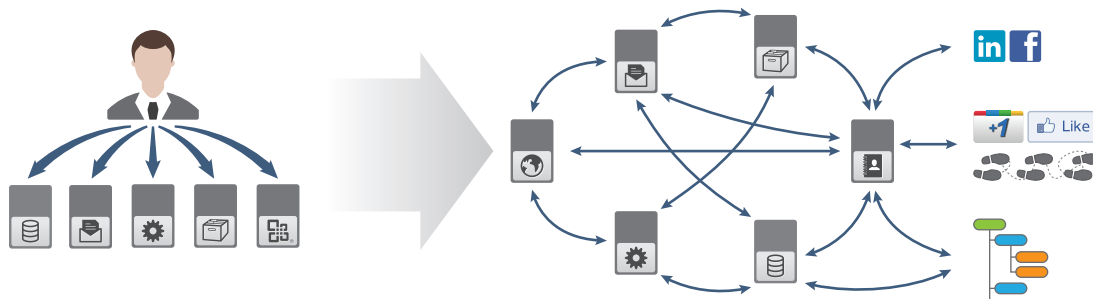
The goal of the Semantic Web is to go well beyond simple machine access to raw data by providing a way of interweaving this data. This process, known as Linked Data, creates a decentralized knowledge base that spans the Web, in which the value of each piece of information is enhanced by the presence of related data. Just as links between pages formed the success of Web 1.0 and links between people led to the explosion of social networks, Linked Data is radically transforming the Internet.

The technology behind LED powers the Web. It is mature and offers daily the proof of its reliability and scalability.

For more information on how the Web has evolved and to gain a better understanding of the benefits and challenges of the Semantic Web, download and read "The future of the Web".

Linking data in the enterprise

The Semantic Web offers major benefits because the relevant technology has been designed to generate a dynamic, scalable description of data, unconstrained by a rigid model; to construct simple links between this data so as to build a giant knowledge base; and to query this base in order to find solutions to complex questions and create new informational objects. These solutions, already in place across the Web, now offer businesses a unique opportunity to transform their information systems by placing data at the heart of their solution.



In a conventional setup, users access separate applications that each manage their own data. Users who must synergize multiple sources of data to resolve a problem are forced to consult each application separately and then undertake the sometimes tedious process of consolidating the results.

At companies that deploy a Linked Data strategy, these internal data sources are interlinked and may even be consolidated with different or comparable external data: data repositories, user traces, social data, statistical data, etc., to create a **truly unified information space**.

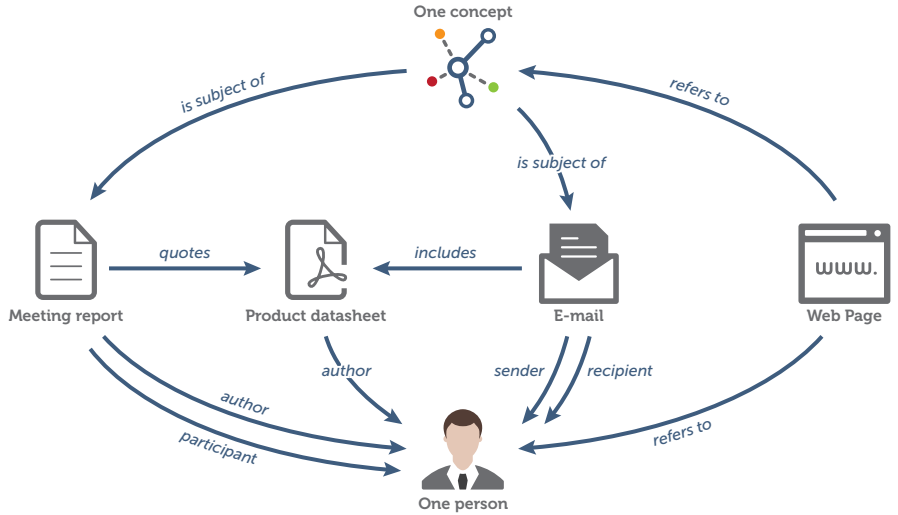
LED consolidates and combines your data so as to yield personalized, contextualized information to aid in your decision-making.

With LED, seamlessly include social data from Facebook or LinkedIn to enhance your sales application (CRM) or boost your marketing analyses.

With LED, overtake the outdated «document centric model». Benefit from truly fine-grained meshed information.

Linked Enterprise Data marks a genuine breakthrough insofar as the data itself is networked, and not just applications or complete silos that sit side by side.

In order to create this dense and coherent information space, we need to model the various data, develop a semantics and interconnect the data using the principles of the Semantic Web. For this we can simply draw on the nature of the source data and generate explicit links from the implicit ones that already exist.

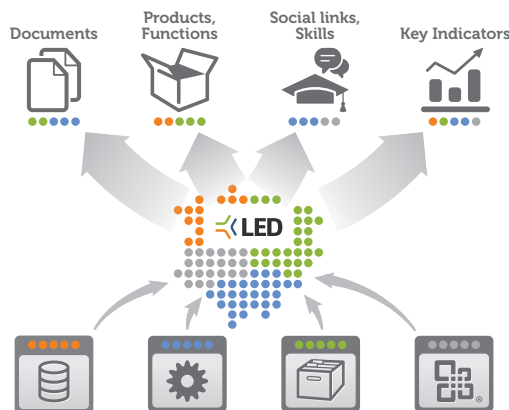


The diagram above illustrates this process using standard business data: the minutes from a meeting identify the author and the meeting participants, as listed in the company directory; an e-mail has a distribution list that can be found in the CRM, and includes as an attachment a marketing document available from the ECM system, that document includes product names taken from the product catalog.

Like the web connects Web pages, Linked Enterprise Data interweaves the data at the most detailed level. These links are like joins in a relational database but across the entire information system, which thus becomes an immense distributed database of company information and knowledge.

We can clearly see how each link generated creates new information: who were the recipients of document X in its version Y? How many employees from the research staff have taken part in project Z? Which project managers have already deployed product P in a project of more than X man-days? Linked Enterprise Data can easily provide an answer to all these questions.

What's more, think of the added power available to, say, an umbrella manufacturer from the automated input of data from Wikipedia or government statistics. It probably knows from its own applications that its distributor X sells N umbrellas per year in Boston. But how many people live in Boston? How many days of rainfall do they get per year? By integrating this information into its computer system, the company has enhanced, more relevant data at its disposal. It transforms itself into a true knowledge base.



Documents from sources, new business data, quantified indicators—all this critical information for decision-making and strategizing is made available to existing applications and users, specifically in the form of business objects that enhance the company's information space.

The efficiency of Linked Enterprise Data

By its nature, **Linked Enterprise Data** is **flexible** and **non-disruptive** to the information system.

LED is flexible in more than one respect

As the source data is being analyzed and target objects are being modelled, LED emphasizes reusability and model sharing. Numerous vocabularies and models have been defined by a variety of communities (experts, occupational sectors, standardization bodies). Unlike relational models that are difficult to capture and adapt, the Semantic Web approach creates models that are natively designed to be shared and expanded to address the specific needs at hand. Linked Data offers a guaranteed capacity to retrieve and share models that can then be easily adapted, enlarged and maintained at low cost, without invalidating the existing model.

In this regard, Linked Enterprise Data is both a horizontal and vertical solution:

- Horizontal, in that it is generic, open, and standards-based. As a result, LED can handle a variety of needs and can be incorporated into the information system as both technology and process.
- Vertical, through its exceptionally powerful capacity to share and reuse object models: LED provides a basis for developing strongly business-oriented strategies. For example, Antidot's LED offering includes a set of standard business objects (360° customer views, product maintenance

and after-sale support, projects, etc.), while some of our partners are working to provide more sector-specific objects (manufacturing, law, etc.).

In technological terms, LED offers the flexibility inherent in data graphs, the biggest advantage of the Semantic Web. Data can be added or new objects created without having to redo the existing IS from scratch.

Thus, Linked Enterprise Data isn't like a cathedral whose construction requires a company to rethink its organization, revamp its information system or anticipate every possible application or scenario from the outset. On the contrary: an LED strategy can begin with a small-scale project that will offer a quick, inexpensive response to an isolated business need, with a demonstrable return on investment. The project proceeds in pragmatic fashion, with an analysis of the available data and ideal use scenarios.

Each new project and requirement will gradually enlarge the information graph, without the need to revise or overhaul the initial models, to yield enhanced value and generate new applications—so the virtuous circle behind Linked Enterprise Data is closed.

With LED, start easily and rapidly with narrow tactical project. And grow large without reinventing.

LED is non-disruptive to the information system

Users enjoy access to the information graph and newly created knowledge objects in two distinct ways:

- For the information system, access is via technical interfaces like Web Services or SPARQL endpoints. Thus, each application can directly retrieve the items it needs in order to enhance its content.
- For users, the search engine is an ideal tool for providing service without modifying existing applications. It can be used to extract selected business objects and to create interfaces and applications quickly and at low cost, without disrupting IS operations or adding complexity.

With LED based on Semantic Web technology, you can exchange data between subsidiaries, business units or partners without having to develop specific APIs.

A Linked Enterprise Data solution provides:

PRODUCTIVITY

create new data and new apps as needed.

SCALABILITY

thanks to its separation of usage, applications and data.

FLEXIBILITY

inherent in data graphs.

CONTINUITY

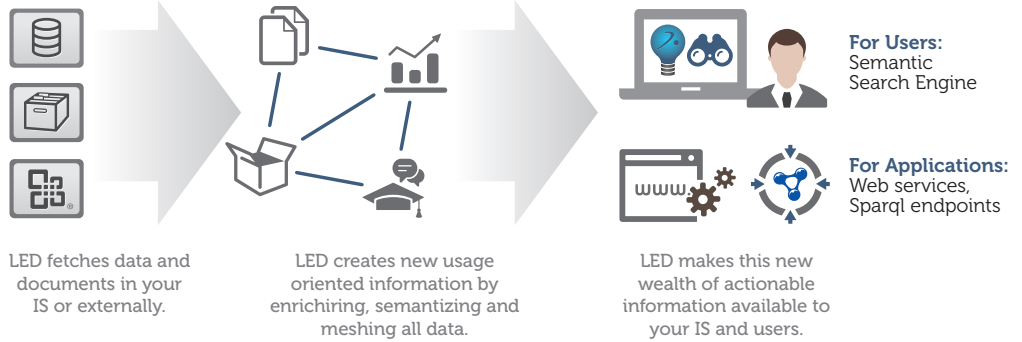
because LED doesn't need IS modification.

REVERSIBILITY

and openness, through its use of W3C standards.

The benefits of Linked Enterprise Data

By federating the content of silos, Linked Enterprise Data (LED) creates a unified warehouse exposing and sharing new knowledge objects appropriate to the context and needs. As a result, LED serves as a Search-Based Application in the fullest sense of the term, since it supplies the open-ended technological building blocks and built-in process that you need to leverage your existing data and create apps your stakeholders want.



Should companies move to Linked Data? In 1995, the question was: "Do we need to create a website?" And in 2000: "Should we revamp our business applications to make them accessible via a browser?" Once again, the answer is clear. Companies that have successfully ridden the wave performed at a higher level thereafter. Those that hesitated have fallen behind. The same will be true of Linked Data, since the power of the Web is such that each breaking wave of innovation transforms the landscape.

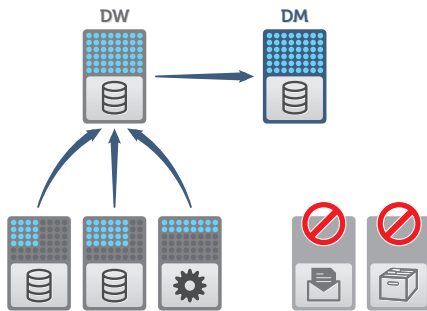
«Leveraging information will continue to fuel business success. But the growth in information volume, velocity, variety and complexity and the new information use cases makes information management infinitely more complex than it has been in the past. In addition to the new sources and the increased demand for multiple context delivery, shareability and reuse, practically all information assets must be available for delivery through varied, multiple and concurrent channels and mobile devices.»

To deal with these new demands, the IT organization needs to dramatically modernize its IT systems, transforming outdated data management infrastructure and replacing it with a more up-to-date and superior information environment able to support an entirely new set of requirements.»

Gartner - sep 2011

4. Comparison with other approaches

Various strategies have been proposed to offset the rigidity of information systems and ensure they can be upgraded in the future. Each addresses a need and meets a specific objective. With hindsight, the lessons learned from experience show that none of them genuinely addresses the fundamental challenges cited earlier.



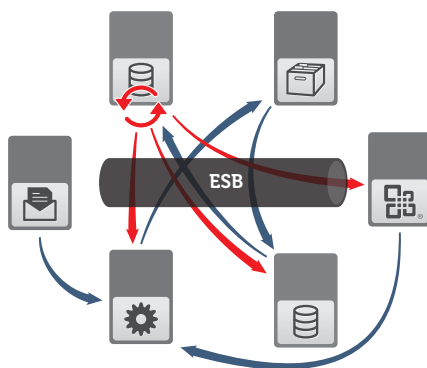
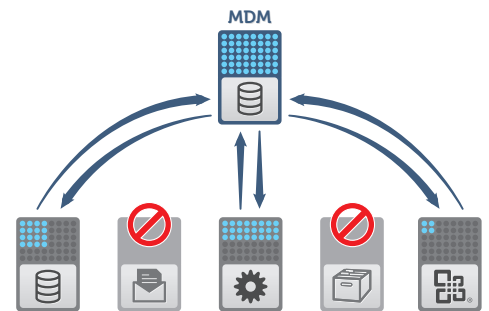
Business Intelligence (BI) offers analytical views and exploratory areas by consolidating quantitative data from various databases. It is valuable for its ability to create informational objects that provide an operational service to the end user.

However, it does not take into account unstructured data which are numerous and often contain important information. Its drawbacks also lie in the lack of scalability since the warehouses (data warehouses and data marts) must be redefined, rebuilt and optimized each time new data is added.

Master Data Management (MDM) tackles the problem of data proliferation and reference data deviation (catalogs, customer information, etc.). MDM addresses a critical business challenge: the unification of a common subset of data, in this approach by creating technical links (synchronization streams) between sources.

MDM creates a new silo in which the selected reference data are duplicated and with which each concerned database must align. The complexity lies in ensuring data synchronization and managing conflicts between sources when data are updated.

However, MDM's difficulty and cost have limited its spread to large companies. In addition, not all data can be integrated into an MDM solution, since it is restricted to databases.

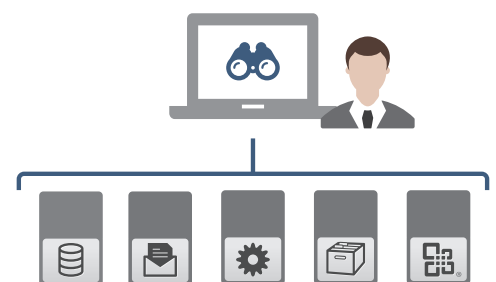


With an urbanization-based approach to information systems, known as SOA (Service-Oriented Architecture), applications can integrate data from other applications. Nonetheless, ERP, CRM, relational databases, intranets, ECM, messaging—all continue to exist in essentially heterogeneous silos, each structured in a particular way and requiring specific access methods. Given the complexity that results from using multiple APIs in combination, maintenance is difficult, and the options are limited to existing data. Enterprise service buses (ESBs) simplify the work involved, but they don't resolve these fundamental biases.

The success of SOA demonstrates the need for a way to integrate and consolidate data from various sources. It has been instrumental in showing that any integration must occur at the level of data rather than applications.

Federated and unified search have been a promising approach for many years, but has now reached its limits: in cases where data sources vary widely in nature, business vocabularies are sparse and user terminology is subjective, even the most powerful search engine can't perform miracles on its own. At best, a federated search engine is able to call up a range of disparate documents (files, e-mail, web pages, database records), deprived of all context.

And when users are faced with a long list of documents to review, even with filters or facets for sorting and refining their search, their business needs are not genuinely being served: reading through a list of results is tedious work that does not fulfill an operational need—unless you are looking for a document that you know exists, and that is rarely the case. Users are more often searching for information rather than a specific document.



Employees can play an active role in improving their company's performance if they have accurate, complete and appropriate information at their fingertips.

Search-Based Application (SBA) use search engines as part of an approach that combines a light BI with document search methods. It has no real technological basis and offer no clear vision or path — SBA was primarily a marketing concept put forth by certain publishers of business search solutions to differentiate their offering of Internet search engines, such as Google. But LED brings SBA the basis it needed to become a technologically coherent and comprehensive solution

The latest approach, known as **Big Data**, involves the ability to process massive quantities of structured or unstructured data. It emerged among the major websites —Twitter, Facebook, Yahoo!— which handle exceptional volumes of data, including tweets, Web pages, photos, etc., numbering in the billions. It uses a technological approach appropriate to a scale of data processing at which typical programming techniques and infrastructure no longer suffice. By itself Big Data does not respond to any specific functional or business expectation. It has to be incorporated into real solutions in order to meet the needs of companies. As such LED can take advantage of Big Data technology to integrate the results of advanced computations with other data and create richer knowledge objects.

Even if they are adopted mostly by large corporations or on a limited basis, the very existence of these solutions reflects a practical need to:

- Pool IS data in order to create information that will make a service operational for users,
- Integrate and distribute data between applications, both internally and externally,
- Establish an information infrastructure that emphasizes agility, scalability and ease of installation.

But meeting these challenges is not simply a matter of technology. First and foremost, we need to change the paradigm by **placing the data at the heart of our approach**. LED shows to be the solution to begin thinking about access to information within an expansive, unified space that receives data from every source in the company.

The table below summarizes the contributions and drawbacks of each of these traditional solutions, compared to the advantages of Linked Enterprise Data:

SOLUTIONS	CONTRIBUTIONS	DRAWBACKS	ADVANTAGES OF LED
BI Business Intelligence	Provides information that is immediately operational. Generates new information as needed.	Complex to install and optimize and unwieldy to upgrade. Limited to quantitative structured data. Slow turnaround.	A LED solution can be expanded without altering the model and mechanisms already in place. Takes all types of data into account, including documents.
MDM Master Data Management	Pooling of data among structured sources.	Significant prior analysis required, technologically complex, costly to upgrade. Limited to databases.	LED is flexible by nature and emphasizes an incremental approach to data integration, with not need to modify existing silos.
SOA Service Oriented Architecture	Allows applications to integrate data from another source.	Difficult to maintain and increasingly complex as the number of sources grows. Only yields data that already exists in other silos.	Unlike peer-to-peer (silo-to-silo) integration, LED provides a comprehensive view of the data as a whole and can be used to create new information.
Federated search	Simple to implement and use, provides comprehensive results.	Document-based approach poorly suited to business needs.	LED goes beyond a document-based framework by creating informational objects and apps contextualized to business objectives.
Big data	Processing of massive amounts of data.	Complex technological approach (at the lower layers) but no operational support in itself.	LED adds semantics to data processing, delivering «Smart Data» rather than just big.

5. Business solutions based on LED

As we have seen, Linked Enterprise Data is at the same time an innovative approach, a set of modern technologies and an effective method to serve an IS strategy fostering agility, leveraging data assets and empowering users with a wealth of information.

But more than technology, LED is a tremendous way to create business solutions that meet the needs of highly targeted areas as diverse as sales, support or human resources ... all critical domains for the company.

This ability to verticalization is directly linked to the essence of LED as the consolidation and data-meshing phases are oriented towards the creation of new knowledge objects by applying rules of reasoning.

The following example illustrates this ability and explains the way it is implemented. This solution has been successfully delivered to a big company who runs 10 business premises clustering thousands of people and hundreds of technological and innovation projects per year. Understanding and fostering relationships between offices and teams was critical

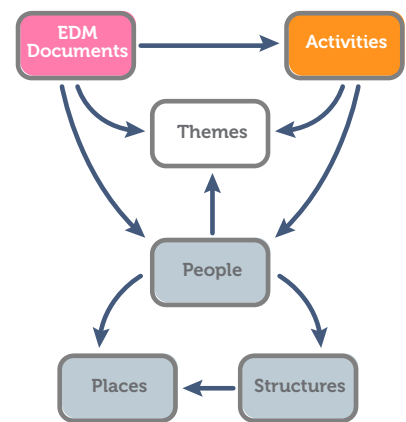
Detection of experts and networks of skills in the enterprise

The objective is to create five original informational objects that meet the challenges of skills management by identifying:

- Employees and their skills
- Structures and groups
- Places: sites, buildings, floors and rooms
- Structures themes or concepts
- Cases and projects grouped under the «activities» term.

Following the linking process, business objects are themselves connected. This makes navigation easier, allowing resource discovery and improving understanding. The diagram gives a macro view of relationships between these objects.

To create this new information, the solution must use data and documents from the following data silos:



People and structures directory



Electronic document management system (EDM)



Portfolio cases management system

1st step: open up data by creating links

Like any other information system, the one of this company has been build by adding successive applications that manage their own closed data silo. Therefore, the first step is to fetch the data contained in the silos, to process it and then convert it in a standard format (called RDF), uppermost taking advantage of existing generic ontologies as target models. Once done, the meshing phase can begin:

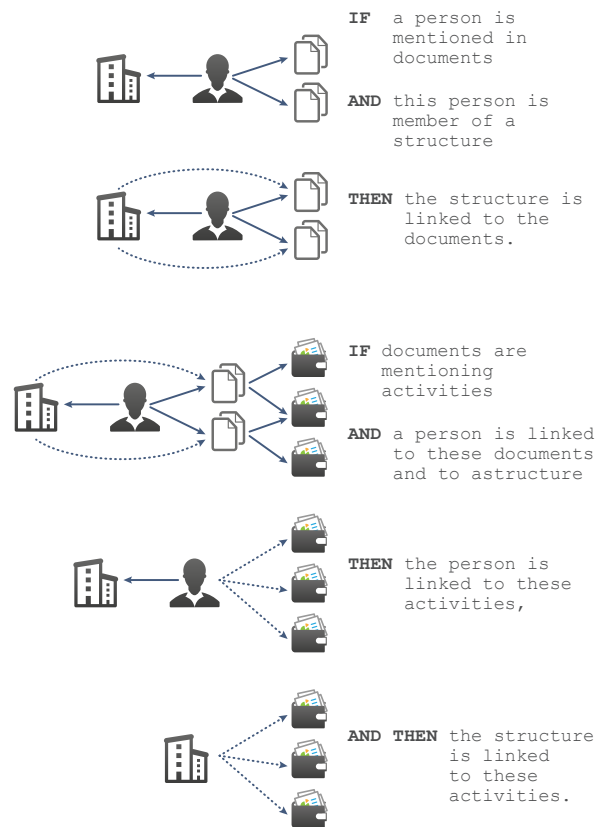
- Some links are immediate because data have « common » fields (if a document is written by someone, then it is the author),
- While others links require more advanced treatments: for example, retrieving the list of participants and topics from the minute of a meeting.



2nd step: produce new information

Once direct links are created, the next step is to densify the linking and generate new relationships by applying specific inference rules to each object type, still driven by target ontology. These inference rules play a key role, because they are an important part of specificities of the project and business objects to be created.

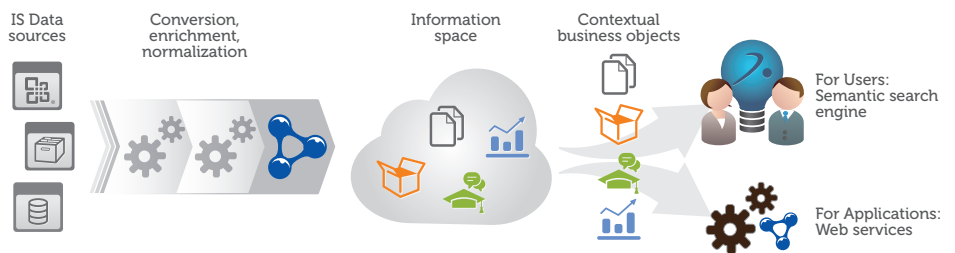
- People:
 - » The person who writes a document on special topics focuses on these topics.
 - » A person who has written a document on an activity is linked to this activity.
 - » Two people involved in the same document are collaborating.
 - » ...
- Places:
 - » If a place includes a person who is interested in a theme then this place is linked to this theme.
 - » If a person is involved in an activity and is located in one place, then the place is linked with this activity.
 - » If a document is written by a person in a place then this place is linked with the themes of this document.
 - » ...
- Activities:
 - » If a person is a member of a structure and is linked to an activity then this activity is linked to the structure.
 - » If a document is linked to an activity and is also linked with a structure then this activity is linked to the structure.
 - » ...
- Structures:
 - » If a document is linked to an activity and its author is a member of a structure then this structure is linked with this activity.
 - » If a people is a member of a structure and is also linked to an activity, then this structure is also linked to this activity.
 - » If a document is linked with a structure then this structure is linked with all topics linked to the document.
 - » ...



3rd step: extract information objects

The processing workflow that makes possible the creation of these new knowledge objects is implemented with Antidot Information Factory (AIF), a solution designed to create and execute advanced data enhancement tasks:

- Data conversion, enrichment and standardization according to various vocabularies,
- New information production by inference,
- Data processing in order to expose it using Web Services.



Eventually, the project helped to automatically build a «network of skills and expertise» for this company: it explicitly describes all existing relationships between people, structures, places and activities, based on implicit information buried in company documents and relying on its business vocabularies. This network is available for users through the Antidot Finder Suite search engine, and to existing or future information system applications via Web Services.

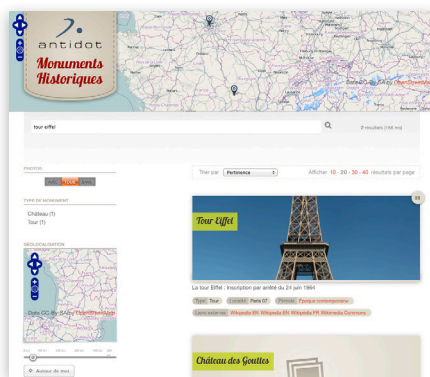
This application dedicated to human resources is just one of the countless ways to use LED in order to rapidly create applications that reuse and enhance existing data

6. Sample projects

In this section we present two applications of LED approach and Web of Data technology in which the circumstances differ, but the common goal was to consolidate and mine data for the purpose of creating value:

- The development of an extensive knowledge base on French historical monuments, by aggregating and making use of seven different sources of data.
- Antidot's new internal search engine, which uses the Linked Enterprise Data technology and approach to provide each employee with a business vision of internal data.

The Historical Monuments Application

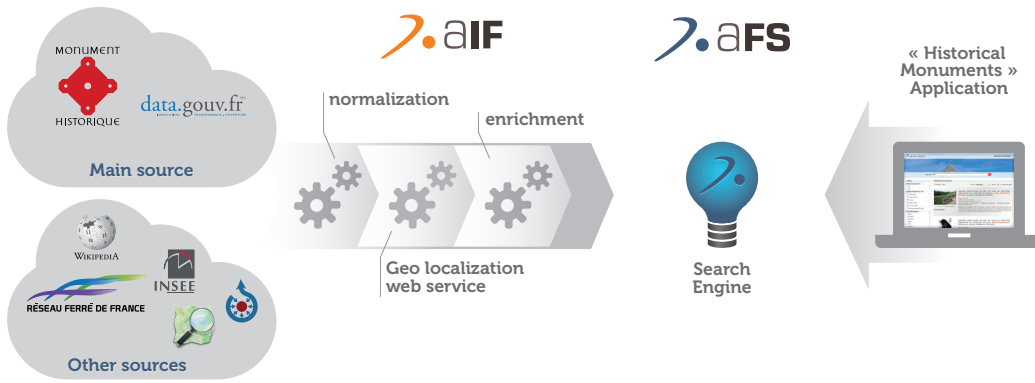


The launch of the data.gouv.fr site on December 5, 2011, followed immediately by further initiatives, was a sign that France's Open Data movement is gathering pace. We have capitalized on this greater access to data by creating an application that demonstrates the value of Linked Data. It interlinks data from a variety of sources, specifically by combining "internal" data with data available on the Internet.

The core data, which in this case we consider the internal data (it could be a product database, points of sale, customer firms, etc.), is the list of buildings in France that are protected by virtue of being Historical Monuments. This data source, available in a CSV file at data.gouv.fr, includes 43,720 monuments.

We have enhanced this database by making use of six sources of open data:

- The list of the 3,065 passenger rail stations in France's National Rail Network, along with their geographical location, as provided in XLS format at data.gouv.fr. This list is used to link each monument with the nearest train station.
- The list of 301 stations in the Paris Metro, along with their location, as provided by OpenStreetMap. We used this source to identify the location of Paris monuments in relation to the nearest Metro station.
- The data in the Official Geographical Code (COG) published by France's National Institute for Statistics and Economic Studies (INSEE). This source, available in RDF format, defines 22 regions, 99 departments and more than 4,000 administrative units within the departments.
- Photos of the historical monuments, as provided by Wikimedia Commons. This source, notably including photos submitted for the Wiki Loves Monuments contest, offers 122,828 photos for 12,586 historical monuments designated by their PA code. This code is generated uniquely for each monument and shown in the core list of monuments.
- Wikipedia's description of the historical monuments, provided by DBpedia. This data source (in RDF format) includes descriptions of 3.64 million objects, including 413,000 sites. It can be accessed directly from the information at Wikimedia Commons.
- The Yahoo! PlaceFinder geolocation service, which can be used to geolocate monuments based on their address (calculation of their latitude and longitude).

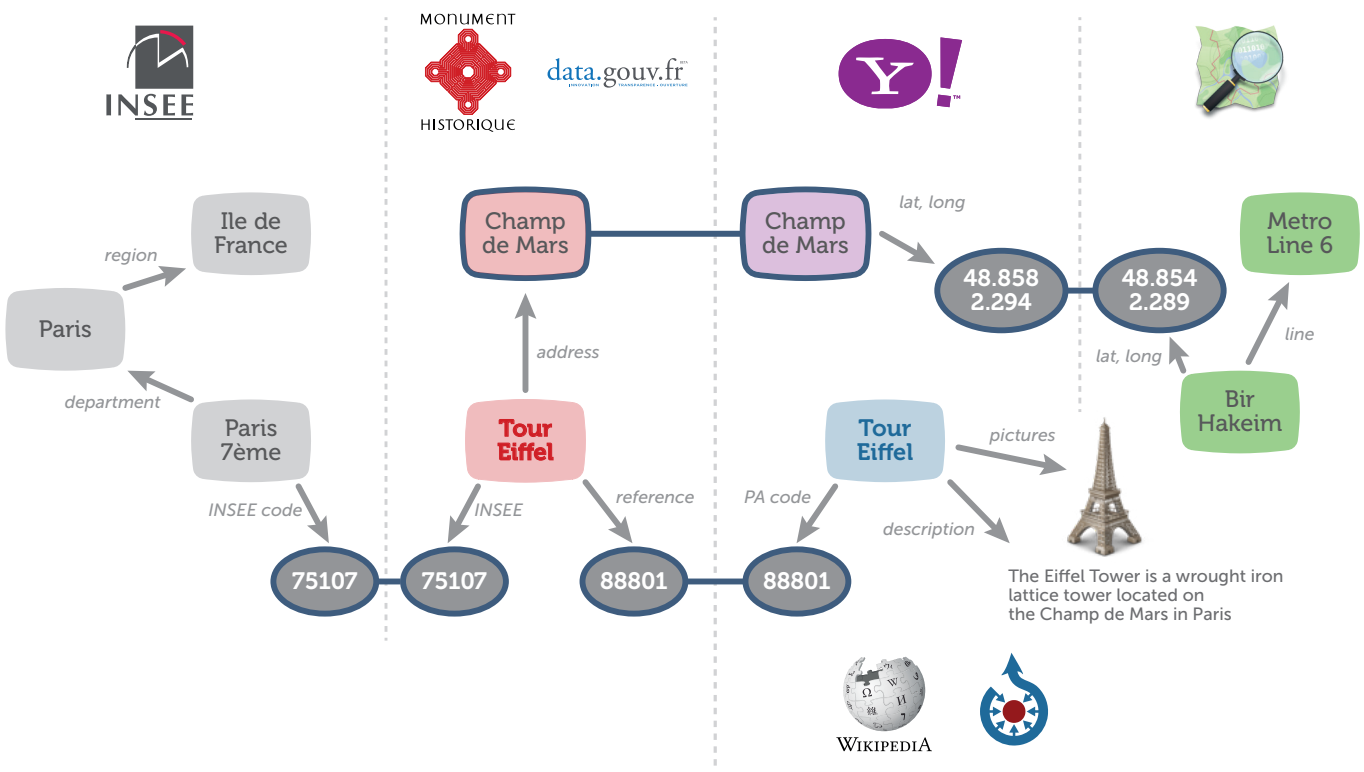


Using the Antidot Information Factory, all this data was collected, cleaned, standardized and then converted into RDF, the data representation format that has been standardized by the W3C for the Web of Data. The core data (the monuments), the list of train stations and the INSEE data were retrieved in file form, while the OpenStreetMap, Wikimedia, Wikipedia and Yahoo! sources were compiled using specific APIs.

The data contained in the core file is fairly cursory. For the Eiffel Tower, for example, the information provided is as follows:

- REF :** PA00088801
- STUD :** Survey HM buildings
- INSEE :** 75107
- TICO :** Tour Eiffel
- ADDR :** Champ-de-Mars
- STAT :** municipally owned
- PPRO :** The Eiffel Tower: recorded by decree of June 24, 1964
- AUTH :** Eiffel Gustave (principal)
- CENT :** 4th quarter 19th century

The diagram below shows how this initial data is enhanced by using the data contained in the supplemental sources.



The resulting target business object is a “monument” whose definition has been enhanced with the following information provided by the supplemental sources:

- geographical location;
- address, city, department and region;
- train stations within a distance of 20 km;
- Metro stations within a distance of 1 km for buildings in Paris;
- an extensive description in several languages;
- historical period;
- type of monument;
- owner;
- photos.

This process is conducted automatically across every database, and the result is a graph that includes more than 4.5 million elements (RDF triples), nearly 450,000 of which were inferred, i.e., created through the application of rules.

This target objects are fed into the AFS search engine. The results are available via a Web application that offers the following search and browsing functions:

- Plain-text searches,
- Filtering by a given region, department or city,
- By type of monument: church, chateau, statue, industrial site,
- By historical period: prehistory, Middle Ages, Renaissance, etc.
- By type of owner: private individual or company, municipality, central government, etc.

with the option of combining all these criteria in the form of “search facets” that are very simple to manipulate.

> **Affiner**

» Photos du monument ?
 Oui 1

» Région / Département / Ville
 Ile-De-France 1

» Type de monument
 Tour 1

» Période historique
 Époque contemporaine 1

» Gares les plus proches
 Champ-de-Mars-Tour-Eiffel 1
 Pont-de-l'Alma 1
 Avenue du Pdt Kennedy 1

» Stations de métro
 Métro 6 1
 ... Bir-Hakeim 1
 ... Passy 1
 Métro 9 1
 ... Iéna 1

Visit the online application to view the results:
<http://labs.antidot.net/demo/monuments>

89



La tour Eiffel : inscription par arrêté du 24 juin 1964

Type **Tour** Localité **Paris 07** Période **Époque contemporaine**

Liens externes [Wikipedia EN](#) [Wikipedia EN](#) [Wikipédia FR](#) [Wikimedia Commons](#)

The Antidot business search engine

This case illustrates how Linked Enterprise Data is the foundation of a new generation of enterprise search engines. Instead of indexing raw files and records, LED creates the information objects that everyone needs at Antidot: sales support, project management...

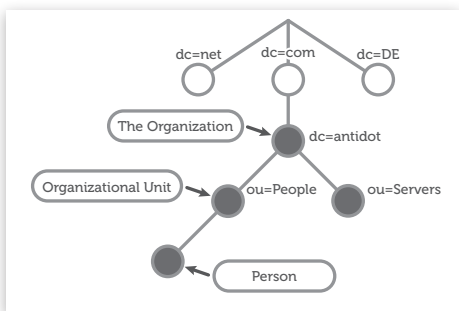
Creating the space

To do this, the Antidot Information Factory (AIF) tool is set to work on the following tasks:

- Capture both structured and unstructured data contained in the various application silos,
- Convert the data: cleaning, standardization, transfer to target format,
- Enhance the data by categorizing it in accordance with classification systems, automatically annotating the data and aligning it with master business data,
- Link the data in order to find synergies and elicit relevant information in the form of business objects,
- Generate the target business objects.

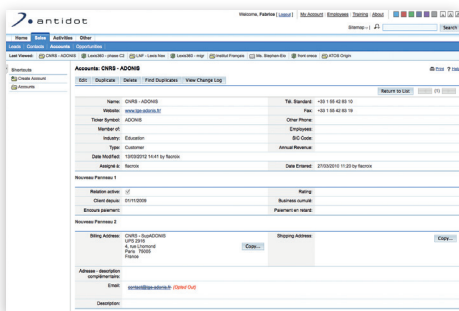
The resulting objects are then indexed and made available using an application based on the Antidot Finder Suite (AFS) semantic search engine.

The table below indicates the sources and information used to create these objects:



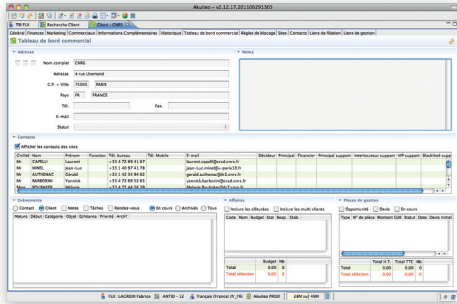
The LDAP directory contains identification and authentication data for Antidot employees as well as partner users and customers who have access to our tools: full name, company, work group, profile, position, e-mail address, telephone number, etc.

All of this information is extracted and used to create an employee database as well as a list of people who have an active professional relationship with Antidot.



The CRM (SugarCRM) tool is used by sales representatives, operational marketing and management to track sales activity, contacts and prospective customers. It contains information on businesses as well as on contacts made at trade shows, sales visits and pre-sales campaigns.

The companies are used to put together a reference list. The individuals listed are also extracted and used to supplement the list of people captured from the directory. This data is consolidated and duplicates are eliminated in order to create a single "personal record".



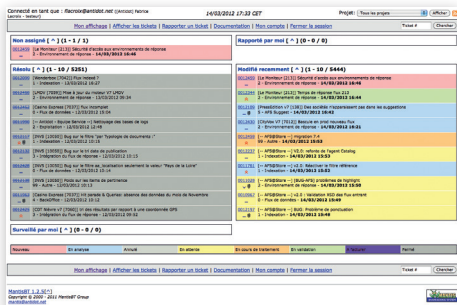
The ERP (Akuiteo) is the enterprise management system. It is used both for customer invoicing and for managing projects in progress: resource allocation, time tracking, production schedules, etc.

By consolidating the data on companies with the CRM data and linking the CRM and ERP data, users gain a perspective that stretches from the pre-sales period to after-sale tracking.

Invoicing data is captured as well. Sales data is cross-referenced with timesheets to generate calculations automatically for each project phase, profiling the activities performed, the time spent and the employees involved.

The SaaS production back office included in the Antidot solution provides a consolidated view of all of the search services managed in SaaS mode within our private cloud. Each customer has access to this back office to oversee and track the operation of its services; Antidot employees use it as well for monitoring and maintenance purposes.

Traffic data, information volumes, indexing dates and an array of indicators are extracted to provide specific information about search service operations for SaaS customers.



The support ticket management (Mantis) is a tool to which all of our customers have access for submitting support requests, reporting incidents or posing questions.

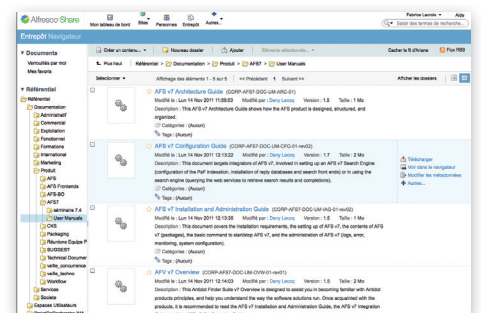
Each ticket is assigned to a project either in development or in production, lists the personnel involved and indicates a product or function.

We use that information to calculate quality indicators by project and customer: number of open tickets, average resolution time, etc. We also calculate indicators for the products or functions that require the most support among customers, and launch a quality control process as needed; in addition, we compile data on the Antidot employees who responded to service requests, by subject matter and customer, to enhance our ability to identify employee know-how.

The ECM (Alfresco) is used to store all of the company's reference documents, ranging from marketing brochures to product manuals and technical release notes.

This ECM tool is notable in several respects. First, it contains an information classification system that reflects Antidot's organization of the data.

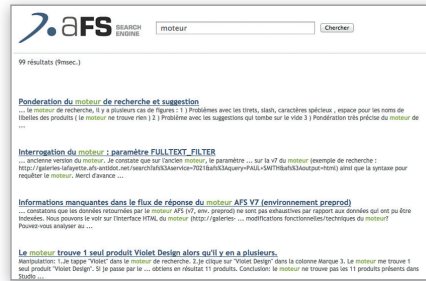
This system is extracted and transformed into a taxonomy. All of the files contained in the various file servers will be categorized on the basis of that organization, so as to provide for unified navigation. The various documents in the ECM are mined to discover references to the product and technologies. These documents are also annotated to reflect the subject matter addressed and the internal or external personnel involved.



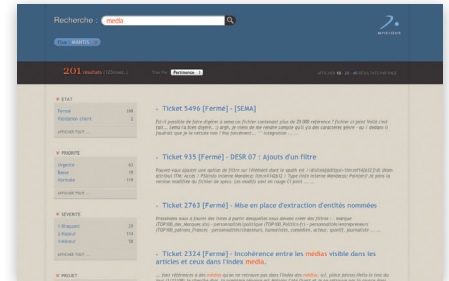
We will not provide a detailed description here of the company's other sources of data: Wiki, internal sites, product and version managers, file servers. These are also captured, analyzed and linked.

Results

The principle of Internet search engines creates little value to users in a business context. Information access solutions that merely reproduce those engines are not efficient, even when indexing all sources and adding filtering or semantic search functions. We have demonstrated this by ourselves using our product AFS to reproduce different types of engine.



In a very basic approach, the user is faced with a long list of documents.



The enterprise search with its semantic, autocompletion, tolerance or filtering features is more comprehensive, but the service is still weak compared to business needs.

LED provides the breakthrough that users expect, in the usage as in the value provided.

By linking data in accordance with Linked Enterprise Data principles, we have created a dense information space from which we have extracted business objects that can be used to model use scenarios and answer specific questions.

Here are three examples:

- The **“customer”** object provides access to a list of projects in development and their timeline, the search services that are active in our cloud or the licenses deployed, the traffic and volume of search services for a customer, indicators regarding current or past support requests, invoicing status, employees involved in the relationship and the most recent contacts made by the sales representative responsible for handling the customer relationship. Naturally, these items will vary considerably depending on the user profile.
- The **“Antidot employee”** object contains a list of the employee’s projects and the customers with whom he or she is in contact; the employee’s projected workload; tasks in progress for the past week and for the two upcoming weeks; scheduled vacation dates; the products, functions and sectors in which the employee has accumulated business expertise; and the technology that he or she has mastered.
- The **“product function”** object describes a module in our offering. Each function includes links to technical documentation, support tickets, bugs and incidents, entries in the knowledge base, customer projects that use the module and employees with expertise that relates to the function. Thus, users can identify customers who will be affected by a modification to the function, and for each support ticket they can automatically link to items in the documentation and entries in the knowledge database. Sales representatives can identify the partners with specific expert know-how.

The screenshot below shows the result when searching for information about a customer: a global virtual record - a knowledge object - is made available, gathering key data from all the IS.

Ongoing projects and business opportunities

Contacts

Support tickets under consideration

Various indicators dynamically calculated from raw data

Production data aggregated from different systems

Key documents from the GED and file servers, sorted and grouped by type: commercial, legal, project management...

The information provided depends on the user profile and the usage scenario. With a single click, the user is led back to the source application where is can act and change the data.

Benefits

The benefits for the company are immediate and obvious. Using a Linked Enterprise Data strategy, the company has aggregated disparate data and given it substance, so that **users can find their way to information they would not typically see** or to which access is difficult. Project managers would only log on rarely to the ERP in order to systematically track a project's profitability; administrative staff would consult the support tool purely to evaluate perceived service quality by our customers.

In the past, providing employees with access to certain applications was a delicate process, because the rights management for these tools was inappropriately clumsy: certain confidential fields could not be blocked. Thus, instead of modifying the application or providing access simply so that information can be read, **we can now provide each person with access to the data he or she needs**, thanks to the business objects created using our LED approach.

Instead of a document search engine that may be relevant but proves largely inefficient in light of the company's operational challenges, the collaborators now have an effective, high-performance tool that gives employees access to the information they need and allows them to navigate among contextual business objects.



Antidot Information Factory is a software solution dedicated to the enhancement and enrichment of data and documents. With AIF, in a few clicks you create processing work-flows by simply assembling modules to:

- Collect data and documents, whether internal or external to the company.
- Cleanse, transform, analyze and semantize them.
- Mesh them and integrate business rules to create a richer information tailored to business challenges.

With more than 60 ready-to-use modules that just need to be configured, you do not have to develop complex programs that are time-consuming and risky to maintain. AIF integrates the Semantic Web approach and technologies, and you can benefit from its power easily and gradually. AIF frees you from technical tasks: you focus on data and value creation. Then AIF reduces the complexity, time and cost of your data projects.

More information:

<http://www.antidot.net/en/products-services/software-products/antidot-information-factory/>

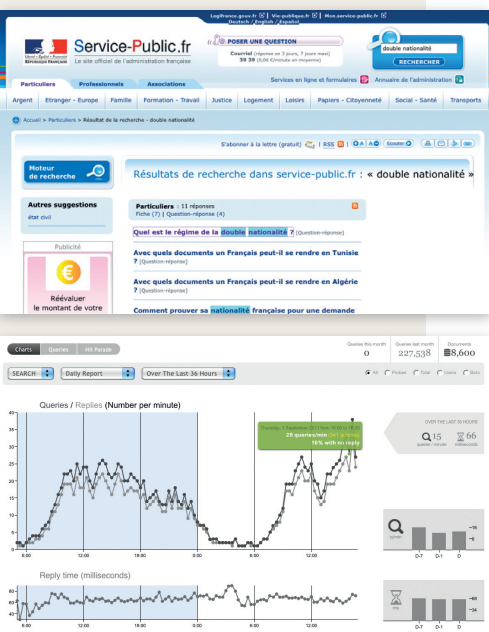


Antidot Finder Suite is the most comprehensive and the most gifted semantic search engine on the market. With its approach based on distributed software agents, AFS offers foremost relevance and performance. Its unique features allow you to create search and information navigation applications that include your business rules for an unmatched efficiency:

- User satisfaction thanks to search accuracy and fluidity of navigation.
- An industrial solution offering rapid deployment, ease of management, performance and scalability.
- Security, compliance with standards and openness for seamless integration with your IS.
- Apps and mobile geolocation, semantic web and linked data ... AFS technological advance supports your creativity.

More information:

<http://www.antidot.net/en/products-services/software-products/antidot-finder-suite/>



About Antidot

Antidot is a software vendor providing information enrichment and access solutions that put data at the heart of the business: search engines, data curation and meshing, collaboration around information. Antidot has developed a strategic vision based on experience in various markets: corporate intranets, e-commerce, publishing, healthcare, institutional websites.

For more information, visit our website www.antidot.net.



Contact

LYON - Headquarters

58 avenue Debourg 69007 Lyon

Tel : +33(0) 4 72 76 03 80

Fax : +33(0) 4 72 76 38 02

PARIS

22 rue du Sergent Bauchat 75012 Paris

LAMBESC - R&D

29 avenue Jean Monnet 13410 Lambesc

Tel : +33 (0) 4 42 63 67 90

Fax : +33 (0) 4 42 28 61 03

www.antidot.net