The Center for Research Libraries

8/22/2006

Archive Profile

LexisNexis (a unit of Reed Elsevier)

By Victoria McCargar, M.A., MLIS, *for the Center for Research Libraries*

Business Operations

Overview

LexisNexis is one of the four major businesses comprised by Reed Elsevier, a global publisher and provider of information solutions with annual revenue of $8.94 billion (December 2005).[*] Other Reed Elsevier entities are Harcourt Education, Elsevier and Reed Business.

The LexisNexis unit, headquartered in Miamisburg, OH, provides business, legal, tax and regulatory information to professional, business and government customers internationally.

LexisNexis North America, which accounts for about 28% of Reed Elsevier's total annual revenue, serves primarily North American legal, regulatory and government markets by providing electronic access to state and federal case law and statutes, court documents and approximately 5 billion searchable documents from more than 35,000 sources, including news producers (newspapers, news magazines, transcripts). This suite of databases is made available to paying subscribers through LexisNexis' flagship online research portal, www.lexisnexis.com.

Central to LexisNexis' mission is its massive search and retrievable capability, based on rigorous indexing and classification operations that are proprietary and highly competitive. Technical infrastructure is centered on supporting one of the most comprehensive set of text databases in the world and the ability to deliver precise search results in a matter of a few seconds to any subscriber in virtually any locale.

Functional Analysis and Discussion of Business Activity

As a subsidiary of a global, publicly held corporation, LexisNexis' annual earnings and revenue are reported as part of Reed Elsevier's annual 20F filing with the U.S. Securities and Exchange Commission, but more granular financial statements for various database

[*] Reed Elsevier reports financial results in British pounds. Dollar figures for the company are calculated at $1.73/£, the exchange rate at the close of Reed Elsevier's fiscal year, Dec. 31, 2005.

products are not available. It is thus not possible to assess financial sustainability of news content exclusively. Where possible this profile addresses the news database sector, but for the most part treats LexisNexis as a whole.

However, LexisNexis' business model—including governance, service and access models, overall funding and especially technical infrastructure—warrants review because it applies to all database content, of which news is just a part. In other words, news content is more or less indistinguishable from legal and governmental databases when it comes to vetting the organization for its ability to maintain critical operations. However, the high standards maintained by LexisNexis to support legal, government and accounting practice trickle down to some extent to benefit news.

LexisNexis is not a preservation repository per se. It is a commercial provider of digital information. But if LexisNexis cannot maintain and deliver information completely and quickly, in the words of a LexisNexis spokesperson, "We are out of business."

In making its products available the company engages in accessioning, maintaining and displaying content—the basics of the OAIS framework. Moreover, a basic functional analysis of its operations reveals that its recordkeeping activities in many ways mimic those of a defined preservation archive. These are an outcome of parent Reed Elsevier's fiduciary responsibility for maintaining its corporate assets in a way that is transparent to investors and shareholders and minimizes risk. In this way, the for-profit model serves to sustain content for the long term: LexisNexis' news content dates back to 1979.

A recent project (2002-ongoing) to convert workflows to XML is also an encouraging development from a preservation standpoint, because it is converting LexisNexis content from complex, aging, proprietary markup to a modern, forward-looking standard. Although this was undertaken for business reasons not having to do with preservation, it should position LexisNexis for more secure data retention over time.

There are a number of areas where LexisNexis' news databases do not meet the criteria for sustainability in the RLG-NARA repository checklist. For example, at no point in the workflow (akin to the archival chain of custody) does LexisNexis assume formal responsibility for preserving a file. It is *licensing* content from producers, not *accepting their deposits.* It maintains content only as long as a license is in force, and will delete it from the databases when that ceases to be the case or as requested by the producer.

Nor is LexisNexis required to authenticate aggregated news data, i.e., to assure that the correct version has been transmitted, and that it is accurate and complete, is the responsibility of the individual producers. (This is not the case with LexisNexis' legal content, which is handled differently in cooperation with courts and government entities).

Each data transaction between LexisNexis and a producer (comprising the Submission Information Package) must be normalized to the aggregator's proprietary syntax, and the syntactical variability from thousands of producers is so large that it is inefficient and not cost effective for LexisNexis to verify the completeness and accuracy of every SIP and its

metadata at the file level. In other words, if a producer inadvertently omitted several paragraphs of an article before or during transmission of the SIP, it wouldn't be discovered by LexisNexis during routine ingestion.

A third area of divergence is in the nature of the electronic files themselves. Aggregated databases like LexisNexis contain some of the oldest born-digital content in the short history of digital preservation: 25 years and older. LexisNexis' success in sustaining content is dependent upon the use of proprietary systems and simple text input—a very narrow platform. They are not subject to the external pressures of format changes and versioning that plague preservation repositories with a richer array of file types. With simple text, considerations like file properties, data characteristics and user needs at the file level are flattened, whereas the user interfaces—highly refined, competitive products and sales tools—are all-important. User access to LexisNexis' interface requires a computer, a recent web browser, a paid subscription or credit card, and enough bandwidth to download a file.

Mission Statement

"LexisNexis is the indispensable partner to legal and professional customers for information-driven services and solutions."

History

LexisNexis' antecedents date to 1967 with the formation of Data Corp., which was acquired in 1968 by Mead Paper to form Mead Data Central. In 1973, Mead attained a technological landmark with the launch of Lexis, the creation of the first commercial search-and-retrieval full-text database to enable more efficient legal research.

Nexis, the news and business information service, was launched in 1979 to provide "archival" news and financial information. Initial content included *The Economist, U.S. News & World Report, Newsweek* and the *Washington Post.* Since that time, the service has grown to become, according to company documentation, "the largest news and business online information service, including comprehensive company, country, financial, demographic, market research and industry reports." Its databases house the international content of thousands of newspapers, magazines, trade journals, industry newsletters, tax and accounting information, financial data, public records, legislative records, and data on companies and their executives.

Mead Data Central was acquired in 1994 by Reed Elsevier, which renamed the unit LexisNexis to reflect its core business. LexisNexis continues to add content to its suite through acquisition of other, smaller aggregators and through strategic alliances with other companies. Current examples appear in the next section.

Recent History and Events

LexisNexis sold its LexisNexis Document Solutions business to Corporation Service Company in 2003. A month later, LexisNexis bought the public records business of Dolan Media. The new business was integrated within the LexisNexis Risk Management operations. In 2003 LexisNexis also purchased Applied Discovery, one of the major providers of electronic discovery services for law firms, large corporations and government agencies in the US. Factiva and LexisNexis Group announced in October 2004 that LexisNexis will provide Factiva's content collection to legal professionals at law firms worldwide.

–LexisNexis reported in April 2005 that 280,000 social security numbers, driver's license numbers and other personal information had been accessed unlawfully in proprietary LexisNexis databases by unauthorized users with access to valid log-ins and passwords. Although no infrastructure or data was compromised, the situation required the company to institute tighter permissions policies and to limit access to that type of information to authorized subscribers such as law enforcement, banks and human resources agencies. Two class actions suits were brought against LexisNexis in connection with the breach and one against Reed Elsevier. The Reed Elsevier suit was sent to mediation in November 2005; the others were in early stages at the time of Reed's 20F filing.

As a general practice, the Reed Elsevier companies have agreements with their data suppliers indemnifying the aggregators from damages arising from the licensed data.

–The landmark April 2001 U.S. Supreme Court decision in *Tasini v. the New York Times* forced the deletion from aggregated databases all identifiable articles and photographs created by freelance writers, and any other news content whose provenance was uncertain.

The decision effectively withdrew digital database content from the protection of the 1978 Copyright Act. For LexisNexis and others, this resulted in the deletion of whole sections from particular producers. LexisNexis attempted to provide notations in its databases where the deleted material had appeared previously, but there was no effort to account for every deleted file, because producers themselves lacked rights information that would allow them to differentiate between files. In the intervening years efforts have been made to restore some of the material where this information could be obtained.

Governance

LexisNexis operates 27 offices in North America, three in Latin America, 11 in Europe, 12 in the Asia-Pacific region, and one in Africa. The largest operations are in North America, with some specialty publications centered in England (non-news).

Andrew Prozes, a U.S.-based managing director of Reed Elsevier, is the global chief executive officer of the LexisNexis Group. Five other chief executive officers (with varying companion titles) oversee LexisNexis sub-units, including (1) U.S. corporate and federal markets (which encompasses the news databases), (2) U.S. legal markets, (3) risk

management, (4) Europe and (5) Asia Pacific. Four executives at the senior vice president level are in charge, respectively, of non-federal and non-corporate "small" law, electronic product development, global production, and technology. The executives and their domains are listed in Appendix I. A list of North American locations is available online at: http://www.lexisnexis.com/about/world/nonflash/northamerica.asp.

Reed Elsevier's major shareholders are institutional investors; the largest, with 9.08% equity, is Capital Group, a leading provider of mutual funds and 401k plans. The others with holdings greater than 3% (disclosed under U.K. securities law) are FMR Corp. (5.2%), Legal & General Group plc (3.9%), Prudential (3.12%) and Barclays (3.02%).

Funding

Parent Reed Elsevier is the umbrella entity for two companies, London-based Reed Elsevier plc and Amsterdam-based Reed Elsevier NV. Reed Elsevier had combined net profit of $647 million in 2005 on revenues of $8.94billion. That compares with combined income of $723 million on sales of $8.32 billion in 2004.

The LexisNexis unit separately enjoyed a 13% increase in revenue from 2004 to 2005, with 2005 sales of $2.54 billion compared to $2.23 billion the prior fiscal year ended Dec. 31. Operating profit for the unit was up 16% year to year, to $377.14 million in 2005 from $325.24 million in 2004. Of total revenue for LexisNexis in 2005, three-fourths, or $1.89 billion, is generated in North America.

LexisNexis' operating margin stood at 23.1% in 2005, a 0.9-point increase from 2004.

Nearly two-thirds of LexisNexis' sales are from subscriptions to its databases. A new area of potential revenue growth, according to the company, is LexisNexis' data hosting and backup services offered to third-party companies. Current customers include the Delaware Secretary of State and Dun & Bradstreet Business Reports. (Further revenue detail by sector is not available.) North American revenue sources break down as follows:

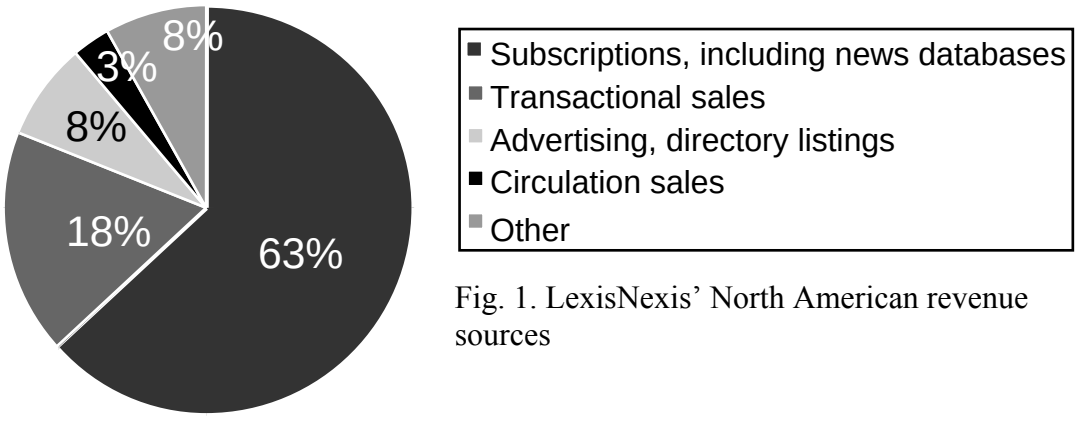

Fig. 1. LexisNexis' North American revenue sources

North American growth in revenue was 15% in 2005 (before acquisitions and sales of units). The growth was attributed to increased demand from law firms for online information (5%). Corporate and federal markets turned in growth of 8% on renewed strength in online news and business, higher volumes of information provided to the U.S. patent and trademark office and increased demand in risk management. Growth due overseas to demand for online information and from new publishing, notably in Europe and Africa, fueled a revenue increase of 7%.

Reed Elsevier reported that total capital expenditures (the benchmark for assessing investment in plant and equipment, including technology) in 2005 were $358.11 million compared with $373.68 million in 2004. The portion allotted for LexisNexis was $164.35 million in 2005, up slightly from $160.89 the prior year.

Risk Factors

As part of its financial reporting obligations under U.S. securities law, LexisNexis' corporate parent, Reed Elsevier, is required to enumerate all identifiable risks that may affect its bottom line. There are several having to do with currency fluctuations, changes to regulations and tax laws and the like that are germane to any multinational corporation. There are others, however, that are relevant to the sustainability of its digital assets, including news databases, insofar as they may limit capital expenditures for upgrades, migration, new products, metadata development and the like:

Risk factor	***Comment***
The competitive environment	*Subject to rapid change, unpredictable; technological change renders products wholly or partially obsolete. Challenges may arise from new technologies and delivery platforms.*
Continued demand for products and services	*Customer may no longer accept products at the prices charged.*
Breaches of security systems or other unauthorized access to databases	*Fraudulent access, already encountered once and corrected, is a constant threat due to the nature of some content in the databases. Methods of hackers change. Exposure is to loss as well as litigation.*

Risk factor	***Comment***
Changes to funding sources such as government and academic subscribers	*Budget considerations in these sectors may adversely affect revenue.*
Intellectual property rights may not be entirely protected	*Reed Elsevier may be subject to challenges, invalidation, or proscription of proprietary rights across different jurisdictions, e.g. the U.S. and E.U. Within the U.S. materials have been withdrawn from the database because of copyright.*

Risk in the competitive environment is of special interest because of the dominance of free search engines like Google, Yahoo and MSN. Reed Elsevier and by extension LexisNexis are threatened by the increasing availability of free information on the World Wide Web. Within the news domain, LexisNexis faces a challenge from publications providing broader access to their own information products at prices lower than what LexisNexis charges subscribers. For now, LexisNexis offers enhanced access through its proprietary search engine, highly developed search terms and metadata management—"added value" that is lacking on individual newspapers' web sites and their own online fee-based searchable archives. The outlook in this arena depends on whether newspapers work to add their own value to their online archives beyond what is currently the norm, and whether current subscribers, in the interest of saving money, are willing to sacrifice the ability to do refined searches in specialized domains.

Newspapers themselves are under increasing pressure from the Google model, which is siphoning off advertising revenue and frequently blamed for the flat to lower earnings in the industry. This could affect LexisNexis adversely by (1) forcing cash-conscious news libraries to reduce or drop their subscriptions, or (2) leading major "premium" newspapers to enhance their own revenue by charging LexisNexis more for licensing their content.

Cumulatively, the risk to revenue translates into reduced working capital for expanding technical infrastructure and developing new markets.

Service and Access Model

Access is through paid subscriptions, either institutional to universities, schools, corporations, law firms and government, or on an individual basis through per-document fees.

Users access a proprietary search interface, either from a web browser or via LexisNexis' proprietary telecommunications network. Power users or trained searchers may access LexisNexis' structured search capability (allowing Boolean searching and proprietary search commands). Others may use natural language searching, the LexisNexis Freestyle™interface, which employs associative retrieval to permit queries in plain English without search logic, including posing the search in the form of a question (this function predates Ask Jeeves by about four years). The newly renovated Nexis™ product is aimed at users seeking access to news titles.

The subscriber models include site licenses, usage rates and simple fee per download, known as LexisNexis a la Carte™.

Documents are delivered via web browser or network, as well as pre-coordinated collections on CD-ROM or via hard copy prints.

LexisNexis does not disclose fees paid to news producers because they are competitive. They vary based on the size and stature of the publication, with bonuses and incentives for timely or early data delivery, for example.

Content Characteristics and Ingest

In its news databases, LexisNexis takes a basic text file and performs a series of enhancements that are central to storage and retrieval, in the process normalizing the files to a LexisNexis proprietary metadata schema in XML. Thus certain characteristics of the original content, such as images, layout, typography, etc., are not preserved in the LexisNexis database.

Besides the comprehensiveness of its data collections, the attraction of LexisNexis to its subscribers is its richly indexed and classified content, enabling precision searching and automated delivery of relevant content to niche markets. At the center of this model to this is how data is processed between ingest and storage (SIP to AIP in the OAIS framework).

At ingest, LexisNexis processes large batches of documents in aggregations of ten thousand files or more.

The initial stage of processing (Data Fabrication) involves the highly proprietary and largely automated indexing and classification of documents using its SmartIndexing™ technology. Over its 25-year history, LexisNexis has developed and refined layers of term identification and weighting algorithms that add search terms to all types of content, including news.

Using statistical text analysis and natural language processing, SmartIndexing indexes tens of thousands of news, company and legislative documents on IBM mainframes daily. Text is processed according to sophisticated "rules" that are constantly monitored and refined by human indexers as well as tuning algorithms.

Words, phrases, names and locations that fit the suite of rules are automatically assigned search descriptors according to LexisNexis' proprietary controlled vocabulary. There are domain-specific vocabularies such as business and law in addition to the more generalized news vocabularies. Specialized term weighting algorithms enable LexisNexis and Reed Elsevier to expand markets by identifying niche information needs and tailoring delivery accordingly—tax-related legal filings updated in real time for accountants, for example.

Quality control occurs as a feedback loop between producer and aggregator, and rigor is determined by the nature of the content. In the case of news, a producer sends the edition files via FTP, which are uploaded and processed as a single batch process. LexisNexis systems and operators monitor the arrival and ingest of the batch, but they do not routinely perform file-level validation (for example, checking for missing stories within the batch or headlines absent from the rest of a text). Premium content (from the most prominent publications) are monitored more carefully than smaller or regional publications. Most news providers, as resources permit, monitor their own content on LexisNexis and work with the aggregator's technicians to resolve problems. Providers are required to give LexisNexis ample notice of changes to their systems that would impact data feeds, typically 60 days.

Stage two (Data Preparation and Update) comprises manipulation of this batched data and metadata submitted at ingest and application of structural markup to content from the licensors. Prior to its XML conversion project, LexisNexis used a homegrown markup known as Variable Input String Format, which dates to 1973.

VISF was difficult to parse into HTML and publication-quality print, and was increasingly cumbersome in the context of parsing thousands of news sources in disparate formats and nonstandard, changeable text structures. The XML concept of document type definition (DTD) allowed LexisNexis a more centralized way of controlling tagsets across different domains, including the all-important indexing operations and to more easily convert files from sources in established or even unknown formats.

The basic structure and metadata associated with a simple text file is as follows:

Attribute groups	***Main elements***
Text structure	*Headline*
	Byline
	Lead (first few paragraphs)
	Body of story
Geographic	*Dateline*
	Country
	State
	City

Classification and indexing	*Named organization*
	Named person(s)
	Subject(s)
Rights statement	*Copyright*
Bibliographic	*Publisher*
	Date
	Section
	Highlight
Related material	*Graphic*
Source information	*Publishing company/organization, industry, product, contact, city, state, country, region*

The final stage of processing (Online Delivery Infrastructure) consists of the systems used to store the XML, search and retrieve the data, and deliver it to customers in a format relevant to how they are accessing the database. LexisNexis added modular stylesheet creation in XML that enabled it to deliver documents via the web, internal telecommunications network or in print by parsing the XML code through the appropriate stylesheet. (The field-to-field data map appears in Appendix II.) Rather than attempt to convert its vast legacy databases to XML, LexisNexis delivers "on the fly" conversion to XML at retrieval time.

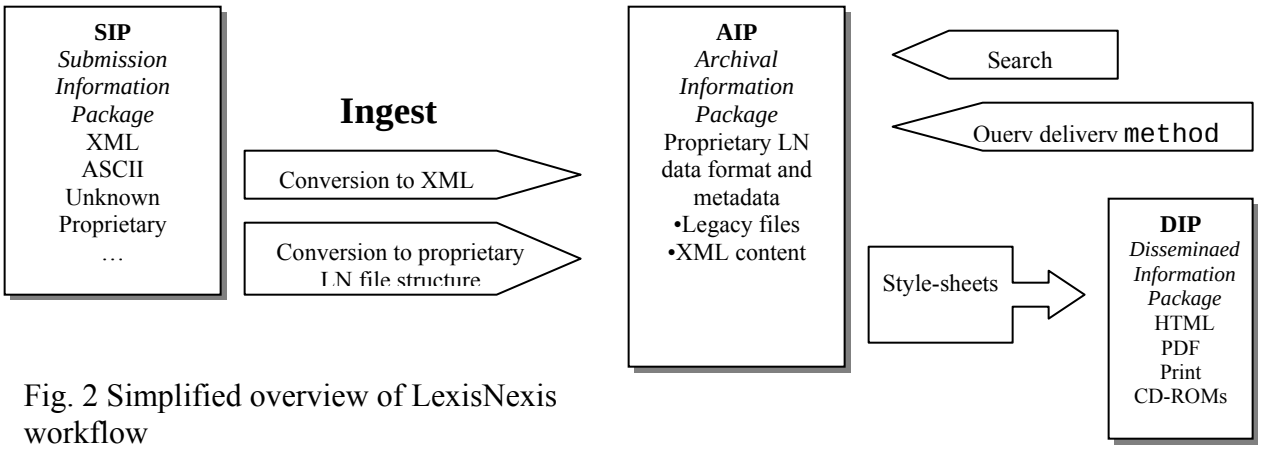

Fig. 2 Simplified overview of LexisNexis workflow

LexisNexis' 35,000-plus sources reside on the same technical platform. While the various user interfaces permit searching of large aggregations of content (either by selecting as many as 50 different sources in a combined search, or submitting queries across numbers of group files containing many documents), there is no interface to allow a search across all 35,000-plus sources in all the domains.

Certain licensors restrict the manner in which their content can be combined with other sources. LexisNexis works to identify ways to aggregate content that will be most useful

to its sectors of users but allows a great deal of flexibility within and across certain domains to optimize recall in a well-structured search.

Technical Architecture and Scale

The scale of LexisNexis dwarfs the normative preservation repositories by a factor of magnitude. The ingest rate is on the order of as many as 18 million documents per week. The systems handles about 700,000 searches per day against a database of approximately 5 billion files. At peak times—during a breaking news event, for example—searches may peak at 1.7 million per day. LexisNexis technicians, in fact, monitor news during the day and add computing resources if necessary to keep up with spikes in demand. The average retrieval time is five seconds. Such computing demands can only be met in a mainframe environment.

The LexisNexis Data Center, located in Miamisburg, Ohio, is one of the largest of its kind in the United States—the size of two football fields. The complex consists of networked servers, software and telecommunications equipment supporting LexisNexis operations, as well as some of the Reed Elsevier business units, including online services, data hosting and backup. It also houses operations in application development, testing, certification and administrative services. The facility is used to host data and backup services for third-party entities as well,.

In 2004, LexisNexis opened a remote data center and development facility in Springfield, OH, about 60 miles away from its Miamisburg site. Each facility serves as a high-access backup and disaster recovery facility for the other.

The Miamisburg location houses 10 large mainframe servers that run 22 multiple virtual server operating systems. In addition, more than 300 mid-range Unix servers and nearly 1,000 multi-processor NT servers support computing resources for LexisNexis worldwide. The networked environment cumulatively provides access to almost 200 terabytes of data capacity, 32 terabytes of which are dedicated to searchable content.

The Springfield center operates over 100 servers and manages more than 100 terabytes of data storage.

LexisNexis' telecommunications network provides different access options to subscribers, including the Internet, private telecommunications services and other online network services. Subscribers number about 2.6 million with either site licenses or on a per-document/per search fee basis.

Linking the telecommunications and computing sections are 18 front-end processors. These operate to provide the capacity to support tens or hundreds of thousands of simultaneous user sessions for the LexisNexis online services, with additional growth capacity.

Multiple levels of redundancy are built into the data center hardware, software and electrical/mechanical systems to provide uninterrupted service in the event that a single component fails. Backup tapes of all databases are also stored at other offsite locations. In the event of a full-scale failure of both the Miamisburg and Springfield data centers, LexisNexis estimates that it could be 80% online within 24 hours using the backup data. The systems are tested and maintained routinely to ensure that they perform properly in case of an emergency. In addition, all critical data is copied and stored off-site and emergency business-resumption plans are tested several times a year.

The data center and telecommunications staffs provide system management and control for the entire complex, using third-party automation software. System management services are provided 365 days a year, 24 hours a day. Additional third-party specialists are also available on site or on call. Architecture is in place to prevent worms, viruses, and hacking attempts. Additional security studies are conducted regularly by third party contractors.

Users (Designated Communities)

As noted above. LexisNexis serves the professional, legal, government and academic communities, mainly in North America (English-speaking) but is expanding rapidly overseas. Some of its databases are used for security and risk management, such as background checks. It is a very broad community of users, but by their nature (law, compliance, finance and personal information), LexisNexis' markets require a high level of data reliability. The company's web site identifies its market segments are as follows:

Legal Sector

- Law firm attorneys
- Corporate counsel
- Electronic discovery
- Electronic document filing
- Information professionals
- Legal research and paralegals
- Solo attorneys
- Law practice management

Public Sector

- Judges and judicial staff
- Law enforcement
- Prisons/corrections institutions
- Revenue and tax agencies
- Government libraries
- Executive and Legislative branches

Risk management industry

- Credit, financial, loan collections
- Fraud management
- Hiring and human resources
- Identify authentication
- Insurance claims
- Risk management
- Special investigators

Academic

- Colleges and universities
- Secondary schools
- Law schools
- Paralegal programs
- Criminal justice programs

Business and private sector

- Accountants
- Business and financial analysis
- Consultants
- Corporate libraries
- Marketing and public relations
- Journalists

News databases in particular are huge resources where the ability to make refined searches across manifold sources is of primary importance. This domain breaks generally into two camps: trained searchers (such as librarians) who use the full (and complex) capabilities of LexisNexis' search interface, and the less-trained and lay users (journalists, students, historians) who use natural language searching to fulfill a specific information need.

LexisNexis structures and maintains its data for the first group and has developed its FreeStyle interface for the second, which employs data structure and metadata from the first in a more intuitive, simpler interface.

For business reasons, LexisNexis is constantly identifying niche groups for whom it can tailor its data delivery (tax law updates to accountants) in the form of pre-aggregated content and pre-structured searches. The LexisNexis interface allows for easy refining of search criteria and filters, which a user or group of users can customize and store for repetitive searching.

Scale / Statistical summary

LexisNexis is one of the largest aggregators of text data in the world.

Searchable database in terabytes	≥ 32 TB
Capacity	200 TB
Number of files	5 billion

Ingest rate	~18 million documents per week
Subscribers	2.6 million
Searches per year (min.)	175 million

Appendix I

Governance

Corporate oversight of the LexisNexis unit is conducted by the following individuals:

Officer	*Title(s)*	*Business Unit*
Andrew Prozes	*Director, Reed Elsevier plc. Global Chief Executive Officer, LexisNexis Group*	*REED ELSEVIER plc*
Kurt P. Sanford	*President and Chief Executive Officer, U.S. Corporate and Federal Markets*	*LexisNexis Group*
Mike Walsh	*President and Chief Executive Officer, U.S. Legal Markets (USLM)*	*LexisNexis Group*
Ann C. Fullenkamp	*Senior Vice President, U.S. Small Law, State & Local Government Markets, U.S. Legal Markets (USLM)*	*LexisNexis Group*
Judy Vezmar	*Chief Executive Officer*	*LexisNexis Group Europe*
Miyuki Suzuki	*President and CEO*	*LexisNexis Asia Pacific*
James M. Peck	*Chief Executive Officer, Risk Management*	*LexisNexis Group*
Allan D. McLaughlin	*Senior Vice President and Chief Technology Officer*	*LexisNexis Group*
—Dave Diemunsch	*Senior Director, Data Center*	*LexisNexis Group*
—Stephanie Singer	*Senior Director, Data Center*	*LexisNexis Group*
Jeff Whittle	*Senior Vice President, Global Electronic Product Development*	*LexisNexis Group*
Ed Gould	*Senior Vice President & General Manager, Global Production*	*LexisNexis Group*

Appendix II

Parent company information

Head Office	*Reed Elsevier plc* *1-3 Strand* *London WC2N 5JR* *United Kingdom*
Description	*Reed Elsevier is a leading global publisher of information and solutions for professional users operating in four core subject areas: science and medicine, law, education and business. The company's products and services are categorized under the following different segments:* *Scientific and medical journals* *Magazines* *Directories* *Business journals* *Legal journals* *Information publications* *Education publications* *Exhibitions* *Reference works* *Books* *CD-ROMs* *Websites and portals*
Phone	*+44 20 7930 7077*
Fax	*+44 20 7930 7077*
Web address	*http://www.reed-elsevier.com*
Financial year end	*December 31*
Employees	*35,600*
SIC	*SIC 2721 Periodicals: Publishing, or Publishing and Printing* *SIC 7375 Information Retrieval Services*
NAICS	*51112, 514191*
New York Ticker	*RUK*
London Ticker	*REL*
LexisNexis Head office	*9443 Springboro Pike* *Miamisburg, OH 45342*
Phone	*+1 937 865 6800*
Web address	*http:// www.lexisnexis.com*

Appendix III

Content characteristics, metadata mapping example

Here is an example of how LexisNexis moved from a proprietary non-intuitive field set (VISF) to XML. The before-and-after example is LexisNexis' indexing control file for news. Data includes structural calls, content calls (Headline, Body, Highlight), some bibliographic data (Section, Publication, Date, etc.), a Copyright declaration and Provider identification information.

#BASES:	#BASES:
#BASE=P801	#BASE=V216
#SEGMENTS:	#ELEMENTS:
#DATE=20	#DATE=DATE0
#HEADLINE=60	#HEADLINE=HEADLINE
#POUND=89	#POUND=SPEC-LIB
#LEAD=119	#LEAD=REAL-LEAD
#BODY=120	#BODY=BODY-1
#CO=160	#CO=LN-CO
#TS=166	#TS=LN-TS
#COUNTRY=188	#COUNTRY=LN-COUNTRY
#ST=185	#ST=LN-ST
#CITY=182	#CITY=LN-CITY
#ORG=163	#ORG=LN-ORG
#PEO=179	#PEO=LN-PERSON
#SUBJ=155	#SUBJ=LN-SUBJ
#IND=171	#IND=LN-IND
#AUDIT=215	#AUDIT=_AUDIT
#PUBSUB=156	#PUBSUB=PUBSUBJECT
#COPYRI=01	#COPYRI=COPYRIGHT
#PUB=02	#PUB=PUB
#DATELINE=100	#DATELINE=DATELINE
#HIGHLT=105	#HIGHLT=HIGHLIGHT
#GRAPH=142	#GRAPH=GRAPHIC
#SECTION=30	#SECTION=SECTION
#BYLN=90	#BYLN=BYLINE
#VENDID2=161	#VENDID2=PUB-COMPANY
#VENDID3=164	#VENDID3=PUB-ORGANIZATION
#VENDID4=167	#VENDID4=PUB-TICKER
#VENDID5=172	#VENDID5=PUB-INDUSTRY
#VENDID6=177	#VENDID6=PUB-PRODUCT
#VENDID7=180	#VENDID7=PUBPERSON
#VENDID8=183	#VENDID8=PUB-CITY
#VENDID9=186	#VENDID9=PUB-STATE
#VENDID10=189	#VENDID10=PUB-COUNTRY
#VENDID11=191	#VENDID11=PUB-REGION
#VENDID12=55	#VENDID12=NAME
#VENDID13=151	

Sources

Callahan, Ron. "Challenges and Reward of Migrating an Electronic Publishing System to XML." Proceedings of XML Conference, Washington D.C., Nov. 17, 2004. http://www.idealliance.org/proceedings/xml04/authors/author19.html.

———— "Adding XML Capability to a Legacy Large-Scale Full-Text Indexing System." Proceedings of XML http://www.idealliance.org/papers/xml02/dx_xml02/papers/05-03-02/05-03-02.pdf

Datamonitor. "Reed Elsevier plc: Company profile" PDF: Reference code 1389, New York and other cities, Datamonitor USA, June 2005.

LexisNexis. Corporate web site, http://www.lexisnexis.com/about/.

Quint, Barbara. "*Tasini* Damage-Reporting Decisions: Today's vendor policy choices will affect customer relations tomorrow." *Information Today* 19(4), April 2002.

Reed Elsevier. Annual report, 20F Filing with the United States Securities and Exchange Commission, Dec. 31, 2005.

Spohr, Cindy, LexisNexis Senior Director, Librarian Relations Group. Personal interviews with author from Miamisburg, OH, April 28, May 2 and May 5, 2006.