LexisNexis
RISK SOLUTIONS

*Linking Liens and
Civil Judgments Data*

Confidently Assess Risk Using Public Records Data with Scalable Automated Linking Technology (SALT)

Table of Contents

Executive Summary

Public record lien and civil judgment data has been used by the financial services industry for decades. Liens and civil judgments help financial institutions assess the creditworthiness of consumers, and inform account receivable strategy to avoid unnecessary and unproductive lawsuits. Given the value and sensitivity of this content, it is constructive to understand how LexisNexis® Risk Solutions collects and links this data to provide the most current and reliable data available in the market.

This paper provides a description of the processes, procedures, and technology that enable LexisNexis to collect this data and reliably link it to the correct consumer, including the company's proprietary Scalable Automated Linking Technology (SALT), cutting-edge big data technology, and vast data resources.

Collecting Liens & Civil Judgments Public Records Data

The ability to provide solutions containing public lien and civil judgment data begins with dependable data collection. LexisNexis Risk Solutions has robust policies and procedures to maximize data reliability and ensure our records accurately reflect the underlying public records. We obtain lien and civil judgment information from government sources, which vary in their systems and policies regarding the availability of public record data. LexisNexis collects public records data from over 3,000 counties, boroughs, and parishes in the United States, representing over 98% of the U.S. population[1]. We utilize digital access methods where viable, in combination with a national network of collection vendors. We collect from each jurisdiction at a frequency of 90 days or less, with a majority of jurisdictions visited on a monthly cycle.

Also, beginning in July 2017, LexisNexis is implementing system upgrades to enable more robust collection performance management across hundreds of data points to discern sources of latency, including vendor collection performance and government record-reporting latency. LexisNexis is particularly focused on the collection of judgment and lien dispositions of paid, vacated or filed in error documents, and continues to invest in audit and collection methodologies to ensure confidence in our comprehensive and timely collection of this information.

The results of these systems and investments is a broad dataset of public records data, including liens and judgments, that is current, reliable, and available for linking to consumers through LexisNexis Risk Solutions' state-of-the-art linking technology.

Extensive Data Coverage – The Basis for Reliable Linking

LexisNexis compiles the largest collection of U.S. consumer identity information available today. We leverage approximately 65 billion public and proprietary records, which are updated regularly. This rich dataset allows LexisNexis to understand the identity information for almost the entire U.S. consumer population (not just a subset of the credit active population), and provides the foundation for our ability to link consumers to public records accurately. This broad coverage enables us to assess the

---

1   Based on census data from July 2015.

uniqueness of different combinations of identity information and to evaluate whether a given record belongs to a given consumer.

Superior Linking Technology

LexisNexis utilizes a proprietary and patented data linking approach called Scalable Automated Linking Technology (SALT) to draw upon our extensive data and turn disparate information into meaningful insights. SALT allows the analysis of large data sets more easily, reliably, and efficiently due to the High Performance Computing Cluster (HPCC) Big Data technology platform. HPCC provides the processing power necessary to run billions of complex statistical analyses and data comparisons. These statistical clustering algorithms run iteratively, learning through repetitive analyses and as incremental data is updated. Core to the linking algorithms is "specificity." Specificity is the measure of uniqueness assigned to each value of data, in each field of the database. Every relevant identity element and combination thereof is assigned a specificity; this allows the system to identify when records have sufficient evidence to confidently match. The automated analysis of data, measuring specificity, informs the matching function for each unique pair of records.

The outcome of the SALT linking algorithms is the assignment of our patented LexID® to each identity, a unique identification number for individuals that is a reliable and secure indicator of an individual. Once created, the LexID is assigned to the set of records associated with each unique individual. LexisNexis performs ongoing analysis of the results of this linking system, periodically auditing a significant sample of the linked identities. Additionally, with each iteration, a sample of newly-linked records is analyzed by a group of engineers and engineering managers with linking expertise. These procedures are fine-tuned over time, as product teams, quality assurance teams, consumers, and customers provide feedback.

Patented LexisNexis Linking Technology Leverages a Unique Identifier

Creating a reliable view of an individual or business.

When we acquire a new record, LexisNexis assigns our own unique identifier, LexID®. To ensure the security of the consumer identities, LexID is not derived from any personally identifiable information, such as a social security number or name.
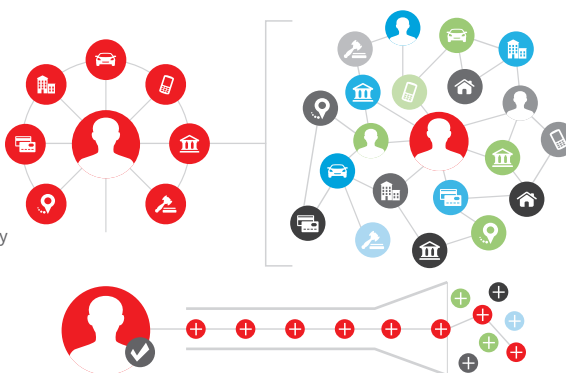
Records with a common LexID are linked together using LexisNexis® Scalable Automated Linking Technology, a proprietary and patented method of linking and clustering data. Identity profiles are continuously updated to ingest new data sets and records.

Having this persistent link across multiple touchpoints and data silos eliminates false positives and builds an extremely comprehensive and accurate representation of the identities that matter to our customers.

Figure 1: *LexisNexis Linking Technology. Our high-precision data linking technology drives linking effectiveness by building extremely comprehensive and precise profiles of consumer identities.*

Leveraging multiple sources and linking these together using proprietary and patented linking technology, LexID delivers a more complete view of individuals, especially those with multiple names and addresses. By combining data integrity with linking algorithms, we make connections and identify relationships with high reliability.

This dynamic process results in a consumer profile that accounts for identity changes over time, since LexisNexis continually updates profiles with new records. By aggregating, cross-linking, and analyzing large volumes of identity data, LexisNexis is able to build complete identity portraits that connect otherwise fragmented information and disparate touchpoints across time. Even when there are misspellings, typographical errors, or other errors in the originating data sources, LexisNexis can associate records with the same individual. We are also able to recognize when someone may have presented himself in different ways over time (i.e., "Alexander Jonathan Marks" vs. "A.J. Marks," "Alex Marks," or "John Marks").

65+ billion
public records distilled down to...

10+ billion
unique name/adress combinations, which
we cross-link, refine, and map to isolate…

300+ million
unique identities, which are further
enhanced to remove deceased and
inactive identities, resulting in…

…a precise view of
279+ million
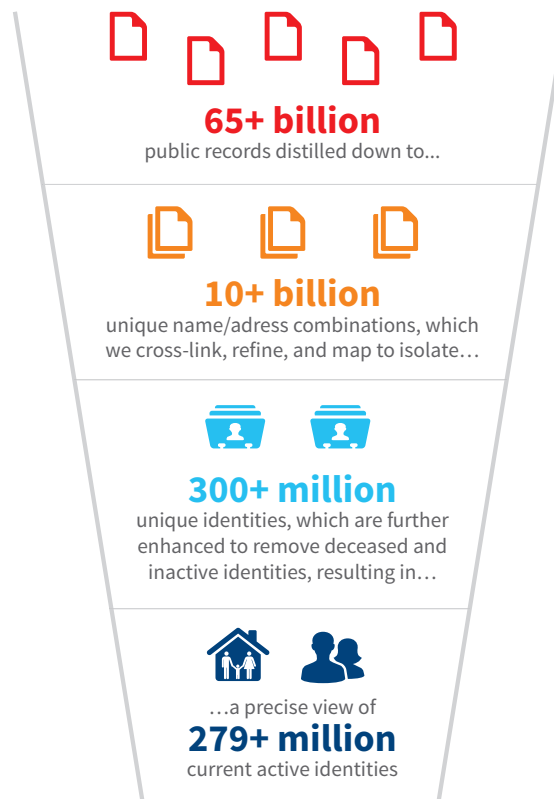current active identities

Figure 2: *Distillation of address data. LexisNexis turns massive raw data into precise, cross-linked, and actionable results.*

Figure 2 graphically depicts what LexisNexis linking technology means for end users. The breadth of LexisNexis public records coverage is demonstrated by the more than 10 billion unique name/address combinations contained within our databases, which

we have cross-linked and mapped to more than 300 million identities – and to which each has been assigned a unique LexID. This linking process affords insight into more than 279 million currently active identities.

Traditional Linking vs. SALT-based Linking

Traditional, deterministic record linkage — also known as data matching — generally uses "if/then" logic to assign a series of confidence factors that indicate the degree to which the content in each field can be considered the same as, similar to, or different from content in an existing consumer file. When this process finds a record that meets the match rules, it links the record. As such, attempts to improve linking accuracy rely in large part on increasing the number of "minimum" fields in source data records in order to establish suitable confidence in the match. In contrast, SALT uses advanced concepts such as term specificity to determine the relevance/weight of a particular field in the scope of the linking process, and proprietary computer algorithms based on the input data rather than the need for hand-coded user rules, which is key to overall linking effectiveness. SALT routines also measure the statistical "strength" of a match, allowing LexisNexis to require strong matches before linking records to existing consumer files. Compared to traditional "if/then" logic, SALT-based linking allows for consideration of a greater number of factors across more dimensions, which increases accuracy and reliability in a statistically significant manner.

The combination of the ability to evaluate the "uniqueness" of a set of input values combined with the logic to prevent a link that is not sufficiently strong allows our linking technology to overcome many of the challenges associated with linking public record data. This includes the scenario when two people with the same name but a different name suffix live at the same address (what is commonly referred to as the "Junior/Senior scenario"). In such cases, our linking technology can often distinguish between such individuals, but in the rare instance when there is not enough data to accurately assign ownership of a record to the right individual, we will not make the link. The end result of the SALT-based approach is a linking technology that uses analytics and machine learning to avoid both false positive and false negative results, and to deliver highly reliable linking outcomes without imposing rigid, hand-coded rules.

Effect of SSN in Linking Lien & Civil Judgment Records

One of the critical aspects of linking liens and judgments records is the ability to match records to consumers reliably without the presence of a Social Security Number (SSN) on the public record. This is critical because approximately 50% of tax lien records and approximately 96% of civil judgment records do not contain a SSN. In order to verify the reliability of our linking technology as it relates to the data available within liens and civil judgments, LexisNexis conducted a study to test the consistency of linking results in the absence of SSN. For this study, a random sample of approximately 26M lien and civil judgment records containing a full SSN or partial SSN (i.e. containing

only the last four digits) was processed twice through our linking processes. In one round of processing, the SSN values in the records were available for use in linking. In the other round of processing, the SSN values on the records were suppressed from the data and not available for use in the linking. Following the two rounds of linking, the results were investigated to determine, for the set of records that were matched to a LexID in both rounds of linking, how often a different LexID was assigned when the SSN was suppressed compared to the LexID assigned when the SSN was not suppressed. As shown in the table below, the results of the study indicate that the presence of a SSN on the source record led to the assignment of a different LexID approximately 0.09% and 0.10% of the time in civil judgments and liens, respectively. We would note that the focus of this study was on the consistency of our linking with and without a SSN, and the different linking results were not reviewed in detail for accuracy – the issue of linking accuracy will be discussed in the next section. However, by examining the differences in LexID assignment with and without a SSN on the record, we can demonstrate the high degree of consistency and robustness of our linking technology.

Public Record Type	Different LexID match with the SSN suppressed (%)
LIENS	0.10%
CIVIL JUDGMENTS	0.09%

Table 1. *Linking study with SSN suppression.*

Assessing Reliability of Liens & Judgments Linking

LexisNexis also conducted a separate study to assess SALT links within lien and civil judgment data. A team of engineers with linking expertise manually reviewed a sample of 1000 liens and judgments records (500 randomly sampled liens and 500 randomly sampled judgments). In the manual review process, the linking engineers reviewed the public record, the assigned LexID, and the other closest LexID matches (e.g. the second and third closest candidates) to confirm that the selected LexID was the appropriate match for the record. Each record being reviewed was labeled "Correct" where the assigned LexID was the appropriate identity for the record, "Incorrect" when the assigned LexID was not appropriate for the record, and "Maybe" when the validity of the match was unclear or ambiguous. Our experience has shown that in the "Maybe" cases, where it is still unclear if the match is correct after deep review, it is conservative to estimate that half of the "Maybe" results are incorrect. Given this assumption, the final accuracy of the sample can be estimated as the sum of the "Correct" results and half of the "Maybe" results divided by the total number of records included in the review. Based on this approach, the linking accuracy for liens and judgments can be estimated at 99.75%, 99% CI [98.90%, 99.94%] (see Table 2).

Match Quality	Lien & Judgment Records
CORRECT	995
MAYBE	5
INCORRECT	0
TOTAL	1000
SAMPLE ACCURACY	99.75%

Table 2. *Results from manual accuracy study of liens and judgments linking.*

Conclusion

LexisNexis Risk Solutions leverages proprietary linking algorithms, cutting-edge big data technology, and vast data resources to reliably link lien and civil judgment public record content to the correct consumer files. This enables our customers to continue to rely on this valuable content to make informed and confident fact-based decisions.

Glossary

Consumer file: A collection of records associated with a given consumer.

Linking: The process of matching a record or inquiry to a unique consumer identity.

Record: A discrete unit of data that has defined values within fields. Fields are the critical data elements used for record linkage.

Specificity: A measure of uniqueness assigned to each value in each field of a database.

For More Information
Call 800.715.0959 or visit lexisnexis.com/CreditRisk.

LexisNexis®
RISK SOLUTIONS