APRIORI

https://www.aprioridata.com

info@aprioridata.com

+1 (202) 779-7000

About Us

Who We Are

Apriori is an international data aggregator with special expertise in finding, acquiring, and processing government issued ID data from many countries and languages. We acquire ID and other large sets of data from verified and reliable sources and make this data suitable for use by government agencies and leading companies worldwide.

We acquire, process, and enhance close to 2 billion records internationally every year in dozens of languages and hundreds of formats. Apriori supplies data for a wide variety of legal purposes, such as age and identity verification for banking, E-commerce, and government functions.

What's Special About Us

Apriori uses the latest and greatest of new technologies combined with time-proven classic technologies to produce world-class data products. We bring together the best experts using the best available software to deliver modern data and software solutions for clients in many industries.

By leveraging our team's many years of hands-on experience and breadth of knowledge in best-of-breed data toolkits, we're able to address today's most complex data-related challenges quickly and efficiently.

We use a precise mix of computer technology with geographic, cultural and linguistic rules to accurately understand, parse, and enhance raw data into finished products that deliver exceptional value to your organization. One of our key skills is acquiring hard-to-get data that is only available in non-Roman character sets, such as Cyrillic or India's Devanagari scripts.

Why Choose Us?

There are many things that set Apriori apart from the competition. Here are just a few.

- We understand your need to automatically acquire, process and analyze billions of records in a short amount of time.
- We consistently deliver accurate and reliable results.
- In the past 15 years, our team has converted many billions of records from hundreds of sources and formats.
- We routinely write complicated software code for data extractions, processing, and task automation.
- We are experts at custom software development exactly suited to your critical needs.

What Technologies We Use

Apriori combines older, proven technologies with modern, cutting-edge tools to address all aspects of data processing. In addition, we've written various tools from scratch to help us achieve specific tasks. Our team is as comfortable writing complicated C# or Java code for Hadoop/Pig jobs on AWS as it is with traditional ETL tools. We always use the best tools for the task at hand.

When dealing with billions of records as we do, the key to success is using the most efficient methods in existence to clean and parse the records, assign geographical coordinates, and index the data for further searches. We're comfortable in both Microsoft and Linux environments, and we do whatever it takes to reach your goals with utmost reliability and efficiency.

https://www.aprioridata.com

info@aprioridata.com

+1 (202) 779-7000

Our Services

Through many years spent finding, acquiring, and processing data from many countries and in many languages, we've gained vast experience at providing usable data to government agencies and companies worldwide. Our data services include bulk licensing of standard data packages and the acquisition of unique data specifically for your business needs.

Bulk Data Licensing

Apriori offers international data to many clients in bulk form for in-house use in various legally allowed activities, such as age and identity verification. We service a variety of industries, including finance, E-commerce, gaming, and government watch lists.

Other clients use our data in conjunction with their own API services and resell the data to their own customers on a per-transaction basis. Letting our experts focus on the nuts and bolts of data acquisition and preparation lets your experts focus on delivering quality products and value-added services to your customers.

Our data offerings are typically updated on a regular schedule and clients are notified via automatic messages when new data becomes available. You're assured of having access to the freshest possible data as quickly as possible.

An example of how data updates work is our recent acquisition of the India Voter Registration Roll. It contains over 1 billion records in 14 languages gathered from all 29 states and 7 Union Territories. The database is updated annually (twice a year for some jurisdictions) with new, cleansed entries and licensees are contacted immediately upon release of the revised database.

International Data Acquisition

We have many years of experience in acquiring national and international databases from a multitude of sources such as voter files, driver and motor vehicle records, citizenship rolls, and many others. Our focus is on acquiring proven, Government-Issued ID datasets that stand out in comparison to private data sources.

The Apriori team is responsible for acquisition of over 2 billion records each year from over one hundred countries. We acquire and process data in multiple languages using in-country experts to help us better understand, process, and enhance the data.

Whether you need data from West Bengal or from the suburbs of Moscow, we have the experience and capability to deliver it to you when you need it most.

Data Processing, Enhancement, and Transliteration

We are using proprietary algorithms to clean, standardize, and enhance customers' data. Our data enhancement tools allow adding and correcting crucial data elements in many languages and for many countries.
In addition to standardizing raw source data, we can significantly enhance the data by adding various data elements based on other data sources. We utilize Postal, Census, Geographical and other available public and government records to add value to your data.

We also offer transliteration services. To ensure the most accurate data rendition possible, we apply machine transliteration algorithms custom built using our extensive background in linguistics and the expertise of in-country language specialists who are native speakers. We can convert all non-English source data into Roman characters by applying our custom linguistic rules and language exceptions.

India Voter Registration Roll

Apriori collects data for our India Voter Registration Roll directly from the India Electoral Authority for each administrative division. We collect data from all of the States and Union Territories in India. The data consists of all registered voters aged 18 years or older and contains over 1 Billion records. We parse, cleanse and normalize the data fields and also add various data elements based on information from India Post. Our database is the most complete and accurate set of India voter data available anywhere.

The Best Database

We collect, process, standardize and enhance over one million individual files in different formats, layouts and languages to produce a centralized digital database of registered voters. The majority of the content is derived from the India Electoral Rolls with additions and crosschecks from India Post address data.

Voter registration is compulsory in India, which means that every citizen 18 years of age or older must register to vote. We collect voter data from the Electoral Authority in each of India's 29 States and 7 Union Territories.

To increase the value of our database, data is aggregated between consecutive years, which allows tracking of more than 1 Billion Indian citizens through address changes, name changes and other updates to voter information.

The Best Data

Data in the India Voter Registration Roll consists of all registered voters 18 years of age or older from all States and Union Territories. It contains Government Issued ID number (EPIC number), name, relation's name, address, age, gender, polling station name, postal code, and other administrative division fields, as well as geographic coordinates.

The EPIC number, or Electoral Photo Identity Card number, is a unique voter ID issued to each registered voter. EPIC numbers stay the same even when a voter changes location and re-registers in a new Constituency, which means that voter information is trackable over time and location.

During the 2014 election, there were 834,101,497 registered voters, including 437,051,538 males and 397,049,941 females. Because we aggregate the data between data releases, a complete history of each voter is provided, including past addresses and any name changes.

We compile data from Final or Draft Electoral Rolls as published by the corresponding Electoral Authority. Full addresses and postal codes are standardized and enhanced with additional government data from India Post. By combining official voter data with India Post data, we ensure accuracy and timeliness superior to that obtainable from the individual sources.

Made Even Better

India raw data comes in an immense variety of layouts and in multiple languages. We parse, cleanse and normalize data fields for easy consumption. For example, we separate voter's relative name and relationship status into mother's, father's or husband's name fields. We also derive an approximate year of birth from the voter's age.

In addition to standardizing raw source data, we significantly enhance the voter roll by adding various data elements based on India Post. We add administrative divisions, such as Taluk, Circle, and District name to the datasets. We also add geographic coordinates to each record based on India PIN (Postal Index Number) Code data. The PIN Code is a 6-digit code of Post Office numbering used by India Post.

Transliteration

The data is available in the language prevalent to a particular State or Territory. There are 14 languages in which the data is published, with the majority being in Hindi and English. If the data is available in multiple languages, we make the data available in all of them.

To ensure the most accurate data rendition possible, we apply machine transliteration algorithms custom built using our extensive background in linguistics and the expertise of in-country language specialists who are native speakers. We also convert all non-English source data into Roman characters by applying our custom linguistic rules and language exceptions.

Database Updates

The data is typically updated on an annual schedule as it becomes available from the State and Union Territories Electoral Authorities. We constantly monitor the data release schedule and acquire, process, and make the updated data available soon thereafter. Data updates are available as either a full database replacement or as a delta update containing changes and additions only. You are notified via automatic process that the newest version of the database is available for download.

Who Uses Our Data

Our database is used for many purposes. For example, a credit reporting agency uses Apriori data to verify their internal data against our cleansed and extended government issued database of registered voters. This cross-reference helps them provide accurate and up to date consumer credit information to their clients.

Another client is a well-known international data aggregator that uses our India data to provide an API solution to its customers in e-commerce and the finance industry on a per transaction basis for incorporation in due-diligence processes.

Information Security Policy

Information contained in India Voter Registration Roll is Public Data It is acquired from Government of India Electoral Authorities and India Post. This data is considered personal information under the IT Act and associated Policy Rules and therefore reasonable security practices and procedures must be applied by any company that gains access to the data through Apriori LLC.