

Rapid Data Profiling

GIVE BUSINESS TEAMS 'FIRST EYES ON DATA'

TECH BRIEF

Data profiling has historically resided in the domain of the technically-minded. However, it would be a mistake to think of data profiling as merely an IT exercise. As data is primarily generated and consumed by business people, data profiling should be considered a line of business practice. Additionally, the complexity of new sources where data content and quality are less known makes business people – who are most familiar with the context and intended state of the data – ideal candidates to profile and improve its quality.

WHAT IS PAXATA RAPID DATA PROFILING?

Paxata addresses the most time-consuming part of data quality projects by providing an intuitive, visual, and interactive application for business users to onboard, profile, and create quality information. To accelerate time to value and rapidly enable business teams, Paxata Rapid Data Profiling is offered in software as a service or a virtual private cloud.

Why Paxata Rapid Data Profiling?

Discover with clicks, not code

Profile the entire data, not samples

Automate data collection and validation

Auto-record profiling steps to create repeatability

Accelerate analytic or data lake projects

INSPECTS COMPLEX DATA FAST

Discover and profile SQL or NoSQL data, Excel, or complex files such as XML, JSON, Avro, and Parquet

Interactivity with large datasets eliminates need for sampling or waiting for batch jobs

INTELLIGENT AND AUTOMATED

Improves productivity with algorithmic intelligence

Leverages machine learning and improves through usage

Automated step recorder enables reuse and collaboration

BUILT FOR BUSINESS CONSUMERS

Visual and interactive experience

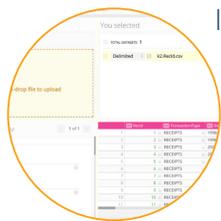
Enables non-technical users to harness the power of big data

Excel-like user interface requires minimal training

“Our team is specifically excited as it will enable us to quickly understand unknown file types and allow us to preview, validate the file content, and select the right fields across multiple data sources, all in the same step.”

NOAH KAYS, DIRECTOR OF CONTENT SUBSCRIPTIONS AT UNILOG

KEY CAPABILITIES



Intelligent Ingest

Quickly ingest data from a wide variety of enterprise sources including complex semi-structure files such as XML or JSON. Paxata intelligently detects your data source type and transforms it into a tabular format easy for point-and-click interaction and profiling.



One Click Profile

With one click, Paxata profiles your data set and generates a scorecard showing data type distribution, field completeness, field length analysis, top/bottom values, leading or trailing string patterns, min/max/range, special character analysis, custom calculations, and regex.



Inspect & Investigate

Paxata's interactive, Excel-like interface allows you to search, explore and discover trends, outliers, and patterns across your entire data set (not just samples). Validate your data visually and and get immediate feedback.



Curate, Remediate, and Repeat

Standardize similar values and misspellings, joins, appends, and overlaps across data sources with smart machine learning recommendations. Paxata self-documents your steps to create repeatability, auditing and governance.



Monitor with Visual Tools

Create, tag, catalog and share governed and trusted data sets with others. Monitor your data profiling and quality checks by publishing to visualization tools such as Tableau, Qlik, and PowerBI.

Paxata Adaptive Information Platform



Paxata Rapid Data Profiling is part of the Paxata Adaptive Information Platform. With powerful machine learning, an intuitive user experience, and a smart distributed architecture, Paxata Adaptive Information Platform offers business analysts the ability to create actionable information in minutes, not weeks or months.

ABOUT PAXATA

Companies around the globe rely on Paxata to get smart about information. Paxata is the pioneer that intelligently empowers all business consumers to transform raw data into ready information, instantly and automatically, with an enterprise-grade, self-service data preparation application and machine learning platform. Our Adaptive Information Platform weaves data into an information fabric from any source and any cloud to create trusted insights. Business consumers use clicks, not code to achieve results in minutes, not months. With Paxata, Be an Information Inspired Business.

Paxata is headquartered in Redwood City, California with offices in New York, Ohio, Washington D.C., and Singapore.