

360-VIEW OF A CONSUMER: DEDUPLICATING AND CONSTRUCTING CONSUMER DATA USING AN IDENTITY GRAPH



THE CHALLENGE

For the last 30 years, a large global digital marketing and technology firm has been collecting consumer data on adults in the United States from hundreds of primary sources to build consumer marketing databases. Because data comes from different sources, variations in data formats, typos, misspellings, missing data, and incorrect information make linking records for the same real world customer an arduous task. As a result, even after attempts to deduplicate customer data, these databases contain records for more than 2 billion distinct consumers (reflecting an estimated 240 million real world individuals). This makes it almost impossible to obtain a 360-view of a consumer, since data about an individual is split across multiple records and products.

With the goal to integrate the intelligence gathered from different data sources and products, Enterprise Knowledge worked with one of our technology partners for graph-based data catalog systems to engineer a solution that would successfully link records across data products that refer to the same individual.

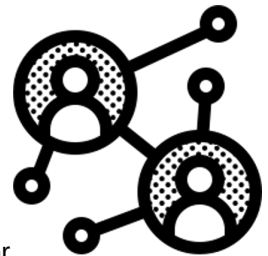


THE SOLUTION

In order to deduplicate the records and associate them to identifiable, unique individuals, EK's team of experts started from the bottom up, conducting an exhaustive exploratory data analysis of each attribute in the data. Using the completeness and uniqueness of data fields, we then prioritized fields for further exploration, allowing us to quickly deliver value by focusing on the data attributes with the highest relevance for matching individuals.

Working with our partner, we developed a data processing pipeline to take messy input data and create a graph of linked records that corresponded to real world individuals. Our pipeline focused on the following steps:

- **Data Cleaning and Normalization** - Our first step involved cleaning and standardizing data to maximize matching opportunities.
- **Graph Construction** - To maximize our potential for matching records, we constructed a knowledge graph, allowing us to match records through the graph based on shared attributes or intermediary nodes.
- **Rules-Based Matching Algorithms** - After looking for explicit matches, EK developed algorithms that would link records pertaining to the same individual.
- **Iterative Validation and Match Quality Improvement** - Every matching algorithm developed was documented, validated, and adjusted to ensure quality results and alignment with business stakeholders. Furthermore, working in an Agile manner, EK was able to both continually build on existing algorithms and develop new rules, increasing the quantity and quality of matches.



THE EK DIFFERENCE

EK's vast experience in knowledge graphs played a key role in delivering a transparent, explainable solution that outperformed the client's existing black box legacy systems. Using an Agile approach, EK was able to maintain alignment with both our partner and the marketing firm's business stakeholders, ensuring that we were able to deliver high value results quickly. In addition to building state-of-the-art graph models, conducting in depth data analysis, and writing detailed technical reports, our data scientists, graph engineers, and analysts collaborated to ensure that all technical terminology and decisions were documented in a business glossary that was accessible to non-technical users. In doing so, EK was able to leverage our knowledge sharing culture, facilitating discussion and collaboration with key stakeholders to ensure the end solution was a true made-to-measure system that solved our client's unique business needs.

360-VIEW OF A CONSUMER: DEDUPLICATING AND CONSTRUCTING CONSUMER DATA USING AN IDENTITY GRAPH

THE RESULTS

By implementing our data pipelines and matching algorithms on the knowledge graph, we managed to reduce the number of unique records by 70% percent, more closely aligning with the target population of 240 million marketable US adults. In doing so, we allowed our client to connect the dots between data that was previously siloed in separate systems, creating a clearer picture of customer behavior and trends. Through collaboration with our technology partner, we continue to fully automate the graph creation and deduplication process. This gives our client the ability to quickly ingest and connect new data, ensuring that the graph, and corresponding business intelligence, will continue to expand.

Enterprise Knowledge (EK) is a services firm that integrates Knowledge Management, Information Management, Information Technology, and Agile Approaches to deliver comprehensive solutions. Our mission is to form true partnerships with our clients, listening and collaborating to create tailored, practical, and results-oriented solutions that enable them to thrive and adapt to changing needs.

Our core services include strategy, design, and development of Knowledge and Information Management systems, with proven approaches for Data and Information Management, Knowledge Graph Implementation in support of NLP, ML, and AI initiatives, Taxonomy Design, Project Strategy and Road Mapping, Brand and Content Strategy, Change Management and Communication, and Agile Transformation and Facilitation. At the heart of these services, we always focus on working alongside our clients to understand their needs, ensuring we can provide practical and achievable solutions on an iterative, ongoing basis.