

# Author Data Management

Straive works with leading information providers, publishers, and aggregators to extract meaningful insights from unstructured content and data.

**600K**  
data points  
captured

**65K**  
unique author  
profiles created

Initial data built  
in less than  
**3 months**

**99.98%**  
accuracy  
achieved

## Business Case

A leading technical professional organization sought an automated solution for processing huge number of author data records from disparate sources and maintain a high quality author database.

One of the world's largest technical professional organization was looking for an automated solution for processing a large number of author data records from multiple sources while also maintaining a high-quality author database for its members.

Author name disambiguation is an important prerequisite for the author level analyses of publication data. For many reasons, identifying unique author entities is important for handling author-related queries and calculating bibliometric and scientometric measures for authors.

The organization selected Straive to deliver standardized and disambiguated author records using its author data management solution.



## The Challenge

The ambiguity of author names is a significant impediment to individual researcher analysis of large scientific publication databases. Within such databases, researchers are typically identified only by their surname and first name initials, as they appear in any given publication. However, hundreds of independent researchers frequently share the same surname and first name initials.

- Non-standardized data and formats
- Inconsistent information presentation
- Inaccurate information extraction

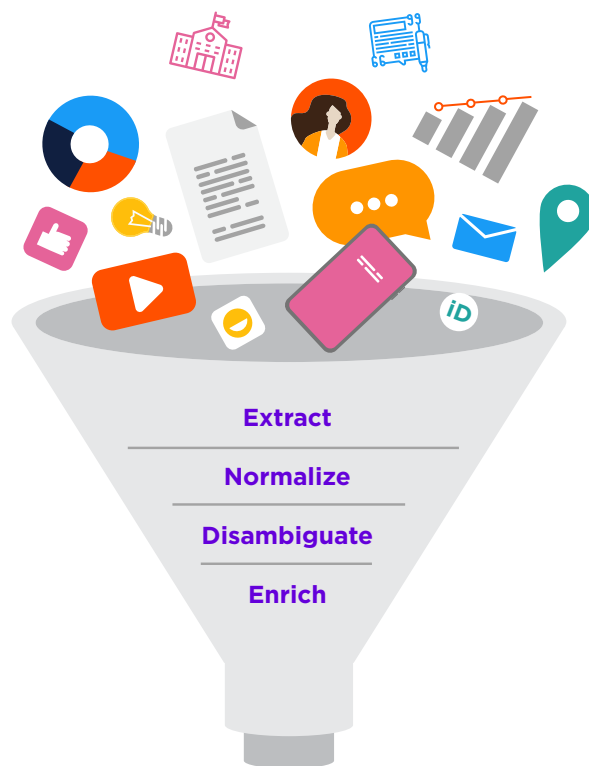
With the exponential growth of scholarly output and its digital dissemination, authors of such output seek increased visibility and recognition for their contributions. The lack of standardization in how their names are represented in various publications, as well as the ambiguity created by multiple authors sharing the same name, create a significant barrier to uniquely identifying an author and aggregating all of their publications.

The customer's primary challenge was that author data was not being captured consistently across all of their publications. Achieving author data clean-up and enhancement required dealing with unstructured data, inconsistent naming conventions, and data variants. Straive deployed its comprehensive author data management solution to extract information about authors from documents, standardize and disambiguate author names, and create a database of unique authors and their profiles.

## The Straive Solution

Straive leveraged its Straive data platform (SDP) to offer the technical professional organization a holistic solution for author data disambiguation.

Data analysts experienced in bibliographic work ensure accuracy and completeness of standardization and disambiguation. Automated algorithms, manual validation, and standardized repositories like ISO standards for countries and states, were applied for accurate standardization of affiliations. Author profiles were further enriched with ORCID IDs through secondary research.



**Output:**  
**Disambiguated Author Profile**

Customizable algorithms for author data disambiguation leverage available metadata about authors, such as email, co-authors, affiliations, areas of interest and more.

Auto-zoning tool for identifying author and affiliation information in the scientific literature



NLP & ML techniques for extracting and parsing author information in diverse formats

## About Straive

Straive is one of the largest and most diverse solution providers to the research content industry. We work with leading players in the industry including commercial publishers, university presses, open access publishers and associations, supporting journals, books, MRWs, and database products/platforms. Our full suite of solutions straddle from manuscript development, peer review management, publishing operations, discovery engine to delivery and customer engagement.

For more information, visit <https://www.straive.com>.

✉ [straiveteam@straive.com](mailto:straiveteam@straive.com)