

Client Profile

A recent class-action lawsuit required a leading tobacco firm to make available to the public approximately 10 million internal documents comprising memos, notes, reports et cetera submitted during the litigation discovery process. These documents were believed to contain evidence of *a priori* knowledge about the harmful effects of tobacco.

The court order was fulfilled by the creation of a public repository of documents in England. The Guildford Archiving Project started out as a means of expanding public access to these documents.

Challenges

Prior to the project, tobacco control researchers had no meaningful way to access the potentially invaluable information contained in the repository. The documents were in paper format, and were not organized in any useful manner. Additionally, our clients had to follow convoluted process to request copies of the documents. It was typical of the tobacco company to delay delivery of such documents and there were instances of several documents containing incriminating evidence being withheld from such deliveries. Managing this process and ensuring access to all documents was next to impossible for this high volume environment.

For the researchers to do a meaningful analysis of the documents it was necessary for them to read every single page before determining the important of the document. Review of the multi-million page paper repository was cost prohibitive if not impossible.

Collaboration between researchers and cross-referencing documents from different sections of the archive was also not possible in this environment.

The Axiom Solution

Axiom was the exclusive solution provider engaged to scan millions of documents (approximately 8 million from the tobacco firm and additional documents that were rolled into the project) and to extract substantial amount of meta-data based on complex business rules. The scanned images were also made searchable using a combination of optical character recognition (OCR) and manual data capture. Each page was tracked using the BATES number and missing pages were identified for each request.

To achieve the highest degree of accuracy, Axiom manually extracted every meta-data field while reviewing every single page. This included identification of pages that contained redacted content, and other important features including handwritten notes. Additional meta-data fields included references to any country, organization, or individual, as well as the author and recipient of each document. This allowed the researchers to effectively isolate documents of interest.

The website is maintained jointly by the London School of Hygiene and Tropical Medicine and UCSF. Tobacco researchers are now able to effectively analyze these tobacco industry documents to study nicotine addiction, cigarette marketing to children and other issues. More than 100 studies using data from industry papers have been published according to lead researchers at the Mayo Clinic.

The Axiom solution was showcased on television and newspapers both in the United States and the UK. Today Axiom's deliverable is part of California's Digital Library finally realizing the intent of the original court order making erstwhile inaccessible documents available to any interested party.

Quote

"[The website] launch marked a milestone in public information relating to smoking. Now people will have access to all this information 24 hours a day and for free. Before, if you really wanted a comprehensive search on a topic you would have to search all 8 million pages. Now you can't hide anything in there, and that is one really important advance."

– Dr. Kelley Lee, Sr Lecturer, London School of Hygiene and Tropical Medicine.

This high-visibility, multi-million dollar project was sponsored by a consortium of organizations.

Effective research and collaboration was impossible in a paper environment.

Manual metadata extraction and OCR facilitated access and search-ability of the documents.

Axiom created a proprietary application incorporating complex business rules for document processing.

Axiom deliverables are now part of the California digital library available online.

Contact:

Aman Ahuja
Business Development Manager
Tel: (925) 946-0267 Ext. 226
aman.ahuja@acg-usa.com