Adam Matthew recognizes the benefits that Data Mining has for new research in the Humanities and Social Sciences and we are committed to enabling these research methods, as endorsed by international law*.

Historically, Adam Matthew has worked with a number of content providers to allow Data Mining of their material. Examples include:

**‘Being Human: Mass Observation and Morale’**  
www.beinghumanfestival.org  
A week-long event, in partnership with the University of Sussex, at which invitees mined data from Mass Observation Online in order to generate visualisations based on the theme of morale in 20th century Britain.

**‘Trading Consequences’**  
www.tradingconsequences.blogs.edina.ac.uk  
A multi-institutional collaboration that uses text mining of four Archives Direct collections to explore international commodity trading in the 19th century British empire.

Librarians may contact Adam Matthew via info@amdigital.co.uk to discuss data mining requests.

* e.g. ‘Fair Use’ provisions in the United States, and ‘The Copyright and Rights in Performance (Research, Education, Libraries and Archives) Regulations 2014
DATA MINING/ TEXT MINING STATEMENT

WHAT IS IT?

The core concept is that computer software applies automated analytical techniques to interrogate data sets for patterns, trends and other useful information that typically would be incredibly labour intensive or difficult to conceive by traditional human research.

One of the key assumptions therein, is that the data is available to the software to carry this out. Indeed, the more controversial aspects of data mining revolve around the practicalities and responsibilities related to making information available to the software, and software “mining” that data from sources in order to analyze it.

In summary:

1. We allow Data Mining/Text Analysis by “Authorized Users” for fair use/academic research.
2. Secure online access to the data via an API* can be provided on submission of an information form (no further permissions or fees required)
3. Data can be extracted from the main collection website by automated software if we are informed about this so we can monitor server performance and reserve the right to restrict this operation if it impacts standard online usage for our customers generally.
4. We are committed, where possible, to apply text analysis and data visualization functionality within our latest products.

* An API is an Applied Programming Interface: A standardized application that provides access to a collection’s metadata and full text data on a stand alone secure server. It is a technical interface not a front end search engine, so someone at the customer end has to have some technical knowledge to set up client software to make search requests of our JSON data.

DO WE ALLOW DATA MINING/TEXT ANALYSIS?

Yes.

However, data mining as an activity is no different from all other usage of our products. It has to conform to all the standard requirements in our licence agreement e.g. it is carried out by Authorized Users under Fair Use academic purposes.

Extract of Standard User Licence Agreement:

Subject to all other provisions of our User Licence Agreement and save for the circumstances (as set out in section III of this Agreement) in which the Licensor’s prior written consent is required, the Licensee and the Authorized Users may use the Licensed Materials to perform and engage in text mining /data mining activities in relation to the Licensed Materials for legitimate academic research and other non-commercial educational purposes, without obtaining the Licensor’s prior written consent.
DATA MINING/ TEXT MINING STATEMENT

ARE THERE ANY RESTRICTIONS ON DATA MINING?

Electronic analysis of data from our products is permitted as outlined above; however there are two key elements that mean we have to have additional processes in place to ensure the following:

1. Performance of live product websites for standard usage are not damaged by automated data mining software crawling online websites.

2. Large volumes Data extracted or full data sets provided from the products are stored in a secure way that does not risk the availability of that data to unauthorized/open usage and therefore risk breaching User Licence agreement.

As a result, any significant automated data extraction or provision of large volumes of data is unauthorized without receiving written request and in offline data supply; permission being granted in writing. As long as suitable assurances as to the purpose and security of the research is assured on completion of a form then this provision will not be unreasonably withheld.

Extract relevant section of standard user licence agreement:

In order to protect the integrity of server performance for the Licensee’s customers, automated extraction of data directly from the Licensed Materials online (for example only, by the use of data mining software) is only permitted after notification to the Licensor for performance monitoring purposes, and if such automatic extraction of data does not affect the performance of the Licensor’s servers. In the event that the Licensor’s servers are negatively impacted, the Licensor reserves the right to decline and prevent access to the Licensed Materials to stop any disruption to the Licensor’s business.

HOW CAN DATA BE MADE AVAILABLE OUTSIDE THE MAIN WEBSITE?

As standard with no permissions:

A Data Mining API
This provides documented API access to the full data set (full text if available but not images) locked down by IP address and API key for security.

Subject to approval that the local storage/hosting provision is secure:

An offline copy of data provided on a hard drive for secure local storage and analysis. Under current agreements this is limited to a 3 year storage period after which time a renewal can be requested or if project complete, the original data (not any research material) deleted.

Extract of relevant part of licence:

On submission to the Licensor of completed form outlined in Appendix A, an offline copy of data from the Licensed Materials for Data/Text Mining purposes can be made available to be securely hosted locally and accessed by Authorized Users. Local hosting for each Data/ Text Mining purpose must not exceed five years unless further written consent is provided by Licensor; after which agreed period the data must be returned or confirmed as destroyed within 15 days.

Licensor and copyright holder of Licensed Materials must be acknowledged in published text analysis research results derived from the Licensed Materials.
DATA MINING/ TEXT MINING STATEMENT

EXAMPLE PERMISSION FORM

Text and Data Mining Information and Permission Request

Applicant name:

Institution:

Contact email:

Contact phone:

Project title:

Name of Adam Matthew product:

Adam Matthew is always interested in supporting research initiatives and learning more about how our products are used. Please provide as much information as you can in response to the following questions:

1. Could you supply a brief overview of the research project, outlining the overall proposal, why you consider this AMD product will be of interest, and initial expectations for data extraction?

2. What is your preferred method for data provision?
   - Client API
   - Offline copy of data set

3. Are you seeking to conduct datamining across other products as part of this research? If so, please provide details.

4. What is your intent regarding dissemination of search results? (i.e., published professionally, made available via the open web, etc.)?

5. What tools do you plan to use to extract the data?
   What tools do you plan to use to analyze these extractions?

6. If you require an offline copy of the dataset, will the original source data be stored/hosted locally by your institution or hosted in the cloud? If so, please answer the following questions in relation to this data.
   a. Where will the data be stored while access is provided?
   b. For what period of time will it be stored?
   c. Who will have access to the data when it is stored?
   d. How will limiting wider access be assured?
   e. What format will the data be in while it is stored?
   f. Does the storage platform share data with any other platforms? If so, with what/in what way?
   g. What measures are in place to assure data security and integrity?
   h. Once the project is completed, how will the original source data be disposed of? How will that outcome be assured?

7. All use of original source data and the results of searching and extracting data therefrom shall be in strict accordance with the terms of the Adam Matthew End User License Agreement and U.S. copyright law. Any other use is prohibited.