Data is a corporate asset that needs to be protected from loss, manipulation and theft. This guide provides an overview of the common methods of **data mining and analytics** and explains how you can use them to guard against fraud.

INBRIEF

A BEGINNER'S GUIDE TO ANTI-FRAUD MEASURES | DECEMBER 2015

FRAUD ADVISORY PANEL

Data mining and analytics

WHAT IS DATA MINING AND ANALYTICS?

Data mining is the process where a single piece of information is extracted from a larger group of information.

Before the internet you would use a telephone directory to find a phone number by searching alphabetically for another person's name such as John Smith (this is called data mining). If there were several John Smiths listed and you knew roughly where he lived, you would use that additional piece of information to focus or refine your search parameters (this is called data matching) to find the number you wanted.

Nowadays, the telephone directory has been replaced by the internet and there are lots of different sources of information. Businesses rely on holding data about their employees, clients, customers, sales inventories and management accounts. Automated search engines such as Google make the process of data mining simple and transparent. While search engines seamlessly retrieve information from the internet, many organisations still have their internal data located in separate systems, or even single-user PCs and these 'data silos' present a different set of challenges.

WHY IS IT IMPORTANT?

Data mining can be useful for analysing operational and transactional information to highlight anomalies that need further examination. In particular, data mining can help to uncover suspicious or unusual gaps, duplicates, patterns, trends, or values which may be due to fraud or error.

It should form a key part of a proactive fraud detection strategy and act as a mechanism for quantifying potential fraud losses.

WHO IS RESPONSIBLE?

Responsibility for data mining and analytics is often widely shared within an organisation.

- Directors and/or trustees (in the case of charities): data mining strategy and policy.
- Heads of finance or internal audit: implementation and oversight of data mining systems, processes and analysis.
- IT department: data extraction.
- Data protection officer: registration with the appropriate authorities (if required) to enable data mining techniques to be used lawfully and ensure ongoing compliance with local data protection and privacy legislation.
- Human resources department: insertion of appropriate wording into employee contracts and staff handbooks and keeping it up to date.

Many organisations outsource their IT function to a third-party service provider. A common pitfall in this approach is that the cost of data extraction is not factored into the service contract. Organisations that wish to implement a regular data mining programme need to ensure that their service provider will provide data extracts on a regular basis at no extra cost.

SOURCES OF DATA

Data exists in the following key locations.

Primary data sources
These consist of the core financial and administrative systems such as human resources, payroll, accounts payable, stock and inventory systems. With the increasing use of 'cloud storage' an organisation may have to negotiate with its service provider to gain access to its primary data sources if these are held externally.

Secondary data sources
These include old 'legacy' systems, archives or backup storage. In addition, systems such as telephone management and billing, access control security, email and internet logs will contain a wealth of data.

Tertiary data sources
These may comprise local end-user developed spreadsheets or access databases, or even manual logs and documentation. For example, while active suppliers will exist in the accounts payable system, details of companies invited to quote for business but who were unsuccessful may only exist as Word documents or in a manual register.

The internet
This can also be a source of external reference data that can be used for data matching. For example, Companies House for information on disqualified directors or the London Gazette for notices of bankruptcies, liquidations, administrators or receivership.

KEY ELEMENTS OF EFFECTIVE DATA MINING

- Define the internal 'data universe' and document which systems hold data that will be analysed to detect fraud. Consider primary, secondary and tertiary data sources as well as the internet.
- Document the time period for any regular data matching processes so that the schedule is commensurate with the risk.
- Assign clear responsibilities and secure senior management buy-in and support.

COMMON METHODS OF DATA MINING

Data mining and analytics can be performed using a number of different systems and platforms.

Spreadsheets
For small sets of data it is possible to use spreadsheets or databases. However, there are a number of fundamental drawbacks to using spreadsheets as a data mining tool, such as:
- the number of records which can be reasonably imported;
- the lack of an audit trail; and

- the automatic reformatting of date and numeric fields.

Specialist audit software tools
There are a number of specialist audit software tools which have the ability to process unlimited numbers of records, with built-in audit logging capabilities and a host of pre-defined functions and formulae. These are often used by forensic accountants and external auditors.

Core back-office Enterprise Resource Management (ERM) systems such as SAP or Oracle can also be used for data mining. There are several considerations to their use.

- While IT staff generally have the necessary software skills to write data matching queries in SAP or Oracle, they usually have a backlog of system development requests requiring agreed priorities and budgets.
- There may be a potential conflict of interest or lack of independence if IT staff are extracting data about the systems they are controlling.

LEGAL CONSIDERATIONS

It is important to comply with data protection and privacy legislation relating to the use of personal data. For example:

- registering with the appropriate data protection authorities such as the Information Commissioner's Office (ICO) and following their guidance;
- inserting appropriate wording into employment contracts on the use of personal data and continuous data mining programmes; and
- ensuring appropriate legal and contractual arrangements are in place in relation to the transfer of data to different jurisdictions and/or third parties.

CONDUCTING DATA MINING

Do:

- ✔ Consider how, when and where data is processed within the organisation. Build and maintain a data library for this purpose.
- ✔ Extract data on a regular basis to test data mining procedures and provide continuous assurance.
- ✔ Verify the data extract is accurate by cross-checking against the source data. Pay specific attention to the number of records, date formats, currency and credit/debit indicators.
- ✔ Be aware that spreadsheets may automatically alter the data format.
- ✔ Independently validate the results. Conduct a 'sanity' check as the initial results may not always be what was expected.
- ✔ Comply with local data and privacy legislation. Conduct a privacy impact assessment in the early stages of any new project and throughout its life cycle. The ICO has a code of practice and assessment template that you can use.
- ✔ Understand the parameters of contracts with third-party IT service providers in relation to data extraction.
- ✔ Keep a contemporaneous audit log of all analytics performed in case they are needed for legal proceedings. A case may be weakened if the results cannot be replicated at some point in future.
- ✔ Ensure that the data being analysed is always kept secure.

Don't:

- ✘ Presume that data extraction is included within contracts with third-party IT service providers.
- ✘ Assume that IT will be able to quickly extract data in an active emergency or that they will be able to interpret the results.
- ✘ Over complicate the process. Consider the audience.
- ✘ Send data and/or the results via email without encryption or password protection.

FURTHER INFORMATION

Available from the **resources** section of our website:

- Fraud detection
- Fraud indicators
- Fraud risk management
- Securing board-level support for anti-fraud measures.

Other resources

- **Information Commissioner's Office**
- **ICAEW (search keyword: data mining)**

ACKNOWLEDGEMENTS

Thanks to Richard Kusnierz (Haymarket Risk Management Ltd) for his valuable contribution to this guide.

SUMMARY OF THE MAIN METHODS

Data volume	Type of solution	Typical costs	Considerations
Small	Spreadsheets and databases	Minimal, included as desktop tools	◆ May be limited in size ◆ No audit log facility ◆ May automatically reformat date fields, truncate or lose leading zeros in numeric fields
Unlimited	Audit software tools such as ACL, IDEA, Arbutus, Tableau	The industry is moving away from software purchase pricing to annual licences	◆ Requires a degree of training ◆ May experience resistance from traditional IT as such an approach allows users the ability to take control of their own data
Unlimited	ERM systems such as SAP or Oracle	Hefty annual licence costs	◆ Requires IT support to access the core data tables and fields, and specialist programming skills ◆ Lack of independence

FRAUD ADVISORY PANEL

Chartered Accountants' Hall
Moorgate Place
London EC2R 6EA UK

T +44 (0)20 7920 8721
E info@fraudadvisorypanel.org
www.fraudadvisorypanel.org

FRAUD
ADVISORY
PANEL

Company Limited by Guarantee Registered in England and Wales No. 04327390
Charity Registered in England and Wales No. 1108863