

CASE STUDY

CLIENT NAME

Anonymous

CLIENT TYPE

Diversified
Commercial Bank

CLIENT LOCATION

Southern Europe

CLIENT SINCE

2006

DATA CLEANSING AND ENRICHMENT: MAJOR EUROPEAN BANK

BACKGROUND

This major European banking client had been collecting internal loss event information for several years, covering actual losses, pending litigation and near misses. Initially, data was collected in an Excel spreadsheet while an internal database was constructed, then all data was migrated into the database during 2006. The data was classified using the Basel II loss event types, causes and business lines. The initial data did not include any descriptive information, although from mid-2006 onwards, new events had some level of description added. Through until the end of 2008, loss event reports were forwarded from operational units to the central operational risk team, who classified the events and captured them into the database. From 2009 onwards, use of the database was rolled out to both various operating units and to country operational risk teams. This introduced a new issue, namely a mixture of languages used for descriptive information.

THE ISSUES

During the annual ICAAP exercise, when the bank's historical experience was required to verify participant's responses, it became apparent to the Head of Operational Risk that the bank had several issues with its internal loss data, including:

- For the purposes of its RCSA, KRI and scenario analysis programmes, the bank had adopted the RiskBusiness Taxonomy, which employed at far greater levels of granularity than was being used for loss event data;



Graci's built-in machine learning capability allows for multiple levels of search criteria to be specified and then used to filter events meeting those criteria

- In trying to analyse potential exposures (RCSA data), historical data (loss events) and current exposures (KRIs), the absence of business process, controls and product or service information in the loss event data precluded accurate cross-referencing;
- As a result of changes in staff (specifically those responsible for loss event classification), evolution and maturity in operational risk thinking and the decentralisation of the loss data collection process, similar events were being classified in very different ways, thus distorting the risk profiles; and
- With over 120,000 loss events, events had a different level of data attributes, some missing descriptions, some not having an allocated cause, different languages used for descriptions and many incorrectly classified.

Given these issues, the Head of Operational Risk invited RiskBusiness to design an approach to resolve the issues while simultaneously adding value to the internal loss data programme. The resulting project plan contained three core phases:

- Phase 1: Taxonomy Migration – mapping the loss event classification structure to the Operational Risk Taxonomy and technically converting the data to a new structure.
- Phase 2: Data Cleansing and Enrichment – analysing existing classification and adding further classification attributes.
- Phase 3: Report Restatement – assessing overall change and communicating it to interesting parties.

PHASE 1: TAXONOMY MIGRATION

The first step was to map all existing classification structures used for internal loss data to the relevant Operational Risk Taxonomy structures, using holding structures labelled “For Cleansing” wherever unclear or invalid structures had been used. Following this, a technical exercise was undertaken by the client’s IT Department, supported by RiskBusiness technical consultants to run a utility against the loss database to modify its contents. At the same time, the entire Operational Risk Taxonomy was loaded into the internal loss database, replacing the previous classification structure.

PHASE 2: DATA CLEANSING AND ENRICHMENT

In the second phase, the bank’s internal loss database was imported into Graci by RiskBusiness, where the data cleansing and enrichment process was undertaken. Graci’s built-in machine learning capability allows for multiple levels of search criteria to be specified and then used to filter events meeting those criteria – examples of combinations could be a specific phrase used in the description, a date range, a business division, an existing cause or risk category, an economic effect, etc – anything which allows similar events to be identified. These filters return different numbers of events, anything from several thousands to individual events.

Once a set of events had been confirmed as all being similar, the assessment team, consisting of RiskBusiness and client staff, reviewed the selection and then re-classified the data set, ensuring that where necessary, additional data attributes were added. In addition to cleaning and augmenting the data, the process provided quality control and provided detailed training for client staff – while also identifying common problems and issues which could subsequently be addressed.

Some issues which were identified during the process included:

- In some cases, issues which were essentially RCSA concerns had been documented as a “near miss” loss event;
- The outcome of certain risk events were often documented as the loss event, rather than the event itself – for example, a lawsuit about some action by the bank is raised as the loss event, not the inappropriate action itself, or the compensation payment to a counterparty rather than the delivery delay which gave rise to an interest claim, etc.;
- Factors such as client type and distribution channel often result in incorrect classification, without the involved products and business processes being taken into consideration;
- Effect types often fail to reflect what the narrative suggests, for example, the narrative describes a client being compensated for fraudulent transactions against their account (restitution), while the effect type is set to “debt write-off”; and
- Similar events can be classified in very different ways by the same individual over a period of time.



Following completion, the RiskBusiness team designed, developed and delivered a detailed training session for all identified loss data collection staff

PHASE 3: REPORT RESTATEMENT

Once the entire data set had been cleansed and enriched, a new set of analytical reports was produced, then the revised dataset was exported and re-imported into the internal loss database.

As a result of the re-classification of many events, the additional of greater levels of detail and the reconsideration of certain classification “rules”, an analysis was undertaken between a set of reports from pre-project and from post-project. A narrative report was then produced by the project team and distributed internally to risk owners to explain the differences. The report and process was also reviewed by the Internal Audit Department.

Following completion, the RiskBusiness team designed, developed and delivered a detailed training session for all identified loss data collection staff, focussing on lessons learned, common pitfalls in classification and in how to best use the Operational Risk Taxonomy.