Data, Information and Knowledge Quality in Retail Security Decision Making

Simone Stumpf
(University College London, United Kingdom
S.Stumpf@cs.ucl.ac.uk)

Janet McDonnell
(University College London, United Kingdom
J.McDonnell@cs.ucl.ac.uk)

Abstract: Knowledge creation and organisational learning are as much about questioning assumptions as they are about harnessing what is already known. We describe a procedure for expressing knowledge, theorising from it, identifying data suitable for testing theories, and the value to a business of the outcomes it produces. This technique, called 'theorise-inquire', supports the validation of knowledge once it is expressed in a shareable form and draws attention to gaps in data and to information quality generally. We illustrate the ideas presented with examples drawn from work with profit protection specialists working in large retail organisations in the UK.
Keywords: knowledge quality, data gaps, repertory grids, data analysis
Categories: H.1, H.2, H.4

1 Introduction

Knowledge management, organizational learning and knowledge creation are as much about questioning assumptions as they are about harnessing what is already known. Experts may present a skewed view of reality based on their experience and prejudices, whilst available data may be not useful in questioning experience. We describe a technique, the 'theorise-inquire' technique, which supports specialists in their business decision making. The technique comprises the identification of interesting conjectures from repertory grids elicited from domain experts [Kelly 55], associating these with appropriate data sources against which they can be tested, identification of data gaps and finally the assessment of the conjectures to suggest a course of action. The technique supports business decision making by supporting the comparison of personal experience with factual data, thus putting businesses in a position to consider organisational knowledge in a broader context.

This work is taking place as part of a project concerned with the capture, representation and sharing of knowledge about dealing with the problems of theft by employees in retail organisations. Theft by employees is not well understood considering the size of the problem. In the UK nearly a third of all retail crime losses can be traced back to theft by someone internal to the company [British Retail Consortium 00]. To make decisions about prediction, counteraction and prevention of crime, businesses need to take into account the complex, uncertain and changing nature of staff theft [Felson and Clarke 98]. Improvements in knowledge management can help companies to address issues of loss prevention by providing greater efficiency in processing information using existing knowledge, and by supporting the

creation of new knowledge to adapt to a changing environment [Nonaka 94]. Our work deals with retail experts whose focus is on profit protection. They have to operate in an environment that is tuned to processes optimised for customer service and not for important support functions, and this creates a difficult challenge for them.

The aim of our work is to start by making experts' knowledge explicit, and then to use this knowledge to steer data analysis to support individual and organizational learning [Argyris and Schön 96]. It could be argued that knowledge should drive what data is collected, whilst data supports the confirmation of knowledge. On the other hand, [Alavi and Leidner 01] argue that knowledge is "personalised information"; tacit knowledge only becomes information if it is expressed in a processed form that can be shared. Taking these two perspectives together, knowledge management systems should allow individuals to process information to gain knowledge whilst at the same time to express their personal knowledge explicitly in a sharable form. We exploit the routes from knowledge to data and from data to knowledge. Repertory grids and associated "maps of expertise" function as a means to personalise information for an expert and to socialise knowledge in an organisation. The theorise-inquire technique supports the identification of data gaps and the validation of knowledge assumptions.

2 Data Gaps

Poor data quality leads to poor decision making [Redman 98]. Data gaps are often understood solely in terms of inaccurate or incomplete data. [Strong et al. 97] have extended this notion by formulating dimensions of data quality that range over intrinsic, accessibility, contextual and representational issues. Here we give examples of how data gaps and data quality issues cause difficulties in loss prevention, an essential but nevertheless marginalised role in the retail sector. Profit protection managers need to access a variety of data sources to inform their decision making. An investigation into the working processes of eight major UK retailers highlighted that the use of Electronic Point of Sales (EPOS) data presents a number of problems to the security support function. We have found that EPOS data often cannot be accessed at all by specialist decision makers; in other cases, security specialists are overwhelmed by the amount of data – some of it irrelevant – which needs to be retrieved and manipulated (a contextual problem). EPOS data needs to be processed, filtered and reorganised to be of any value as information that is understandable to profit protection managers (a representational problem). This has led to the proliferation of exception reports which filter and aggregate EPOS transactions. Even when they are readily available we find that exception reports themselves need further data associated with them to give sufficient context to inform investigations. As part of an investigation, information about employees needs to be linked to EPOS data, the way that employees are represented differs in these systems; this is a representational problem concerned with consistency.

Decision making about profit protection is not adequately served by data gathered to underpin the (customer-focussed) core functions of the business; thus, organisational security work proceeds in a setting which is not even "data rich, information poor". Support functions routinely lose out in the trade-offs during the

design of corporate databases; the data gaps are gaps by design. If we focus on supporting business activities with data we need to add to the data quality issues identified by [Strong et al.] a pre-requisite category to deal with whether data is available at all (an existence dimension) and whether it is in a form that is suitable for automated processing (a digitisation dimension). For example, profit protection managers make use of the knowledge that a fraudulent refund often occurs within a short timeframe of the original sale with which it is associated; in many EPOS databases the link between refund and sale is not captured in the data recorded about till interactions. They also use hand-written records as an important source of clues; these hand-writing clues are not available from electronically recorded data.

Specialist business support units do capture supplementary data to assist their particular data needs. Loss prevention units are no exception; these data resources need to be carefully designed, integrated and maintained to overcome data quality issues like those we have described. A business case has to be made for every extra demand for data collection and management. A major outcome of the theorise-inquire technique, which we describe next, is the identification of data previously overlooked by specialists or the organisation as a whole which may be an important resource for managing security functions.

3 The Theorise-inquire Technique

The 'theorise-inquire' technique proceeds through 4 stages [see Fig. 1]. First, tacit expert knowledge is made explicit – and therefore sharable – through the use of repertory grids as a conversational device. We provide means for organising and analysing repertory grids to derive information on potentially interesting features which allows business specialists to express, explore, clarify and refine their understanding of these features whilst using their wealth of tacit experience to decide what is interesting or what could make an impact on a business process; these are expressed as theories. Features involved in the theories are then associated with appropriate data, either by providing a mapping to data sources or identifying data gaps that need to be addressed. Finally, knowledge is tested by applying data analysis techniques to factual data and comparing the result with experts' theories.

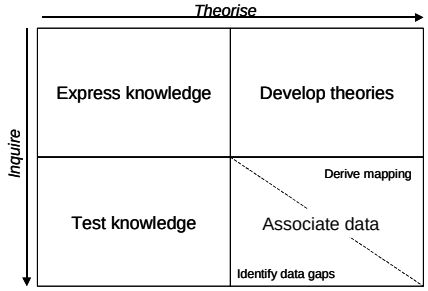

Figure 1: The theorise-inquire technique

3.1 Expressing Knowledge

We capture knowledge about stereotypical situations in very small data sets. To develop these data sets, we make use of the repertory grid technique. The repertory grid technique has been used extensively as a knowledge acquisition tool in expert systems [Boose 88] [Gaines and Shaw 93a] [Gaines and Shaw 93b]; we therefore do not describe it in any detail here. The technique functions as a conversational tool to make tacit expertise explicit and therefore amenable to inspection by knowledge users. As the technique is applied, individual situations that the expert has experienced (elements) are compared to draw out distinctions (constructs) between them. Distinctions are expressed as dimensions of contrasting poles. Each element is rated as to its location between the poles of each construct; ratings appear in the cells of the grid. Whilst distinctions are usually expressed as scalar dimensions, a rating can include binary, continuous and categorical data types [Boose 88].

An extract of a repertory grid for stock theft is given in [Fig. 2]. Elements, in this example cases of stock theft by employees, are shown in columns whereas construct poles are shown to the left and right of the matrix. In [Fig. 2] we use a five-point scale for all constructs for simplicity. A rating of 1 is associated with the construct pole on the left, a rating of 5 with the construct pole on the right.

1	Walford	Homerton Village	Hackney Marshes	Bluetrench	Strikeby	Lowsound	5
Good management compliance	5	5	5	3	3	2	Poor management compliance
Low store total net revenue	4	3	3	1	1	3	High store total net revenue
Permanent employee	1	1	1	1	1	5	Contracted employee
Area manager's base store	1	1	5	5	5	5	Not area manager's base store
No previous burglary	1	2	1	1	2	5	5 or more previous burglaries
No previous robberies	1	2	4	1	1	2	4 or more previous robberies
Immediate closure of audit (within period)	4	4	4	2	2	1	Prolonged audit closure (6 months or more)
Offender has family working in business	1	1	5	5	1	1	Offender has no family working in business
Keyholder offender	1	1	1	1	1	5	No keyholder offender
Offender Long service	1	2	2	1	5	5	Offender Short service
Offender Management	1	1	1	1	5	1	Offender Sales Person
High staff turnover (>30% p.a.) in store	1	1	1	4	4	2	Low staff turnover(<20% p.a.) in store
Inventory declaration available and justified	5	5	5	1	5	5	Inventory declaration not available or justified
High management turnover (>20% p.a.) in store	3	1	1	4	3	2	Low management turnover (<10% p.a.) in store
Suspect temporary promotion	5	1	1	5	5	1	Suspect not temporarily promoted
Activity evident after trading hours	1	1	1	1	1	1	No activity outside trading hours
good audit result/ low discrepancies	5	5	5	3	3	5	poor audit result / high discrepancies
low shrinkage store	5	5	4	2	3	5	high shrinkage store
Store risk level high	1	1	1	1	1	1	Store risk level low
low level of autoreplenishments	5	5	4	5	2	5	high level of autoreplenishments
Offender low transfer rate store to store (1 store per year)	5	5	5	1	1	3	offender high transfer rate store to store (4+ per year)
low volume rental product loss	5	5	5	2	3	5	high volume rental product loss
store without manager	1	1	5	1	1	5	store with manager
Offender low appraisal rating	2	3	3	4	2	2	offender high appraisal rating

Figure 2: Example repertory grid

We have helped each retailer to develop repertory grids containing staff theft knowledge for different sorts of offence, e.g. refund fraud, theft of stock, etc. Initial sessions typically extracted repertory grids that contained between 6 and 10 cases of theft or fraud as elements and between 20 and 55 characteristics of these cases as constructs. Follow-on sessions refined the grids until security experts were satisfied; the final grids usually included more cases and fewer characteristics than those produced at initial sessions.

3.2 Developing Theories

Expertise expressed in a repertory grid functions as a tool to think with by encouraging reflection on knowledge. Explicit knowledge can be manipulated, reorganised and analysed to allow the expert to learn new insights about their domain and develop theories. To develop a theory experts inspect the grid and identify constructs and relationships that they believe to be important or interesting. Experts can use analysis techniques to develop "maps" that reflect their theories about some aspects of their domain. The results of analyses are particularly interesting to specialists when they surprise them by challenging preconceived views or by providing new information. Grids can be treated as small data sets, where constructs can be viewed as attributes and elements as records. In principle, many data analysis techniques can be directed at grid data. Commonly used techniques for analysing grids to produce 'maps of knowledge' include decision trees or rules [Gaines 89], hierarchical clustering [Shaw 80] and principal components analysis [Slater 76].

Since we are interested in identifying data gaps and steering analysis of data sources like EPOS data and personnel records we have extended the range of analyses that are available to experts [Stumpf and McDonnell 02a]. For example, we have developed analysis techniques that are based on contrasts between elements, constructs and construct ratings, such as group contrasts [Stumpf and McDonnell 03]; these help profit protection managers in developing theories. Group contrasts are an analysis technique that allows an expert to explore and interpret repertory grids interactively to find significant contrasting relationships between attributes and test these against data sources.

Several of our retailers have developed the theory that employees that have responsibility for keeping store keys are much more likely to be involved in stock theft as a result of work on producing their repertory grids. To a casual reader of the grid in [Fig. 2] this may seem obvious, to the retail companies themselves this is new and important information that was previously hidden amongst the interaction of a large set of staff theft characteristics. A more complex theory concerns the relationship between the number of previous burglaries and the incidence of stock theft by employees in a store. Work is ongoing to develop theories on offender and store profiles that are important in refund fraud and stock theft. The theories resulting from analysis of repertory grids are used as a basis for identifying both data gaps and further tests which can be carried out on existing databases.

3.3 Associating Theories with Data

To allow strategic decision making and policy design in a corporation, it is necessary to move from personal knowledge to organisational knowledge. Repertory grids only carry personal meaning and suffer from restrictions due to the small data sets from which they are constructed (in our work, repertory grids typically characterise 10 cases through 50 attributes). Organisational knowledge rests on adequate generalisation from these small data sets and a formalisation of data requirements.

In this phase of the technique, experts look for attributes in organisational data sets that match features derived from a repertory grid and attempt to apply a

mapping between them. In the simplest case, it would be possible to map all features to existing data. Most commonly, in practice, one or more data quality problems are identified during the mapping attempt. For example, all retailers we have worked with have identified issues of data quality in existing business databases. These issues range from incomplete data to data that is not gathered at all, as we have described above for example, those that occur in the use of EPOS and personnel records where we find accessibility, contextual and representational problems.

Where conjectures cannot be tested because the mapping fails, gaps need to be filled by reconsidering the business priorities for collecting data. An assessment needs to be made of the gains to be made in loss investigation versus the costs of data collection. Where steps have to be taken to address the data problems, constructs in a repertory grid can be used to inform data requirements. In this case, constructs taken from a repertory grid have formed a rough-cut structure for gathering further data.

3.4 Testing Knowledge

Where information from data sources has been associated with constructs and elements, the expert's theory can be tested by reference to a wider sample. To apply a test, data sources are analysed using the same technique as the one applied to the initial repertory grid analysis to generate a theory. To support the testing of knowledge we use group contrasts that compare results from both repertory grids and data sources [Stumpf and McDonnell 03]. Testing assesses the quality of knowledge by comparing personal experience and factual data; a pre-requisite to modifying operational procedures which are sanctioned on an organisational level. For example, the theory of key holder being much more likely to be involved in stock theft than non-key holders was tested against data available from a security database. The theory was confirmed and the way that retailers investigates stock theft has been modified.

In the absence of large-scale data sets on an organisational level, promising conjectures can be tested by one-off field investigations. This is the approach taken to explore the relationship between burglaries at stores and the amount of theft of stock by staff and further serves to underpin a business case for policy change.

On contrasting the theory and the test – the former derived from repertory grids, the latter from existing data sources – it would be tempting to give greater credence to the test with a larger data set. If this is all we get, it could be argued that we are merely showing how representative or otherwise the small set of cases used in the repertory grid are of cases overall. However, these tests are as much about questioning assumptions as they are about harnessing what is already known. A theory that is rejected still has its use in this context as an opportunity for an expert to reflect on their knowledge and revise their (mis)conceptions. Furthermore, our investigation raises an interesting question for the businesses involved when a theory and its test resulting from existing data sources diverge widely. Neither appear to be perfect: experts may present a skewed view of reality based on their experience and prejudices, however it is also possible that it is the existing data that is suffering from data quality problems. In either case, the comparison draws businesses' attention to issues based on evidence rather than speculation.

4 Conclusion

We have presented our work with retail security specialists investigating staff theft. The 'theorise-inquire' technique was described; this procedure enables experts to express and capture knowledge, develop theories, associate these theories with data, identify data gaps and test theories to establish and improve organisational knowledge. Using this technique supports business decision making by specialists and organisations as a whole by allowing the comparison of theories based on experience with factual data.

However, there are some potential pitfalls. Seeing elements and constructs as records and attributes invites criticism about the sample size a grid is based on. Such criticism displays an impoverished understanding of repertory grids and detracts from the value of the technique: the repertory grid technique originates from expressing the way that an individual construes experiences. However, the quality of a grid is very sensitive to the conversational approach used to elicit it in the first place and, associated with this, the quality of 'items of experience' which are used to construct a grid. Thus, great care must be taken to choose stereotypical elements that cover the range of the domain under investigation [Gaines and Shaw 93b]. The issues that remain in our work are that counterexamples to staff theft are not readily available; hence, conceptions of what is not suspicious need to be reintroduced at some stage.

We find that the knowledge an organisation holds cannot easily be shared across organisations although, in the case of our work, organisations are both willing and anxious to learn from each other [Stumpf and McDonnell 02b]. Once grids are in a form that supports organisational learning – within one organisation – we find that the most interesting insights are not easily transferable to others. It suggests that the sorts of grids, theories and tests which promote learning within retailers are not those which promote 'best practice' exchanges between organisations.

Results of data analysis techniques are difficult to interpret by domain experts. Whilst we have identified a number of analysis techniques that are useful, work remains to make the results of these analyses understandable to security experts without mediation by others.

Loss prevention makes an important contribution to retail profits, however, these knowledge-intensive support activities suffer to a more extreme degree from the tension between experts' intuition and factual evidence than work that supports core business activities where feedback from data to a business strategy is more readily available. A study of knowledge management in these areas highlights problems arising from the tension more generally. The 'theorise-inquire' technique helps to make visible issues associated with data, information and knowledge quality, which is the first step in quantifying their impacts on an organisation.

References

[Alavi and Leidner 01] Alavi, M., Leidner, D. E.: "Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues"; MIS Quarterly, 25, 1 (2001), 107-136

[Argyris and Schön 96] Argyris, C., Schön, D.A.: "Organizational Learning II: Theory,Method, and Practice"; Wokingham / Addison-Wesley (1996)

[Boose 88] Boose, J.: "Uses of repertory grid-centred knowledge acquisition tools for knowledge-based systems"; International Journal of Man-Machine Studies, 29 (1988), 287-310

[British Retail Consortium 00] British Retail Consortium: "8th Retail Crime Survey"; London, UK (2000)

[Felson and Clarke 98] Felson, M., Clarke, R.: "Opportunity Makes the Thief: Practical Theory for Crime Prevention"; Police Research Series: Paper 98, Home Office, London, UK (1998)

[Gaines 89] Gaines, B.R.: "An ounce of knowledge is worth a ton of data"; Proceedings 6th International Workshop on Machine Learning, (1989), 156-159.

[Gaines and Shaw 93a] Gaines, B.R., Shaw, M.L.G.: "Eliciting knowledge and transferring it effectively to a knowledge-based system"; IEEE Transactions on Knowledge and Data Engineering, 5, 1 (1993), 4-14

[Gaines and Shaw 93b] Gaines, B.R., Shaw, M.L.G.: "Knowledge acquisition tools based on personal construct psychology"; Knowledge Engineering Review, 8, 1 (1993), 49-85

[Kelly 55] Kelly, G. A.: "The Psychology of Personal Constructs"; WW Norton & Company, New York, (1955)

[Nonaka 94] Nonaka, I.: "A Dynamic Theory of Organizational Knowledge Creation"; Organization Science, 5, 1 (1994), 14-37

[Redman 98] Redman, T.C.: "The Impact of Poor Data Quality on the Typical Enterprise"; Communications of the ACM, 41, 2 (1998), 79-82

[Shaw 1980] Shaw, M.L.G.: "On Becoming a Personal Scientist"; Academic Press, London, (1980)

[Slater 76] Slater, P.: "The Measurement of Intrapersonal Space by Grid Technique"; Wiley, London (1976)

[Strong et al. 97] Strong, D., Lee, Y., Wang, R.: "Data Quality in Context"; Communications of the ACM, 40, 5 (1997), 103-110

[Stumpf and McDonnell 02a] Stumpf, S.C., McDonnell, J.T.: "Taking off the blinkers and creating knowledge: decision support for retail security specialists"; 7th International Psychology and Crime Investigation Conference, University of Liverpool, 12-14 June, (2002)

[Stumpf and McDonnell 02b] Stumpf, S.C., McDonnell, J.T.: "Is there an argument for this audience?"; 5th International Confeernece on Argumentation, Amsterdam, 25-28 June, (2002)

[Stumpf and McDonnell 03] Stumpf, S.C., McDonnell, J.T.: "Using repertory grids to test data quality and experts' hunches"; Workshop on Theory and Applications of Knowledge Management (TAKMA'03), Prague, 1-5 September, (2003)